Self-perceptions about academic achievement: Evidence from Mexico City[☆]

Matteo Bobba [a,*], Veronica Frisancho [b]

[a] *Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France*
[b] *Research Department, Inter-American Development Bank, Washington DC, United States*

ARTICLE INFO

ABSTRACT

A growing body of evidence suggests that people exhibit large biases when processing information about themselves, but less is known about the underlying inference process. This paper studies belief updating patterns regarding academic ability in a large sample of students transitioning from middle to high school in Mexico City. The analysis takes advantage of rich and longitudinal data on subjective beliefs together with randomized feedback about individual performance on an achievement test. On average, the performance feedback reduces the relative role of priors on posteriors and shifts substantial probability mass toward the signal. Further evidence reveals that males and high-socioeconomic status students tend to process new information on their own ability more effectively.

1. Introduction

Recent work on social psychology suggests that self-assessments of individual traits are often flawed in substantive and systematic ways. For instance, it is often argued that people tend to hold rather favorable views of their abilities — both in absolute and relative terms (Moore and Healy, 2008; Dunning et al., 2004).

Upwardly biased self-views may have important economic consequences. Some studies, for example, show that managers tend to have more faith in their firms than is warranted (Daniel et al., 1998; Malmendier and Tate, 2005, 2008). When individuals invest in human capital early in life, biased beliefs about academic skills may be related to mistakes and mismatches in schooling choices that are hard to reverse, with potentially long-lasting consequences for labor market outcomes. A recent strand of the economics of education literature has identified important updating effects on beliefs

and choices due to the provision of information about individual academic performance.[1] Much less is known on the different pathways through which exposure to informative achievement signals may correct biased priors among youth on the verge of important schooling decisions.

This paper aims at partly filling this gap by taking advantage of longitudinal data on students' beliefs about their own academic ability and exogenous variation in exposure to individualized performance feedback. We first document the level of pass-through of the ability signal and the degree of persistence of priors in the distribution of posterior beliefs. Next, we dissect the updating process by proposing an updating model that flexibly explores the relationship between individual priors and posteriors. The model estimates shed light on the channels through which different individuals process information about their own ability.

The evidence is drawn from a field experiment embedded in the context of the centralized assignment mechanism that allocates students into public high schools in the metropolitan area of Mexico City. We design a mock version of the admission exam that we give to a large sample of potential applicants and reveal individual scores to a randomly chosen subset of test takers. We elicit repeated probabilistic statements about performance beliefs in the actual admission test over a discretized support to generate longitudinal measures of subjective expectations that are tightly linked to immediate and high-stakes schooling decisions. Baseline data collected before administering the mock test reveal that there are large discrepancies between students' prior expectations and their actual test scores, with relatively more upwardly biased and more uncertain beliefs among low-performing students.

We first estimate the average changes in the individual distribution of beliefs about academic achievement across students with and without access to the feedback on individual performance in the mock test. The feedback halves the relative role of priors on posteriors and increases the probability mass in the interval containing the signal by 17 percentage points. We further show that high-performing students exhibit higher signal pass-through than their lower-performing counterparts, while higher-socioeconomic (SES) students have relatively less persistent priors when compared to lower-SES students. Overall, these updating patterns may mask some changes along the individual distribution of beliefs that are difficult to capture in a simple reduced-form empirical framework.

We thus propose and estimate an updating model that allows us to detect more nuanced patterns of updating behavior. In this framework, probabilistic weights characterize the mapping between priors and posteriors in each interval of the support of the elicited belief distributions. These conditional probabilities can be estimated using the longitudinal variation in beliefs before and after the provision of performance feedback to the students in the treatment group. Estimation results confirm that the performance feedback represents an informative signal that spurs changes in the probability mass allocated to each interval of the prior densities. However, the results also point to a limited effect of the signal due to the relatively strong persistence of priors. The estimates reveal systematic differences in updating behaviors along the test score distribution, especially around the two tails. For instance, students who get the lowest-valued signals only partly incorporate them into their posteriors, whereas those who get the highest-valued signals still allocate some density to the lowest interval of the prior distribution. Further sub-group analysis reveals that socio-demographic characteristics are related to how students behave when processing new information about their own academic ability. Males and higher-SES students experience signal pass-through to a greater extent (i.e., they allocate a greater transition weight to the interval that contains the signal) and, overall, they tend to exhibit the most effective updating patterns when compared to other sub-groups of students.

Earlier studies document updating patterns that are more or less consistent with the Bayes rule (El-Gamal and Grether, 1995; Zafar, 2011). However, recent advances in behavioral economics confirm that agents systematically depart from the Bayesian updating benchmark (Benjamin, 2019; Fuster et al., 2019). In particular, (Wiswall and Zafar, 2015) document evidence on asymmetric updating in beliefs about earnings: self-beliefs seem to be more responsive to information when priors are below beliefs about population earnings.

Experimental evidence reveals that individuals exhibit particularly large biases when processing new information and forming perceptions about personal traits and skills. For instance, Eil and Rao (2011) use the quadratic scoring rule to repeatedly elicit beliefs about intelligence and beauty and find that agents' posteriors are less predictable and less sensitive to signal strength after receiving negative feedback. Burks et al. (2013) obtain an alternative test by combining cross-sectional data on beliefs and actual ability measures among truck drivers and also reject the null of Bayesian updating. In the context of an online experiment about relative performance on an IQ test, Mobius et al. (2011) find evidence of asymmetric updating and over-weighting of positive signals.

A number of papers have proposed various models of non-Bayesian inference that depart more or less radically from Bayes rule. For example, Rabin and Schrag (1999), Compte and Postlewaite (2004) and Koszegi (2006) slightly modify Bayes' rule by allowing decision-makers to discard negative feedback about themselves. Such parsimonious models preserve much of the predictive power of Bayesian updating but might be too ad hoc and restrictive to explain systematic deviations from Bayesian inference. At the other extreme, Akerlof and Dickens (1982), Brunnermeier and Parker (2005)

---

[1] Stinebrickner and Stinebrickner (2012, 2014) document substantial updating effects on beliefs and major outcomes for college students. Azmat and Iriberri (2010), Elsner and Isphording (2017) and Azmat et al. (2019) study the role of students' ordinal rank and feedback about relative performance on study effort and academic performance. Bergman (2015) and Dizon-Ross (2019) focus instead on whether and how information asymmetries between parents and their children affect schooling investments. In a companion paper (Bobba and Frisancho, 2019), we show that providing students with information about their own ability changes school choices and assignment patterns across schools, thereby reducing high-school drop-out.

and Benabou and Tirole (2002) allow agents to optimally choose subjective beliefs. While these models can help explain some of the patterns in the data, it is hard to imagine a single framework that governs the widely heterogeneous patterns of updating behavior across different individuals.

Our work also departs from the Bayesian benchmark as we flexibly estimate the updating parameters that govern changes in individual belief distributions about own ability after exposure to an informative signal. The results of this study may inform the design of policy interventions aimed at disseminating information in education markets. In particular, our estimates provide direct evidence on whether and how the patterns of belief updating systematically vary depending on individual traits such as academic ability, gender, and SES. These heterogeneous effects point towards the need of complementary strategies of information provision that may potentially enhance the degree of signal extraction among certain sub-groups.

2. The feedback provision experiment

2.1. Context and experimental design

Access to public schools at the upper-secondary level in Mexico City is regulated by a centralized assignment mechanism known as the COMIPEMS admission system (by its Spanish acronym). In 2014, the year in which the experiment and the data collection took place, over 238,000 students were placed in the 628 participating public high schools. Overall, the assignment system accounts for roughly three-quarters of high school enrollment within the metropolitan area.

The application process starts by the end of the last year of middle school. Ninth graders receive all relevant information about the process through a booklet containing important dates and detailed instructions of the application process as well as information about the available high school programs, including the curricular track or modality (general, technical, or vocational) and the corresponding cut-off scores for the past three years. To participate in the admission process, students submit a registration form and a socio-demographic survey, as well as a rank-ordered list of up to 20 preferred schools early in the calendar year. Placement in a given school is solely determined by students' submitted choices and their scores in a single standardized achievement exam, which takes place in July, after registration and towards the end of the school year.[2] The timing of the events throughout the admission process implies that the submitted school rankings – which partly determine the observed sorting patterns across schools – are sensitive to students' subjective expectations of their own performance in the admission exam.

