What you know can't hurt you (for long): A field experiment on relative performance feedback in higher education*

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

November 3, 2015

Abstract

This paper studies the effect of providing feedback to college students on their position in the distribution of grades using a randomized control experiment. This information was updated every six months during a three-year period. In the absence of treatment, students' underestimate their position in the distribution of grades. The treatment improves significantly the students' self-assessment. We find that treated students experience a significant decrease in their educational performance, as measured by their accumulated GPA and number of exams passed. This, however is short lived. Students catch up in subsequent periods, and on receiving repeated updates on their relative position, there are no further impacts. Interestingly, the provision of information improves students' self-reported satisfaction with the course, measured by survey responses taken after information is provided but before students take their exams.

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

---

*We would like to thank César Alonso, Brian Jacob, Imran Rasul, Javier Ruiz-Castillo, Natalia Zinovyeva and participants in presentations at Aalto University, LSE, CSEF, University of Essex, Université de Paris I, Norwegian School of Economics, Universidad Carlos III de Madrid, Nagoya University, Universidad de Navarra, Institute of Fiscal Studies, Warwick University, East Anglia University, Workshop on Experimental Economics at Granada, the Institute for Economic International Studies, Zurich University and Lancaster University for their useful comments. We also acknowledge the support of Isabel Gutiérrez and the administration of Universidad Carlos III for 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 ECO2012-34581). Nagore Iriberri acknowledges financial support from Ministerio de Economía y Competición (ECO2012-31626), Departamento de Educación, Política Lingüística y Cultura del Gobierno Vasco (IT869-13). All remaining errors are our own.

†Queen Mary, University of London and Centre for Economic Performance, LSE. E-mail: g.azmat@qmul.ac.uk.

‡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

Universities are a key element of the human capital accumulation process. In OECD, the average proportion of individuals aged 25-34 with tertiary education increased from 26% in 2000 up to almost 40% in 2012. A natural consequence of an increase in the proportion of university educated individuals is a reduction in the signalling value of the university degree (Spence 1973). The university market has adapted this change in various ways. For instance, the competition to get into the most selective universities has become stronger (Hoxby 2009). Moreover, university students reacted to the increase in the number of their peers by striving to perform well, such that they could "stand out from the crowd" and improve their employment opportunities (Hoxby 2009). In parallel with increased university entry, there has been increased competition between universities, given the higher propensity of students to exercise choice. Universities work hard to improve or establish a good reputation to attract the best students and, quite often, to attract funding. One way in which reputation is measured is through rankings, which are nationally or internationally well-known. An important component of university rankings is student satisfaction: National Student Survey (NSS) in the UK or American National Survey of Student Engagement (NSSE) in the US. Results from those surveys show that there is an increasing demand on universities to provide students with more feedback on their performance. Williams and Kane (2009) for example show that "assessment and feedback routinely score less well than other course-related aspects of the student experience and have done so for many years."

Despite the importance of university education and its consequences, few studies have explored the technology of student performance.[1] In this paper, we focus on the role of providing relative performance feedback on subsequent performance, and student satisfaction. Students are provided with information regarding their relative standing –namely, their decile rank– with respect to other students in their cohort. On

[1]Existing studies have focused on aspects such class size (Bedard and Kuhn 2008), peer effects (Sacerdote 2001), class attendance (Crede, Roch and Kieszczynka 2010) and teaching methods (Emerson and Taylor 2004, or Gok 2011).

completion of their studies, students may be often aware, or at least better informed, of how they compare to their classmates. However, large cohort sizes or a lack of transparency on grades information, often mean that students are unaware of their relative standings during the course of their studies, even though this information may be relevant for many of their choices within university. For instance, the information might be relevant to decide on courses or majors, how much effort to exert, the kind of employers or jobs to target, and so on. It is thus important to understand the link between the provision of information and educational outcomes.

We conduct a randomized controlled trial over four years (2009-2013) in a large Spanish university to study the effect of providing relative performance feedback information on educational outcomes. A cohort of approximately 1,000 students enrolled in various degrees were randomly assigned into treatment and control groups. Students in the control group, as per usual, receive only information on their own performance. Students in the treatment group were additionally provided with access to information on their decile position in the distribution of performance of their cohort. We follow students throughout their four year degree programs. Students undertake exams every six months, at the end of each semester. Relative performance feedback is provided to students in the treatment group for the first time at the end of the first semester of their second year of study and is updated every six months until the end of the fourth (and final) year of the degree. We analyze how the intervention affects students' awareness of their relative position, their performance and their (self-reported) effort and satisfaction.

