I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12945

Self-Perceptions about Academic Achievement: Evidence from Mexico City

Matteo Bobba
Veronica Frisancho

JANUARY 2020

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 12945

Self-Perceptions about Academic Achievement: Evidence from Mexico City

Matteo Bobba
University of Toulouse Capitole and IZA

Veronica Frisancho
Inter-American Development Bank

JANUARY 2020

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Self-Perceptions about Academic Achievement: Evidence from Mexico City*

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 paper 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, especially those attending relatively better schools, tend to process new information on their own ability more effectively.

JEL Classification: C93, D80, D83, D84, I24

Keywords: information, subjective expectations, academic ability, Bayesian updating, overconfidence, secondary education

Corresponding author:
Matteo Bobba
Toulouse School of Economics
University of Toulouse Capitole
21 Allée de Brienne
31015 Toulouse
France
E-mail: matteo.bobba@tse-fr.eu

* We are grateful to the Editor, the anonymous referees, John Rust, and the participants of the 2018 CESifo Workshop on Subjective Expectations for very useful comments and suggestions that greatly improved the paper. We thank the Executive Committee of COMIPEMS, Ana Maria Aceves and Roberto Peña of the Mexican Ministry of Education (SEP), *Fundacion* IDEA, C230/SIMO, and Maria Elena Ortega for the field-work activities. Financial support from the *Agence Francaise de Developpement* (AFD) and the Inter-American Development Bank (IDB) is gratefully acknowledged. This study is registered in the AEA RCT Registry and the unique identifying number is: AEARCTR-0003429.

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 either undo or reinforce biased priors. A better understanding of the dynamics of subjective beliefs in education markets is obviously key for the design and targeting of policy interventions aimed at disseminating information among youth on the verge of important schooling decisions.

This paper aims at partly filling this gap by taking advantage of rich and longitudinal data on students' beliefs about their own academic ability and the introduction of exogenous variation in exposure to individualized performance feedback. We first document the average effect of the ability signal on 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

[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]; 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.

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 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 and high-socioeconomic (SES) students seem more effective at processing new information about their own academic ability in terms of higher signal pass-through and lower prior pass-through, respectively. 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, once appropriately parameterized, 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 substantial changes in the probability mass allocated to each interval of the prior densities. The estimates also reveal some systematic differences in updating behaviors along the test score distribution, and 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 a relatively large probability weight 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, especially those who attend better schools, allocate a greater weight towards 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 self-beliefs about earnings: self-beliefs seem to be more responsive to information when prior beliefs are below beliefs about population earnings. More related to our findings, 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. They 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]; 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]; 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 heterogenous patterns of updating behavior across different individuals.

Our work represents one of the first attempts to depart from the Bayesian benchmark and points toward a more nuanced characterization of updating behaviors that hinges upon specific changes in the shape of the individual belief distributions in response to an informative signal. The analysis may inform future theoretical models of updating behavior that go beyond the Bayesian vs. non-Bayesian dichotomy. Our findings reveal stark differences in the updating patterns across individuals, which may be relevant for the design of policy interventions aimed at disseminating information about individual academic skills.

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 the 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 potential that is familiar and timely, and that contains relevant information about their own ability.

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

[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 public or private 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.

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 demographic characteristics, initial credentials (GPA in middle school or admission exam score), and assignment outcomes.

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.

2.2 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 were 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].

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, the treatment did not generate differential attrition patterns by treatment arm (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. Fortunately, with the exception of households' socioeconomic status (SES), all the other relevant variables used in the empirical analysis

[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.

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.

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).

Figure 1 reports the average frequencies of the prior distributions (i.e., beliefs before giving the mock test) for different values of 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. Yet, overconfidence in prior beliefs seems to be 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.

Figure 2 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

[5]Table A.1 in the 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.

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. To track the relative importance 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:

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

where $d^{'}_{vij}$ and d_{vij} denote the individual densities of the posterior and the prior distributions, respectively, for each discrete interval $v = \{1, 2, ...5\}$ 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, which may reflect the extent of noise in the belief elicitation process and/or the arrival of concomitant ability signals. Provided that the individual fixed effect (η_i in equation 1) effectively control for students' unobserved ability, 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 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 expect that $\beta_1 < 0$. One benchmark case is a situation of complete pass-through of the signal (score in the test): $\gamma_1 = 1$, i.e., irrespectively of the location of the prior densities, all the probability mass of the posterior distribution is allocated to the interval where the signal is located.

More standard approaches often rely on summary statistics to characterize individual belief distributions (such as the mean). Focusing on interval 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 data.[6]

3.2 Estimation Results

Table 3 reports the OLS estimates of the regression model (1). In the absence of personalized feedback, students assign a probability mass of roughly two-thirds to the priors when forming their 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 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 generates an important update of posterior beliefs,

[6]Table A.2 in the 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 (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.

but signal pass-through is far from complete: the signal induces an increase of 17 percentage points in the probability mass of the posterior belief distribution associated with the interval in which the mock test score lands.

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 Figures 1-2). As such, they may also process differently new and informative signals about their own 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 dependence on priors is much stronger among more academically prepared students, with a 50 percent increase in the probability mass allocated to priors when compared to less prepared students. Importantly, 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 boost 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 are 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 heterogenous effects in updating based on students' socio-demographic characteristics. First and foremost, Column 1 tests for any gender differences. 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 high-SES students and, consequently, we should observe a lower pass-through of the score in the mock test. On the other hand, high-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.

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. 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.

