What you don't know... Can't hurt you? A natural field experiment on relative performance feedback in higher education *

Ghazala Azmat† Manuel Bagues‡ Antonio Cabrales§ Nagore Iriberri¶

June 1st, 2018

Forthcoming in *Management Science*

Abstract

This paper studies the effect of providing feedback to college students on their position in the grade distribution by using a natural field experiment. This information was updated every six months during a three-year period. We find that greater grades transparency decreases educational performance, as measured by the number of exams passed and GPA. However self-reported satisfaction, as measured by surveys conducted after feedback is provided but before students take their exams, increases. We provide a theoretical framework to understand these results, focusing on the role of prior beliefs, and using out-of-trial surveys to test the model. In the absence of treatment, a majority of students underestimate their position in the grade distribution, suggesting that the updated information is "good news" for many students. Moreover, the negative effect on performance is driven by those students who underestimate their position in the absence of feedback. Students who overestimate initially their position, if anything, respond positively. The performance effects are short lived - by the time students graduate, they have similar accumulated GPA and graduation rates.

Keywords: Relative performance feedback, ranking, natural field experiment, school performance.

*We would like to thank César Alonso, Brian Jacob, Imran Rasul, Roland Rathelot, Javier Ruiz-Castillo, Natalia Zinovyeva, participants in multiple presentations and the editor and three anonymous referees for their useful comments. We also thank Ryan Kendall for his excellent research assistance and we acknowledge the support of Isabel Gutiérrez, Paloma Arnáiz, Elena Zamarro y Luis Losada from Universidad Carlos III in conducting this study. Manuel Bagues acknowledges financial support from Ministerio de Economía y Competición (ECO 2012-31358). Antonio Cabrales acknowledges financial support from Ministerio de Economía y Competición (ECO 2012-34581). Nagore Iriberri acknowledges financial support from Ministerio de Economía, Industria y Competitividad and Fondo Europeo de Desarrollo Regional (ECO2015-66027-P MINECO/FEDER), Departamento de Educación and Política Lingüística y Cultura del Gobierno Vasco (IT869-13). All remaining errors are our own.

†Sciences Po and Centre for Economic Performance, LSE. E-mail: ghazala.azmat@sciencespo.fr.

‡Aalto University and IZA, Helsinki, Finland. Email: manuel.bagues@aalto.fi.

§University College London. E-mail: a.cabrales@ucl.ac.uk.

¶University of the Basque Country UPV/EHU, IKERBASQUE, Basque Foundation for Science. E-mail: nagore.iriberri@gmail.com.

1 Introduction

Providing feedback on individual performance is common practice, both in labor as well as in educational settings. Feedback is often provided on absolute performance or relative to some relevant reference group. A commonly held belief across organizations has been that more information will aid accountability. However, the actual response to information is likely to be complicated by a number of underlying features – making it unclear whether one can conclude that more information is always good for future performance. From a theoretical point of view, besides the associated incentives, individuals' preferences and their prior beliefs about their relative standings are likely to feed into their reaction to feedback, such that more transparency can actually increase or reduce performance.

Using a natural field experiment (Harrison and List, 2004) in an educational setting, this paper studies the dynamic effects of providing relative performance feedback, on both individual performance and satisfaction, shedding light on an important underlying mechanism – that of prior beliefs. A cohort of approximately 1,000 students enrolled in various degrees were randomly assigned into treatment and control groups. The experiment is carried over four years –following students through their full degree programs– in a large Spanish university.[1] Students in the control group, as per usual, receive only information on their own performance, while students in the treatment group are provided with their decile rank with respect to other students in their cohort. Relative performance feedback is provided to treated students for the first time at the end of the first semester of their second year of their program and is updated every six months until the end of the fourth (and final) year of the degree. Additionally, the timing is such that, after receiving feedback but before undertaking exams, students respond to surveys on course satisfaction. Finally, an important feature of our study is that we conduct an out-of-trial survey on students' prior beliefs, which allows us to understand the mechanism through which the treatment affects performance and

[1]An important merit of natural field experiments is that they bypass selection issues due to participation constraint, since all students in the cohort participate in the experiment (Al-Ubaydli and List, 2015)

satisfaction.

The academic performance of students in the treatment and control groups in the pre-treatment year (first year of degree) is similar. Once the treated students are provided with their rank information, we observe a significant decrease in their performance relative to those in the control group. In particular, during their second year, treated students pass, on average, 0.4 fewer course modules than students in the control group. Their GPA in that year falls by around 0.20 standard deviations. Overall, our results suggest that increased grade transparency has a detrimental effect on academic performance. Interestingly, satisfaction levels increase during this period. In the pre-treatment period, treated and control students report similar levels of satisfaction. However, after the provision of feedback in the second year – which is before exams are taken – treated students report a higher level of satisfaction.

To understand the mechanism that reconciles the effects of feedback on performance and satisfaction, we provide a theoretical framework that focuses on students' prior beliefs. The framework aims to understand the conditions under which relative performance feedback will have an impact on performance and the direction of its effect. In particular, we show that two issues are vital to understanding the sign of the effects of information provision. First, the type of objective functions that individuals maximize and, in particular, whether they maximize only their own performance, or whether they also care about their relative standing in a competitive manner. This may be because either relative performance materializes into different outcomes or individuals inherently care about relative standing in performance. Second, and more important, we consider individuals' prior beliefs regarding their, and others', abilities. This helps to understand, whether, in the absence of information, they were underestimating or overestimating their relative position. In particular, in a context where individuals have competitive preferences, learning that one's ability, with respect to others', is higher than initially believed, leads to lower effort provision, and conversely, learning that one's relative ability is lower than initially believed, when overestimating, leads to greater exertion of effort.

To test our theoretical mechanism, we use data from an out-of-trial student survey on prior beliefs of their relative standing. The average student makes an error in her prediction of 22 percentiles, indicating that students are indeed initially unaware of their relative standing. Moreover, contrary to the commonly held belief that people are generally overconfident about their ability, in our study the average student underestimates her relative ranking by 18 percentiles. This pattern is consistent with previous studies that have shown that over- and under-confidence vary systematically, and there tends to be under-confidence on difficult tasks (Moore and Cain, 2007). We show that students who are predicted to underestimate their relative rank are those who react negatively to the information. Students who are predicted to overestimate their relative position perform better, although the effect is not statistically significant. For students whose prior beliefs are predicted to be correct (i.e., for whom the feedback does not provide additional information), the effect is close to zero.

Since the university attracts highly-able students, the underestimation of rank may, in part, be due to students' feeling less able than his or her peers. Measuring students' priors allows us to understand their role in explaining the response to relative performance feedback. While there is some recent work on how individuals update their beliefs (Eil and Rao, 2010, Mobius et al., 2011, Ertac, 2011), little attention has been devoted to studying the role that prior beliefs play in the reaction to relative performance feedback.

Since we follow the students until the completion of their degree program, we can look beyond the immediate impact of treatment and study the long-run effects of the provision of feedback. We find that the provision of feedback information has a short-lived effect on academic performance. In particular, students whose performance deteriorated in response to the feedback information retake failed exams at the end of the academic year and close the gap in grade with respect to the control students. Although the accumulated GPA is still lower at the end of the academic year, on graduation, the performance of the treatment and control groups —as measured by the likelihood of graduating or the average accumulated GPA at end of the degree—

is statistically indistinguishable. The out-of-trial survey conducted in the final year of study suggests that, as students get closer to program completion, they are better informed of how they compare to their classmates. Among the control group, the prediction error falls to 15 percentiles, suggesting that there is learning over time. Another potential explanation might be the existence of information spillovers.

Our study relates to the empirical literature on the effect of relative performance feedback, both in educational and workplace settings.[2] The findings in the literature have largely been mixed. In an educational setting, focusing on younger, pre-college aged students, Azmat and Iriberri (2010), Tran and Zeckhauser (2012) and Katreniakova (2014), all find positive effects of the provision of relative performance feedback. In a workplace setting, observing other's work or getting relative performance feedback, Mas and Moretti (2009) and Blanes-i-Vidal and Nossol (2011) find a positive effect on performance, while Barankay (2011) finds a negative effect and Blader et al. (2015) find positive or negative effects, depending on the context. These mixed results in the literature suggest that the response is not straightforward. Our study helps to shed light on the potential mechanism at play. Unlike previous studies, we account for individuals' beliefs prior to the provision of information.

In a lab setting, studies have also investigated the role of feedback under various incentive schemes, showing its importance (Ertac, 2005; Eriksson et al., 2009; Hannan et al., 2009; Khunen and Tymula, 2012; Charness et al., 2013; Gerhards and Siemer, 2014; Azmat and Iriberri, 2016; Gill et al., 2016). These studies do not elicit beliefs on ranking position, with the exception of Khunen and Tymula (2012), who, under flat-rate incentives, find that those who rank lower than expected increase effort and those who rank higher than expected reduce effort, although the overall effect is positive.

We focus on a particular setting – that of higher education – where the demand for more feedback has been increasing among students. Williams and Kane (2009), for example, show that "assessment and feedback routinely score less well than other

[2]The effect of interim absolute performance feedback has also been studied in the literature. For example by Bandiera et al. (2015), where they find that the provision of feedback has a positive effect on students' subsequent test scores.

course-related aspects of the student experience and have done so for many years." However, the results of the study extend to other types of organizations, where greater transparency is becoming the norm. We provide a comprehensive analysis of the role of feedback on traditional outcomes, such as performance, which helps to understand why individuals might respond differentially.

The paper is organized as follows. Section 2 presents the theoretical framework. Section 3 describes the institutional background and the design of the experiment, as well as the additional surveys we conducted in the field. Section 4 presents the results. Finally, Section 5 concludes.

2 Theoretical Framework

We introduce a theoretical framework to understand agents' reactions to feedback on relative performance. We show that two elements are crucial when predicting a particular direction in the change of effort - one is the motivational component and the other is the informativeness of the feedback relative to agents' prior beliefs. At the end of the section, we propose a hypothesis on how differential reaction to feedback depends on individuals' prior beliefs. This is then tested in the natural field experiment, which is carried out at a university. A more formal description of the model is provided at the end of the paper in Appendix B.

Assume that the individuals' utility consists of two elements. One is a cost function of effort and the other is composed of two main motivational drivers that enter in an additive way. The first of these drivers is an output function of the individual's effort and ability, while the other is a competitive function. This competitive function depends on individuals' own effort and ability, as well as the efforts and abilities of other individuals. We can interpret this as the probability of winning a "prize" (which could, of course, be psychological reward)[3].

The output driver is a function of individuals' effort and ability, which we assume are complementary (i.e., the marginal output of effort is increasing in ability). The

[3]See for example Charness, Masclet and Villeval (2013).

competitive driver is a function of own, as well as other individuals', effort and ability. Individual's own higher ability or effort makes it more likely that she wins the prize, while others' higher effort or ability makes her winning less likely. In the competitive driver component, we specify three more relations. First, own effort and others' effort are strategic complements[4]. Second, the marginal product of own effort is increasing in the ability of others. And third, own effort and own ability may be complements or substitutes. Note that relative performance feedback can be informative about own and others' ability. With this utility function, we can make predictions about the effects of the provision of feedback.

If feedback informs the decision maker that others' ability is lower than expected, it suggests that she was initially underestimating her relative position. In which case, the reaction function for the effort of agent will shift down from the effect on the competitive motivation. This is because marginal product of effort is increasing in others' ability. If everyone lowers their estimate of the ability of opponents, the equilibrium choice of effort will decrease for all. This is because of the strategic complementarity between own effort and others' efforts.

If, however, feedback reveals that own ability is higher than initially believed, then the effect is more complicated. On the one hand, the reaction function for effort should shift up. This is because of the complementarity of own effort and ability in the output driver component. However, the reaction function for effort could shift down. This is because own ability and effort, in the competitive motivation, could be substitutes. In turn, the impact of feedback could increase or decrease effort. It would depend on the relative sizes and signs of the effect of the output and competitive drivers on effort. If the individual impact is the same for everyone, the equilibrium choice of effort for everyone goes in the same direction. This is due to the strategic complementarity of own and others' efforts. The final effect depends on the weight assigned to maximizing own output versus having a high standing within the cohort. A high weight for own output increases effort upon learning that own ability is better than expected. The

[4] As it is assumed in Benabou (1993), Calvò-Armengol et al. (2009), and Albornoz et al. (2017).

opposite can happen with a high weight for relative cohort standing if ability and effort are substitutes in the competitive component.

The overall effect of feedback, therefore, depends on the prior knowledge of own ability, versus the knowledge of others' ability. If we assume that information about others' ability is the only novelty and that individuals care for the competitive component, then a positive (negative) surprise about others' ability would lead to an unambiguous decrease (increase) in effort. If, however, information on own ability is the novelty, then the effect would be ambiguous.

This theoretical framework highlights the key determinants for agents' effort choices. One determinant is the different motivations in utility and another is the expectations individuals have before the provision of information. A final important determinant is whether feedback is informative of own ability or others' ability. The framework also makes salient that the average treatment effect of providing relative feedback could take any direction. It could be positive, negative or zero. The framework allows for different responses and, at the same time, provide guidance about what effects we expect to find. In particular, in our setting, it seems likely that knowledge of one's own ability is more precise than the knowledge of others' ability. This is because peers are new for most students. Thus, the feedback will make individuals update their priors of others' ability rather than their own ability. It also seems reasonable to assume that students have strong competitive motives since grades in a university serve as a signal of ability to potential employers and to graduate school admissions officers. Moreover, although some students will have a motivation for better grades, many will have an even stronger desire to perform better than others. If this is the case, the dominant force will be the one that shifts effort up or down in the presence of a negative or positive surprise. We can, therefore, test the following hypothesis:

Hypothesis: individuals who receive positive news (who were underestimating their relative position in the absence of information) reduce their output, while individuals who receive negative news (who were overestimating their relative position in the absence of information) increase their output.

In the sections that follow, we will describe our empirical application, which will test this hypothesis.

3 Background and Experimental Design

In our study, we follow a cohort of around 1,000 students, who enrol at University Carlos III in Madrid, Spain, in 2009. Over three years, 2010 to 2013, we conducted a natural field experiment. Students were randomly allocated into treatment and control groups at the start of their second year and treatment was carried out every six months until the end of the fourth year. In the control group, students receive information on their own performance (as is the norm). In the treatment group, students additionally receive information on their relative performance.

The majority of students pursue their degree in Spanish, but a small minority do so in English. Our study includes only students enrolled in the Spanish track on four four-year degree programs – Business, Economics, Finance, and Law – and one six-year degree – Business and Law. Two of these degrees – Business and Business and Law – are held simultaneously in two different locations, the Northern and the Southern campuses. We study both, such that our study involves students in seven different degree-locations. Table A1 summarizes the allocation into treatment and control groups by campus and degree program.

In this section, we explain the most relevant features of the university system, the design of the experiment, as well as the timing of the experiment.