We find that 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 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 have a lower accumulated GPA (0.05 standard deviations of the mean). They also complete, on average, 0.4 less course modules than students in the control group (0.1 standard deviations of the mean). Thus, providing students with information about their relative position in the cohort has an immediate detrimental effect on their academic performance.

An important advantage of our study is that we follow the students until the completion of their degree. This allows us to study the dynamic effects of providing feedback information, such that we are able to look beyond the immediate impact of treatment. We find that the provision of feedback information has a short-lived effect on academic performance. In particular, we show that students whose performance deteriorated in response to the feedback information catch up most of the difference in grades with respect to the control relatively fast. At the end of the academic year, students are given the opportunity to resit failed exams. We find that after this period, treated students complete the same number of course modules passed as students in the control group. Although the accumulated GPA is still lower at that time, by the time students graduate the performance –as measured by the likelihood to graduate or the average accumulated GPA at end of degree– of the treatment and control groups is statistically indistinguishable. In the third year, when students have the opportunity to elect different modules, we do not observe any difference across groups in the courses selected. For example, the degree of difficulty of the courses chosen is the same in both groups.

We further investigate the impact of the intervention on student satisfaction. Each semester –before sitting exams– students complete a survey on teaching satisfaction. In the pre-treatment period, we find that the treated and control students report similar levels of satisfaction. Interestingly, after the provision of feedback (for the first time) but before sitting exams, treated students report a *higher* level of satisfaction. In the same survey, students also report their exerted effort (hours of study per week). There is no statistically significant difference in self-reported effort between students in the treated and control groups neither before nor after the treatment. This is at odds with the *lower* actual performance when the students take exams, since one would expect some change in effort for this to take place. However, note that this is *self-reported* effort, whereas we measure *actual* performance.

The paper provides a model to help understand the mechanism that drives the effect of feedback information on performance and satisfaction. Two assumptions are necessary. The first assumption requires that initial knowledge of own ability is more precise than knowledge of others' ability. This seems reasonable in our setting since cohort sizes are large and most peers are new. The second assumption requires that concerns for relative standing are stronger than the desire to reach an absolute goal in terms of grades. This too is realistic in a university setting, where important rewards, such as internships, the possibility to study abroad and so forth, are awarded based on the basis of relative performance. Under these two assumptions, our theoretical framework predicts agents should decrease effort under certain conditions related to the students' beliefs. In particular, if students learn that their relative standing is higher than they expected, as it happens in our case, they should *decrease* their effort levels. Contrary to the commonly held belief that people are in general overconfident about their ability, it has been shown that over and underconfidence vary systematically, and there tends to be underconfidence for difficult tasks (Moore and Cain 2007).

To verify these conditions about beliefs on relative standing, we conducted several surveys among students. The surveys suggest that, prior to the intervention, students were relatively uninformed about their position in the ranking and, in general, they underestimate their ranking. Initially the average student self-reported position is 18 percentiles lower than her true position. The intervention improves significantly the information available to students in the treatment group. At graduation this gap has decreased to 6 percentiles among students in the treatment group, but is still around 11 percentiles in the control group. Given that we know that feedback information improves satisfaction, it is reasonable to think that the "good news" provided to the treated group is the cause of this improved satisfaction.

Other studies have examined the impact of feedback on relative performance in the field. In an educational setting, Azmat and Iriberri (2010) show that the performance of high school students improved notably when, due to a change in the IT of the school, the report card added information on the average grade obtained by students in the class. Similarly, Tran and Zeckhauser (2012) find that Vietnamese students increase their effort and perform better in an English course when provided with their rank position. Katreniakova (2014) conducted an experiment on the impact of feedback on relative performance in 53 Ugandan schools. The provision of feedback improves students' performance, particularly when financial or reputation rewards are also present. None of these studies elicits information on students' beliefs about their relative position. Also, they all focus, differently from us, on pre-tertiary education, and do not looked at the long-term impacts. For a workplace setting, Blanes-i-Vidal and Nossol (2011) find that workers increase their effort after they start to receive feedback on their relative performance, when their pay was related to output. Blader, Gartenberg and Prat (2015), on the other hand, show that the provision of feedback may have a negative or a positive effect on the performance of truck drivers, depending on whether they have undergone a program that was intended to build a more team-oriented environment. Finally, Barankay (2011) in a three-year randomized control trial shows that the provision of feedback has a negative effect on the performance of furniture salespeople. In his setting, unlike in ours, information is removed for the treated group, rather than provided. All salespeople received regularly feedback on relative performance prior to the trial. Then, a random (treated) group of the salespeople stops receiving this information. Again, neither of these studies control for individuals' beliefs prior to the provision of information. A possible explanation for these mixed results is that in different contexts agents may hold different priors about their relative performance or possibly different objective functions.