We design and implement a field experiment that provides students with individualized feedback on their academic skills during their transition from middle to high school. We administer a mock version of the admission test, communicate individual scores to the treatment group, and elicit probabilistic statements about performance beliefs in the admission test. In this setting, the score in the mock exam provides students with a signal about their own academic skills.

Among the universe of public middle schools in the Mexico City metropolitan area, we focus on those located in neighborhoods with high or very high poverty levels, since the students therein are less likely to be exposed to previous signals about their individual performance in the admission exam.[3] We further restrict the sample to schools with a large mass of potential applicants to the school assignment mechanism as measured by their relative contribution to the pool of applicants in the year 2012. Even though we focus on less advantaged students, Table 1 shows that our final sample is largely comparable to the general population of applicants in terms of basic socio-demographic characteristics, initial academic credentials (GPA in middle school), preparation and support through the assignment mechanism, and personality traits.

The final sample is comprised of 90 schools distributed across 12 strata, which are defined by four geographic regions and terciles of school average performance among ninth-graders in a national standardized achievement test (ENLACE, 2012). We randomly pick one ninth-grade classroom in each sampled school to participate in the experiment. Treatment assignment is randomized within strata at the school level. As a result, 44 schools are assigned to a treatment group in which we administer the mock exam and provide face-to-face feedback on performance (see Section 2.2), while 46 schools are assigned to a control group in which we only administer the mock exam, without providing information about the test results.

---

[2] A deferred-acceptance matching algorithm (see, e.g., Pathak, 2011) with priorities defined by the individual scores in the admission exam is used to assign students to their most preferred schooling option with available seats. Whenever ties occur, participating institutions agree on whether admitting all tied students or none of them. Applicants who are not placed by the algorithm can request admission to schools with available seats in a second round of the assignment process or search for a seat in private schools or public schools with open admissions outside the system. Whenever applicants are not satisfied with their placement, they can request admission to another school in the same way unassigned applicants do. All in all, the matching algorithm discourages applicants to remain unplaced and/or list schools they will ultimately not enroll in. About 10 percent of the students in our sample do not apply for the COMIPEMS assignment mechanism. Among those who participate in the admission system, 11 percent remain unplaced and only 2 percent are admitted through the second round of the matching process.

[3] Recent evidence from the United States documents that less privileged students tend to be relatively more misinformed when making educational choices (Hastings and Weinstein, 2008; Avery and Hoxby, 2012). Administrative data from the 2012 edition of the assignment system shows that, on average, 33 percent of applicants took a preparatory course before submitting their schooling choices. This share ranges from 44 to 12 percent across schools in neighborhoods with low and high levels of poverty, respectively.

Table 1
Applicants' characteristics.

Sample	All applicants		Experiment		Adj. Mean difference
statistic	Mean	SD	Mean	SD	*P*-value
Works	0.273	0.446	0.315	0.465	0.152
Indigenous student	0.041	0.199	0.098	0.297	0.001
Disabled student	0.113	0.317	0.141	0.348	0.159
Scholarship in middle school	0.112	0.315	0.108	0.310	0.881
Grade retention in middle school	0.134	0.340	0.142	0.349	0.316
GPA (middle school)	8.130	0.894	8.147	0.843	0.477
Lives with both parents	0.746	0.436	0.789	0.408	0.000
Mother with college degree	0.117	0.321	0.053	0.224	0.004
Asset-based SES index	−0.040	1.904	0.012	1.836	0.000
Exam score in COMIPEMS	70.99	21.17	64.91	19.78	0.000
Took preparatory courses	0.536	0.499	0.508	0.500	0.358
Guidance through application	0.977	0.151	0.977	0.149	0.257
Plans to go to college	0.808	0.394	0.724	0.447	0.001
Does not give up	0.882	0.323	0.879	0.326	0.445
Tries his/her best	0.760	0.427	0.720	0.449	0.259
Finishes what he/she starts	0.756	0.430	0.713	0.452	0.079
Works hard	0.766	0.423	0.723	0.448	0.101
Number of observations	203,121		2293		205,414

NOTE: The 'Mexico City' sample consists of all applicants in the centralized assignment in the year 2014 from the Mexico City metropolitan area who were assigned through the matching algorithm. The 'Experiment' sample consists of the sample of students selected from the population above according to the criteria specified in Section 2.1. The P-values reported in the last column refer to the adjusted differences in means between the two samples after controlling for municipality fixed effects.

2.2. Beliefs elicitation and data

Beliefs are measured in two survey rounds, both before and after the application of the mock test. The mock test was administered a few days after the baseline survey and the score obtained was provided to the treatment group during the follow-up survey, which took place a few weeks before the beginning of the registration period for the school assignment process. Beliefs among students in the treatment group are collected twice during the follow-up survey, both before and after the delivery of the performance feedback. To accurately measure probabilistic statements about individuals' achievement in the test, the elicitation process in both survey rounds relied on visual aids (Delavande et al., 2011). We explicitly linked the number of beans placed in a cup to a probability measure, where zero beans correspond to a zero probability event and 20 beans indicate that the student believes the event will occur with certainty. The survey question eliciting beliefs reads as follows (authors' translation from Spanish):

"Suppose that you were to take the COMIPEMS exam today, which has a maximum possible score of 128 and a minimum possible score of zero. How sure are you that your score would be between … and …"

When asking this question, surveyors provided students with a card divided into six discrete intervals of the score and then asked them to allocate the 20 beans across the intervals so as to represent their perceived chances of scoring in each bin. When delivering the individual scores in the mock exam, surveyors showed a personalized graph with two pre-printed bars: the average score in the universe of applicants during the 2013 edition of the school assignment mechanism and the classroom-average score in the mock test. During the interview, a third bar was plotted corresponding to the student's individual score in the mock test. The surveyors' interactions with the students were always private, without the interference or presence of other students or school staff. This minimizes issues related to social image concerns when reporting subjective beliefs (Ewers and Zimmermann, 2015; Burks et al., 2013).

The mock test that we use to measure academic achievement was designed by the institution in charge of the official test in order to mimic its structure, content, level of difficulty, and duration (three hours). The test comprises 128 multiple-choice questions worth one point each, without negative marking, covering a wide range of subjects that correspond to the public middle school curriculum (Spanish, mathematics, social sciences and natural sciences) as well as mathematical and verbal aptitude sections.[4] The linear correlation in our sample between performance in the mock exam and the actual exam is 0.82. In turn, the linear correlation between a freely available (but possibly noisier) measure of ability such as middle school GPA and the score in the mock test is 0.45. The mock test score is also a strong predictor of high-school outcomes, such as GPA and graduation on time, even after controlling for GPA in middle school (Bobba and Frisancho, 2019).

---

[4] Thirteen questions related to the curriculum material that had not been covered by the time the mock test was administered were not graded. Before providing feedback about individual performance in the test, we normalize raw scores in the 115 valid questions to correspond to the 128-point scale.

Table 2
Summary statistics and randomization check.

	Control (1)	Treated (2)	Treated-control (3)
Administrative data:			
Exam score	64.931	64.883	0.235
	(19.647)	(19.926)	[1.168]
GPA in middle school	8.138	8.157	−0.003
	(0.851)	(0.834)	[0.050]
Scholarship in middle school	0.103	0.113	0.007
	(0.304)	(0.316)	[0.015]
Grade retention in middle school	0.148	0.136	−0.005
	(0.356)	(0.342)	[0.021]
Does not skip classes	0.965	0.975	0.011
	(0.183)	(0.157)	[0.010]
Plans to go to college	0.729	0.718	−0.014
	(0.445)	(0.450)	[0.021]
Disabled student	0.139	0.142	0.001
	(0.346)	(0.350)	[0.017]
Indigenous student	0.094	0.101	0.011
	(0.292)	(0.302)	[0.015]
Lives with both parents	0.784	0.795	0.010
	(0.412)	(0.404)	[0.018]
Works	0.324	0.306	−0.021
	(0.468)	(0.461)	[0.021]
Mother with college degree	0.052	0.055	0.002
	(0.222)	(0.227)	[0.011]
Father with college degree	0.092	0.098	0.007
	(0.290)	(0.298)	[0.015]
High SES (asset index)	0.484	0.518	0.024
	(0.500)	(0.500)	[0.025]
Number of observations	1192	1101	2293
Survey data:			
Mock exam score	58.772	60.752	1.654
	(15.618)	(16.403)	[1.075]
Mean beliefs at baseline	74.388	74.449	0.015
	(14.422)	(14.404)	[0.955]
SD beliefs at baseline	18.056	17.624	−0.526
	(8.287)	(8.328)	[0.455]
Previous mock exam with feedback	0.139	0.174	0.030
	(0.346)	(0.379)	[0.036]
Male	0.469	0.497	0.024
	(0.499)	(0.500)	[0.017]
Number of observations	1318	1226	2544

Note: Columns 1 and 2 report means and standard deviations (in parentheses). Column 3 displays the OLS coefficients of the treatment assignment indicator and the standard errors (in brackets), which are clustered at the middle school level.