3.1 Educational Institution

In Spain, access to university (full-time, undergraduate) is based on applicants' *entry grade*, which is calculated as a weighted average of their high school accumulated GPA (60%) and the grade obtained on a standardized exam known in Spanish as *Selectividad* (40%). University Carlos III offers the most selective degrees in the region according to the required minimum entry grade.[5]

An academic year includes two 14-week terms. The first term takes place from September to December, with exams taken in January. The second term takes place from February to April, with exams taken in May. Students that fail to pass an exam on either of the two terms have the opportunity to retake that exam in June.

Each week students attend one lecture and one tutorial. The assignment of students to lecture and tutorial groups is based on the first letter of their surname.[6] As an illustration, Figure A1 depicts how students enrolled in 2010 in the 1st year of the Business degree at the Southern campus were distributed across groups. For instance, students whose surname initial began with "A" or "B" were assigned to tutorial group number 74 and lecture group "74-75-76" (which combines tutorial groups 74, 75 and 76). In the Spanish context, surname order is uncorrelated with socio-economic status or academic performance, and as a result, performance across groups tends to be balanced.

All courses in the 1st and 2nd year of the degree are compulsory. Courses in the 3rd and 4th year of the degree are mostly elective. In each course, the final grade is usually a weighted average of the grade obtained in the end of term exams (60%), midterm evaluations (20%) and group presentations/assignments (20%). The end of term exam is usually the same for different groups in the same subject.

Students' permanence at the university is subject to certain requirements. During their first year at Carlos III, students must pass at least two courses. By the end of their second year, they must have passed every first-year course. Finally, they cannot fail the same exam more than three times. If any of these conditions is not satisfied, students cannot pursue their studies.[7]

After each semester, students receive information on the grades that they have ob-

[5]Information on minimum entry grades is available at `http://portal.uc3m.es/portal/page/portal/acceso_universidad/notas_corte_pc/notas_corte_09_10/notasmadrids09.pdf`, retrieved on April 30 2015.

[6]The only exception are second-year students in the English track. This is why we do not consider these students in our analysis and restrict our attention to students in the Spanish track.

[7]More detailed information is available on the university's webpage `http://portal.uc3m.es/portal/page/portal/conocenos/nuestros_estudios/normativa_09/Permanencia)`, retrieved on February 11 2015.

tained in each subject. The university summarizes this information through an official measure of accumulated grade point average (AGPA), which students can also access at any point in time on the university intranet.[8] Students do not receive information on their position in the distribution of AGPAs, relative to other students, or about the AGPA of any other student.

Students are not explicitly rewarded for their relative performance, except for a prize given to the best student in the cohort at the time of graduation.[9] Nonetheless, relative performance might still be relevant. For instance, performing relatively well might lead to better placement on study-abroad programs, receiving a good reference letter that would help to obtain an internship or a better labor market outcome.

3.2 Experimental Design

The experiment was conducted using the cohort of students who entered university in the Fall 2009 and who were registered in at least one second-year course in Fall 2010.[10] We assigned students to the treatment and control groups using a randomized block design. In Table A1, we describe how we assigned students into these groups. Overall, we consider seven blocks on the basis on the location (North campus and South campus) and degree of students (five degree programs). Within each block, we randomly assign one lecture group to the treatment group and the rest to the control group. The table highlights that in the degree-locations that had two groups, one group was randomly assigned to the control and the other to the treatment. However, in some cases, the degree-locations had three groups. For these groups, we decided to randomly assign two groups to the control and one to the treatment. As a result of the random draw, seven lecture groups including 354 students were assigned to the

[8] The university calculates the accumulated grade point average summing the grades obtained by the student, modified with a penalty for the number of times an exam is taken, and dividing this sum by the total number of courses taken. There is no penalty if the exam for the course is taken only once. If the student failed once, the course grade is multiplied by 0.95, by 0.90 if the student failed twice, and so forth.

[9] This prize, known as *premio extraordinario*, is awarded by the Ministry of Education.

[10] This condition excludes approximately 10 percent of the 2009 cohort, generally students who were expelled because they did not manage to pass at least two courses during the first year.

treatment group and ten groups including 623 students to the control group.[11]

The time line of the experiment is shown in Figure 1. The first intervention was done in December of 2010 - before the end of semester exams, which took place in January. Similarly, the second intervention was done in April 2011, before the second semester exams - in May. We followed the same timing in the third and fourth year of the program. Specifically, treated students received an email message from a corporate account stating the following:[12]

> This email is part of a pilot project of academic assessment management. If you want to see your average grade, and your relative position in terms of average grade among the students that started the degree the same year you did, you can do it by clicking **here**

After logging in with their university login and password, students would obtain access to a screen that displays their own GPA and their position in the distribution of grades, as measured in deciles (Figure 2).

In addition to the natural field experiment, we conducted our own surveys and gathered information from surveys conducted by the university administration: (i) course evaluation surveys completed by all students, (ii) a survey on students' knowledge of their relative position in the distribution of grades, administered to a sample of 2nd year students, who were not affected by the intervention, and (iii) a similar survey administered to a sample of graduating students belonging both to the experiments' treatment and control groups.

The teaching evaluations asked students about their satisfaction with the course and the grading standards, as well as their self-reported effort. In terms of timing,

[11] A few students were enrolled in several lecture groups. They were assigned to the group where they attended the majority of their courses.

[12] The control group did not receive an email. Therefore, as in many interventions in which the treatment is the reception of an email, whether the effect is due to the reception of email or to the content of the email cannot be easily disentangled. However, since - among the treated - we find a differential treatment effect based on priors, the effect is more likely to be due to the content of the email.

these surveys were conducted in December and then in April - after treated students received relative performance feedback but before taking exams. Note that we focus only on teaching surveys in 2010-2011, and not for the later years. This is because, as is typical with teaching evaluations, they are performed anonymously and the information is only available at the group level. In contrast to the second year where all classes are compulsory, students choose elective courses in the third and fourth and can, therefore, be in groups that mix treated and control students. In turn, we cannot identify teaching evaluations from treated and non-treated students for the third and fourth academic years.

The other two surveys measure students' knowledge of their relative standing, both prior to and after the treatment. The first is conducted among second year students in the later cohort and the second is conducted among the students in the experiment as they approach the end of their program. These belief-elicitation surveys were non-incentivized. One reason for this choice is that, in the survey that was conducted in the 4th year, the use of incentives would give a clear monetary advantage to the treated group.[13]

3.3 Individual Characteristics and Balance-check

Table 1 provides information on the individual predetermined characteristics of the 977 students - at the end of their first academic year - who participated in the intervention and it compares the treatment and control groups. Just over half of the students are women, and most students are Spanish (97 percent). A large majority of students attended high school before entering university - only 5 percent have a vocational training background. Approximately two-thirds of the students come from the Madrid region, and within this region, most come from the center of Madrid (31 percent).

[13]Schotter and Treviño (2014) and Murad, Sefton and Starmer (2016) highlight that the effect of the use of incentives is still an object of study in experimental economics and that more studies are needed in order to conclude that they encourage truth-telling and/or precision more than not incentivizing. Schotter and Treviño (2014) show that paying a flat-rate versus not paying at all does not seem to make a differential effect in belief elicitation. Closer to our setting, Murad, Sefton and Starmer (2016) find that non-incentivized measures reproduce 'hard-easy effect' (overconfidence in easy tasks and underconfidence in hard tasks) and that incentivized measures produce general underconfidence, which is reduced but not eliminated.

Approximately 22 percent come from municipalities located in the southern part of the region, an area that tends to be less affluent.

In their first year, students usually take a total of 11 courses within their program. All courses in the first year (and second year) are compulsory. From Table 1, we see that, relative to their "Entry Grade" - based on national-level tests taken at the end of high school - students experience a significant decrease in their grades during their first year in university (as expressed by GPA). While the average entry grade into the university is 7.24 (out of 10), the average GPA at the end of the first year is equal to 6.02 (out of 10), which implies a decrease of roughly one standard deviation. As shown in Figure A2, grades shift down along the entire distribution. It is important to point out that this university is a relatively selective university, such that the entry grade is high relative to other universities. The fall in grade might reflect different grading standards with respect to the entry level exams. Hence, the additional information that we provide on rank, might help students understand this better - especially early in their university program.

Student' satisfaction, hours of study and grading satisfaction are obtained through teaching evaluations.[14] Students are relatively satisfied with the quality of the courses they received before the intervention took place. On a scale from 1 (not at all) to 5 (very satisfied), students' average assessment is equal to 3.8. They are slightly less satisfied with the fairness of grading; again using a scale from 1 to 5, the response answer is 3.6. The teaching evaluations also provide (self-reported) information on study time. Students devote between 4 and 7 study hours to each subject in each week.[15] Taking into account that there are typically 5 or 6 courses per term, this implies that on average students spend approximately 32 hours studying each week, which combined with class attendance, amounts to approximately 50 hours per week

[14] Teaching evaluations are collected by the university administration twice per year. During academic year 2010-2011, students completed their 1st term teaching evaluations in December, before the intervention took place.

[15] Hours of study takes value 1 if the individual studied less than one hour per week; value 2 for between one and four hours of studying; value 3 for four to seven hours of studying; value 4 for seven to ten hours of studying; and value 5 for more than ten hours of studying.

in college-related work.[16]

We formally test whether these predetermined characteristics are balanced across the treatment and control groups using the following regression:

$$X_{s,d,g} = \alpha + \beta Treatment_{d,g} + \boldsymbol{Z_d \lambda} + \epsilon_{s,d,g} \tag{1}$$

where $X_{s,d,g}$ refers to a given predetermined characteristic of student s, enrolled in degree d and tutorial group g. $Treatment_{d,g}$ takes value one if the student is exposed to the treatment and the equation also includes a set of degree fixed effects (Z_d). As expected, the two groups are very similar in terms of their demographic characteristics and their academic performance before the intervention took place (Table 1, upper and middle panel).

Similarly, we formally test whether the treatment and the control groups are balanced on satisfaction, self-reported effort and satisfaction in grading before the intervention took place using the following regression:

$$Y_{c,g,d} = \alpha + \beta Treatment_{c,g,d} + \boldsymbol{X_c \gamma} + \boldsymbol{Z_d \lambda} + \epsilon_{c,g,d} \tag{2}$$

where $Y_{c,g,d}$ represents some average self-reported measure in course c (e.g. Econometrics I), tutorial group g (e.g. group 72) and degree d (e.g. Business at the Southern Campus). The regression includes a set of course fixed effects ($\boldsymbol{X_c}$) and degree fixed effects ($\boldsymbol{Z_c}$). As expected, students in the treatment and control groups report very similar values before the intervention in terms of their overall satisfaction with courses, the fairness of the grading and the hours of study (Table 1, lower panel). Finally, we also look at the number of students who complete the teaching evaluation per class and find that the response rates do not differ significantly across the treated and the control groups.

[16]According to survey information provided by teachers, the attendance rate at lectures is approximately 80% (Information available at `https://portal.uc3m.es/portal/page/portal/calidad/Resultados_encuestas_a_alumnos_y_profesores/00_Informe_1_cuatrimestre_2012_2013.pdf`, retrieved on April 30, 2015). Each course includes four hours of weekly lectures, which implies that a student enrolled in 5.5 courses who attended 80% of lectures would spend 18 hours weekly in class.

Overall, of 15 observable characteristics, in no dimension is the difference significant at the 5%, and in two dimensions the difference is significant at 10% (Table 1, column 4). An F-test confirms that it is not possible to statistically reject that the assignment was random.

4 Empirical Analysis

We analyze the impact of the intervention in four different steps. First, we examine the informational impact of the intervention. We measure students' prior beliefs before the intervention, using the out-of-trial survey, which will be important to understand the sign of the relative performance feedback effect, as guided by the theoretical framework. Then, we verify whether treated students actually accessed the link that they received by email. We finish testing the informational impact on students' knowledge after the treatment at the end of the fourth year. Second, we study the impact on students' performance and satisfaction, both in the short and long run. Third, we explore whether the impact of the treatment depends on subjects' prior beliefs. Finally, we also report some heterogeneity analysis on the relative performance feedback effect.

4.1 Students' Prior Beliefs about Relative Position and the Impact of Treatment on Beliefs

4.1.1 Students' Prior Beliefs

The intervention provides treated students with information on their position in the grade distribution at the beginning of their second year. The impact of this treatment depends crucially on the information that was available to students before the intervention. We investigate students' knowledge about their position in the grade distribution, absent of any intervention, using information from another cohort of students, that is, an out-of-trial sample of students. We conducted a survey among a group of students from the 2010 cohort at the beginning of their second year (November 2011). The survey was administered during the lecture of a compulsory course and in total 57

Economics students participated.[17] We decided not to conduct this survey among students belonging to the treated cohort (2009 cohort) to avoid the introduction of any confounding effects that might affect their subsequent performance.

Students were asked to answer privately the following question:[18]

> When you enrolled in this degree one year ago, your cohort included N students. If we were to rank all students in this cohort by their accumulated grade point average (AGPA) such that number 1 is the student with the highest AGPA and number N is the student with the lowest AGPA, in which position do you think you would be?

The answers are reported in Figure 3. The x-axis reports the actual position of the student in the ranking, normalized between 0 (lowest grade) and 1 (highest grade) among students who enrolled in Economics in Fall 2009. The y-axis provides information on their self-reported relative performance, normalized in a similar way. Most observations lie far away from the diagonal, reflecting that students are uninformed. Moreover, students tend to underestimate their position in the distribution of grades. The average student makes an error in her prediction of 22 percentiles and tends to underestimate her relative ranking by 18 percentiles. One possible explanation for this systematic divergence might be related to the difference in grading between high school and university. Although both are graded out of 10, the standards are different, such that (on average) they experience a sharp decline in grades as shown by the grades distribution in Table A2. However, they may not have understood that it affects all students, not only themselves.

To obtain a better understanding of which students underestimate their position in

[17] Specifically, we surveyed students enrolled in Game Theory, Degree in Economics, groups 63, 64, 68, and 69. A total of 21 people did not attend lecture on the day of the survey. All attending students except one participated in the survey.

[18] N was equal to 300, which corresponds to the number of students who enrolled in 2010 in the Economics degree offered by Universidad Carlos III at its Southern Campus

the distribution and which ones overestimate it, we estimate the following equation:

$$Y_s = \alpha + \boldsymbol{X_s}\beta + \epsilon_s, \tag{3}$$

where Y_s refers to the difference between the self-reported and the actual relative ranking. The dependent variable takes positive values when students overestimate their own ranking and negative otherwise. The set of independent variables $\boldsymbol{X_s}$ includes gender, entry grade, and performance during the 1st year. As shown in Table A2, underestimation is relatively stronger among women, among students with low high school grades, and among students who received relatively higher grades during their first year in university. These observable characteristics explain approximately 50 percent of the variation in the gap between students' self-reported ranking and their actual position.

Overall, this analysis shows that the average student underestimates her initial position, suggesting that there is room for students to learn about their relative ranking. Moreover, according to the theoretical framework, the provision of feedback should indeed affect negatively students' performance as long as the treatment reaches the students.