In a lab setting studies have investigated the role of feedback information under various conditions. Eriksson et al. (2009) in a real-effort experiment finds that while although feedback does not affect performance, it increases the mistake rate of the worst performing agent. Hannan, Krishnan and Newman (2009) and Azmat and Iriberri (2014) study the effect of feedback provision under different incentive schemes. Hannan et al. (2009) find that, while performance improves under piece-rate incentives, it decreases in a tournament setting and is unchanged under fixed-rate incentives. Similarly, Azmat and Iriberri (2014) find that performance improves under piece-rate and is unchanged under flat-rate. They also find that the provision of feedback information increases inequality in satisfaction when performance is related to pay (piece-rate) but not when it is independent of pay (flat-rate). Under flat-rate incentives, Charness, Masclet and Villeval (2013) provided with their rank in the session and Gerhards and Siemer (2014) provide information regarding to be the best performers. They find that individuals choose higher effort when this information is privately and publicly provided. As in our setting, Khunen and Tymula (2012), study the role of beliefs when providing feedback information. Under flat-rate incentives, they find that those individuals who rank lower than expected increase effort and those who rank higher than expected reduce effort, where the overall effect is positive.

Our study focuses on the provision of feedback to students in high education, on the interaction between beliefs and feedback, as well as on performance and satisfaction. Furthermore, the long running of the study, allows us to understand the consequences of repeatedly providing feedback information. The evidence suggests that providing information on relative performance to college students does not necessarily improve their performance and might even have a negative impact. More precisely, the impact of the treatment might depend crucially on students' priors and on their preferences. In the case of the college students analyzed here, learning that they were doing better than expected had on average a negative impact on their performance.

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 carried out in the field. Section 4 presents data. Section 5 shows the empirical analysis and finally, Section 6 concludes.

2 Theoretical Framework

From a theoretical perspective the impact of relative performance feedback on effort is ambiguous. Agents' reaction depends on the agents' prior beliefs about own and others' ability, the new information inferred from the feedback, as well as on the agents' inherent motivations. For example, if ability is complementary to own effort for the purpose of achieving a particular outcome, positive (negative) news about own ability will make individuals work more (less). In addition, agents might care about their relative standing, showing a ”competitive” motivation in their preferences, perhaps because corporate recruiters or graduate school admissions officers value relative on top of absolute performance. If that is the case, learning that others' ability is lower (higher) than initially thought could make agents exert a lower (higher) level of effort.

We introduce a theoretical model that includes different drivers of motivation, and where ability and effort are complements, to help interpret the possible reactions to the provision of relative performance feedback. We show that both the different motivations as well as the informativeness of the feedback relative to agents' prior beliefs, are crucial when predicting a particular direction in the change of effort.

Let the utility of an individual depend on her output, F, where output is a function of individual's effort x_i and ability θ_i in a complementary fashion, and $0 < \delta < 1$ is a constant.

$$F(x_i, \theta_i) = (\theta_i x_i)^\delta$$

Given the complementarity between x_i and θ_i, the marginal output of effort x_i is increasing in ability θ_i

$$\frac{\partial F(x_i, \theta_i)}{\partial x_i \partial \theta_i} = \delta^2 (\theta_i x_i)^{\delta - 1} > 0 \tag{1}$$

Assume further that individuals have a “competitive” motivation in their preferences, so that their utility also depends on the relative standing in the group. For example, the individuals are competing for a prize and the probability that individual i wins the price is given by the expression

$$G(x_i, \theta_i, x_{-i}, \theta_{-i}) = \left(1 - e^{-(\theta_i x_i - \theta_{-i} x_{-i})}\right) \tag{2}$$

where clearly 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 it less likely.

Note that own effort and others' effort are strategic complements in $G(.)$ since

$$\frac{\partial G(x_i, \theta_i, x_{-i}, \theta_{-i})}{\partial x_i \partial x_{-i}} = \theta_i \theta_{-i} e^{-(\theta_i x_i - \theta_{-i} x_{-i})} > 0 \tag{3}$$

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

$$\frac{\partial G(x_i, \theta_i, x_{-i}, \theta_{-i})}{\partial x_i \partial \theta_{-i}} = \theta_i x_{-i} e^{-(\theta_i x_i - \theta_{-i} x_{-i})} > 0 \tag{4}$$

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

$$\frac{\partial G(x_i, \theta_i, x_{-i}, \theta_{-i})}{\partial x_i \partial \theta_i} = (1 - \theta_i x_i) e^{-(\theta_i x_i - \theta_{-i} x_{-i})} \tag{5}$$

depends on whether $\theta_i x_i$ is smaller or bigger than 1.