The results presented in this Section confirm that the performance feedback constitutes a rather informative signal about performance in the admission exam, as it led to substantial changes in the weights attached to the priors and the signal itself in the process of belief updating. In contrast, the mere fact of taking the test has very little consequences on the formation of posterior beliefs for the students in our sample. The heterogeneity analysis reveals that the updating patterns do not seem to vary systematically by gender or initial uncertainty about priors. However, the exposure to performance feedback has a stronger negative effect on the average weight assigned to prior beliefs among high SES students, suggesting that they are somehow processing more effectively the information contained in the score of the mock test. The exposure to performance feedback further induces differential changes in the updating process depending on the realization of the score in the mock test, with high scorers revealing a slightly larger pass-through of the signal.

The empirical framework employed here is well-suited for detecting average differences

[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.

between the students who are exposed to performance feedback and those who are not. The estimated treatment effects on both prior pass-through and signal pass-through denote the average correlation between the probability masses allocated across the different intervals of the support of the distribution and hence cannot detect more nuanced changes occurring within specific segments of the belief distributions. In the next section, we propose an alternative empirical framework aimed at unpacking the relationship between prior and posterior beliefs along the entire range of possible realizations of the ability signals.

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 $s \in \{1, ...S\}$. 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 $z \in \{1, ..., Z\}$. 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:

$$\pi_i(s|z) = \frac{f(z|s)\pi_i(s)}{\sum_{s'=1}^{S} f(z|s')\pi_i(s')}. \tag{2}$$

The likelihood functions $f(z|s)$ 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.[8] 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., $\pi_i(s|z) = \pi_i(s)$ $\forall s \in \{1, ...S\}$. 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 likelihood functions with a flexible logit specification:

[8] 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.

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

where $\sum_z f(z|s;\theta) = 1 \; \forall s \in \{1, ...S\}$ 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 equation (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 treated students' longitudinal variation in their individual belief distributions elicited during the follow-up survey, both before and after the provision of the signal.

Estimation of the likelihood functions can also be undertaken by further conditioning equations (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

Figure 3 displays the values of the estimated likelihood functions depicted in equation (3).[9] 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 largest share of the probability mass 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. Accordingly, we only observe a monotonic decline in the estimated likelihoods around the interval in which the signal lands when the signal's value corresponds to the second and third score intervals. This evidence may help explain the substantial degree of overconfidence observed in prior beliefs, particularly among lower-performing students (see Figure 1).

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

[9]Table A.3 in the Appendix reports the estimated values of the likelihood functions associated with Figure 3 along with the bootstrapped standard errors. Tables A.4, A.5, A.6 and A.7 report estimates and standard errors for the different sub-samples analyzed in this sub-section.

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.

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 may be able to uncover. Indeed, Figure 4 shows that, unlike female students, males tend to systematically assign the largest weight to priors in the same interval of the signal, and the decline in the estimated likelihoods of the neighboring intervals is monotonic. In addition, males 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. The Wald test based on the difference between the estimated likelihood functions across the male and female sub-samples strongly rejects the hypothesis of equal parameters (χ^2=53.63, p-value=0.0001).

Figure 5 reports the estimated likelihood functions by SES. Consistent with the results discussed in Section 3.2, high-SES students seem to internalize more the ability signal provided through the performance feedback. High SES students 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 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 high-valued signals. Among those who get the highest score, the likelihood function allocated to the 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 (χ^2=36.94, p-value=0.012).

The heterogeneous patterns observed by SES in this and the previous section motivate further analysis based on students' background. More specifically, we evaluate the hypothesis that a relatively more stimulating school environment may enable students to better internalize achievement signals, partly alleviating the observed differences in updating behavior by SES. Based on the school-average scores in the national standardized achievement test that we use for stratifying the randomization in our sample (see Section 2.1), we classify the 44 schools assigned to the treatment group into two sub-groups depending on their average score relative to the median in the sample.

Figure 6 presents the estimation results by both SES and school quality. Notice that quality is broadly defined since higher average scores may reflect better peers due to ex ante differences across schools or higher value-added due to, for instance, more effective teachers, or both. The set of best-behaved likelihood functions is identified among high-SES students in relatively better-performing schools (see Panel c). Students in this group allocate the highest weight to the interval in which the signal lies, and they monotonically assign lower weights to intervals that are further away from the score. We also observe less erratic patterns in the extremes of the score distribution. Another sub-group that exhibits more or less accurate updating patterns is the one composed of students from low SES in relatively worse-performing schools (see Panel b). These students exhibit similar updating patterns when compared to better-off students in relatively better performing schools, except for a spike in the density allocated to the lowest interval for those who get a score in the highest interval. In turn, high-SES students in worse-performing schools and low-SES students in better-performing schools seem to internalize less effectively the information contained in the score of the mock test. Panels (a) and (d) show that the students in these two groups exhibit a spike in the estimated likelihood function of the lowest interval for those who get a score in the highest interval. In addition, those with the score in the lowest interval tend to largely discard the information provided since the likelihood functions tend to increase, as opposed to decrease, over that range of the support of the score distribution.

This last set of results points to a possible complementarity between household and school resources. However, these results should be interpreted with some caution. First, there may well be sorting across middle schools by unobserved student types that in turn correlates with individuals' ability to process the performance feedback provided by the treatment. Second, each sub-sample relies on a reduced number of observations to estimate the model, which explains why the combined updating patterns by students' SES and school types are not statistically different from each other (p-value of Wald tests are 0.545 for below-average SES across school types and 0.938 for above-average SES across school types).[10]

[10] In the sample of treated schools, high-SES individuals make up 42 percent of the students in worse-performing schools and 64 percent of the students in better-performing schools.

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 across the individuals in our sample 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 therein. Providing individualized feedback on academic performance substantially tilts the individual posterior distributions toward the realization of the signal and reduces the relative role of the priors in the updating process. 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 associated pass-through is far from complete.