4.1.2 Do Students Access the Information?

The treatment provides students the possibility of obtaining information regarding their relative ranking, as students in the treatment group receive an email with a link to a personalized webpage where they can find feedback on their relative performance.

As part of the design, we can observe whether students indeed accessed the information and the number of times they did so. 72 percent checked this information at least once. The average student checked the ranking four times during the duration of the treatment. As shown by Figure A3, there is quite a lot of variation, as 23% of students check the information only once and about 10% of them checked the information 8 times or more during the three years of treatment. Also, as shown by Figure A4, about 25% of students check during one or two years and a slightly lower proportion checks during all three years. Finally, as shown in Figure A5, the probability of checking is strongly correlated with the position in the ranking. In the top quartile, nearly 90 percent of students accessed the information, while in the bottom quartile, less than half did. This result is confirmed by the regression analysis, shown in Table 2, as for different dependent variables higher ranked students are more likely to check for the information. Female students are also slightly more likely to check, but the difference is only marginally significant once we account for the ranking (Table 2).

We cannot determine why some students do not check the information. Some individuals might not read emails from corporate accounts, while others perhaps read the email but prefer not to learn about their position in the ranking. One-third of the students that did not check their ranking were expelled from the university at the end of their second year due to their failure to meet the permanence requirements. It is possible that these students were not active students at the time of the intervention. While our study is probably not well suited to explain the reasons for information acquisition, we can offer some speculation as to how these results might fit with the theoretical framework outlined in Section 2. Suppose accessing the information has a cost (either cognitive from processing the material, or psychological from fearing bad news[19]). Then, if the competitive motive for effort is dominant as we assume in Section 2 and is consistent with our results, then it is natural than the individuals with better priors about their position in the ranking have more incentives to check it. After all, if the competitive motivation arises from a desire to win a "prize," those who have a reasonable chance to win the prize stand to gain more from accessing the information.

4.1.3 Learning and Information Spillovers over Time

The intervention was designed to minimize information spillovers, but it is still possible that students from the control group received some information from treated students (Duflo and Saez, 2003). Students in both groups might also increase their knowledge of their position in the distribution over time, independent of the intervention.

[19] As per the anticipatory utility theory, see, e.g., Caplin and Leahy (2001).

To study this issue, we surveyed a sample of students from the treatment and the control groups concerning their relative ranking three years after the beginning of the intervention. The survey was conducted at the end of the undergraduate thesis presentation, which is the final requirement that students satisfy before graduation.[20] The sample includes 97 students from Economics, Business and Finance degrees. Four students did not reply to the survey. Note that, by construction, the sample of students surveyed is not a random sample of all students. Students in the upper part of the grade distribution are over-represented.[21]

The information displayed in Figure 4 reveals two interesting patterns. First, students in the control group have more accurate information on their relative performance at the end of their 4th year than students at the beginning of their 2nd year, before the intervention took place. The average error has decreased from 22 percentiles to 15 percentiles (Table A3). This improvement might potentially reflect learning over time or information spillovers. Unfortunately we cannot disentangle these two hypotheses. Second, at the end of their 4th year students in the treatment group are significantly better informed than students in the control group. The average error among treated students is equal to 9 percentiles compared to 15 percentiles in the control group. Note also that students in the treatment group do not perfectly predict their position in the ranking. This might be due to several factors. First, students were asked about their exact position in the ranking, while the intervention provided access only to their position in terms of decile. Second, the survey was conducted after the final exams but before students could access information on their final ranking; the last update of the ranking information took place shortly after we conducted the survey. Third, a few students in this group (less than 10%) had never checked the information provided. Finally, some students may have forgotten their position in the ranking.

Overall, the survey information suggests that, while initially the intervention gen-

[20]To prevent (treated) students from having access to the information provided, they were not allowed to access the internet during the survey.

[21]43% of surveyed students belong to the treated group and the rest to the control group. We have also checked if students are balanced across control and treated groups in terms of gender and High School entry grade. They are balanced in terms of gender and students in the control group show slightly higher High School entry grade than students in the treated group, significant at the 5%.

erated a large gap in terms of how well informed are students in the treatment and the control group, over time the control group catches up. Therefore, the effect of relative performance feedback is likely to differ in the short run, when the treatment group experiences an information shock, and in the long run, when the control group is learning relatively faster.

4.2 Feedback effect on Academic Performance, Satisfaction and Effort

We estimate the impact of feedback on academic performance, satisfaction and effort using the following regression:

$$Y_{s,d,g,t+i} = \alpha + \beta Treatment_{d,g} + \boldsymbol{Z_d \lambda} + \boldsymbol{X_{s,d,g,t}} + \epsilon_{s,d,g,t+i}, \tag{4}$$

where $Y_{s,d,g,t+i}$ represents the performance of student s, enrolled in degree d and tutorial group g, in the academic term $t+i$, and t refers to the time of the intervention. $Treatment_{d,g}$ takes value one if the student belongs to the treatment group. Given that there was not full compliance, β provides information on the *intention-to-treat* effect. The equation also includes a set of degree fixed effects ($\boldsymbol{Z_d}$), as well as, $\boldsymbol{X_{s,d,g,t}}$ a set of predetermined individual characteristics listed in subsection 3.3. To account for the potential existence of common shocks, we report standard errors clustered at the lecture group level.

Below, we report separately the short-term effects (second year) and the longer-term effects (third and fourth years).

4.2.1 Short-term effects

Table 3 provides information on the effects of the intervention during the regular exam period of the second year. The provision of relative performance feedback has a negative impact on the performance of students in the treatment group. On average, students in the treatment group passed 0.36 (0.1 of a standard deviation) fewer exams during the regular exam period, a difference that is significant at the 6 percent level. The average grades, as measured by students' GPA, decrease by 0.20 standard deviations, significant at 1 percent level. This negative impact on performance is consistent with the main hypothesis as the average student is expected to underestimate her position in the relative performance and therefore, will be getting positive news.

In the short run, we can also analyze students' satisfaction and self-reported effort using equation (2). Recall that the evaluations are conducted after students receive the information but before they do their exams. We cannot observe these two variables at the individual level, but we can exploit the information on teaching evaluations available at the course and group level. According to the theoretical framework, the good news received by the relative performance feedback makes the student more satisfied and work fewer hours. We do find some evidence for this – treated students show approximately one-third of a standard deviation higher satisfaction with the teaching. However, we do not observe any significant impact on self-reported effort. Both the treated and control groups report that they study between four and seven hours weekly per course. One possible explanation is that the treatment was not strong enough to move students' effort beyond these boundaries, or alternatively, that students do not feel that they are putting in less effort (despite their later lower performance), which might be a response to the positive news they receive.

The last column of Table 3 presents the adjusted p-values for multiple hypothesis testing, using the Holm-Bonferroni correction. As expected, p-values are generally larger, but the results on *GPA* and *satisfaction* remain significant at the 5% even after the standard error correction. The impact on *Exams Passed* is not anymore significant at the standard levels.

4.2.2 Longer-term effects

Table 4 shows the treatment effect on performance in the longer term. We look at the academic performance of students in the retake exams that take place at the end of the second year, where they have the opportunity to pass exams that they have failed. We also look at the third and fourth academic years, as well as at the overall longer term outcome variables. Unfortunately, we can only look at performance measures and not satisfaction measures.

As shown in Table 4, we find no evidence for any treatment effect on performance in the long run. The negative impact found in the short run seems to disappear by the summer retakes, as treated students tend to catch up at the number of passed exams. By the third year, we do not observe any significant differences in the number of exams passed or the average GPA. In addition to academic performance, we also consider the probability of dropping out or graduating in four years. Overall, we find no evidence that the relative performance feedback had any significant impact in the long run.

The findings of the short- and longer-term analysis are summarized in Figure 5, which displays the evolution of the educational performance of the treatment and the control group over time. While the performance of the two groups was similar during the first year - before the intervention took place - the performance of the treatment group declines during the second year. During the third and the fourth years we observe a rapid attenuation of this short-run effect and, by the time of graduation, there are no significant differences.

Although the theoretical framework is silent on the long versus short run differential effect, there are several possible explanations for why the treatment has a decreasing impact over time. It may be partly related to the design of the intervention, which provides a relatively large amount of information initially but has decreasing informational content over time. Students receive information on their position in the ranking in terms of their accumulated GPA and, as a result, the influence of each additional term on their position in the ranking decreases over time. As shown in Figure A6, while 45% of students experienced a variation in their ranking at the beginning of their 2nd year, at the end of the 4th year only 25 percent of students experience any such variation. An alternative explanation is that, as shown in the previous section, the informational gap between treated and control students increases sharply when the intervention is first implemented but then it tends to decrease over time.

4.3 Differential Treatment Effect Depending on Prior Beliefs

The theoretical framework indicates that the impact of feedback on relative performance depends on subjects' preferences and their prior information. Within the proposed framework, the observed negative impact on performance is consistent with a context where students care about their relative performance but, on average, they tend to underestimate it.

Beyond this average effect, the theory suggests that the observed pattern should be driven by the set of students who underestimated their position. Instead, students who overestimated their position may react by increasing their performance. We do not have direct information on the prior beliefs of students who participated in the intervention, but we infer whether a given student was positively or negatively surprised by the feedback on relative performance using the information provided by the survey that was conducted during the second year among a group of students who were not affected by the treatment. We estimate the following equation:

$$Y_s^{self-reported} = \alpha + \beta Y_s^{true} + \boldsymbol{X_s}\gamma + \epsilon_s \tag{5}$$

where $\boldsymbol{X_s}$ includes students' performance during their first year, gender and entry grade. We use these estimates to predict the type of news that students are expected to receive when they obtain access to the ranking information (see Table A2). We classify students into three groups, according to whether the actual ranking and the predicted ranking lie within the same decile (*no news*), the actual ranking is larger than the predicted one (*positive news*), or vice versa (*negative news*). Using this methodology we infer that 644 students are expected to underestimate their position in the distribution, 142 have an accurate estimate, and 180 overestimate it. Figure 6 depicts the distribution of these three groups according to the actual relative performance of students.

We regress equation (4) separately for these three groups of students, using as the two performance outcome variables, *Exams Passed* and *GPA*, during the regular exam period in the second year. According to our estimates, students who, according to our estimations, receive "positive" news, pass 0.47 fewer exams during their second year, and their GPA is 0.21 standard deviations lower, relative to comparable students in the control group (Table 5, columns 2-4). These effects are statistically significant at the 5% level.[22] The treatment has virtually no effect on students who are expected to have correct priors regarding their position in the ranking. However, students receiving "negative" news pass 0.26 more exams during their second year, although this effect is not statistically significant. Overall, these estimates are consistent with the hypothesis that the impact of information depends crucially on the students' prior beliefs.

4.4 Heterogeneity: Gender, High-School Grades and First Year Grades

We have also examined the potential existence of differential treatment effects based on other observable characteristics reported in Table 1, such as gender, performance in the first year and High-School entry grade.

As shown in Table 6, we do not observe any significant difference in the impact of the treatment across any of these dimensions. If anything, the impact of the treatment on educational performance is slightly stronger among female students, among students with a lower performance in their first year in university, and among students with a lower High-School entry grade.[23]

4.5 Robustness and Additional Results

We consider several extensions and robustness checks. First, we report the estimates from an instrumental variables (IV) specification. Second, we examine three additional ways in which the provision of feedback might affect performance: through changes

[22]We are not correcting standard errors for multiple testing in this table because the analysis is driven by the theoretical prediction that the treatment should (i) have a negative impact for individuals that underestimate their ranking, (ii) no effect for individuals who know their position in the distribution, and (iii) a positive impact for individuals that overestimate their position in the distribution.

[23]Similarly, we also observe that the effect is stronger among students who experience a larger decline in their grades relative to their High School grades.

in grading standards, students' participation in feedback surveys and the choice of electives. Finally, we report the post-study probability (PSP) that our findings reflect a true effect (Maniadis et al. 2014).

4.5.1 IV Specification

Given that not all students in the treatment group accessed the information, we use the (random) assignment to the treatment group as an instrument for receiving feedback information. As expected, the point estimates are slightly larger (Table A4). These estimates provide information on the local average treatment effect, under the assumption that individuals in the treatment group were only affected by the provision of additional information, but they did not react somehow to the reception of the email.

4.5.2 Grading Standards

There are several factors that, in this context, limit the potential impact of the treatment on grading standards. First, the same teacher typically grades exams for both the treated and control groups, and whether a student is treated or non-treated is not known to the teacher. Thus it is not possible for the teachers to condition on the treatment when they are grading the exams. Also teachers do not explicitly grade on a curve. In Figures A7 and A8 we plot the grades distribution by course and by different groups for a particular course (in both cases, we look over time). We see that these vary substantially across different groups for a specific course, as well as across different courses, suggesting that teachers do not grade on a curve.

A further piece of evidence for there not being curving is that, in the subjective teaching evaluations, we do not find that perceived fairness in grading changed across the treated and the control students. As shown in Table A5, both groups report after the intervention statistically similar values in term of fairness of grading, *Grading Satisfaction*, indicating that students did not perceive any changes in grading practices.[24]

[24] A related problem would arise if the performance of the treatment groups affected the grading in the control groups, which would lead to a violation of the stable unit treatment value assumption (SUTVA). In this case, the observed gap in performance might overestimate the magnitude of the effect.

Based on this evidence, we conclude that it is unlikely that variations in grading standards are the main driving force behind the observed effect.

4.5.3 Participation in Teaching Evaluations

Not all students participate in teaching evaluations. During the period of our study, only around one third of students completed them. If the provision of feedback information affects which students reply, it might introduce a bias in the information provided by teaching evaluations. While we cannot observe the identity of respondents, we can observe the total number of respondents. As shown in Table A5, there is no significant difference in the response rate of students in the treatment and the control group, suggesting that the treatment did not affect the composition of the sample.

4.5.4 Choice of Electives

During the third and fourth years, students can choose elective courses. A potential way to improve their relative position in the ranking would be to enrol in elective courses where grades tend to be higher, either because grading standards are more lenient or because the value added by the lecturer is larger.

To obtain a proxy for the grades that students may expect to obtain in each elective course, we collected information on the grades received by students in these courses during the two previous years. Overall, we observe 26,119 grades in 168 courses. Using this information, we estimate the following equation:

$$Grade_{c,s} = \alpha + \boldsymbol{C_c}\boldsymbol{\beta} + \boldsymbol{S_s}\boldsymbol{\gamma} + \epsilon_{c,s}, \tag{6}$$

where $Grade_{c,s}$ reflects the grade obtained by student s in course c, and $\boldsymbol{C_c}$ and $\boldsymbol{S_s}$ are two vectors of course and individual dummies, respectively. The vector of coefficients $\boldsymbol{\beta}$ captures the average grade that students enrolled in each course obtain, conditional on their performance in other courses.