Total utility is given by:

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

Relative performance feedback can be informative about own ability as well as about others' ability.

Assume first that relative performance feedback informs the decision maker that others' ability, θ_{-i} is lower than expected, and thus that they were underestimating their relative position. Then, the reaction function for effort of agent $x_i(\theta_i, x_{-i}, \theta_{-i})$ will shift down from the effect on the competitive motivation (eq. 4). And if everyone lowers their estimate of the ability of opponents, given the strategic complementarity between own effort and others efforts (from eq. 4), then the equilibrium effort x_i^* will go down for everyone.

Assume next that relative performance feedback reveals that own ability θ_i is higher than initially thought. Then the effect is more complicated. On the one hand, from the complementarity of own effort and ability in $F(.)$, the reaction function for effort should shift up (see eq. 1), but since the relationship between own ability and effort in the competitive motivation $G(.)$ could be one of substitutability, the reaction function for effort could shift down (if $\theta_i x_i > 1$, see eq. 5). Then, if the shift of the reaction function is the same for everyone (up or down depending on the relative sizes and signs of effects on $F(.)$ or $G(.)$), the strategic complementarity of own and others' efforts should shift the equilibrium choice of effort for everyone in the same direction, up or down, as the individual reaction functions. People with a high relative desire for maximizing their own output versus having a high standing within the cohort ($\alpha >> \beta$) could increase effort after learning their relative position is better than expected, whereas people with a high relative desire for standing within the cohort ($\beta >> \alpha$), and a value for $\theta_i x_i > 1$ (so that own effort x_i and own ability θ_i are substitutes in $G(.)$) could decrease effort after learning their relative position is better than expected.

The final effect therefore depends on the prior knowledge of own ability θ_i, versus the knowledge of others ability θ_{-i}. If information about θ_{-i} is the only novelty, the effect would be an unambiguous decrease in effort, provided $\beta > 0$. If information about θ_i is the novelty, the effect would be ambiguous.

This theoretical framework shows that different motivations in the utility, the expectations individuals have prior to the provision of information, and whether feedback informs about own ability or about others' ability, are important determinants of effort choices, which can lead to different reactions in effort. Note however, that the fact that the framework allows for different responses does not mean that the model does not provide guidance as to what effects we should find. Particular directions for the effect depend on particular types of information. For example, it is natural to expect that knowledge of own ability θ_i, is more precise than knowledge of others ability θ_{-i}, particularly in a university, where all peers are relatively new for most students. Therefore, the feedback will make individuals update their knowledge of others' ability than their knowledge of their own ability.

In terms of the motivations it seems likely that students have strong competitive motives. The grades in a university serve as a signal of ability to potential employers and to graduate school admissions officers. This means that although some students will have an intrinsic motivation to have better grades, it is likely that many of them will have an even stronger desire to do well with respect to 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.

3 Background and experimental design

We conducted a Randomized Control Trial over four years (2009-2013) at University Carlos III in Madrid, Spain. The university offers several four-year and six-year degrees in three different campuses. The majority of students do their degree in Spanish but a small minority do it in English. Our study involves students enrolled in the Spanish track of four of these four-year degrees - Business, Economics, Finance, Law - and one six-year degree - Business and Law.[2] Two of these degrees, Business and Business and Law, are held simultaneously in two different locations, the Northern and the Southern campuses. The study therefore involves students in seven different degree-locations.

In the control group students receive information on their own performance (as is the norm). In the treatment group, students also receive information on their relative performance. We examine how the treatment affects students' educational performance and their satisfaction. Below we explain the most relevant features of this university and the design of the experiment.

3.1 Educational Institution

In Spain access to university degrees is based on applicants' *entry grade*, which is calculated as a weighted average of their High School accumulated GPA (60%) and the grade obtained in a standardized exam known in Spanish as *Selectividad* (40%). University Carlos III offers the most selective degrees in the region according to the

[2]The choice of degrees and campuses was based on data availability and size. We did not consider degrees where there is only one lecture group.

required minimum entry grade.[3]

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 chance to resit 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 their surname initial.[4] As an illustration, Figure 1 shows how students enrolled in 2010 in the 1st year of the Business degree in the Southern campus were distributed across groups. For instance, students whose surname initial was "A" or "B" were assigned to tutorial group number 74 and lecture group "74-75-76" (which combines tutorial groups 74, 75 and 76). As we show below, 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 tend to be optional. 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/assignment (20%). The end of term exam is usually the same in different groups of the same subject.