Further heterogeneity analysis reveals that male and higher-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 Figure 5), thereby providing a channel through which the mean level of overconfidence is larger for men than for women.

To the best of our knowledge, the heterogeneity found by SES is novel and potentially relevant for the design of policy interventions aimed at disseminating information on individual academic skills. We also provide suggestive evidence on some patterns of complementarity between household and school resources, as students from relatively more favorable socio-economic backgrounds who are enrolled in high-performing schools, as measured by average scores in standardized achievement tests, seem to be the sub-group with the most accurate updating behavior.

One general 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. However, the elicitation of beliefs over a discretized support pursued here may be portable across different and possibly richer datasets.

References

Akerlof, G. A. and Dickens, W. T. [1982], 'The Economic Consequences of Cognitive Dissonance', *American Economic Review* **72**(3), 307–19.

Avery, C. and 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. and Iriberri, N. [2019], 'What you don't know... can't hurt you? a natural field experiment on relative performance feedback in higher education', *Management Science* **65**(8), 3714–3736.

Azmat, G. and Iriberri, N. [2010], 'The importance of relative performance feedback information: Evidence from a natural experiment using high school students', *Journal of Public Economics* **94**, 435–452.

Barber, B. and Odean, T. [2001], 'Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment', *The Quarterly Journal of Economics* **116**(1), 261–292.

Benabou, R. and Tirole, J. [2002], 'Self-Confidence And Personal Motivation', *The Quarterly Journal of Economics* **117**(3), 871–915.

Benjamin, D. J. [2019], Chapter 2 - errors in probabilistic reasoning and judgment biases, *in* B. D. Bernheim, S. DellaVigna and D. Laibson, eds, 'Handbook of Behavioral Economics - Foundations and Applications 2', Vol. 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, North-Holland, pp. 69 – 186.

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

Bordalo, P., Coffman, K., Gennaioli, N. and Shleifer, A. [2019], 'Beliefs about Gender', *American Economic Review* **109**(3), 739–773.

Brunnermeier, M. K. and Parker, J. A. [2005], 'Optimal Expectations', *American Economic Review* **95**(4), 1092–1118.

Burks, S. V., Carpenter, J. P., Goette, L. and Rustichini, A. [2013], 'Overconfidence and social signalling', *The Review of Economic Studies* .

Buser, T., Niederle, M. and Oosterbeek, H. [2014], 'Gender, Competitiveness, and Career Choices', *The Quarterly Journal of Economics* **129**(3), 1409–1447.

Compte, O. and Postlewaite, A. [2004], 'Confidence-Enhanced Performance', *American Economic Review* **94**(5), 1536–1557.

Daniel, K., Hirshleifer, D. and Subrahmanyam, A. [1998], 'Investor Psychology and Security Market Under- and Overreactions', *Journal of Finance* **53**(6), 1839–1885.

Delavande, A., Giné, X. and McKenzie, D. [2011], 'Eliciting probabilistic expectations with visual aids in developing countries: how sensitive are answers to variations in elicitation design?', *Journal of Applied Econometrics* **26**(3), 479–497.

Dizon-Ross, R. [2019], 'Parents' Beliefs About Their Children's Academic Ability: Implications for Educational Investments', *American Economic Review* **109**(8), 2728–2765.

Dunning, D., Heath, C. and Suls, J. M. [2004], 'Flawed self-assessment: Implications for health, education, and the workplace', *Psychological Science in the Public Interest* **5**(3), 69–106.

Eil, D. and Rao, J. [2011], 'The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself', *American Economic Journal: Microeconomics* **3**(2), 114–138.

El-Gamal, M. and Grether, D. [1995], 'Are People Bayesian? Uncovering Behavioral Strategies', *Journal of the American Statistical Association* **90**(432), 1137–1145.

Elsner, B. and Isphording, I. [2017], 'A Big Fish in a Small Pond: Ability Rank and Human Capital Investment', *Journal of Labor Economics* **35**(3), 787–828.

Ewers, M. and Zimmermann, F. [2015], 'Image And Misreporting', *Journal of the European Economic Association* **13**(2), 363–380.

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

Hastings, J. S. and Weinstein, J. M. [2008], 'Information, School Choice, and Academic Achievement: Evidence from Two Experiments', *The Quarterly Journal of Economics* **123**(4), 1373–1414.

Koszegi, B. [2006], 'Ego Utility, Overconfidence, and Task Choice', *Journal of the European Economic Association* **4**(4), 673–707.

Malmendier, U. and Tate, G. [2005], 'Ceo overconfidence and corporate investment', *The Journal of Finance* **60**(6), 2661–2700.

Malmendier, U. and Tate, G. [2008], 'Who makes acquisitions? CEO overconfidence and the market's reaction', *Journal of Financial Economics* **89**(1), 20–43.

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

Moore, D. and Healy, P. [2008], 'The trouble with overconfidence', *Psychological Review* **115**(2), 502–517.

Niederle, M. and Vesterlund, L. [2007], 'Do Women Shy Away From Competition? Do Men Compete Too Much?', *The Quarterly Journal of Economics* **122**(3), 1067–1101.

Pathak, P. [2011], 'The Mechanism Design Approach to Student Assignment', *Annual Review of Economics, Annual Reviews* **3**(1), 513–536.

Rabin, M. and Schrag, J. L. [1999], 'First Impressions Matter: A Model Of Confirmatory Bias', *The Quarterly Journal of Economics* **114**(1), 37–82.

Reuben, E., Wiswall, M. and Zafar, B. [2015], 'Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender', *The Economic Journal* **127**, 2153–2186.

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

Stinebrickner, R. and Stinebrickner, T. R. [2012], 'Learning about Academic Ability and the College Dropout Decision', *Journal of Labor Economics* **30**(4), 707–748.