Using this information, we calculate the average grade associated with the elective courses chosen by students in the treatment and the control groups, and we normalized this variable to have mean zero and standard deviation one. We then compare the choices of students in the treatment and the control group estimating equation (4). Table A5 shows that students in the treatment group tend to select elective courses with slightly higher grades (0.03 standard deviations), but the difference is not statistically significant.

4.5.5 Power Calculation and Post-study Probability

As Maniadis et al. (2014) point out, the probability that a declaration of a research finding is true depends, not only on the statistical significance of the estimate, but also on the power of the exercise and on the available prior information about the validity of the hypothesis.

In order to assess the power of the short-term analysis reported in Table 3, we consider several plausible scenarios about the potential magnitude of the underlying effect. In the most conservative scenario, we consider that the treatment might affect the outcome variable by 5% of a standard deviation, in an intermediate scenario we consider that it might affect the outcome variable by 15% of a standard deviation and, in the less conservative scenario, we consider a 25% effect. We calculate the power for each one of these three scenarios taking into account the structure of the cluster randomized design, the standard deviation of the outcome variable in the control group, and the observed intra-class correlation within each cluster.[25] Our study would be generally underpowered in the most conservative scenario. If the impact of the treatment was equal to 5% of a standard deviation, our field experiment would be able to detect only with 6% to 19% probability, depending on the variable.[26] However, if the true effect had a magnitude of 15%, which is more in line with the impact of information treatments observed in other studies, the study would be able to detect with a 90% probability its impact on GPA and with a 86% probability its impact on the number of exams passed, although it would be still rather underpowered to detect the impact

[25]We use for the calculation the *Stata 15* command *power twomeans*.

[26]The power is equal to 6% for the number of hours studies, 7% for satisfaction, 17% for exams passed and 19% for GPA.

on satisfaction (power=30%) and on the number of hours studied (power=18%).[27]

Following Maniadis et al. (2014), we calculate the *post-study probability* (PSP) for the effect of the treatment on students' GPA and satisfaction, the two variables that are significant at the 5% level (see Table 3, column 6). The PSP indicates the probability that our research findings are true for a given prior belief about their validity. We conduct this calculation taking into account that this is the first field experiment that studies the impact of relative performance feedback in a context where subjects underestimate their performance. As shown in Figure A9, the analysis is most informative about the impact of the treatment on the GPA. For instance, given a prior belief that there is a 20% probability that the treatment decreases GPA by 15%, the observed evidence would increase this belief from 20% to 82%. However, the analysis is less informative about the impact of the treatment on satisfaction, reflecting the lower power available in this case. Given a 20% prior belief, the evidence raises the posterior belief only to 60%.

5 Conclusions

We study the role of relative performance feedback in a higher education setting, where there has been an increasing demand to provide students with more feedback on their performance. We elicit beliefs from students concerning their relative position and find that students in our study are uninformed about their rank in their cohort and that they tend to underestimate their position in the distribution of grades. We randomly assigned some students into a treatment that gave them access to information regarding their relative position in the distribution of grades. The treatment was effective in informing students of their rank compared to a control sample who were not given access to this information and who remained relatively uninformed and underestimated their rank. We found that providing feedback on students' relative performance had a negative impact on their performance. This effect is driven by those students who

[27]For instance, Azmat and Iriberri (2010) find an effect of 14.8% of a standard deviation and in Blanes-i-Vidal and Nossol (2011) the impact is around 14.5% of standard deviation.

underestimate their position in the absence of feedback. By regularly providing access to this information to the treatment group over the course of their studies, there is, however, no further impact on their performance and the treated students catch up. In addition to the effect on academic performance, we find a positive effect on self-reported student satisfaction with the quality of the courses. This was, perhaps, a response to the positive surprise concerning their own ranking.

Our study also highlights a number of important considerations regarding providing individuals with feedback and raises a number of interesting questions that are relevant to policymakers and organizations. First, our results suggest that the impact of relative performance feedback may depend crucially on individuals' prior information. Nonetheless, given that our analysis on the role of priors relies on observational evidence, further work is needed exploiting exogenous variations in subjects' expectations about their relative performance. Second, the timing of the information is relevant. We showed that the impact of the treatment is confined to the first time that the students receive the information. If the information had been provided in the final year of study, or after graduation, the impact could have been different. This therefore raises the question of the optimal timing of information release. If, for example, the program lasts three or four years, as in the case of an undergraduate degree, the optimal timing might be different than that for an MBA lasting just one or two years. Third, the reference group might matter. The individuals in our study compare themselves to the cohort to which they belong, and thus the individuals' reference group is unchanged over time. This might be one reason for the lack of response to feedback beyond the first time individuals received information. If the reference group changed over time, then the information may once again have an impact. Fourth, the coarseness of the information provided may play a role. We provided very detailed information, in particular, individuals learned about the decile to which they belonged. If they were only informed of whether they were above or below an average student, or if they were given the exact percentile, the response might have been different. Again, there may an optimal information partition to provide (Diamond, 1985). Fifth, the incentives associated with relative performance could change the response to the information. In our setting, there was no explicit immediate reward within the university for ranking high versus low. However, there might be implicit or longer term rewards associated with rank. Finally, whether information feedback is provided privately or publicly may have significant impact. In our case, it was provided privately. If such a ranking were made public, there might be some consequences because of status seeking, even in the absence of explicit rewards within the organization.

References

[1] Albornoz, Facundo, Antonio Cabrales, and Esther Hauk (2017), "Occupational choice with endogenous spillovers," Barcelona GSE Working Paper 972.

[2] Al-Ubaydli, Omar and John List (2015), "Do Natural Field Experiments Afford Researchers More or Less Control than Laboratory Experiments?," *American Economic Review*, Vol. 105(5), pp. 462.

[3] Azmat, Ghazala and Nagore Iriberri (2010), "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students," *Journal of Public Economics*, Vol. 94(7-8), pp. 435–452.

[4] Azmat, Ghazala and Nagore Iriberri (2016), "The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction," *Journal of Economics and Management Strategy*, Vol. 25(1), pp. 77-110.

[5] Bandiera, Oriana, Valentino Larcinese and Imran Rasul (2015), "Blissful ignorance? A natural experiment on the effect of feedback on students' performance," *Labour Economics*, Vol. 34, pp. 13-25.

[6] Barankay, Iwan (2012), "Rank Incentives: Evidence from a Randomized Workplace Experiment", The Wharton School, mimeo.

[7] Bedard, Kelly, and Peter Kuhn (2008), "Where class size really matters: Class size and student ratings of instructor effectiveness," *Economics of Education Review*, Vol. 27(3), pp. 253-265.

[8] Benabou, Roland (1993), "Workings of a city: location, education, and production," *Quarterly Journal of Economics*, Vol. 108(3), pp. 619-652.

[9] Blader, Steven, Claudine Madras Gartenberg and Andrea Prat (2015), "The Contingent Effect of Management Practices," Columbia Business School Research Paper No. 15-48.

[10] Blanes i Vidal, Jordi and Mareike Nossol (2011), "Tournaments without Prizes: Evidence from Personnel Records," *Management Science* Vol. 57(10), pp. 1721–1736.

[11] Calvò-Armengol, Antoni, Eleonora Patacchini and Yves Zenou (2009), "Peer effects and social networks in education," *Review of Economic Studies*, Vol. 76(4), pp. 1239-1267.

[12] Caplin, Andrew, and John Leahy (2001), "Psychological expected utility theory and anticipatory feelings," *Quarterly Journal of Economics*, Vol. 116(1), pp. 55-79.

[13] Charness, Gary, David Masclet, and Marie Claire Villeval (2013), "The Dark Side of Competition for Status," *Management Science*, Vol. 60(1), pp. 38-55.

[14] Credé, Marcus, Sylvia G. Roch, and Urszula M. Kieszczynka (2010) "Class attendance in college a meta-analytic review of the relationship of class attendance with grades and student characteristics," *Review of Educational Research* Vol. 80, pp. 272-295.

[15] Diamond, Douglas W. (1985), "Optimal release of information by firms," *Journal of Finance* Vol. 40, pp. 1071-1094.

[16] Duflo, Esther and Emmanuel Saez (2003), “The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment,” *Quarterly Journal of Economics*, Vol. 118 (3), pp. 815-842.

[17] Eil, David, and Justin M. Rao (2011), “The good news-bad news effect: asymmetric processing of objective information about yourself,” *American Economic Journal: Microeconomics*, Vol. 3.2, pp. 114-138.

[18] Emerson, Tisha, and Beck A. Taylor (2004), “Comparing student achievement across experimental and lecture-oriented sections of a principles of microeconomics course,” *Southern Economic Journal*, Vol. 70(3), pp. 672-693.

[19] Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval (2009), “Feedback and incentives: experimental evidence,” *Labour Economics*, Vol. 16, pp. 679-688.

[20] Ertac, Seda. “Does self-relevance affect information processing? Experimental evidence on the response to performance and non-performance feedback.” *Journal of Economic Behavior & Organization*, Vol. 80(3), pp. 532-545.

[21] Ertac, Seda (2005), “Optimal Information Revelation in the Presence of Social Comparisons: Theory and Experiments,” mimeo.

[22] Gerhards, Leonie and Neele Siemery (2014), “Private versus public feedback: The incentive effects of symbolic awards,” mimeo.

[23] Gill, David, Zdenka Kissova, Jaesun Lee and Victoria Prowse (2016), “First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision,” mimeo.

[24] Gok, Tolga (2011), “Comparison of student performance using web-and paper-based homework in large enrollment introductory physics courses,” *International Journal of the Physical Sciences*, Vol. 6(15), 3740-3746.

[25] Harrison, Glenn W., and John A. List (2004), “Field experiments,” *Journal of Economic Literature*, Vol. 42.4, pp. 1009-1055.

[26] Hannan, R. Lynn, Ranjani Krishnan and Andrew H. Newman (2008), "The effects of disseminating relative performance feedback in tournament versus individual performance compensation plans," *Accounting Review*, Vol. 83(4), pp. 893-913.

[27] Hoxby, Caroline M. (2009), "The Changing Selectivity of American Colleges," *Journal of Economic Perspectives*, Vol. 23(4), pp. 95-118.

[28] Katreniakova, Dagmara (2014), "Information, Aspirations and Incentive to learn: A Randomized Field Experiment in Uganda," CERGE-EI, mimeo.

[29] Khunen, Camelia N. and Tymula Agnieszka (2012), "Feedback, Self-Esteem, and Performance in Organizations," *Management Science*, Vol. 58(1), pp. 94-113.

[30] List, John A., Azeem M. Shaikh, and Yang Xu (2016), "Multiple hypothesis testing in experimental economics," NBER Working paper n. 21875.

[31] Maniadis, Zacharias, Fabio Tufano, and John A. List. (2014), "One swallow doesn't make a summer: New evidence on anchoring effects," *American Economic Review* Vol. 104.1, pp. 277-290.

[32] Mas, Alexandre and Enrico Moretti (2009), "Peers at work," *American Economic Review*, Vol. 99(1), pp. 112-145.

[33] Mobius, Markus, Muriel Niederle, Paul Niehaus, and Tanya Rosenblat (2014), "Managing self-confidence," mimeo.

[34] Moore, Don A., and Daylian M. Cain (2007), "Overconfidence and underconfidence: When and why people underestimate (and overestimate) the competition," *Organizational Behavior and Human Decision Processes*, Vol. 103, pp. 197-213.

[35] Murad, Zahra, Martin Sefton, and Chris Starmer (2016), "How do risk attitudes affect measured confidence?," *Journal of Risk and Uncertainty* Vol. 52(1), pp. 21-46.

[36] Sacerdote, Bruce (2001), "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *Quarterly Journal of Economics*, Vol. 116(2), pp. 681-704.

[37] Schotter, Andrew, and Isabel Trevino. "Belief elicitation in the laboratory." Annual Review of Economics Vol. 6, pp. 103-128.

[38] Spence, Michael (1973), "Job market signaling," *Quarterly Journal of Economics*, Vol. 87, pp. 355-374.

[39] Tran, Anh, and Richard Zeckhauser (2012), "Rank as an inherent incentive: Evidence from a field experiment," *Journal of Public Economics*, Vol 96, pp. 645–650.

[40] Williams, James, and David Kane (2009), "Assessment and feedback: Institutional experiences of student feedback, 1996 to 2007." *Higher Education Quarterly*, Vol. 63, pp. 264-286.

Figure 1: Time-line of the Experiment

2009-2010: No intervention

2010-2011, 2011-2012 and 2012-2013: Two-time intervention during three academic years