Students' permanence in 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.[5]

Students receive regularly information on the grades that they have obtained in each subject. The university summarizes this information through an official measure

[3] 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.

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

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

of Accumulated Grade Point Average (AGPA), which students can also access at any point in time in the intranet of the university.[6] Students do not receive information about their position in the distribution of AGPAs, relative to other students, or about the AGPA of any other students.

Students are not explicitly rewarded for their relative performance, except for a prize given to the best student in the cohort.[7] Nonetheless, relative performance might be relevant. For instance, many students enroll in the Erasmus exchange program, typically during their third or fourth year. Whether students are admitted to the program or not is based on their performance in a language exam and their position in the distribution of grades. The relative position of students in the distribution of grades might also play a role when students apply for an internship, typically during the last year of the degree, or later after graduation, when they enter the labor market.

3.2 Experimental Design

The intervention was restricted to students who had entered the university in Fall 2009 and who were registered in at least one second-year course in Fall 2010. This condition excludes approximately 10% of the 2009 cohort, in general students who were expelled because they did not manage to satisfy one of the permanence requirements: passing at least two courses during the first year.

Students' were assigned randomly to the treatment or to the control group based on the lecture group in which they were enrolled in.[8] We selected randomly one of the 432 different possible assignments. The set of possible assignments was subject to the constraint that there is one treated group per degree-location. As a result of the random draw, 623 students were assigned to the treatment group and 354 to the control group. Table 1 shows the distribution of students to the control and the treatment group by

[6]The university calculates the accumulated grade point average adding up the grades obtained by the student, modified with a penalty for the number of times the 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, twice by 0.90 and so on.

[7]This prize, known as *premio extraordinario* is awarded by the ministry of education upon graduation.

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

degree and campus.

The intervention starts in early December of 2010 and it concludes three years later, at the end of the fourth academic year. During this period students in the treatment group were granted access to feedback on their relative performance every six months. More precisely, treated students received every six months an email message from a corporate account saying:

> 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 get access to a screen where they can observe their own AGPA and also their position in the distribution of grades, measured in deciles (Figure 2).

We also collected information from three different surveys: (i) teaching evaluations filled by students, which are collected by the university (ii) a survey about students' knowledge of their relative position in the distribution of grades, to a sample of 2nd year students, who were not affected by the intervention, (iii) a similar survey to a sample of graduating students belonging both to the treatment and the control groups. On the one hand, teaching evaluations will be useful as they provide measures of student satisfaction as well as their self-reported effort. On the other hand, the surveys on students' knowledge of their relative performance will be useful to measure students' prior knowledge on their relative standing, both prior to the treatment as well as after the treatment.

Note that students receive information about their position in the ranking in terms of their AGPA. Given that by construction the influence of each additional course on their ranking decreases overtime, students' position in the ranking varies increasingly less over time. As shown in Figure 6, 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% of students experience any such variation.

4 Baseline characteristics and balance check

4.1 Individual characteristics

Table 2 provides information on the individual predetermined characteristics of the 977 students who participated in the intervention. A little over half of the students are women, and practically all of them are Spanish. In general they attended previously a High School, only 5% have a vocational training background. Around two thirds of the students come from the Madrid region and, within this region, most of them come from the center of Madrid (31%). Approximately 22% come from municipalities located in the Southern part of the region, an area which tends to be the less affluent.

Students experience a significant decrease in their grades during their first year in university relative to the grades that they used to obtain in secondary education. While the average entry grade into the university is 7.24 (out of 10), the average AGPA at the end of the first year is equal to 6.02, which implies a decrease of roughly one standard deviation. As shown in Figure 3, grades shift down along the whole distribution.

In relative terms, the average student in our sample is placed in percentile 54, relative to all students who also registered in the same degree the previous year. This figure is slightly higher that 50, reflecting that approximately 10% of first year students failed to satisfy the permanence requirements.

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

$$X_{s,d,g} = \alpha + \beta Treatment_{d,g} + \boldsymbol{Z_d \lambda} + \epsilon_{s,d,g} \tag{6}$$

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. Out of 14 observable characteristics, in no dimension the difference is significant at the 5% and in two dimensions the difference is significant at the 10% (Table 2, columns 4). An F-test confirms that it is not possible to statistically reject that the assignment was random.