We interviewed 3,001 students in the baseline survey, and 93 percent of them took the mock test. The number of students surveyed in the follow-up survey was 2,839. In order to document updating patterns we need complete data on prior beliefs, signals, and posterior beliefs and hence we drop students with incomplete survey records, which yields a final sample of 2,544 students. Since the provision of performance feedback took place during the follow-up survey, there is no differential attrition by treatment assignment (P-value $= 0.549$).

The survey data are complemented with individual-level administrative records from both the registration form and the assignment process in itself, which allow us to observe admission exam scores, cumulative GPAs in middle school, socio-demographic information, and other individual characteristics such as personality traits and study habits. About 10 percent of the students in our final survey sample do not apply to the COMIPEMS system and thus do not fill the registration form. With the exception of households' socioeconomic status (SES), all the other relevant variables used in the empirical analysis come from survey records.[5]

Table 2 provides basic descriptive statistics and shows that the clustered-block randomization design we implemented was successful in achieving balance between the treatment and control samples along a variety of individual characteristics as well as household socio-demographic variables.

---

[5] Table A.1 in the Web Appendix checks the share of missing records for SES by treatment status and tests the differences across them. The estimates reported in Column 1 show that 21 percent of students in the control group have no records on SES, and this share is not affected by the exposure to performance feedback. Column 2 documents that after conditioning on those who apply to the COMIPEMS system the share of missing records falls to 13, and it still does not differ across treatment status.

2.3. Descriptive evidence on beliefs

The elicited distributions of beliefs about test performance seem well-behaved. Using the 20 observations (i.e., beans) per student, a normality test (Shapiro and Wilk, 1965) is rejected for only 11 percent of the individual baseline distributions in the sample. As few as 6 percent of the respondents place all the beans in one interval of the grid, which suggests that the chosen discretization of the support of the admission exam is not too coarse for the vast majority of the applicants in our sample. Still, there are relatively fewer observations in the upper tail of the score distribution, in line with average mock and admission exam scores around 59 and 65 points out of 128, respectively. Thus, throughout the analysis, we merge the last two intervals into one (85–128).

Fig. 1 reports the average frequencies of the prior distributions (i.e., beliefs before giving the mock test) conditional on the score in the mock test. While there seems to be quite a lot of dispersion in the priors for each discrete interval of the score, there is a clear shift in the probability mass towards higher-valued intervals as the score increases. This visual pattern indicates some degree of accuracy in the subjective expectations elicited in the survey, especially among higher-performing students. Among those with lower scores in the mock test, the probability mass allocated to each interval is slightly increasing (rather than decreasing) along the support of the score.

Students seem to be quite uncertain about their own academic skills, as a large share of them allocate positive density mass to intervals far away from their actual score in the mock test. For each interval of the score, Fig. 2 plots the share of students that place at least one bean in each interval of the support of beliefs. At baseline, over a third of the students who later score in the highest interval assign positive probability to the event that they could score in the lowest score interval. A similar but even stronger pattern emerges among the lowest scorers: while 60% of them believe that they could indeed score between 0 and 40, almost 70% of them assign some density mass to the highest score interval. Thus, the variance of beliefs is decreasing in the mock exam score.

Fig. 1 also points out that overconfidence in prior beliefs is ubiquitous in the sample. Irrespectively of the value of the score, students tend to assign the largest probability mass to the highest interval of the score. Fig. 3 provides an alternative way to characterize the degree of overconfidence in our sample by plotting the relative share of students who would receive "good" or "bad" news depending on whether or not the score in the mock test belongs to an interval that is above the one corresponding to the median of their prior distribution. The figure shows that, indeed, very few students (8 percent) in our sample would receive good news about their performance relative to their prior expectations. The presence of students who would receive good news is concentrated in the highest segments of the test score distribution, confirming that beliefs are, on average, more accurate among better-performing students.

3. Experimental evidence on belief updating

3.1. Empirical framework

The random assignment of individualized feedback across the students in our sample allows us to measure the impact of the signal on posterior beliefs. Only if we show that the signal is salient and effectively alters beliefs, we can move onwards to examine the specific updating patterns that the sample exhibits. To track the relative importance of the signal vis-a-vis that of priors, we require two contemporaneous measures of beliefs for students in the treatment and in the control group, both before and after the delivery of the performance feedback. We make use of the posteriors measured at follow-up (which are collected after the delivery of the performance feedback for the treatment group) and the priors measured at baseline.

We estimate linear regressions of the following form:

$$d'_{vij} = \beta_0 d_{vij} + \beta_1 d_{vij} \times T_j + \gamma_0 I(z_{vij} = v) + \gamma_1 I(z_{vij} = v) \times T_j + \eta_i + \epsilon_{ij}, \tag{1}$$

where d'_{vij} and d_{vij} denote the individual densities of the posterior and the prior distributions, respectively, for each discrete interval $v = \{1, 2, \ldots, 5\}$ in the support of the exam score. The indicator function $I(z_{vij} = v)$ takes the value of one if student i's score in the mock test lies within interval v and zero otherwise, while the variable T_j takes the value of one for any school j randomly assigned to receive performance feedback and zero otherwise. The term η_i captures individual-specific constant terms and ϵ_{ij} is the usual error term, which is two-way clustered at the individual and school level.

The parameters β_0 and γ_0 capture the relative effects of baseline priors and the score, respectively, in the formation of the posteriors for students in the control group, who do not receive the performance feedback. In particular, β_0 measures the average degree of persistence in beliefs between the two survey rounds while γ_0 measures the extent to which students can infer something about their performance in the admission test by simply taking the mock test.