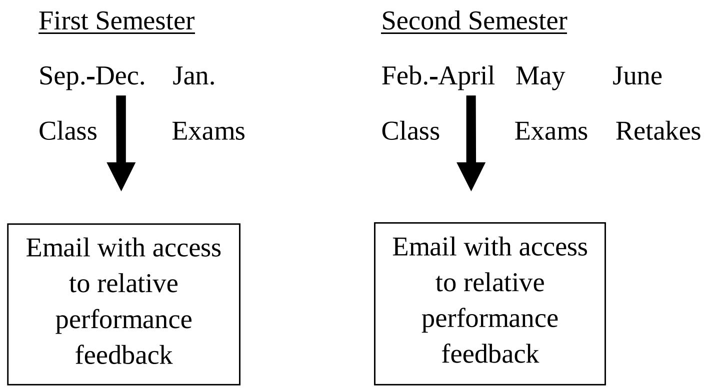
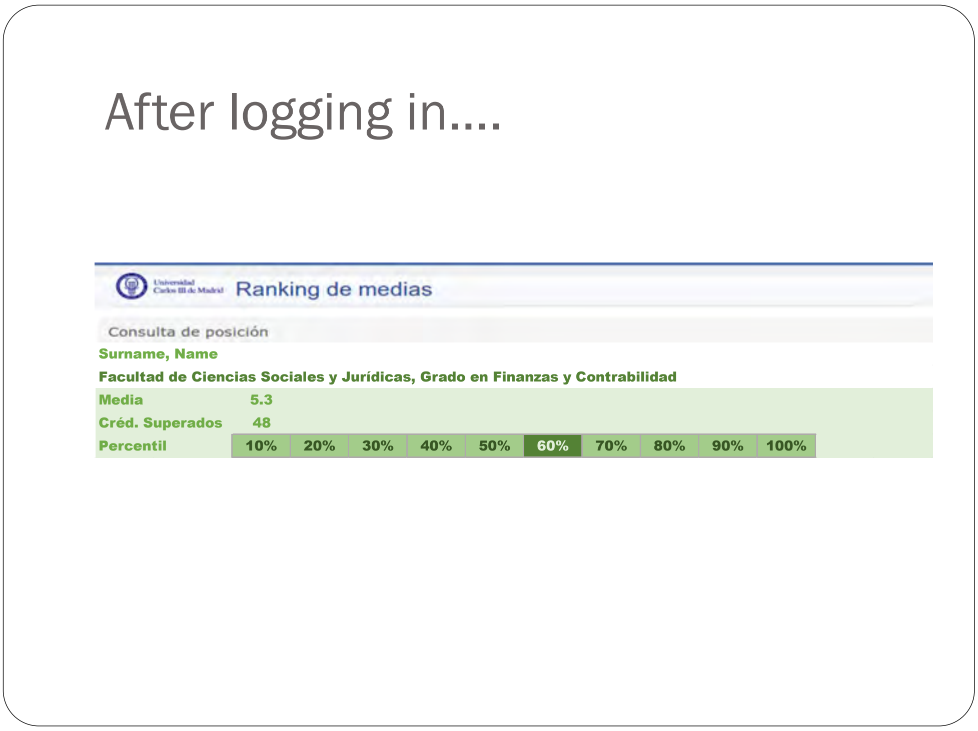
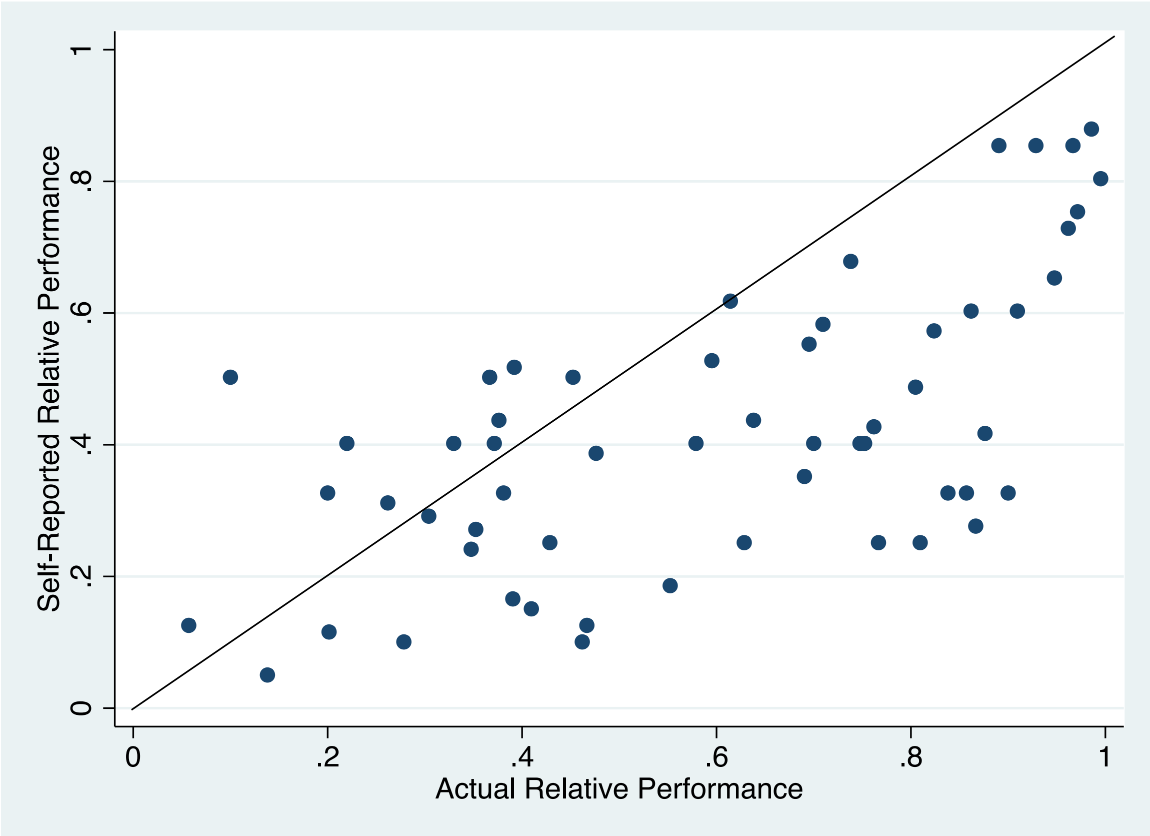


Figure 2: Feedback on Relative Performance



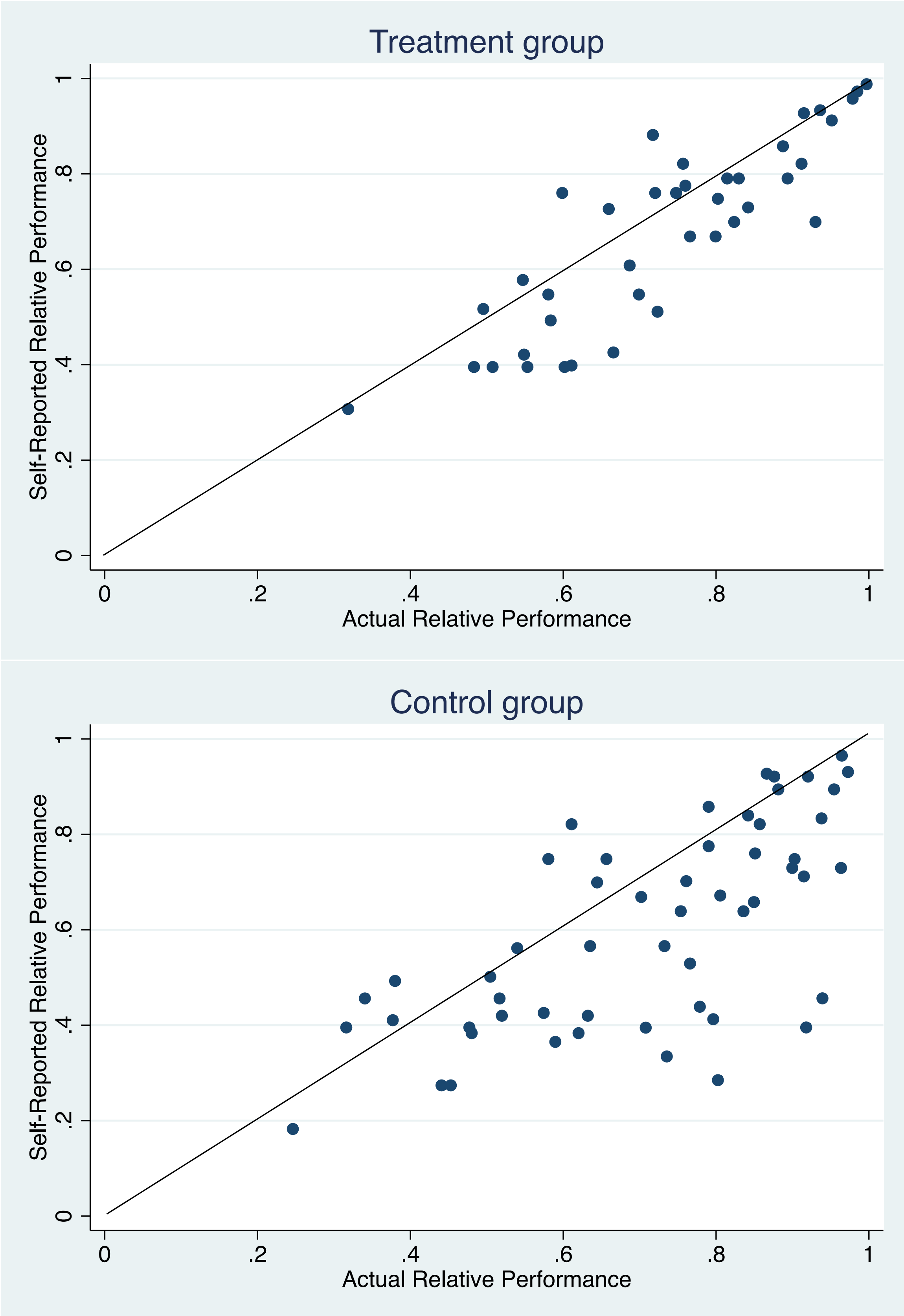
Note: The figure shows the screen treated students accessed to when clicking in the link received in the email.

Figure 3: Expected and Actual Relative Performance Second Year (out-of-trial) Students



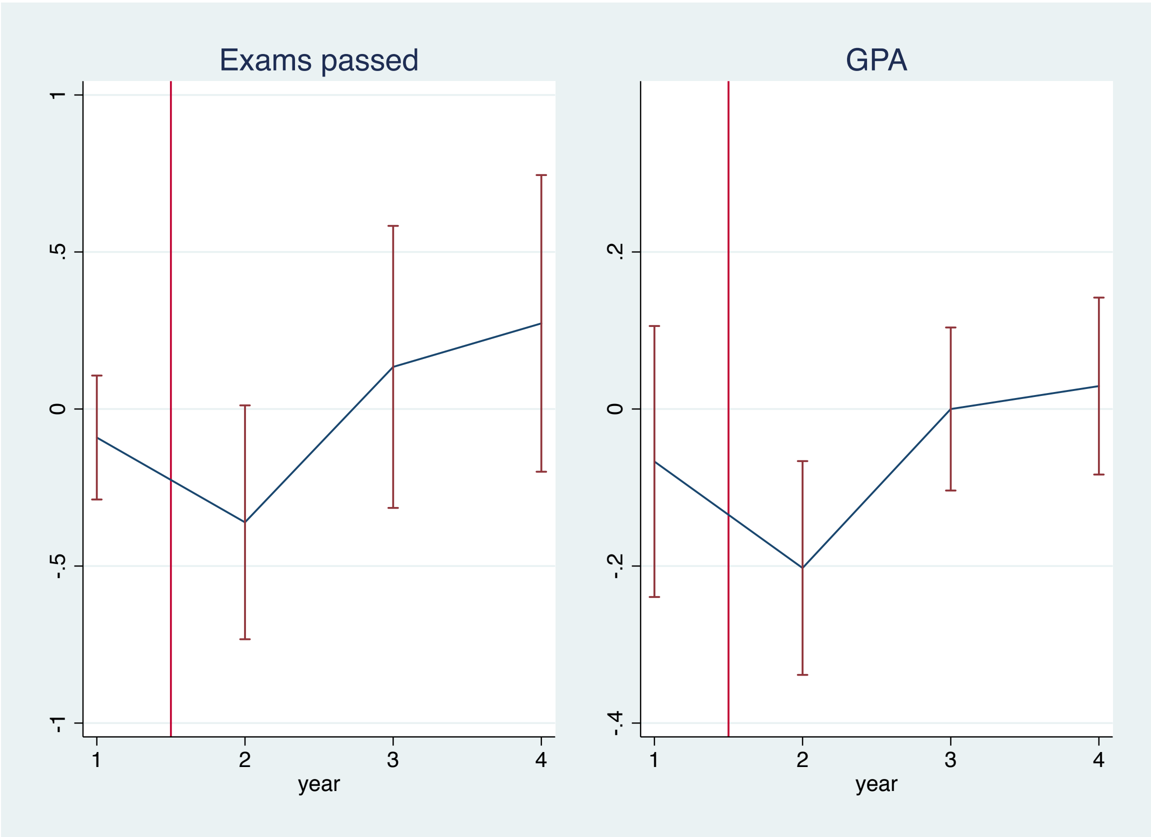
Note: The figure includes information from 57 second year Economics students, class of 2014, who were surveyed in November 2011, at the beginning of the second year. The x-axis reports the actual position in the ranking, normalized between 0 (lowest grade) and 1 (highest grade) among students who enrolled in the same degree in Fall 2009. The y-axis provides information on the self-reported relative performance, normalized in a similar way.

Figure 4: Expected and Actual Relative Performance after Treatment



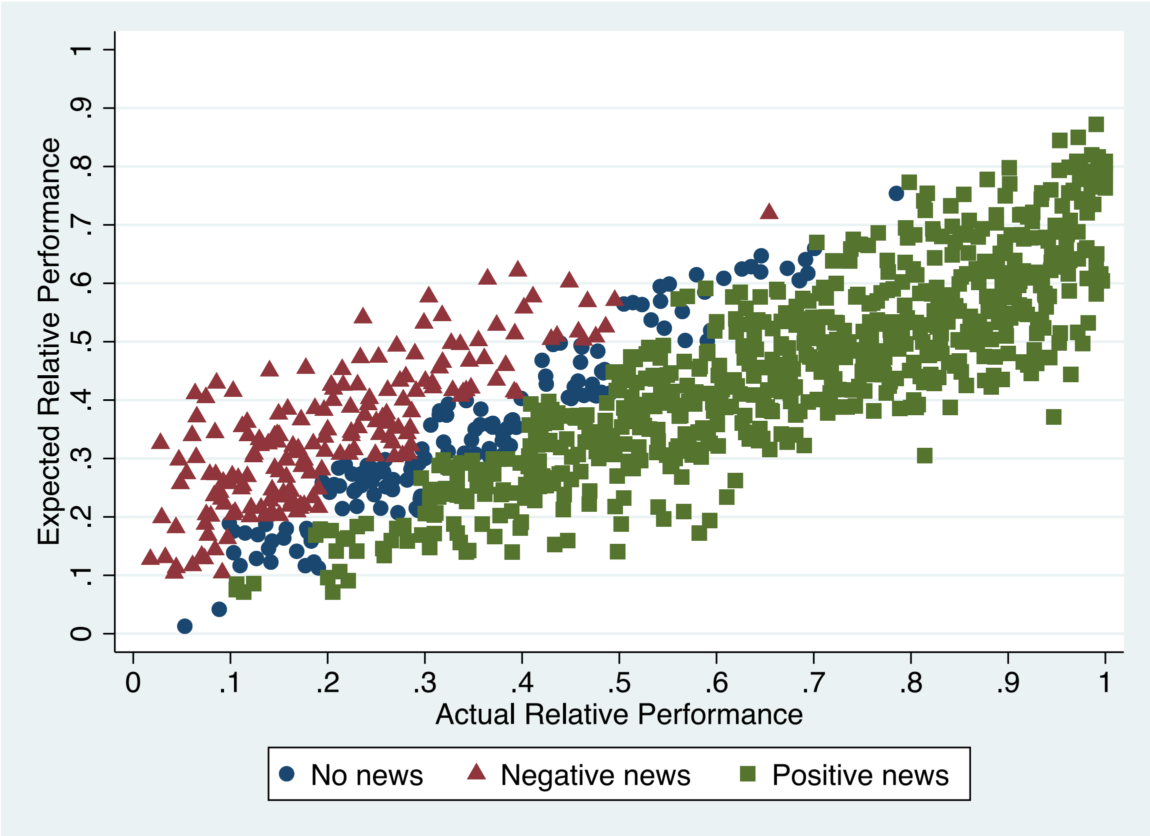
Note: The figure includes information from 93 students in Economics and Business who were surveyed in the summer of 2013, at the time of graduation. The upper (lower) panel includes students in the treatment (control) group. The x-axis reports the actual position in the ranking, normalized between 0 (lowest grade) and 1 (highest grade), relative to students from the same cohort. The y-axis provides information on the self-reported relative performance.

Figure 5: Short- and Longer-term Impact



Note: The above figures display the corresponding point estimates and 95% confidence intervals from the analyses reported in Tables 1, 3 and 4.