4.2 Teaching evaluations

Students were relatively satisfied with the quality of the courses they receive before the intervention took place (Table 3, upper panel).[9] In 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 and 5, the average answer is 3.6. The teaching evaluations also provide (self-reported) information on study time. Students devote each week roughly between 4 and 7 study hours to each subject.[10] Taking into account that there are typically 5 or 6 courses per term, this implies that on average students spend each week approximately 32 hours studying, which combined with class attendance, implies that the average student devotes around 50 hours a week to college related work.[11]

We verify whether the treatment and the control group were similar in these dimensions before the intervention took place using the following regression:[12]

[9]Teaching evaluations are collected by the University administration twice a year. During the academic year 2010-2011, students completed their 1st term teaching evaluations before the intervention took place, in late november, and they completed their 2nd term teaching evaluations after they had received feedback on their relative performance, but before they had received the results of the exams of the second term.

[10]This information is only available at the group level. 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.

[11]According to survey information provided by teachers, the attendance rate to lectures is around 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 sitting in class.

[12]Teaching evaluations are anonymous, so we cannot match the teaching evaluations to the students in our sample. But given that we know the tutorial group the teaching evaluations belong to, we can assign teaching evaluations to the treatment and the control group based on the tutorial group during the academic year 2010-2011, when students are registered in compulsory 2nd year courses.

$$Y_{c,g,d} = \alpha + \beta Treatment_{c,g,d} + \boldsymbol{X_c}\boldsymbol{\gamma} + \boldsymbol{Z_d}\boldsymbol{\lambda} + \epsilon_{c,g,d} \tag{7}$$

where $Y_{c,g,d}$ stands for 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 in the Southern Campus). The regression includes a set of course fixed effects ($\boldsymbol{X_c}$) and degree fixed effects ($\boldsymbol{Z_c}$).

As shown in columns 3 and 4, students in the treatment and the control group report before the intervention very similar values in terms of their overall satisfaction with courses, the fairness of the grading and the hours of study.

4.3 Students' prior information on relative performance

The intervention provides treated students information on their position in the distribution of grades 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 distribution of grades, absent of any intervention, using information from another cohort of students. We conducted a survey among a group of students of 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.[13] We decided not to conduct this survey among students belonging to the treated cohort (2009 cohort) in order to avoid the introduction of any confounding effects that might perhaps affect their performance later on.

Students were asked to answer privately the following question:[14]

Unfortunately we cannot match the information during the third and the fourth academic years, when most courses are elective.

[13]More precisely, we surveyed students enrolled in Game Theory, Degree in Economics, groups 63, 64, 68, 69. 21 people did not attend the lecture the day of the survey. All attending students except one participated in the survey.

[14]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 in its Southern Campus

> When you enrolled one year ago in this degree 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 4. 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 below the diagonal, reflecting that students tend to be uninformed. Moreover, students underestimate their position in the distribution of grades. The average student makes an error in her prediction of 22 percentiles and she tends to underestimate her relative ranking by 18 percentiles. One possible explanation for this systematic divergence is that students may have not realized that the sharp decline in grades that they experience during their first year at university affects all students, and not only themselves.

To get a better understanding of which students underestimate their position in the distribution and which ones overestimate it, we estimate the following equation:

$$Y_s = \alpha + \boldsymbol{X_s}\beta + \epsilon_s, \tag{8}$$

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 4, underestimation is relatively stronger among women, among students with low High School grades, and among students who during their first in university managed to receive relatively higher grades. These observable characteristics explain around 50%

of the variation in the gap between students' self-reported ranking and their actual position. Overall, this analysis shows that there is room for students learning about their relative ranking and that the provision of feedback should indeed change students' underestimation.

5 Empirical analysis

We analyze the impact of the intervention in different steps. First, we verify whether treated students actually access the link that they received by email. Second, we examine whether the intervention managed to have a long-lasting differential impact on the information available to students in the treatment and the control groups. Third, we study the impact on students' performance. Fourth, we study the effect on students' satisfaction. Finally, we discuss some robustness checks.

5.1 Do students access the information?

The treatment consists of the possibility to find out 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 got access to the information as well as the number of times they accessed to the information. 72% of students checked this information at least once. The average student checked four times the ranking during the duration of the treatment. As shown in Figure 5, the probability to check is strongly correlated with the position in the ranking. In the top quartile almost 90% of students accessed the information, in the bottom quartile less than half did. Female students are also slightly more likely to check, but the difference only marginally significant once we take ranking into account (Table 5).

Unfortunately we cannot disentangle why some students do not check the information. Some individuals might not read emails from corporate accounts, some others perhaps read the email but prefer not to find out about their position in the ranking. One third of students that did not check their ranking were expelled from the university at the end of their second year due to the unfulfillment of the permanence requirements. It is possible that these students were not active students at the time of the intervention.