The parameters β_1 and γ_1 have a causal interpretation as they rely on the random variation in the treatment assignment. Estimates of β_1 and γ_1 measure the differential effects of baseline priors and the signal value, respectively, in the formation of posteriors for students who receive performance feedback (treated group) when compared to students who do not receive any feedback on their performance (control group). To the extent that the score in the mock test conveys relevant information regarding individual performance in the actual admission test, we expect $\gamma_1 > 0$. Insofar as students' prior expectations appear quite inaccurate relative to their observed performance (see Section 2.3), we also
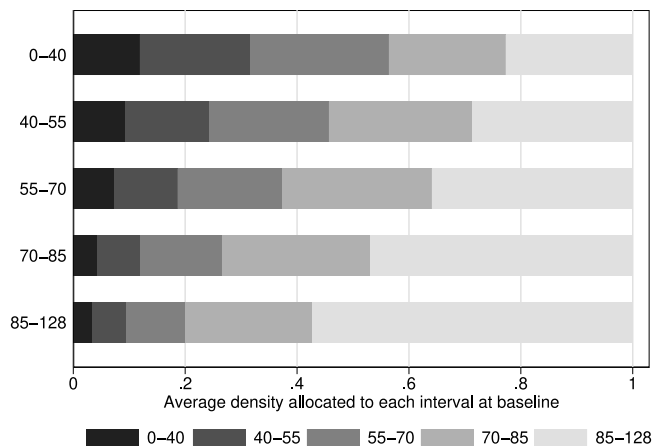
Fig. 1. Distribution of prior beliefs conditional on the score in the mock test. NOTE: For all students scoring in a given interval of the distribution of the mock test (vertical axis), the figure reports the average density allocated to each interval in the baseline survey (horizontal axis).
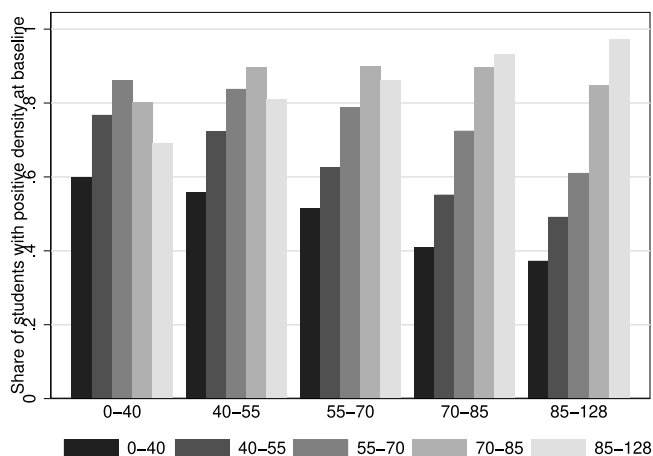
Fig. 2. Share of students allocating positive probability mass to each interval conditional on the score in the mock test. NOTE: For all students scoring in a given interval of the distribution of the mock test (horizontal axis), the figure reports the share of students allocating at least one bean (1/20 of probability mass) to each interval in the support of the belief distribution at baseline.
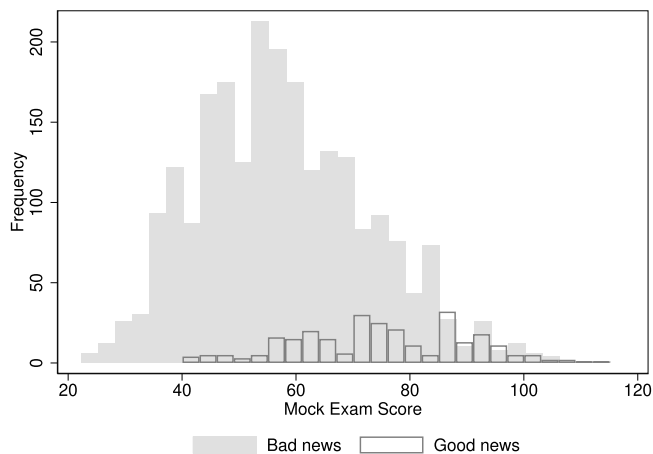
Fig. 3. Empirical density of the score in the mock test by Good/Bad news. NOTE: "Good news" is defined as to whether or not the individual scores in the mock exam lie in an interval that is above the one corresponding to the median of the baseline belief distributions.

Table 3
Belief updating.

Dependent variable:	Density of posterior in each interval		
	(1)	(2)	(3)
Density of prior	0.621***	0.487***	0.609***
	[0.032]	[0.033]	[0.034]
Signal	0.010*	−0.028***	0.013
	[0.006]	[0.007]	[0.009]
Density of prior ×Feedback	−0.342***	−0.363***	−0.333***
	[0.042]	[0.052]	[0.045]
Signal ×Feedback	0.169***	0.142***	0.183***
	[0.015]	[0.020]	[0.021]
Density of prior ×[Signal > Median]		0.228***	
		[0.043]	
Density of prior ×Feedback ×[Signal > Median]		0.001	
		[0.058]	
Signal ×[Signal > Median]		0.061***	
		[0.012]	
Signal ×Feedback ×[Signal > Median]		0.046**	
		[0.022]	
Density of prior ×High uncertainty			0.064*
			[0.037]
Density of prior ×Feedback ×High uncertainty			−0.047
			[0.049]
Signal ×High uncertainty			−0.005
			[0.010]
Signal ×Feedback ×High uncertainty			−0.028
			[0.022]
Number of observations	12 720	12 720	12 720
Number of schools	90	90	90
Number of students	2544	2544	2544
R-squared	0.275	0.293	0.276

NOTE: * significant at 10%; ** significant at 5%; *** significant at 1%. OLS estimates with student fixed effects using interval-level data. Standard errors clustered at the school and student-level are reported in brackets.

expect $\beta_1 < 0$. One benchmark case is a situation of complete pass-through of the signal (score in the test): $\gamma_0 + \gamma_1 = 1$, i.e., irrespectively of the location of the prior densities, the delivery of the signal moves all the probability mass of the posterior distribution to the interval where the signal is located.

More standard approaches tend to rely on specific moments (such as the mean) in order to characterize individual belief distributions. Focusing on interval-level data allows us to flexibly explore the relationship between priors and posteriors, as we only impose linearity between the probability mass allocated to a given interval in the support of the test score across survey rounds. For example, we allow for non-linearities in the relationship between mean priors and mean posteriors. An underlying assumption of this approach is that students can accurately assign densities to each interval, ruling out uncertainty or mistakes in calculating the number of beans allocated to each bin. To minimize mistakes, we carefully developed and implemented a protocol that included visual aids and practice examples that worked well in our sample (see Section 2.2). Nevertheless, we did not elicit the level of uncertainty when answering the survey questions on beliefs and hence cannot rule out its existence in the interval-level data.[6]

3.2. Estimation results

Table 3 reports the OLS estimates of the regression model (1). Column 1 shows that in the absence of personalized feedback, priors account for roughly two-thirds when forming posterior beliefs. At the same time, the estimated coefficient of the score shows that, on average, the experience of taking the test on its own induces a very small effect in the updating process. The lack of signal pass-through among students in the control group reflects the fact that taking the mock test is

---

[6] Table A.2 in the Web Appendix presents the OLS estimates of a variant of the regression model (1), which considers the mean and the standard deviation of beliefs about academic achievement as alternative outcome variables. Assuming a uniform distribution within each interval of the score, mean beliefs are constructed as the summation over intervals of the product of the mid-point of the bin and the probability assigned by the student to that bin. The standard deviation of the distribution of beliefs is the square root of the summation over intervals of the product of the square of the mid-point of the bin and the probability assigned to the bin minus the square of mean beliefs. The estimates reported in Column 1 of Table A.2 show that the score in the mock test is positively correlated with the mean and negatively correlated with the dispersion of posterior beliefs. The performance feedback reduces the dependence of mean posteriors on mean priors by 63% and induces a three-fold increase in the weight attributed to the signal, which is qualitatively consistent with our preferred estimates based on interval-level data (see Column 1 in Table 3). The estimates reported in Column 2 of Table A.2 further document that the feedback has no pass-through in the dispersion of the posterior distributions. However, it does reduce the dependence on priors by 20 percent.

indeed a very weak and/or noisy signal. Students in the treatment group reduce the relative role of priors on posteriors by more than half. On average, the performance feedback leads to reduced persistence of priors on posterior beliefs, but signal pass-through is far from complete. The signal induces a 17 percentage-point increase in the probability mass of the posterior belief distribution associated with the interval of the score in the mock test. Significant persistence of priors remains even after the provision of personalized feedback, with an estimated coefficient of 0.28 ($\approx 0.62 - 0.34$) for the treatment group.

Students may react differently depending on the value of the realized score in the mock test. A possible asymmetry in the updating process along that dimension may partly explain the presence of overconfidence in the baseline assessments of own skills observed in our sample. The descriptive evidence discussed in Section 2.3 suggests that higher-performing students have more accurate perceptions of their own skills (see Figs. 1–3). As such, they may also process differently new and informative signals about their own test performance. Column 2 in Table 3 presents heterogeneous updating estimates according to whether or not the score in the mock test is above or below the sample median. The results show that the performance feedback triggers a more pronounced response among high scorers in terms of signal pass-through. While students with below-median scores increase the probability mass allocated to the interval of the score by 14 percentage points, students with above-median scores experience an additional 5 percentage points increase in the density allocated to this interval.

Column 3 in Table 3 focuses on the role of initial uncertainty about priors in updating. In a Bayesian updating framework, noisier priors lead to greater pass-through of the signal. As before, we classify students into two groups depending on whether or not the individual standard deviation of beliefs at baseline is above or below the sample median. The results do not support the presence of differential effects of the performance feedback on posteriors by the level of initial uncertainty in beliefs.

Table 4 further explores heterogeneous effects in updating based on students' socio-demographic characteristics. First and foremost, Column 1 tests for any gender differences in signal pass-through. Several studies show that males tend to be more overconfident than females (Barber and Odean, 2001; Bordalo et al., 2019; Buser et al., 2014). We thus expect males to pay less attention to the performance feedback by exhibiting lower signal pass-through into the posterior distribution. The results from our sample do not find empirical support for any such effect. Both signal pass-through and prior pass-through are indistinguishable between male and female students.