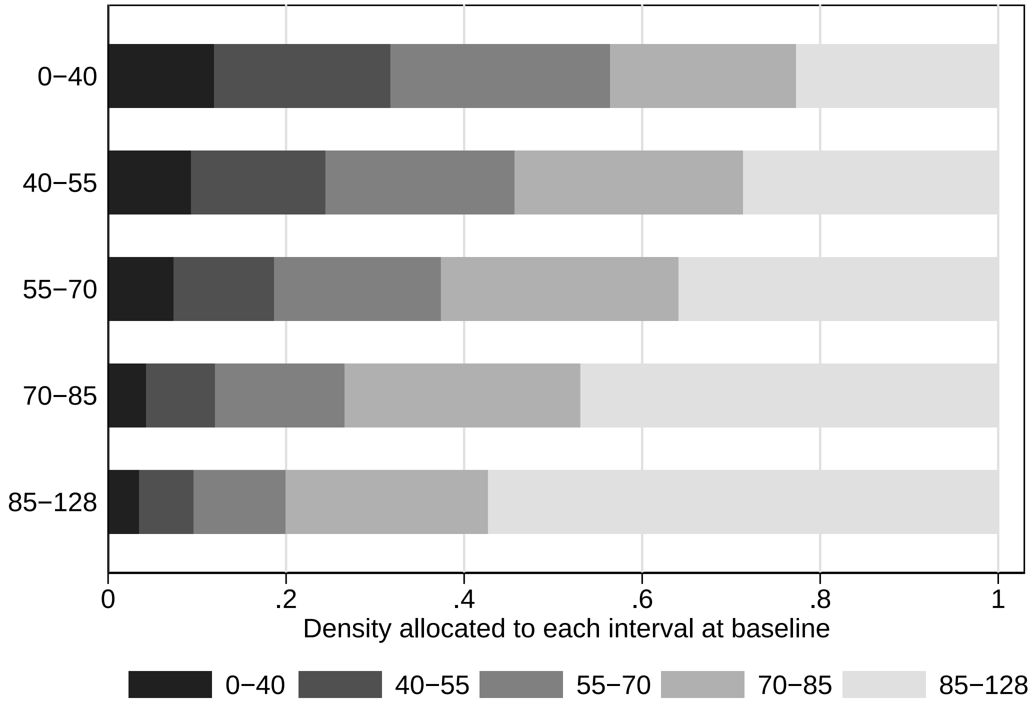
Stinebrickner, R. and Stinebrickner, T. R. [2014], 'A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout', *Review of Economic Studies* **81**(1), 426–472.

Wiswall, M. and 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.