Figure 6: Predicted Difference between Expected and Actual Relative Performance



Note: The figure includes information on the actual ranking of the 977 individuals who participated in the intervention and on their expected ranking, according to their observable characteristics. The red group (triangles) includes individuals who expect a higher ranking than their actual one, the blue group (points) includes individuals with accurate expectations, and the green group (squares) includes individuals who are expected to underestimate their relative ranking.

Table 1: Balance-Check

	1	2	3	4
	All		**Treated-Control**	
	Mean	St. Dev.	Difference	p-value
		Demographics		
Female	0.54	0.50	0.03	0.43
Foreigner	0.03	0.18	-0.00	0.71
High School	0.95	0.21	-0.02	0.17
Entry Grade	7.24	0.99	-0.10*	0.07
Subjects taken	4.89	0.78	-0.06	0.23
Geographic origin:				
Central Madrid	0.31	0.46	-0.01	0.81
Western Madrid	0.11	0.32	0.01	0.72
Southern Madrid	0.22	0.41	0.05*	0.07
Other regions	0.30	0.46	-0.04	0.19
		Performance		
Exams passed	3.70	1.49	-0.09	0.37
GPA	6.02	1.36	-0.07	0.45
		Teaching evaluations		
Satisfaction	3.87	0.76	0.01	0.89
Hours of study	2.92	0.45	0.13	0.13
Grading satisfaction	3.56	0.67	-0.00	1.00
Response rate	0.41	0.19	0.02	0.68

Notes: The table provides information on 977 students that took part in the intervention, except variable *Entry Grade* which is available only for 966 students. Column (3) reports the difference between the treatment and the control group, conditional on degree. Column (4) reports the p-value of this difference. *Exams passed* provides information for the second term of the first year and *GPA* is measured at the end of the first year. The lower panel of the table includes information from teaching evaluations that were completed in Fall of academic year 2010-2011, before the intervention took place. The sample includes 182 tutorial groups. *Satisfaction* provides information on students' self-reported satisfaction with the overall quality of each course, coded in a scale from 1 (not at all) to 5 (very satisfied). *Hours of Study* provides information on the number of hours studied weekly. Hours of study takes value 1 if the individual studied less than an hour per week; 2, between one and four hours; 3, four to seven hours; 4, seven to ten hours and 5 more than ten hours. *Grading Satisfaction* reports the average satisfaction with the grading, also coded in a scale from 1 (not at all) to 5 (very satisfied). *Response rate* reports the share of students who complete the teaching evaluation survey. This information is available for 149 groups.

Table 2: Who Checks the Information?

	1	2	3	4
Dependent variable:	**Checked**		**# checks**	**# years**
Female	0.106**	0.079*	0.065	0.092
	(0.047)	(0.045)	(0.075)	(0.103)
True rank		0.585***	1.409***	1.814***
		(0.097)	(0.160)	(0.221)
Entry grade		-0.047	-0.070	-0.119
		(0.034)	(0.057)	(0.078)
Constant	0.665***	0.708***	0.762**	1.225**
	(0.034)	(0.229)	(0.380)	(0.523)
Observations	354	347	347	347

Note: The regression includes information from 354 students who were assigned to the treatment group. The dependent variable in (1) and (2) is a dummy that takes value one if the students checked at least once the information. The dependent variable in (3) measures the (log) number of times students checked the information, and the dependent variable in (4) measures the number of years in which the student checked the information. Robust standard errors in parenthesis. *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$

Table 3: Short-term Impact

	1	2	3	4	5	6
	All		**Treated vs. Control**			
					P-value	
	Mean	St. Dev.	Difference	St. Error	Unadjusted	Adjusted for MHT
Exams passed	7.75	3.83	-0.36	0.18	0.057	0.114
GPA	5.80	1.36	-0.20	0.06	0.006	0.024
Satisfaction	3.63	0.85	0.30	0.11	0.010	0.030
Hours of study	3.00	0.48	0.15	0.10	0.131	0.131

Note: Rows 1 and 2 include information on performance measures measured at the student level (N=977). Rows 3 and 4 provide information measured at the level of course and tutorial group (N=165). The variable *Exams passed* refers to the number of exams passed during the regular exam season in academic year 2010-2011 (January and May). *GPA* refers to the Grade Point Average. *Satisfaction* provides information on students' self-reported satisfaction with the overall quality of each course, coded in a scale from 1 (not at all) to 5 (very satisfied). *Hours of Study* provides information on the number of hours studied weekly. Hours of study takes value 1 if the individual studied less than an hour per week; 2, between one and four hours; 3, four to seven hours; 4, seven to ten hours and 5 more than ten hours. Column 3 reports the main estimates from equation (4), and each cell corresponds to a different regression where the independent variable is a dummy that takes value one if the student was part of the treatment group and the dependent variable is indicated on the left-hand side of the table. All regressions include controls for a set of individual predetermined characteristics. Column 5 reports standard errors clustered at the lecture group level in parenthesis, and column 6 reports Holm-Bonferroni standard errors corrected for multiple hypothesis testing (MHT).

Table 4: Longer-term Impact

	1	2	3	4	5	6
	All		Treated vs. Control			
					P-value	
	Mean	St. Dev.	Difference	St. Error	Unadjusted	Adjusted for MHT
			End of Second year: Retakes			
Retakes passed	1.12	1.25	0.19	0.12	0.15	0.15
GPA after retakes	6.04	1.24	-0.07	0.03	0.06	0.12
			Third year			
Exams passed	8.07	4.06	0.13	0.21	0.54	1.00
GPA	6.26	1.23	0.00	0.05	1.00	1.00
Retakes passed	0.98	1.28	0.05	0.09	0.61	1.00
GPA after retakes	6.14	1.25	-0.04	0.03	0.15	0.60
			Fourth year			
Exams passed	6.69	4.41	0.27	0.22	0.24	0.96
GPA	6.71	1.39	0.03	0.05	0.59	1.00
Retakes passed	0.68	1.11	0.02	0.05	0.72	0.72
GPA after retakes	6.30	1.27	-0.03	0.04	0.37	0.74
			Overall			
Exams passed	25.82	11.54	0.32	0.50	0.54	1.00
Accumulated GPA	6.30	1.27	-0.03	0.04	0.37	1.00
Dropout	0.15	0.36	-0.01	0.02	0.74	0.74
Graduation in 4 years	0.51	0.50	0.02	0.02	0.45	1.00

Note: Columns 1 and 2 include information on 977 students that took part in the intervention for performance measures (*Exams passed*, *GPA*, *Retakes passed* and *GPA after retakes*). The variable *Exams passed* refers to the number of exams passed during the regular exam season (January and May). *GPA* refers to the Grade Point Average. Variable *Retakes passed* refers to exams passed during the retake season (June). *GPA after retakes* refers to accumulated Grade Point Average grade up to that point in the Degree. Dropout takes the value of 1 if the student dropped the Degree. Graduation in 4 years takes the value of 1 if the student completed the degree in 4 years. Column 3 reports the main estimates from equation (4), and each row corresponds to a different regression where the independent variable is a dummy that takes value one if the student was part of the treatment group and the dependent variable is indicated in the first column. All regressions include controls for a set of individual predetermined characteristics. Column 4 reports standard errors clustered at the lecture group level in parenthesis. Column 5 reports standard errors clustered at the lecture group level in parenthesis, and column 6 reports Holm-Bonferroni standard errors corrected for multiple hypothesis testing.

Table 5: Differential Treatment Effect Based on Prior Beliefs

	1	2	3	4
		News		
Dependent variable:	All	Positive	No News	Negative
Exams passed	-0.36*	-0.47**	0.16	0.26
	(0.18)	(0.22)	(0.32)	(0.59)
GPA	-0.20***	-0.21**	-0.14	-0.11
	(0.06)	(0.08)	(0.14)	(0.18)
N	966	644	142	180

Note: The dependent variable is the number of exams passed and GPA during the regular exam period of the 2nd year. All regressions include controls for gender, nationality, entry grade, academic background, academic performance during the first year at university and geographical origin. Standard errors are clustered at the lecture group level. Significance levels: *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$

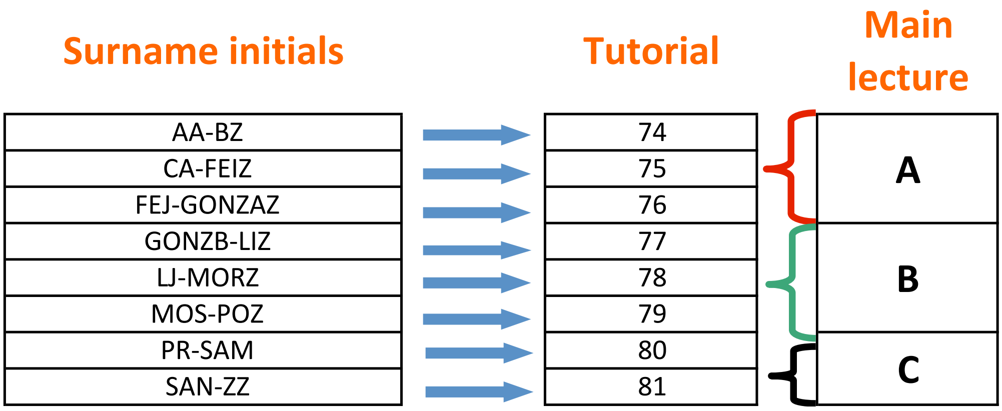
Table 6: Heterogeneity Analysis

	1	2	3	4	5	6
	Gender		**1st year grades**		**HS grades**	
Dependent variable:	Female	Male	Low	High	Low	High
Exams passed	-0.46	-0.18	-0.33*	-0.18	-0.60**	-0.13
	(0.28)	(0.22)	(0.16)	(0.22)	(0.21)	(0.23)
P-value (unadjusted):	0.47		0.54		0.05	
P-value adjusted for MHT:	0.94		0.54		0.15	
GPA	-0.19**	-0.21**	-0.20***	-0.14	-0.22***	-0.19**
	(0.07)	(0.09)	(0.06)	(0.10)	(0.07)	(0.08)
P-value (unadjusted):	0.81		0.53		0.62	
P-value adjusted for MHT:	0.81		1.00		1.00	
N	521	445	435	531	482	479

Note: The dependent variable is the number of exams passed (upper panel) and GPA during the regular exam period of the 2nd year (lower panel). All regressions include controls for gender, nationality, entry grade, academic background, academic performance during the first year at university and geographical origin. Standard errors are clustered at the lecture group level. The third and the fourth row of each panel provide information on the p-value of a test of equality of coefficients, unadjusted and adjusted for multiple hypothesis testing. Significance levels: *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$

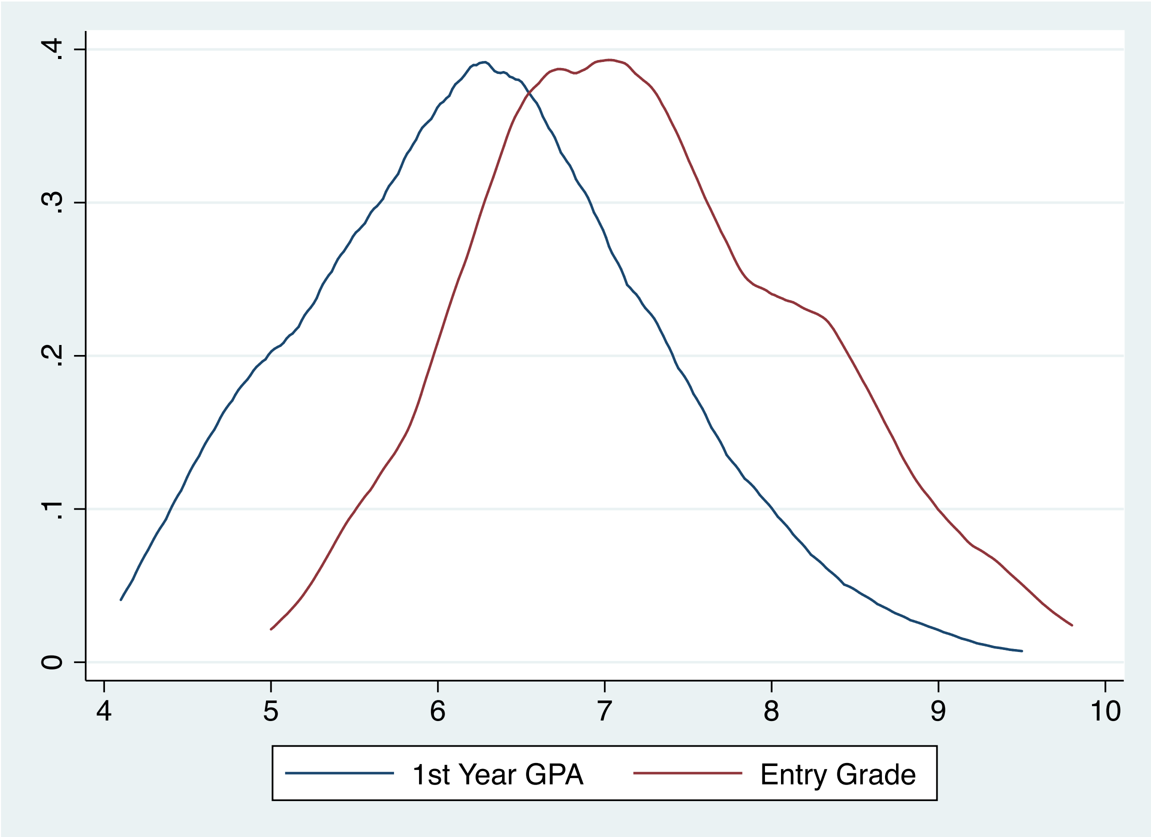
A Figures and Tables

Figure A1: Assignment to Tutorial and Lecture Groups



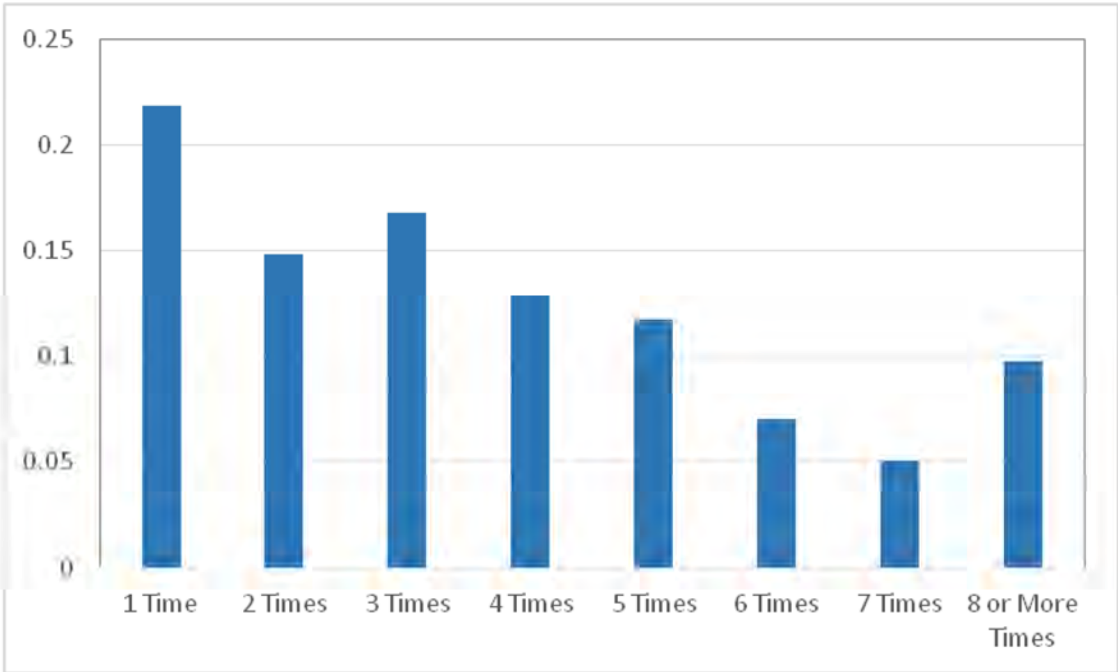
Note: This assignment corresponds to 1st year students, Business Administration, Getafe, Spanish track, 2010.