In Column 2 we perform the same test across sub-samples of students with low and high socioeconomic status (SES), as measured by an asset index that is the first factor of a principal component analysis on households' ownership of a large set of durable goods.[7] Previous empirical evidence on updating along that specific dimension is scant, and ex ante predictions are unclear. On one hand, wealthier families are likely to provide greater support to their children in preparation for the admission exam. If this were the case, we would expect prior beliefs to be closer to true ability for higher-SES students and, consequently, we should observe a lower pass-through of the score in the mock test. On the other hand, higher-SES students could be more responsive to the score in the test simply because they are more effective at interpreting and processing information about their own academic skills. The estimation results do not find support for differences in the pass-through of feedback provision by SES. However, we find that higher-SES students who are exposed to the performance feedback tend to rely less on priors in forming their posteriors when compared to their lower-SES counterparts.[8]

We finally check if access to specific resources for the preparation of the admission exam, in the form of previous exposure to other mock tests, may at least partly explain the observed heterogeneous responses by SES in terms of prior pass-through.[9] Column 3 in Table 4 shows that the performance feedback has no differential impact on the formation of posteriors among students with previous exposure to other mock tests when compared to students with no prior test exposure. Hence, differential exposure to comparable signals does not seem to be the main channel through which SES plays a role in the updating patterns uncovered in Column 2 of Table 4.

All in all, the results in this Section confirm that the performance feedback constitutes a rather informative signal about performance in the admission exam as it led to significant changes in the weights attached to the priors and the signal itself in the process of belief updating. However, signal pass-through is quite limited and significant persistence of priors remains even after providing personalized performance feedback. The heterogeneity analysis reveals that signal pass-through does not seem to vary systematically by gender, SES, or initial uncertainty about priors, and it modestly increases among students with higher scores in the mock test.

Even though signal pass-through seems to be homogeneous by SES, students who are relatively better-off in terms of parental backgrounds seem to process more effectively the information contained in the score of the mock test. Exposure to performance feedback has a stronger negative effect on the persistence of prior beliefs among higher-SES students.

---

[7] The goods included in the principal component analysis are: telephone, television, washing machine, refrigerator, microwave oven, internet, cable television, tablet, computer, automobile, and water and sewage connection.

[8] Figure A.1 in the Web Appendix shows that the distributions of both overall and math performance in the test do not differ systematically by students' SES. This evidence is difficult to reconcile with the view that higher SES students are also more able and hence they update more sophisticatedly.

[9] Higher-SES students are more likely to have benefited from previous mock exam experiences: while only 22% of the students from lower-SES have previously taken a mock exam, over a third of the students in the higher-SES group have done so.

Table 4
Belief updating by individual characteristics.

Dependent variable:	Density of posterior in each interval		
	(1)	(2)	(3)
Density of prior	0.611***	0.580***	0.574***
	[0.037]	[0.041]	[0.032]
Density of prior ×Feedback	−0.344***	−0.297***	−0.321***
	[0.047]	[0.051]	[0.045]
Signal	0.007	0.002	0.014**
	[0.008]	[0.010]	[0.007]
Signal ×Feedback	0.172***	0.180***	0.163***
	[0.017]	[0.022]	[0.018]
Density of prior ×Male	0.022		
	[0.040]		
Density of prior ×Feedback ×Male	0.004		
	[0.058]		
Signal ×Male	0.008		
	[0.013]		
Signal ×Feedback ×Male	−0.006		
	[0.022]		
Density of prior ×High SES		0.131***	
		[0.039]	
Density of prior ×Feedback ×High SES		−0.123**	
		[0.055]	
Signal ×High SES		0.009	
		[0.012]	
Signal ×Feedback ×High SES		−0.009	
		[0.025]	
Density of prior ×Previous mock			0.163***
			[0.039]
Density of prior ×Feedback ×Previous mock			−0.070
			[0.057]
Signal ×Previous mock			−0.016
			[0.013]
Signal ×Feedback ×Previous mock			0.029
			[0.030]
Number of observations	12 720	9895	12 720
Number of schools	90	90	90
Number of students	2544	1979	2544
R-squared	0.276	0.300	0.279

NOTE: * significant at 10%; ** significant at 5%; *** significant at 1%. OLS estimates with student fixed effects using interval-level data. Standard errors clustered at the school and student-level are reported in brackets. The difference in the number of students and, in turn, in the number of observations in Column 2 with respect to the other Columns is due to missing values in the SES index, which are unrelated to the exposure to Performance Feedback (see Table A.1 in the Appendix).

This suggests that, despite a similar perception of signal strength across sub-groups, higher-SES students differentially incorporate the content of the signal in shaping their posterior beliefs. This last piece of evidence hints towards potential heterogeneous updating processes by sub-groups. In the next Section, we explore this avenue under a more flexible updating model that allows us to better characterize the transition from priors to posteriors along the distribution of the realized score in the mock test.

4. A model of belief updating

4.1. Empirical framework

Let $\pi_i(s)$ denote the prior probability assigned by student i to a test score in interval s, where $s \in \{1, \ldots, S\}$. Students take the mock test and are provided with their score z_i as a signal of their performance in the admission test. Signals and priors have the same support and hence the former can be discretized to mimic the intervals of the latter so that $z \in \{1, \ldots, Z\}$. By Bayes rule, we can compute the posterior belief of what the student will get on the admission exam after observing the score on the mock exam:

$$\pi_i(s|z) = \frac{f(z|s)\pi_i(s)}{\sum_{s'=1}^{S} f(z|s')\pi_i(s')}, \tag{2}$$
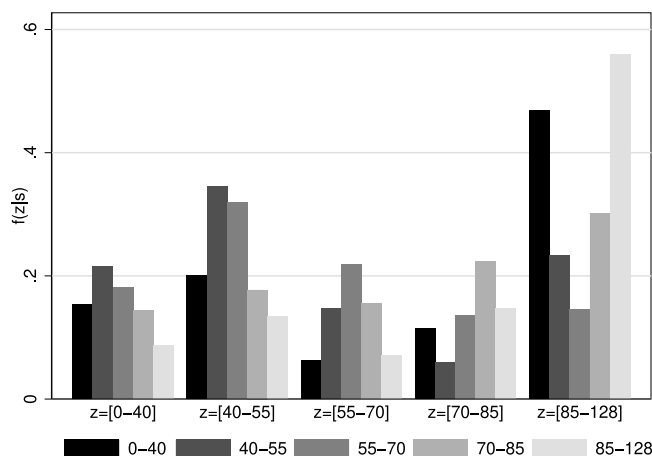
Fig. 4. Estimated conditional probabilities $f(z|s)$. NOTE: For different discrete values of the score in the mock test (signal) z, each bar in the figure reports the conditional probability $f(z|s)$, as defined in Eqs. (2)–(3), estimated by NLS using the downhill simplex (Nelder–Mead) algorithm to minimize the sum of the square of the differences between the LHS and the RHS of Eq. (2). Sample of ninth graders in schools that belong to the treatment group. See Table 5 for the full set of estimates along with the bootstrapped standard errors.

where $f(z|s)$ denote the conditional probability that a student who expects to get a score in interval s in the admission test will score in interval z in the mock test.[10] For each realization of signal z the model yields one $f(z|s)$ in each interval s of the support of the test score, which fully characterize the process of belief updating. For instance, if $f(z|s) = \frac{1}{S}$ for a given z, the signal is non-informative and does not generate any impact on posteriors – i.e., $\pi_i(s|z) = \pi_i(s) \ \forall s \in \{1, \ldots, S\}$. In turn, if the signal is perfectly informative about performance in the admission test, then students believe that their score in the admission test will fall in the interval of the realized score in the mock test with probability one – i.e., $\pi_i(s|z) = I(s = z)$, or complete signal pass-through.