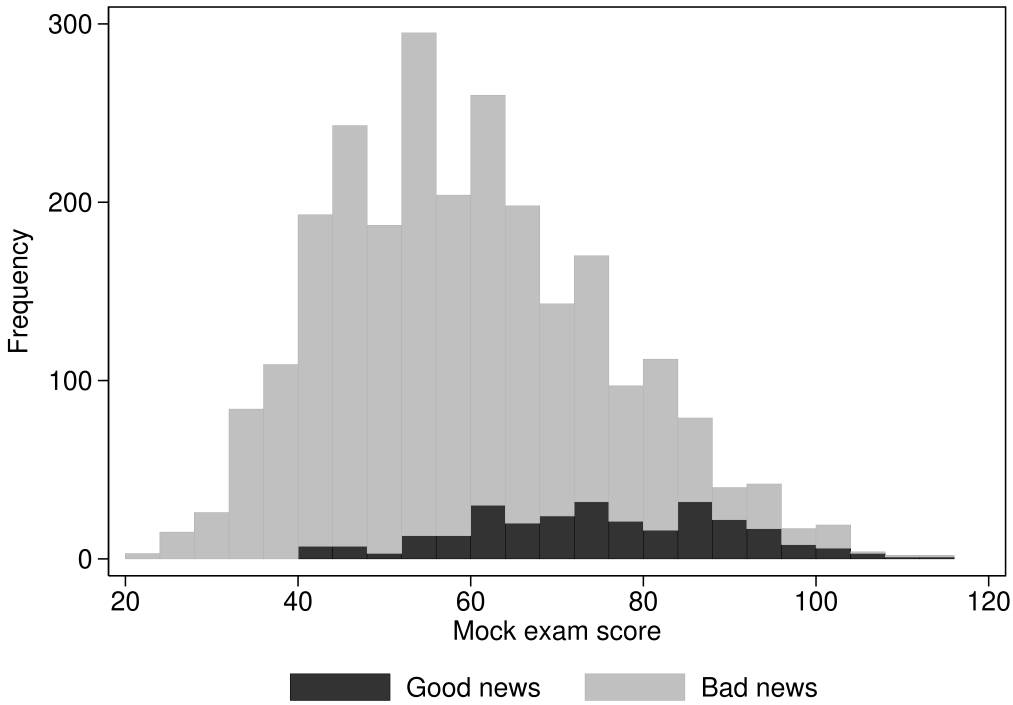
Figures and Tables

Figure 1: Distribution of Prior Beliefs by the Score in the Mock Test



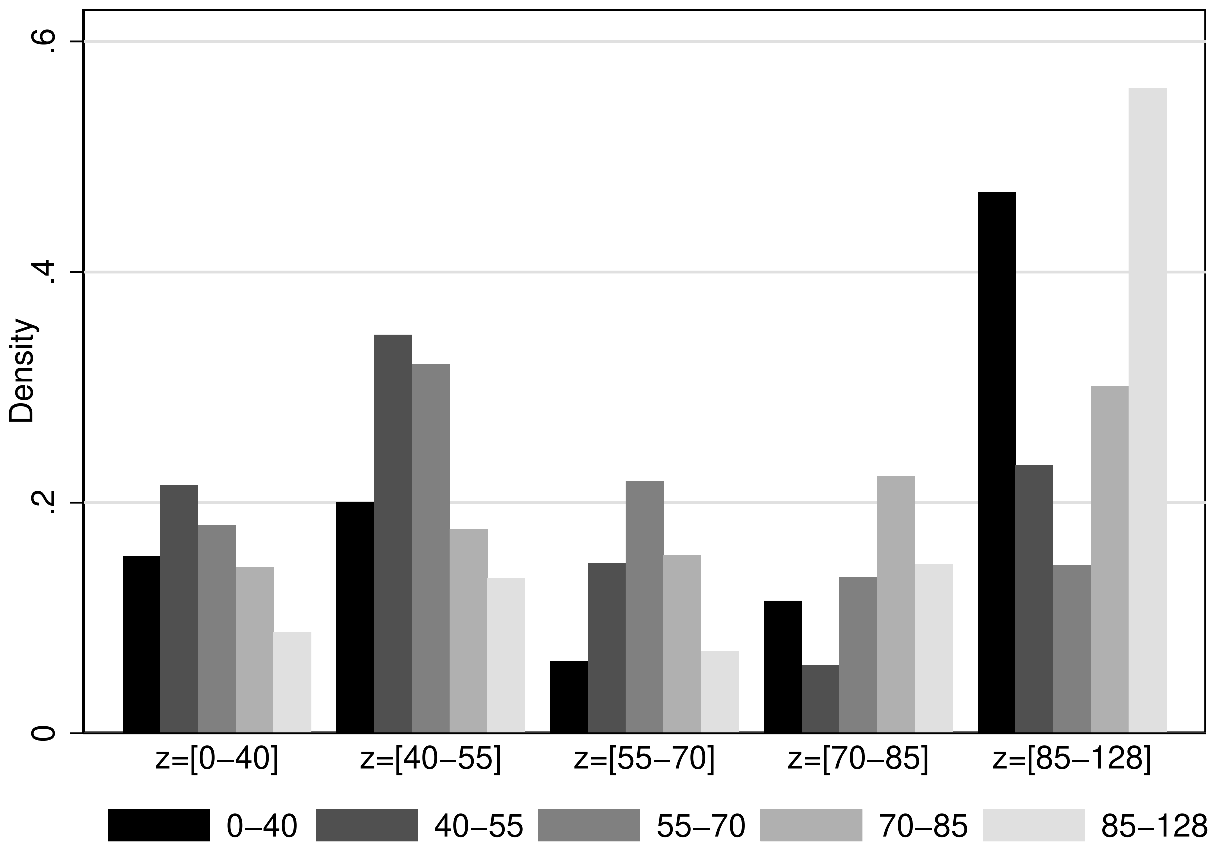
NOTE: The Figure reports the relative frequencies of elicited beliefs in the baseline survey in the horizontal axis conditional on the score in the mock test (vertical axis).

Figure 2: Empirical Density of the Score in the Mock Test by Good/Bad News



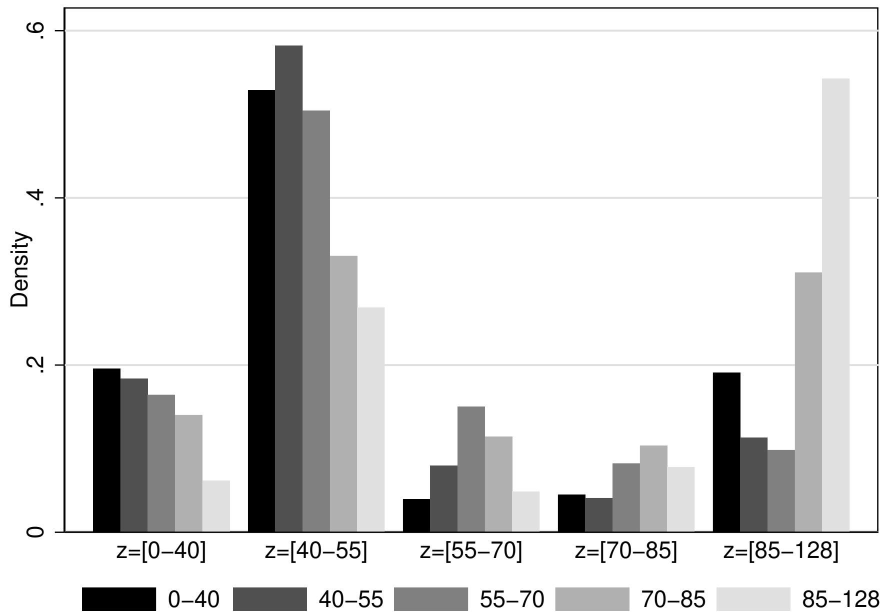
NOTE: "Good news" is defined as to whether or not the individual scores in the mock exam lie in an interval that is above the one corresponding to the median of the baseline belief distributions.