5.2 Learning and information spillovers

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. Students in both groups might also increase over time their knowledge about their position in the distribution, independently of the intervention.

To study this issue, we surveyed a sample of students from the treatment and the control groups three years after the intervention about their relative ranking. The survey was conducted at end of the undergraduate thesis presentation, which is the last requirement that students satisfy before graduation.[15] The sample includes 97 students from Economics, Business and Finance degrees. Four students did not reply to the survey. By construction the sample of students who was surveyed is not a random sample of all students. Students in the upper part of the grade distribution are over-represented.

The information displayed in Figure 7 reveals two interesting patterns. First, compared to students at the beginning of their 2nd year, at the end of their 4th year students have more accurate information about their relative performance. The average error has decreased from 22 percentiles to 12 percentiles. Second, students in the treatment group are significantly better informed than students in the control group. The average error is equal to 9 percentiles among students in the treatment group and equal to 15 percentiles among students in the control group (Table 6).

For students in the control group, this improvement might potentially reflect learning over time or potential information spillovers. Unfortunately we cannot disentangle these two hypothesis. Note also that students in the treatment group do not per-

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

fectly 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. Last, some students may have forgotten their position in the ranking. Overall, we find that students, even in the absence of any intervention, improve their knowledge about their relative ranking. However, more importantly, we see that the intervention indeed made a differential change in students' knowledge about their ranking, decreasing the gap between their expected and true position in the ranking.

5.3 Feedback effect on academic performance

We estimate the impact of feedback on academic performance. We compare the performance of all individuals in the treatment and the control groups (*intention-to-treat effect*) and we also report estimates from an instrumental variables (IV) estimation, where we instrument access to the feedback information using the random assignment to the treatment group.

5.3.1 Intention-to-treat effect

Table 7 provides information on students' academic performance during the three years that lasted the intervention. The intervention took place in Fall of the second year. During the regular exam season of their second year students take on average eleven exams and they pass approximately eight. Students have the chance in June to resit exams that they had failed. During the second year resit season, students on average take around three exams and pass one of them. The number of exams taken and passed during the third and the fourth year is slightly lower. By September of their fourth year approximately half of the students in our sample have managed to graduate and

15% had dropped out, typically during their second year.[16]

We compare the performance of the treatment and the control group using the following regression:

$$Y_{s,d,g,t+i} = \alpha + \beta Treatment_{d,g} + \boldsymbol{Z_d \lambda} + \epsilon_{s,d,g,t+i}, \tag{9}$$

where $Y_{s,d,g,t+i}$ stands for 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 is exposed to the treatment, and the equation also includes a set of degree fixed effects ($\boldsymbol{Z_d}$).To account for potential existence of common shocks, we report standard errors clustered at the tutorial group level (45 groups). In columns 3 and 4 of Table 7 we report the estimates from equation (9), and in columns 5 and 6 we report results from a specification which also includes the set of predetermined individual characteristics $\boldsymbol{X_{s,d,g,t}}$ listed in subsection 4.1. As expected, the latter estimates are statistically similar but they are slightly more precise.

We do not observe any impact on the number of exams taken by students during the regular exam period that year. On the other hand, the performance of the treatment group is significantly worse. On average, students in the treatment group passed 0.36 (9% of a standard deviation) fewer exams during the regular exam period, a difference which is significant at the 5%. Rows 3 and 4 provide information about resits, which are scheduled in June. Students in the treatment group take 0.34 more resits, reflecting their higher failure rate during the year, and they manage to recover half of the gap. During the third and the fourth years there are no significant differences in performance between the treatment and the control group. If anything the performance of the treatment group is slightly better and, by the end of the fourth year, there are no significant difference between students in the treatment or the control group in terms of the number of exams passed, the dropout rate, time to graduation or the accumulated grade point average. In sum, the treatment group experiences a short-term negative

[16]This calculation excludes 200 students who were enrolled in the Business and Law degree, which has a six-years length.

impact on performance but in the longer term the gap disappears.

5.3.2 Instrumental variables

Not all students in the treatment group accessed the information (Table 5). We also conduct the analysis using an instrumental variable (IV) strategy where we use the (random) assignment to the treatment group as an instrument for accessing the information. The point estimates from the IV exercise are slightly larger but overall the results are statistically similar (Table 8).

The interpretation of these IV estimates depends on the mechanism that explains why some students in the treatment group did not access the ranking information. On the one hand, if those who did not access the information is because they did not receive or did not read the emails, the IV estimates provide information on the average treatment effect on the treated. On the other hand, some students may have read the email but they may have preferred not to obtain information on their relative performance. In this case, the treatment may have affected them even if they did not access the information, and the IV estimates would not have a straightforward interpretation.