Assuming that students have homogeneous expectations of the realization of the score in the mock test within each interval s, we can parameterize the conditional probability with a logit specification:

$$f(z|s; \theta) = \frac{e^{\theta_{sz}}}{1 + \sum_{z'=2}^{Z} e^{\theta_{sz'}}}, \tag{3}$$

where $\sum_z f(z|s; \theta) = 1 \ \forall s \in \{1, \ldots, S\}$ after imposing $\theta_{s1} = 0$ as an arbitrary identification normalization. The $S \times Z$ matrix of parameters Θ can be consistently estimated by applying a Non-Linear Least Squares (NLS) estimator on Eq. (2) using the observed priors $\pi_i(s)$ and posteriors $\pi_i(s|z)$ elicited in our survey. Since performance feedback is only provided to the treated group, we focus on the longitudinal variation in the individual belief distributions of treated students elicited during the follow-up survey, both before and after the provision of the signal. Conditional on receiving signal z, each estimated $f(z|s)$ represents the relative weight of the probability mass allocated to interval s in the prior distribution when constructing the posterior distribution. Thus, these probability weights provide an accurate characterization of the degree of persistence of priors along the ability distribution.

The estimation procedure can easily be extended by further conditioning Eqs. (2) and (3) on observed student types, which can be characterized by the realizations of one or more discrete covariates observed in our dataset such as gender and SES.

4.2. Estimation results

Fig. 4 displays the values of the estimated conditional probabilities depicted in Eq. (3). Table 5 reports the same coefficients of Fig. 4 along with the bootstrapped standard errors. Three broad patterns emerge. First, it is confirmed that the performance feedback provides an informative signal, as students systematically shift the weight in their initial priors towards the value of the signal. There is substantial probability mass assigned to the other intervals, though, which is qualitatively consistent with the evidence of incomplete signal pass-through reported in Section 3.2 (see Table 3).

Second, there is some evidence of asymmetric updating along the distribution of test scores. Although most students allocate the highest weight to priors in the same interval of the signal, this pattern does not hold for students who score in the lowest interval (0–40), where the corresponding weight is smaller compared to those assigned to the second or third interval. This is particularly worrisome as priors are already biased upwards among these students, who allocate

---

10 The fact that the mock test occurs before the actual admission exam does not compromise the definition of the $f(z|s)$ as well-defined hypothetical conditional probabilities.
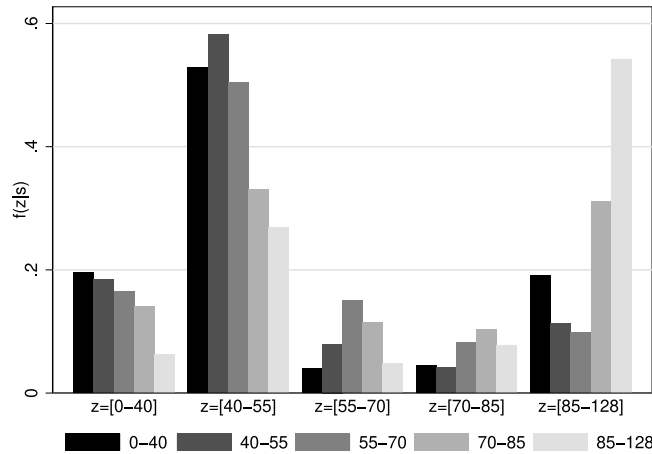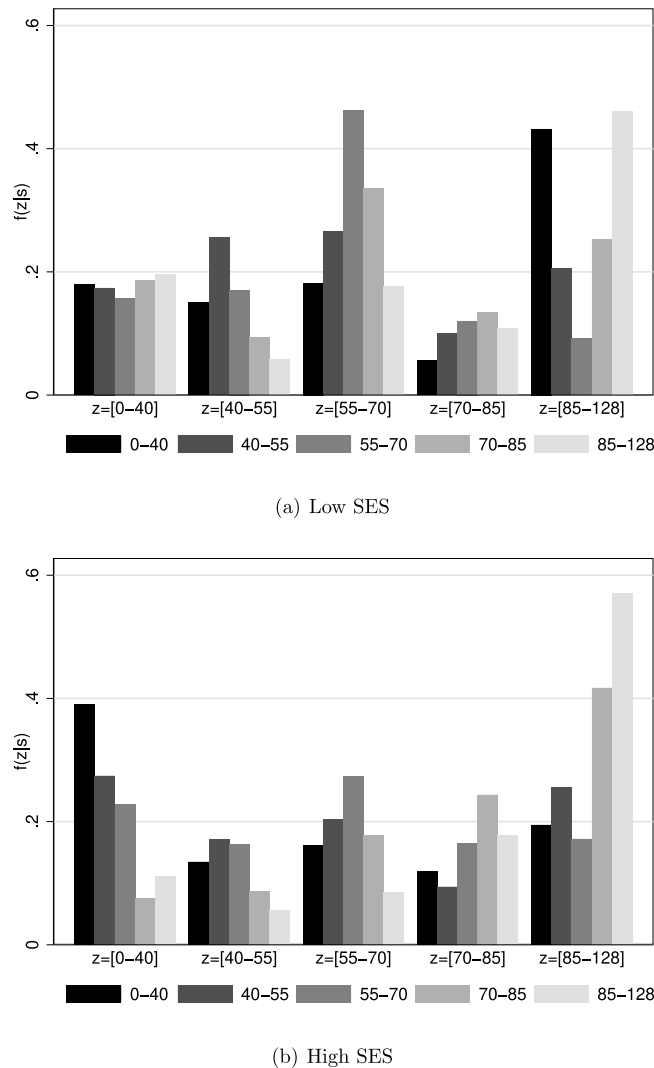
(a) Male
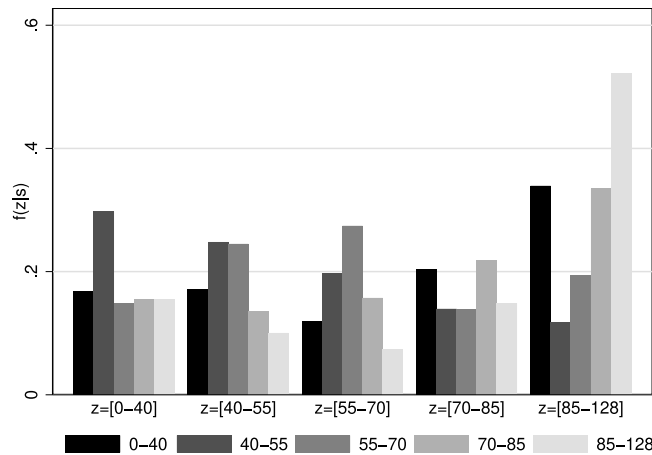
(b) Female

Fig. 5. Estimated conditional probabilities $f(z|s)$ by gender. Noᴛᴇ: For different discrete values of the score in the mock test (signal) z and gender sub-group, each bar in the figure reports the conditional probability $f(z|s)$, as defined in Eqs. (2)–(3), estimated by NLS using the downhill simplex (Nelder–Mead) algorithm to minimize the sum of the square of the differences between the LHS and the RHS of Eq. (2). Sample of ninth graders in schools that belong to the treatment group. See Table 6 for the full set of estimates along with the bootstrapped standard errors.

very low average density to this interval to begin with (see Fig. 1). Accordingly, we only observe a monotonic decline in the estimated conditional probabilities around the interval in which the signal lands when its value corresponds to the second and third intervals. This evidence may help explain the substantial degree of overconfidence observed in prior beliefs, particularly among lower-performing students (see Fig. 3).

Third, students who get signals in the highest interval (85–128) are somehow "prudent" in the extent of their update. While they do allocate a large probability mass to the highest interval of the prior distribution, they also disproportionately assign a relatively high weight to the lowest interval. In fact, the second interval also receives a larger weight when compared to the one allocated to the middle interval. The resulting U-shaped pattern in the updating process for high-performing students moderates the potential shift to the right in the distribution of posteriors. Notice that the average density allocated to these intervals among the highest scorers is quite limited (see Fig. 1), and hence the estimated spikes in the transition weights for the lowest and highest interval of the score have a limited impact in shaping posterior beliefs.