Figure 3: Estimated Likelihood Functions $f(z|s)$

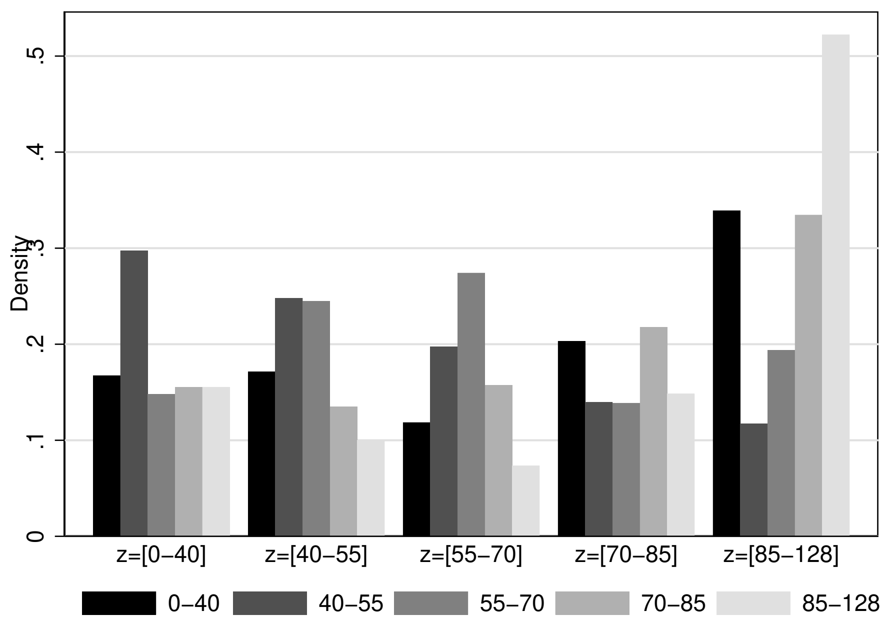


Note: For different discrete values of the score in the mock test (signal) z, each bar in the figure reports the conditional probability $f(z|s)$, as defined in equations (2)-(3), estimated by NLS using the downhill simplex (Nelder-Mead) algorithm to minimize the sum of the square of the differences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

Figure 4: Estimated Likelihood Functions $f(z|s)$ by Gender



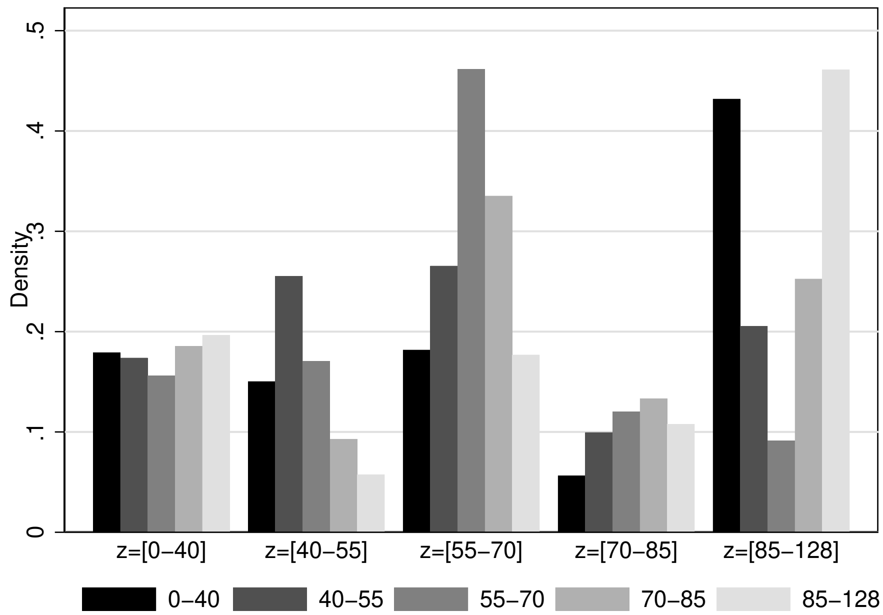
(a) Male



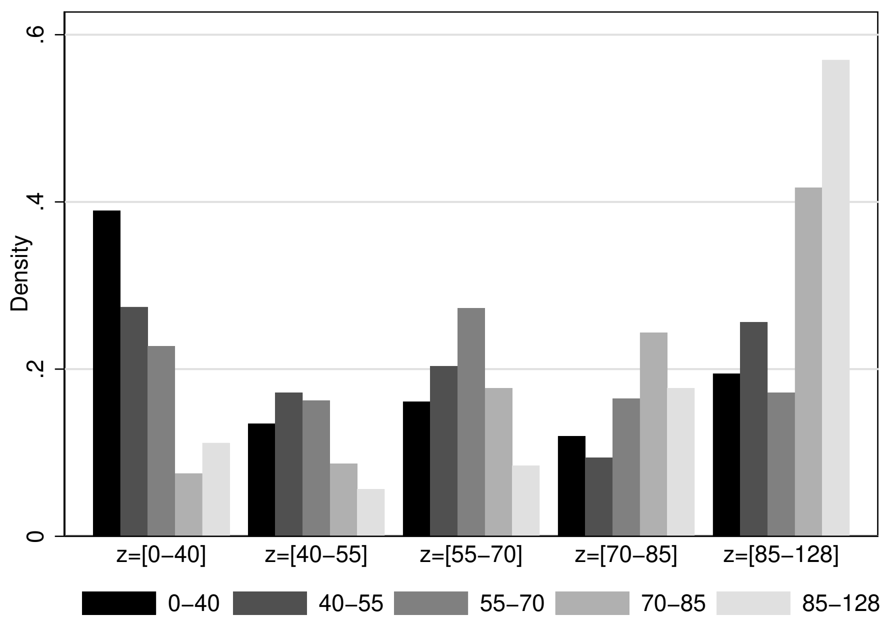
(b) Female

NOTE: 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 equations (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 equation (2). Sample of ninth graders in schools that belong to the treatment group. See Table A.4 in the Appendix for the full set of estimates along with the bootstrapped standard errors.