Figure A2: Entry Grade and 1st Year Grades in College



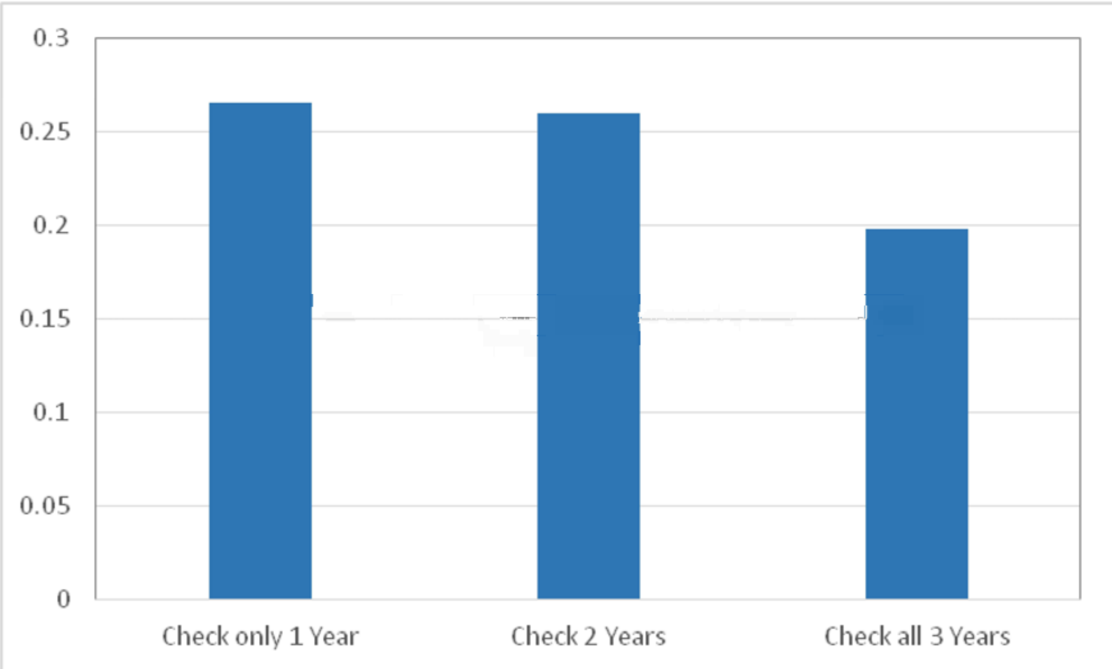
Note: Grades distribution of entry grades and 1st year grades in college. Both grades lie between 0 and 10, 5 being the passing grade.

Figure A3: Share of Individuals who Check the Ranking, by Number of Times



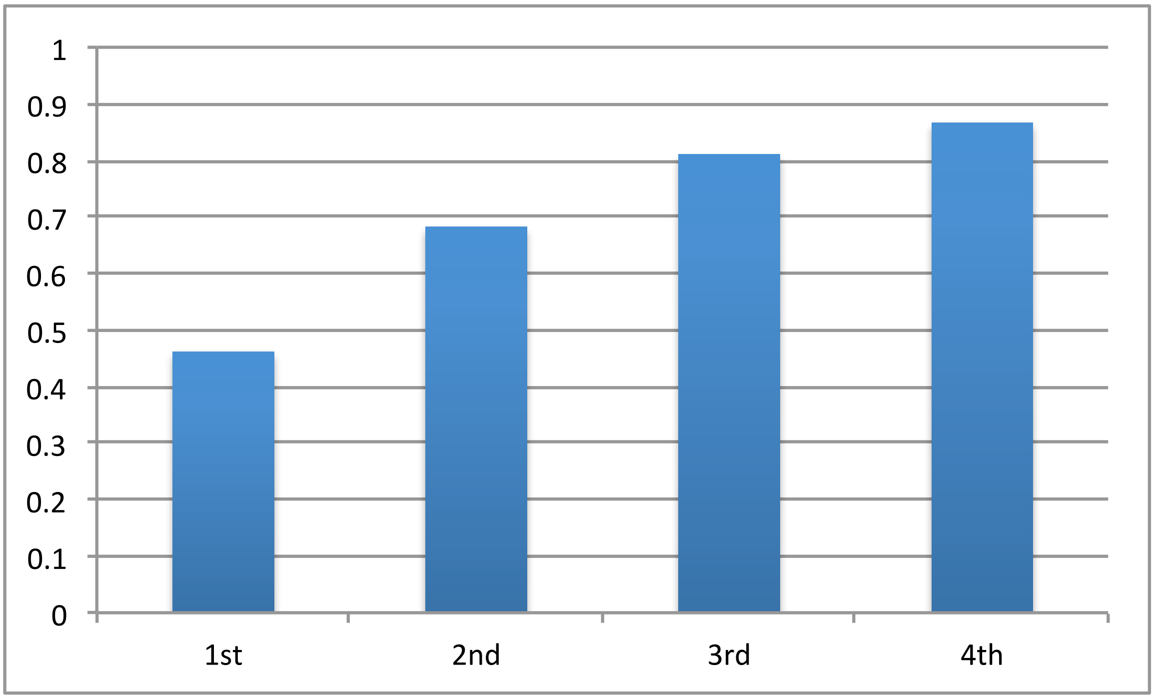
Note: Each bar reflects the proportion of people who check once, twice, three times, up to 8 times or more the relative performance information.

Figure A4: Share of Individuals who Check the Ranking, across the Three Academic Years



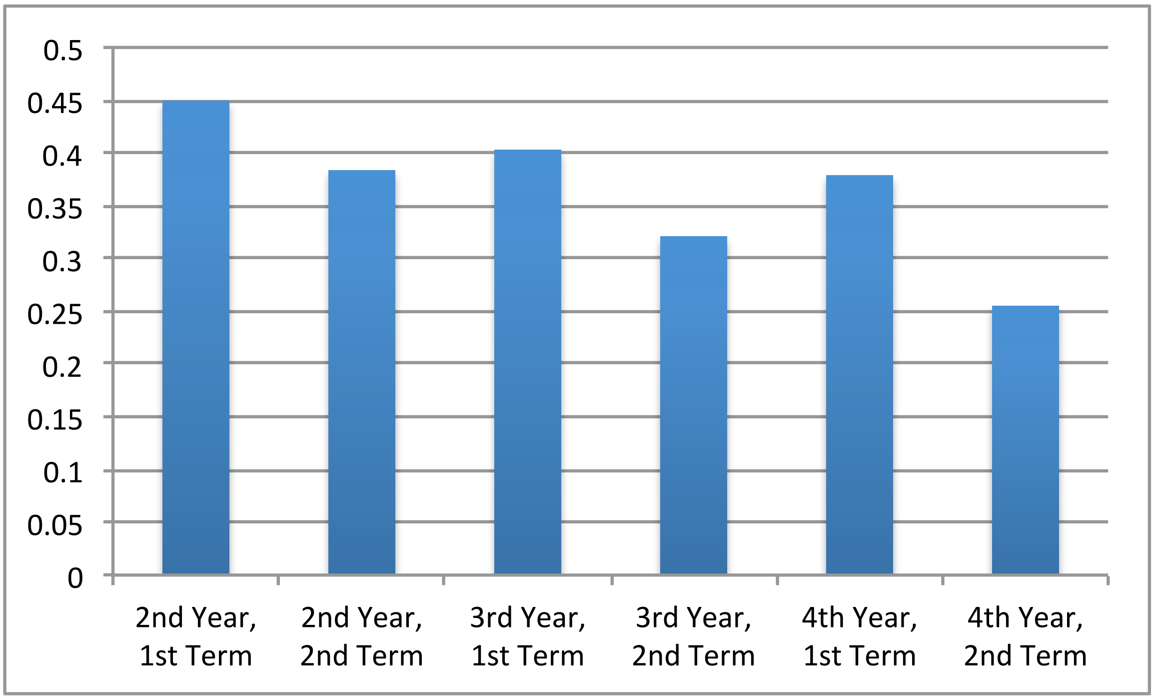
Note: Each bar reflects the proportion of people who checks at least once the relative performance information during one year only, during two years or all three years.

Figure A5: Share of Individuals who Check the Ranking, by Quartile



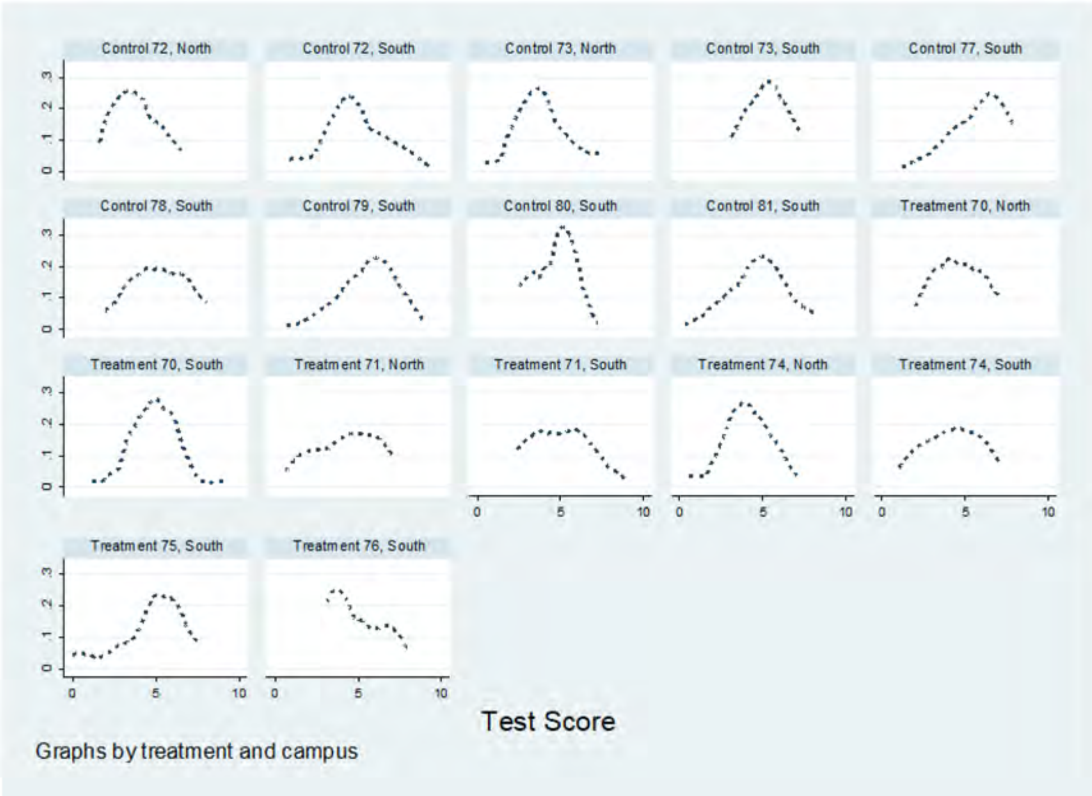
Note: Each bar reflects the proportion of people who checks at least once the relative performance information, by Quartile.

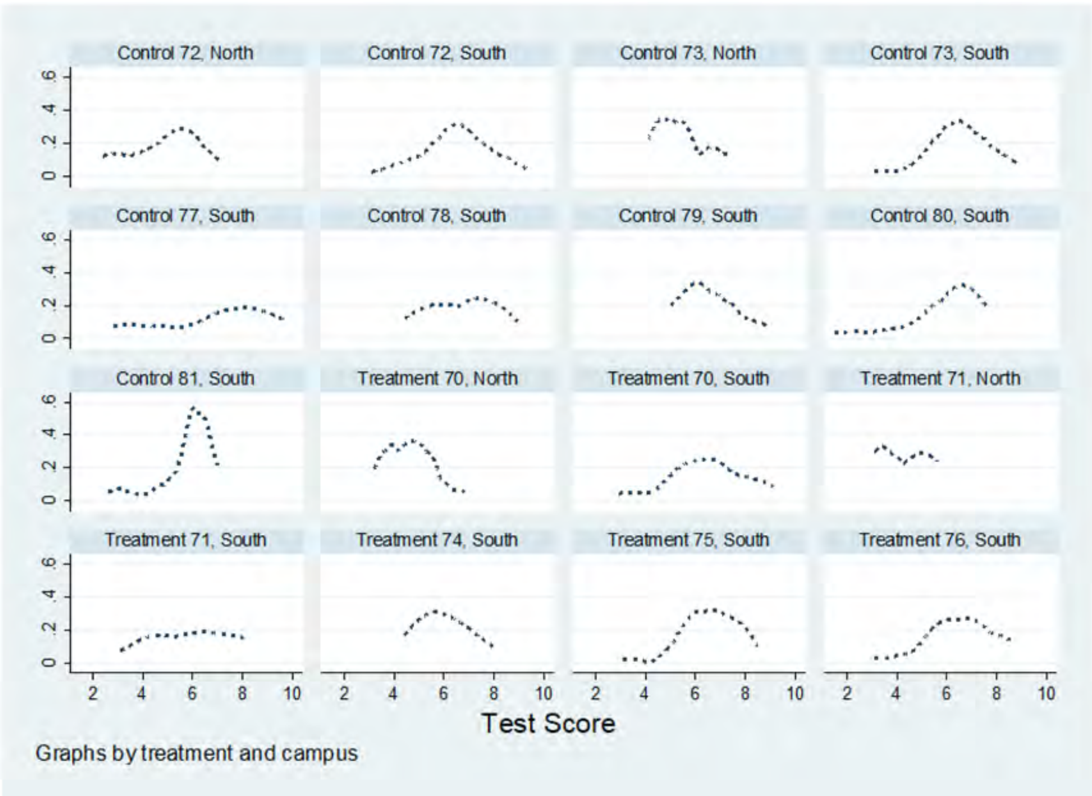
Figure A6: Information over Time



Note: Each bar reflects the proportion of people who experienced mobility from one term to the next in terms of their decile in the relative distribution. For instance, approximately 45% of individuals were placed in a different decile at the end of the 1st term of their 2nd year relative to their position at the end of the 1st year.