In order to better understand these updating patterns, we re-estimate the model (2)–(3) by sub-groups of students defined by socio-demographic characteristics. We start by splitting the sample by gender. The evidence discussed in Section 3.2 (see column 2 in Table 4) may hide some heterogeneous transition patterns for different values of the signal that the analysis in this section should be able to uncover. Indeed, Fig. 5 and Table 6 show that, unlike female students, male students tend to systematically assign the largest weight to priors in the same interval of the signal, and the decline

(a) Low SES

(b) High SES

Fig. 6. Estimated conditional probabilities $f(z|s)$ by SES. NOTE: For different discrete values of the score in the mock test (signal) z and SES sub-group, each bar in the figure reports the conditional probability $f(z|s)$, as defined in Eqs. (2)–(3), estimated by NLS using the downhill simplex (Nelder–Mead) algorithm to minimize the sum of the square of the differences between the LHS and the RHS of Eq. (2). Sample of ninth graders in schools that belong to the treatment group. See Table 7 for the full set of estimates along with the bootstrapped standard errors.

in the estimated weights of the neighboring intervals is monotonic. In addition, male students who receive a score in the highest interval seem to display a less pronounced U-shaped pattern in the updating process when compared to their female counterparts. All in all, we uncover some stark differences in the updating patterns by gender, with male students being relatively more accurate than females in interpreting the ability signal. Females seem to have more persistent priors, which diffuse the effect of signal provision on posteriors. The Wald test based on the difference between the estimated conditional probabilities across the male and female sub-samples strongly rejects the hypothesis of equal parameters (χ^2=53.63, p-value=0.0001).

Fig. 6 and Table 7 report the estimated conditional probabilities by SES. Consistently with the results discussed in Section 3.2, high-SES students seem to internalize more the ability signal provided through the performance feedback. They systematically assign the largest share of the probability mass to priors in the same interval of the signal. The relative weight assigned to higher-valued priors tends to monotonically increase with the value of the signal received. The updating process for the best performing students in the high-SES group is not U-shaped, as a very small weight is allocated to the lowest interval of the priors. In turn, the results suggest that updating behavior among low-SES individuals is much more erratic. These students tend to discard the information derived from signals in the lowest interval, as the weights allocated to their priors are more or less uniform across intervals. They also seem to extract relatively less information from higher-valued signals. Among those who get the highest score, the estimated weight allocated to the

Table 5
Estimated Conditional Probabilities $f(z|s)$.

Prior ($\pi_i(s)$)	Signal (z)				
	0–40	40–55	55–70	70–85	85–128
0–40	0.153	0.201	0.062	0.115	0.469
	(0.069)	(0.086)	(0.059)	(0.099)	(0.146)
40–55	0.215	0.346	0.148	0.059	0.233
	(0.057)	(0.114)	(0.073)	(0.060)	(0.101)
55–70	0.181	0.320	0.219	0.136	0.145
	(0.032)	(0.093)	(0.102)	(0.055)	(0.079)
70–85	0.144	0.177	0.155	0.223	0.301
	(0.035)	(0.050)	(0.076)	(0.090)	(0.085)
85–128	0.088	0.135	0.071	0.147	0.560
	(0.051)	(0.042)	(0.040)	(0.084)	(0.113)

NOTE: NLS estimates using the downhill simplex (Nelder–Mead) algorithm to minimize the sum of the square of the differences between the LHS and the RHS of Eq. (2). Standard errors calculated with 50 bootstrap replications are reported in parentheses. Sample of ninth graders in schools that belong to the treatment group. Priors are measured at follow up, before signal delivery. Posteriors are measured at follow up, after the delivery of the signal.

Table 6
Estimated conditional probabilities $f(z|s)$, by gender.

(a) Sample: Males

Prior ($\pi_i(s)$)	Signal (z)				
	0-40	40–55	55–70	70–85	85–128
0-40	0.196	0.529	0.040	0.045	0.191
	(0.063)	(0.138)	(0.047)	(0.137)	(0.181)
40–55	0.184	0.582	0.080	0.041	0.113
	(0.049)	(0.182)	(0.058)	(0.091)	(0.134)
55–70	0.164	0.504	0.150	0.082	0.098
	(0.042)	(0.109)	(0.089)	(0.096)	(0.078)
70–85	0.140	0.331	0.114	0.104	0.311
	(0.042)	(0.086)	(0.077)	(0.117)	(0.090)
85–128	0.062	0.269	0.049	0.078	0.542
	(0.047)	(0.073)	(0.039)	(0.130)	(0.122)

(b) Sample: Females

Prior ($\pi_i(s)$)	Signal (z)				
	0-40	40–55	55–70	70–85	85–128
0-40	0.168	0.171	0.119	0.203	0.339
	(0.084)	(0.087)	(0.077)	(0.141)	(0.150)
40–55	0.297	0.248	0.198	0.140	0.117
	(0.067)	(0.106)	(0.100)	(0.103)	(0.090)
55–70	0.148	0.245	0.274	0.139	0.194
	(0.049)	(0.105)	(0.148)	(0.087)	(0.081)
70–85	0.155	0.135	0.157	0.218	0.335
	(0.048)	(0.063)	(0.093)	(0.124)	(0.098)
85–128	0.155	0.100	0.074	0.148	0.522
	(0.056)	(0.063)	(0.053)	(0.137)	(0.146)

NOTE: NLS estimates using the downhill simplex (Nelder–Mead) algorithm to minimize the sum of the square of the differences between the LHS and the RHS of Eq. (2). Standard errors calculated with 50 bootstrap replications are reported in parentheses. Sample of ninth graders in schools that belong to the treatment group. Priors are measured at follow up, before signal delivery. Posteriors are measured at follow up, after the delivery of the signal.

first interval is very large and close in magnitude to the one corresponding to the highest-valued interval. The Wald test rejects the hypothesis of equal parameters across the two sub-samples defined by SES ($\chi^2 = 36.94$, p-value $= 0.012$).

All in all, the model estimates presented here reveal markedly heterogeneous updating patterns that could not be detected under the empirical framework discussed in Section 3. The additional structure imposed by the updating framework and the functional form assumptions allow us to accurately estimate the degree of priors' persistence in the formation of posterior beliefs, conditional on receiving feedback and the value of the signal itself. We uncover interesting differences by gender and SES that suggest that females and lower-SES students may not fully incorporate an informative signal due to the relatively higher persistence of prior beliefs.

Table 7
Estimated conditional probabilities $f(z|s)$, by SES.

(a) Sample: Low SES

Prior ($\pi_i(s)$)	Signal (z)				
	0-40	40–55	55–70	70–85	85–128
0-40	0.179 (0.063)	0.150 (0.083)	0.182 (0.050)	0.057 (0.142)	0.432 (0.171)
40–55	0.174 (0.060)	0.255 (0.126)	0.266 (0.063)	0.100 (0.106)	0.205 (0.096)
55–70	0.156 (0.043)	0.171 (0.089)	0.462 (0.099)	0.120 (0.107)	0.091 (0.092)
70–85	0.186 (0.041)	0.093 (0.057)	0.335 (0.073)	0.133 (0.136)	0.253 (0.108)
85–128	0.196 (0.055)	0.058 (0.043)	0.177 (0.043)	0.108 (0.174)	0.461 (0.152)

(b) Sample: High SES

Prior ($\pi_i(s)$)	Signal (z)				
	0-40	40–55	55–70	70–85	85–128
0-40	0.390 (0.078)	0.134 (0.135)	0.161 (0.066)	0.120 (0.160)	0.195 (0.179)
40–55	0.274 (0.061)	0.172 (0.117)	0.204 (0.069)	0.094 (0.103)	0.256 (0.123)
55–70	0.228 (0.047)	0.163 (0.111)	0.273 (0.081)	0.165 (0.101)	0.171 (0.107)
70–85	0.075 (0.046)	0.087 (0.081)	0.177 (0.053)	0.243 (0.134)	0.417 (0.120)
85–128	0.112 (0.064)	0.056 (0.092)	0.085 (0.031)	0.177 (0.140)	0.570 (0.141)

NOTE: NLS estimates using the downhill simplex (Nelder–Mead) algorithm to minimize the sum of the square of the differences between the LHS and the RHS of Eq. (2). Standard errors calculated with 50 bootstrap replications are reported in parentheses. Sample of ninth graders in schools that belong to the treatment group. Priors are measured at follow up, before signal delivery. Posteriors are measured at follow up, after the delivery of the signal.

5. Conclusion

We use a large-scale field experiment to study belief updating in a setting where beliefs are tightly linked to high-stakes choices and outcomes. We repeatedly elicit probabilistic statements about performance expectations in an achievement test using bean counts over a discretized support. Such a task appears a priori challenging, yet our approach turns out to be intuitive and accessible for the age group that the intervention targets. We complement the resulting longitudinal measures of subjective beliefs with randomized exposure to performance feedback on an achievement test.