Figure 5: Estimated Likelihood Functions $f(z|s)$ by SES



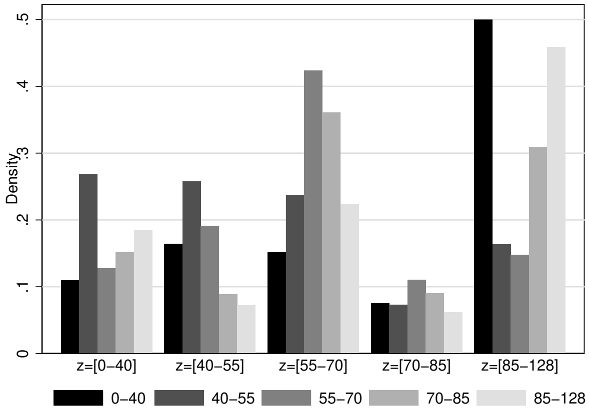
(a) Low SES



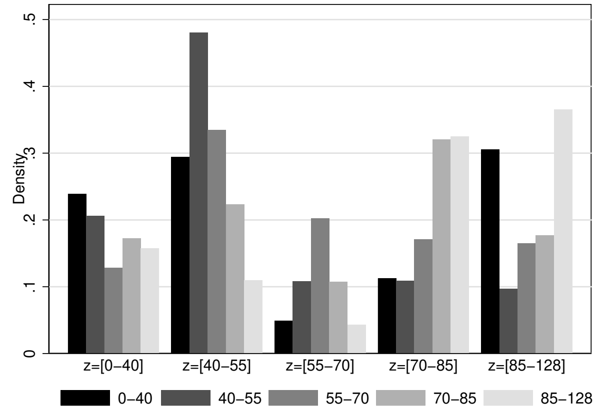
(b) High SES

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 equations (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 equation (2). Sample of ninth graders in schools that belong to the treatment group. See Table A.5 in the Appendix for the full set of estimates along with the bootstrapped standard errors.

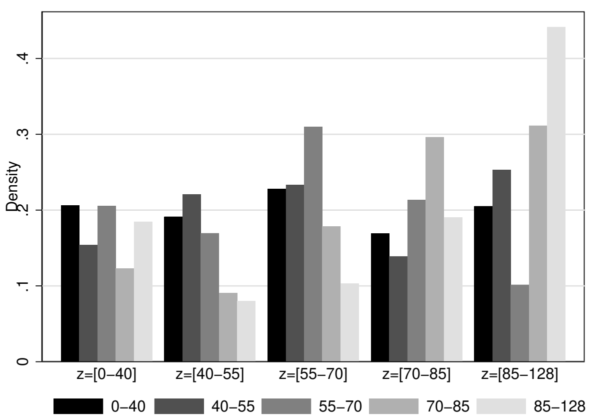
Figure 6: Estimated Likelihood Functions $f(z|s)$ by SES and School Quality



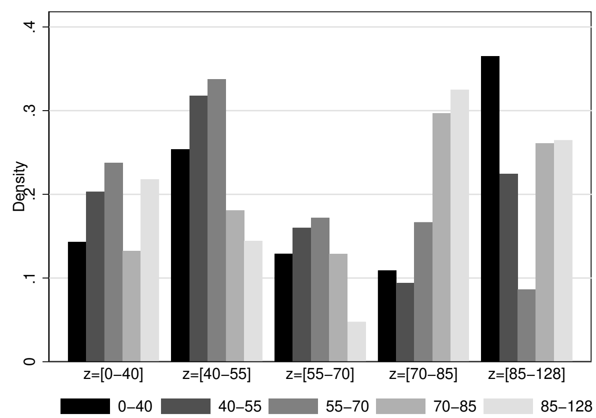
(a) Top School and Low SES



(b) Bottom School and Low SES



(c) Top School and High SES



(d) Bottom School and High SES

Note: For different discrete values of the score in the mock test (signal) z and genderXSES sub-group, each bar in the figure reports the conditional probability $f(z|s)$, as defined in equations (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 equation (2). Sample of ninth graders in schools that belong to the treatment group. See Tables A.6 and A.7. for the full set of estimates along with the bootstrapped standard errors.

Table 1: Comparing Population and Sample

Sample	Mexico City		Experiment	
Statistic	Mean	SD	Mean	SD
Student Characteristics				
Works	0.273	0.446	0.315	0.465
Indigenous student	0.041	0.199	0.098	0.297
Disabled student	0.113	0.317	0.141	0.348
Scholarship in Middle School	0.112	0.315	0.108	0.310
Grade retention in Middle School	0.134	0.340	0.142	0.349
Plans to go to college	0.808	0.394	0.724	0.447
GPA (middle school)	8.130	0.894	8.147	0.843
Lives with both parents	0.746	0.436	0.789	0.408
Mother with college degree	0.117	0.321	0.053	0.224
Father with college degree	0.189	0.391	0.095	0.294
Assignment Outcomes				
Exam score	70.99	21.17	64.91	19.78
Academic Track	0.605	0.489	0.481	0.500
Distance from school of origin (Km)	7.052	6.267	9.716	4.744
Number of observations	203,121		2,544	

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.

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. Strata dummies are included in all OLS specifications but they are not reported for space constraints.

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 above Median		0.228***	
		[0.043]	
Density of Prior × Feedback × Signal above Median		0.001	
		[0.058]	
Signal × Signal above Median		0.061***	
		[0.012]	
Signal × Feedback × Signal above 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	12720	12720	12720
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 individual-interval level data. Standard errors clustered at the school and student-level are reported in brackets.

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	12720	9895	12720
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 individual-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).

A Appendix

Table A.1: Share of Missing Data in Definition of SES

	Full Sample	COMIPEMS Participants
	(1)	(2)
Feedback	0.019	0.014
	[0.021]	[0.018]
Number of Observations	2544	2293
Number of schools	90	90
Mean in Control	0.21	0.13
R-squared	0.01	0.01

NOTE: OLS estimates with student fixed effects using individual-interval level data. Standard errors clustered at the school and student-level are reported in brackets.