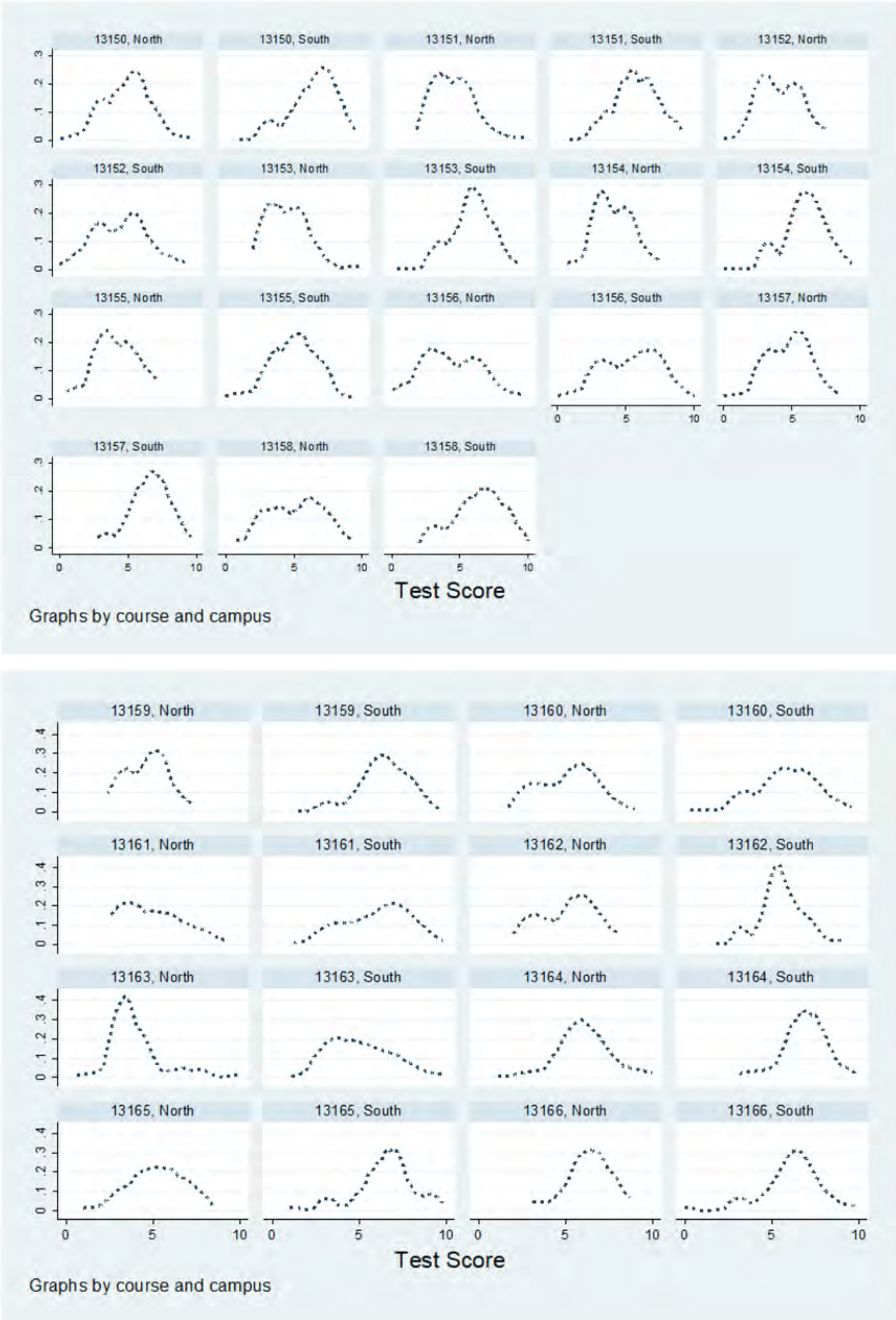
Figure A7: Grade Distributions of Microeconomic in 2009 and Macroeconomics in 2010





Note: Test score distributions for different courses and groups in 2009 and 2010. All test scores are from the Business Degree, North and South Campuses.

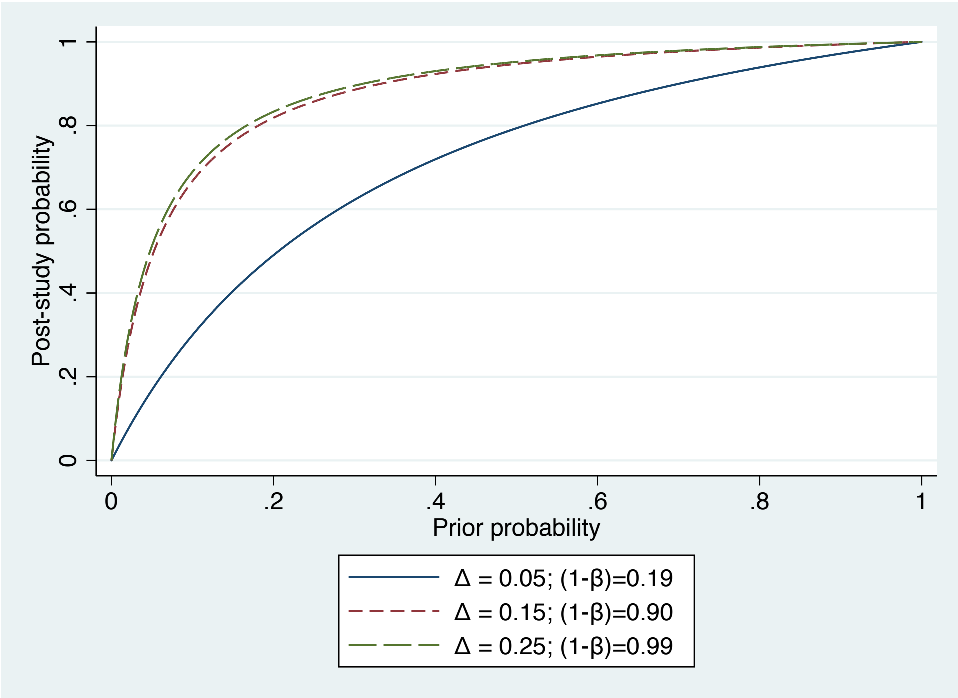
Figure A8: Grade Distributions of all courses in 2009 and 2010



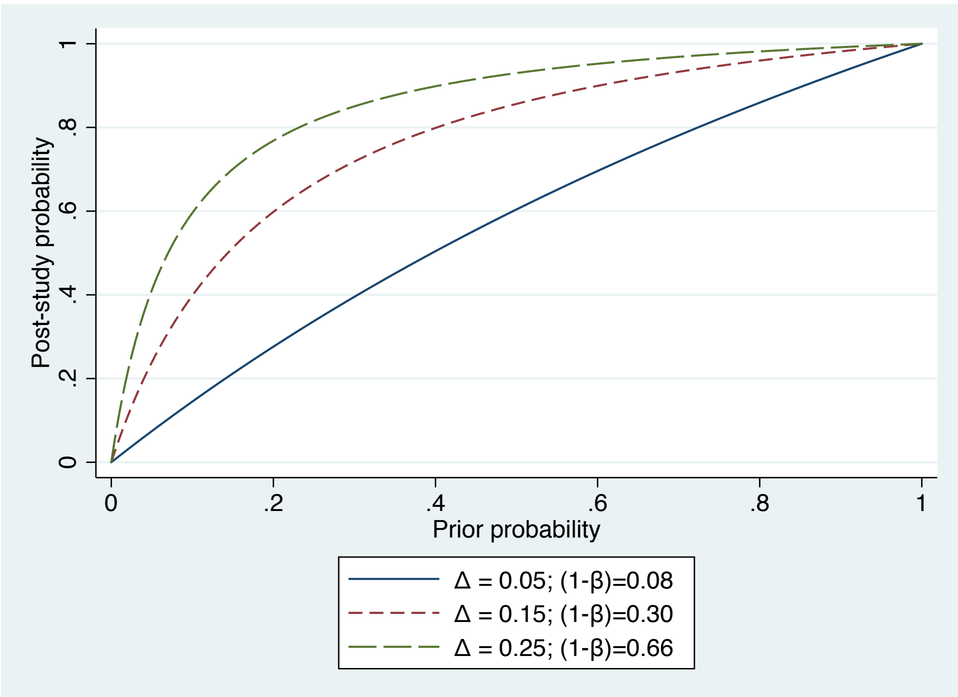
Note: Test score distributions for different courses and groups in 2009 and 2010. All test scores are from the Business Degree, North and South Campuses.

Figure A9: Post-Study Probability

(a) GPA



(b) Satisfaction



Note: The figure provides information on the *Post-Study Probabilities* that the treatment affects the different outcome variables reported in Table 3, following the methodology proposed by Maniadis et al. (2014). The post-study probability depends on the significance of estimates, the potential strength of the effect (Δ), and the prior probability assigned to this effect. For this calculation we rely on significance levels reported in Table 3, column 6. In order to calculate the power of the estimation (1-β), we consider several possible scenarios relative to the potential strength of the effect (Δ), measured in standard deviations, and researchers' priors (π). For instance, the upper figure indicates that, given an initial prior that there is a 20% probability that providing feedback on relative performance increases students GPA by 15% of a standard deviation (and a 80% probability that it has no effect), the post-study probability would be around 82%.

Table A1: Assignment to Treatment

	Southern Campus		Northern Campus	
	Treatment	Control	Treatment	Control
Finance and Accounting	36 (1)	59 (1)		
Economics	47 (1)	187 (2)		
Business	60 (1)	121 (2)	40 (1)	35 (1)
Law	60 (1)	132 (2)		
Law and Business	50 (1)	49 (1)	61 (1)	40 (1)

Note: Each cell includes information on the number of students assigned to each group and, in parentheses, on the number of lecture groups.

Table A2: Expected vs. Actual Relative Performance - Before the Treatment

Female	-0.07
	(0.04)
True rank	0.34***
	(0.11)
Entry grade	0.11***
	(0.03)
Constant	-0.49**
	(0.21)
Adj. R-squared	0.48
N	52

Note: The regression includes information from 52 students who were surveyed at graduation. The dependent variable is the absolute difference between the self-reported position in the ranking and the actual one, normalized between 0 and 1. Robust standard errors in parenthesis. *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$

Table A3: Expected vs. Actual Relative Performance - After the Treatment

	1	2
Treatment	-0.050**	-0.048*
	(0.025)	(0.025)
Female		0.048**
		(0.022)
True rank		0.023
		(0.070)
Entry grade		-0.004
		(0.023)
Constant	0.143***	0.129
	(0.019)	(0.142)
Adj. R-squared	0.055	0.073
N	93	93

Note: The regression includes information from 93 students who were surveyed at graduation. The dependent variable is the absolute difference between the self-reported position in the ranking and the actual one, normalized between 0 and 1. Robust standard errors in parenthesis. *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$

Table A4: Impact on Academic Performance - IV Estimates

	1	2	3	4
	Regular Exams		**Retakes**	
	Exams Passed	GPA	Exams Passed	GPA
Second year	-0.49*	-0.27***	0.26	-0.09*
	(0.24)	(0.09)	(0.17)	(0.04)
Third year	0.18	0.00	0.07	-0.06
	(0.29)	(0.06)	(0.13)	(0.04)
Fourth year	0.37	0.04	0.03	-0.04
	(0.31)	(0.07)	(0.08)	(0.05)

Note: Each cell reports the result of a different IV regression on the sample of 966 students that took part in the intervention and for whom there is information available on their predetermined characteristics. The independent variable is a dummy variable that takes value one if the student accessed the information on relative performance, instrumented by being assigned to the treatment. The first two rows provide information for the 2nd academic year, the second two rows for the 3rd academic year, and the last two rows for the fourth academic year. The first two columns report information from exams passed and GPA during the regular period (January and May). Columns (3) and (4) provide information from retakes (June). The dependent variable in columns (1) and (3) is the number of exams passed. The dependent variable in columns (2) and (4) is the GPA. All regressions include a control for academic performance during the first year and degree fixed effects. Standard errors clustered at the lecture group level in parenthesis. *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table A5: Other Outcome Variables: Grading Satisfaction and Grading Elective Courses

	1	2	3	4
	All		**Treated-Control**	
Dependent variables:	Mean	St. Dev.	Difference	St. Error
Grading satisfaction	3.15	0.82	0.12	0.12
Response rate	0.29	0.20	-0.04	0.04
Grading elective courses	0	1	0.03	0.03

Note: Columns 1 and 2 include information on 977 students that took part in the intervention. *Grading Satisfaction* reports the average satisfaction with the grading, also coded in a scale from 1 (not at all) to 5 (very satisfied). *Response rate* measures the proportion of students who participated in teaching evaluations in the second term of academic year 2010 in each group. *Grading elective courses* is a measure of the grades that students obtained in the previous two years in the elective courses selected by the students. Column 3 reports the main estimates from equation (4), and each row corresponds to a different regression where the independent variable is a dummy that takes value one if the student was part of the treatment group and the dependent variable is indicated in column 1. Column 4 reports standard errors clustered at the lecture group level. *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

B Theoretical Model

Total utility of an individual is given by three main components. First, an output driver that depends on own effort, x_i, and ability, θ_i. Second, a competition motivational component that depends on own effort, x_i, and own ability, θ_i, as well as, others' efforts, x_{-i}, and abilities, θ_{-i}. And third, the cost function of own effort:

$$\alpha F\left(x_i, \theta_i\right) + \beta G\left(x_i, \theta_i, x_{-i}, \theta_{-i}\right) - C\left(x_i\right).$$

The output driver, $F\left(x_i, \theta_i\right)$, is a function of individual's effort and ability in a complementary fashion:

$$\frac{\partial F\left(x_i, \theta_i\right)}{\partial x_i \partial \theta_i} > 0 \tag{B.1}$$

The competitive driver, $G\left(x_i, \theta_i, x_{-i}, \theta_{-i}\right)$, is a function where a higher talent θ_i or effort x_i of individual i makes it more likely that she wins the prize, while a higher talent θ_{-i} or effort x_{-i} of opponents makes her winning less likely. We assume own effort and others' effort are strategic complements in $G\left(.\right)$

$$\frac{\partial G\left(x_i, \theta_i, x_{-i}, \theta_{-i}\right)}{\partial x_i \partial x_{-i}} > 0, \tag{B.2}$$

and that marginal product of own effort x_i in the competitive motivation function is increasing in the ability of others θ_{-i}

$$\frac{\partial G\left(x_i, \theta_i, x_{-i}, \theta_{-i}\right)}{\partial x_i \partial \theta_{-i}} > 0, \tag{B.3}$$

but in terms of the competitive motivation, own effort x_i and own ability θ_i may be complements or substitutes

$$\frac{\partial G\left(x_i, \theta_i, x_{-i}, \theta_{-i}\right)}{\partial x_i \partial \theta_i} \lesseqgtr 0. \tag{B.4}$$