The data show that prior beliefs about academic achievement of the ninth-grade students in our sample are relatively inaccurate when compared to an actual achievement measure, especially for those who do not perform very well. Providing individualized feedback on academic performance tilts the individual posterior distributions toward the realization of the signal and reduces the relative role of the priors in the updating process, but signal pass-through is far from complete. A second set of results based on a simple model of updating behavior confirms that the feedback provided with the experiment induces changes in the probabilistic weights allocated to each interval of the prior densities, but the persistence of priors on posteriors plays a predominant role.

Further heterogeneity analysis reveals that male and high-SES individuals are relatively more effective at processing an informative signal about their own academic skills. The result on gender may shed light on previous findings from the experimental literature on gender differences in overconfidence (Niederle and Vesterlund, 2007; Reuben et al., 2015). The U-shaped updating pattern found among high-performing female students moderates the potential shift to the right in posteriors when compared to male students (see Fig. 6), thereby providing a channel through which boys end up being relatively more optimistic than girls about their own academic skills.

To the best of our knowledge, the heterogeneous updating patterns found between high-SES students and low-SES students are novel and potentially relevant for the design of policy interventions aimed at disseminating information on individual academic skills. The strong persistence of prior beliefs and the rather erratic updating behavior among low-SES individuals point toward the possible limitations of simple and cost-effective ways of delivering information to this sub-population of students. Effectively undoing biases in subjective beliefs and realigning educational aspirations among disadvantaged students may require contemplating policy tools that provide more in-depth feedback such as tutoring and career counseling (see, e.g., Carlana et al., 2017).

One broader lesson from our findings is that characterizing movements in the entire belief distribution, rather than some of its moments, may reveal some nuanced patterns in updating behaviors that are key to understand how individuals process and internalize new information. Our analysis also features some limitations, as it relies on a very short panel of individual observations (albeit with a large cross-sectional dimension) and one controlled information shock to study the dynamics of individual beliefs. The elicitation of beliefs over a discretized support pursued here may be portable to longer panel datasets, which potentially feature multiple spells of exposure to information shocks.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2020.06.009.

References

Akerlof, G.A., Dickens, W.T., 1982. The economic consequences of cognitive dissonance. Amer. Econ. Rev. 72 (3), 307–319.

Avery, C., Hoxby, C., 2012. The Missing "One-Offs": The Hidden Supply of High Achieving, Low Income Students. NBER Working paper No. 18586.

Azmat, G., Bagues, M., Cabrales, A., Iriberri, N., 2019. What you don't know…can't hurt you? A natural field experiment on relative performance feedback in higher education. Manag. Sci. 65 (8), 3714–3736.

Azmat, G., Iriberri, N., 2010. The importance of relative performance feedback information: Evidence from a natural experiment using high school students. J. Publ. Econ. 94, 435–452.

Barber, B., Odean, T., 2001. Boys will be boys: Gender, overconfidence, and common stock investment. Q. J. Econ. 116 (1), 261–292.

Benabou, R., Tirole, J., 2002. Self-confidence and personal motivation. Q. J. Econ. 117 (3), 871–915.

Benjamin, D.J., 2019. Errors in probabilistic reasoning and judgment biases. In: Bernheim, B.D., DellaVigna, S., Laibson, D. (Eds.), Handbook of Behavioral Economics - Foundations and Applications 2. In: Handbook of Behavioral Economics: Applications and Foundations 1, vol. 2, North-Holland, pp. 69–186 (Chapter 2).

Bergman, P.L.S., 2015. Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment, CESifo Working Paper Series 5391, CESifo Group Munich.

Bobba, M., Frisancho, V., 2019. Perceived Ability and School Choices, TSE Working paper No. 16-660.

Bordalo, P., Coffman, K., Gennaioli, N., Shleifer, A., 2019. Beliefs about Gender. Amer. Econ. Rev. 109 (3), 739–773.

Brunnermeier, M.K., Parker, J.A., 2005. Optimal expectations. Amer. Econ. Rev. 95 (4), 1092–1118.

Burks, S.V., Carpenter, J.P., Goette, L., Rustichini, A., 2013. Overconfidence and social signalling. Rev. Econom. Stud.

Buser, T., Niederle, M., Oosterbeek, H., 2014. Gender, competitiveness, and Career Choices. Q. J. Econ. 129 (3), 1409–1447.

Carlana, M., La Ferrara, E., Pinotti, P., 2017. Goals and Gaps: Educational Careers of Immigrant Children, CEPR Discussion Papers 12538, C.E.P.R. Discussion Papers.

Compte, O., Postlewaite, A., 2004. Confidence-enhanced performance. Amer. Econ. Rev. 94 (5), 1536–1557.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. J. Finance 53 (6), 1839–1885.

Delavande, A., Giné, X., McKenzie, D., 2011. Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design? J. Appl. Econometrics 26 (3), 479–497.

Dizon-Ross, R., 2019. Parents' beliefs about their children's academic ability: Implications for educational investments. Amer. Econ. Rev. 109 (8), 2728–2765.

Dunning, D., Heath, C., Suls, J.M., 2004. Flawed self-assessment: Implications for health, education, and the workplace. Psychol. Sci. Publ. Interest 5 (3), 69–106.

Eil, D., Rao, J., 2011. The good news-bad news effect: Asymmetric processing of objective information about yourself. Am. Econ. J. Microecon. 3 (2), 114–138.

El-Gamal, M., Grether, D., 1995. Are people Bayesian? Uncovering behavioral strategies. J. Amer. Statist. Assoc. 90 (432), 1137–1145.

Elsner, B., Isphording, I., 2017. A big fish in a small pond: Ability rank and human capital investment. J. Labor Econ. 35 (3), 787–828.

Ewers, M., Zimmermann, F., 2015. Image and misreporting. J. Eur. Econom. Assoc. 13 (2), 363–380.

Fuster, A., Perez-Truglia, R., Wiederholt, M., Zafar, B., 2019. Expectations with Endogenous Information Acquisition: An Experimental Investigation, NBER Working Paper 24767.

Hastings, J.S., Weinstein, J.M., 2008. Information, school choice, and academic achievement: Evidence from two experiments. Q. J. Econ. 123 (4), 1373–1414.

Koszegi, B., 2006. Ego utility, overconfidence, and task choice. J. Eur. Econom. Assoc. 4 (4), 673–707.

Malmendier, U., Tate, G., 2005. CEO overconfidence and corporate investment. J. Finance 60 (6), 2661–2700.

Malmendier, U., Tate, G., 2008. Who makes acquisitions? CEO overconfidence and the market's reaction. J. Financ. Econom. 89 (1), 20–43.

Mobius, M.M., Niederle, M., Niehaus, P., Rosenblat, T.S., 2011. Managing Self-Confidence: Theory and Experimental Evidence, NBER Working Papers 17014, National Bureau of Economic Research, Inc.

Moore, D., Healy, P., 2008. The trouble with overconfidence. Psychol. Rev. 115 (2), 502–517.

Niederle, M., Vesterlund, L., 2007. Do women shy away from competition? Do men compete too much?. Q. J. Econ. 122 (3), 1067–1101.

Pathak, P., 2011. The mechanism design approach to student assignment. Annu. Rev. Econ., Annu. Rev. 3 (1), 513–536.

Rabin, M., Schrag, J.L., 1999. First impressions matter: A model of confirmatory bias. Q. J. Econ. 114 (1), 37–82.

Reuben, E., Wiswall, M., Zafar, B., 2015. Preferences and biases in educational choices and labour market expectations: Shrinking the black box of gender. Econ. J. 127, 2153–2186.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52 (3–4), 591–611.

Stinebrickner, R., Stinebrickner, T.R., 2012. Learning about academic ability and the college dropout decision. J. Labor Econ. 30 (4), 707–748.

Stinebrickner, R., Stinebrickner, T.R., 2014. A major in science? Initial beliefs and final outcomes for college major and dropout. Rev. Econom. Stud. 81 (1), 426–472.

Wiswall, M., Zafar, B., 2015. How do college students respond to public information about earnings? J. Hum. Capital 9 (2), 117–169.

Zafar, B., 2011. How do college students form expectations? J. Labor Econ. 29 (2), 301–348.