Table A.2: Belief Updating on the Mean and the SD of Beliefs

Dependent Variable:	Mean Posterior	SD Posterior
	(1)	(2)
Mean prior	0.616***	
	[0.027]	
SD prior		0.587***
		[0.031]
Signal	0.105***	-0.038***
	[0.019]	[0.009]
Mean prior × Feedback	-0.388***	
	[0.032]	
SD prior × Feedback		-0.133***
		[0.035]
Signal × Feedback	0.361***	-0.005
	[0.037]	[0.010]
Constant	21.611***	9.232***
	[1.803]	[0.704]
Number of Observations	2544	2544
Number of Schools	90	90
R-squared	0.463	0.368

NOTE: * significant at 10%; ** significant at 5%; *** significant at 1%. OLS estimates with strata fixed effects using individual level data. Standard errors clustered at the school level are reported in brackets. 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 (SD) 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.

Table A.3: Estimated Likelihood Functions $f(z|s)$

	Signal (z)				
Prior ($\pi_i(s)$)	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 equation (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 A.4: Estimated Likelihood Functions $f(z|s)$, by Gender

(a) Sample: Males

	Signal (z)				
Prior ($\pi_i(s)$)	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

	Signal (z)				
Prior ($\pi_i(s)$)	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 equation (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 A.5: Estimated Likelihood Functions $f(z|s)$, by SES

(a) Sample: Low SES

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

(b) Sample: High SES

	Signal (z)				
Prior ($\pi_i(s)$)	0-40	40-55	55-70	70-85	85-128
0-40	0.390	0.134	0.161	0.120	0.195
	(0.078)	(0.135)	(0.066)	(0.160)	(0.179)
40-55	0.274	0.172	0.204	0.094	0.256
	(0.061)	(0.117)	(0.069)	(0.103)	(0.123)
55-70	0.228	0.163	0.273	0.165	0.171
	(0.047)	(0.111)	(0.081)	(0.101)	(0.107)
70-85	0.075	0.087	0.177	0.243	0.417
	(0.046)	(0.081)	(0.053)	(0.134)	(0.120)
85-128	0.112	0.056	0.085	0.177	0.570
	(0.064)	(0.092)	(0.031)	(0.140)	(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 equation (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 A.6: Estimated Likelihood Functions $f(z|s)$, by School Quality for Low SES Students

(a) Sample: Top School

	Signal (z)				
Prior ($\pi_i(s)$)	0-40	40-55	55-70	70-85	85-128
0-40	0.109	0.164	0.152	0.075	0.500
	(0.084)	(0.093)	(0.046)	(0.111)	(0.181)
40-55	0.269	0.258	0.238	0.073	0.164
	(0.080)	(0.115)	(0.071)	(0.096)	(0.129)
55-70	0.128	0.191	0.423	0.110	0.148
	(0.064)	(0.077)	(0.125)	(0.151)	(0.145)
70-85	0.151	0.089	0.361	0.090	0.309
	(0.057)	(0.046)	(0.139)	(0.142)	(0.162)
85-128	0.184	0.072	0.223	0.062	0.459
	(0.081)	(0.049)	(0.105)	(0.123)	(0.170)

(b) Sample: Bottom School

	Signal (z)				
Prior ($\pi_i(s)$)	0-40	40-55	55-70	70-85	85-128
0-40	0.239	0.294	0.049	0.113	0.305
	(0.074)	(0.107)	(0.071)	(0.153)	(0.180)
40-55	0.206	0.480	0.108	0.109	0.097
	(0.057)	(0.137)	(0.068)	(0.100)	(0.109)
55-70	0.128	0.335	0.202	0.171	0.165
	(0.044)	(0.109)	(0.091)	(0.092)	(0.101)
70-85	0.173	0.223	0.107	0.321	0.177
	(0.052)	(0.079)	(0.053)	(0.116)	(0.113)
85-128	0.157	0.110	0.043	0.325	0.365
	(0.053)	(0.043)	(0.026)	(0.203)	(0.191)

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 equation (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 A.7: Estimated Likelihood Functions $f(z|s)$, by School Quality for High SES Students

(a) Sample: Top School

	Signal (z)				
Prior ($\pi_i(s)$)	0-40	40-55	55-70	70-85	85-128
0-40	0.206	0.191	0.228	0.169	0.205
	(0.103)	(0.188)	(0.071)	(0.096)	(0.211)
40-55	0.154	0.221	0.233	0.139	0.253
	(0.100)	(0.177)	(0.076)	(0.070)	(0.135)
55-70	0.205	0.170	0.310	0.213	0.102
	(0.060)	(0.137)	(0.105)	(0.084)	(0.119)
70-85	0.123	0.091	0.179	0.296	0.311
	(0.053)	(0.105)	(0.061)	(0.114)	(0.117)
85-128	0.185	0.080	0.103	0.190	0.442
	(0.056)	(0.118)	(0.044)	(0.101)	(0.149)

(b) Sample: Bottom School

	Signal (z)				
Prior ($\pi_i(s)$)	0-40	40-55	55-70	70-85	85-128
0-40	0.143	0.254	0.129	0.109	0.365
	(0.070)	(0.082)	(0.068)	(0.131)	(0.173)
40-55	0.203	0.318	0.160	0.094	0.225
	(0.054)	(0.095)	(0.074)	(0.086)	(0.127)
55-70	0.238	0.337	0.172	0.167	0.086
	(0.054)	(0.084)	(0.072)	(0.092)	(0.100)
70-85	0.132	0.181	0.129	0.297	0.261
	(0.046)	(0.057)	(0.047)	(0.136)	(0.134)
85-128	0.218	0.144	0.048	0.325	0.265
	(0.072)	(0.049)	(0.025)	(0.209)	(0.207)

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 equation (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.