5.3.3 Effort

The treatment affected negatively students' performance during the second year. In principle, this should reflect a decrease in their effort. However, we do not observe any significant impact on students' self-reported effort. We run equation (7) using as a left-hand side students' self-reported effort. As shown in the lower panel of Table 3, both treated and control groups tend to report that they study between three and seven hours weekly per course. One possible explanation for this puzzling result is that perhaps the treatment was not strong enough to move students' effort beyond these boundaries.

5.3.4 Heterogeneity analysis

Are all students equally affected by the provision of information on relative performance? We consider several sources of heterogeneity.

First, we consider the type of information that students have received. We do not have direct information on the priors of students that participated in the intervention, but we can try to infer whether a given student was positively or negatively surprised by the feedback on relative performance exploiting 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:

$$Y_s^{self-reported} = \alpha + \beta Y_s^{true} + \boldsymbol{X_s}\gamma + \epsilon_s \tag{10}$$

where $\boldsymbol{X_s}$ includes students' actual ranking, gender and entry grade. We use these estimates to predict the type of news that students are expected to receive when they get access to the ranking information (see Table A1). We classify students in 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*), *et 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 8 shows the distribution of these three groups according to the actual relative performance of students.

We regress equation (9) separately for these three groups of students, using as dependent variable the number of exams passed during the second year in the regular exam period. According to our estimates, students who, according to our estimations, receive 'positive' news, pass 0.47 fewer exams during their second year, relative to comparable students in the control group. The treatment has virtually no effect on students who are expected to have correct priors about their position in the ranking. On the other hand, students receiving 'negative' news pass 0.26 more exams during their second year, although this effect is not statistically significant (Table 9, columns 2-4).

Overall, these estimates are consistent with the hypothesis that impact of information depends crucially on the students' priors.

Second, in columns 5-10 we examine the impact of the treatment according to the gender of students, their grades in High School and their performance during their first year in university. The impact is slightly larger in the case of women and students with low high school grades. We do not find any differential effect according to whether individuals are above or below the median during their first year in university.

5.4 Satisfaction

The provision of feedback on relative performance has a short-term negative impact on the performance of students. This effect is driven by students who, according to their observable predetermined characteristics, are expected to receive positive news about their position in the distribution. To get a better understanding of the underlying mechanism, we investigate how the treatment affects students' satisfaction.

We cannot observe students' satisfaction at the individual level, but we can exploit the information provided by teaching evaluations. The satisfaction of the treated group is significantly larger than the satisfaction of the control group (approximately one third of a standard deviation), suggesting that students' satisfaction increases when they learn that their relative performance is substantially better than expected (see lower panel of Table 3).

5.5 Robustness checks

We consider two alternative ways in which the treatment may have affected students' performance.

5.5.1 Grading standards

An alternative way in which grades can change is through changes in teachers' grading standards. In Carlos III university teachers do not explicitly grade on a curve but, nonetheless, we cannot discard that the performance of students somehow affects grading standards. For instance, some teachers may unconsciously become more benevolent in their grading if they realize that the overall performance of a certain group of students is relatively lower. This would introduce an attenuation bias in our results. To investigate this issue, we compare the information provided by students through the teaching evaluations. After the intervention both groups report statistically similar values in term of fairness of grading, indicating that students did not perceive any changes in grading practices (Table 3, lower panel).

A related problem would arise if the performance of the treatment groups affects the grading in the control groups, leading 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.

5.5.2 Choice of electives

During the third and the fourth year, students can choose elective courses. A potential way to improve the relative position in the ranking would be to choose elective courses where grades tend to be higher. Students may enroll in courses with high grading standards or in courses where the value added provided is higher, leading also to higher grades.

Ir order to obtain a proxy of 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:

$$Grade_{c,s} = \alpha + \boldsymbol{C_c}\boldsymbol{\beta} + \boldsymbol{S_s}\boldsymbol{\gamma} + \epsilon_{c,s}, \tag{11}$$

where $Grade_{c,s}$ reflects the grade obtained by student s in course c, and $\boldsymbol{C_c}$ and $\boldsymbol{S_s}$ are respectively two vectors of course and individual dummies. 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 to the elective courses chosen by students in the treatment and the control groups, and we normalized this variable to have mean equal to zero and standard deviation equal to one. We compare the choices of students in the treatment and the control groups using (9). Students of the treatment group tend to select elective courses with slightly higher grades (0.03 standard deviations), but the difference is not statistically significant (last row in Table 7).

6 Conclusions

In this paper we study the role of relative performance feedback information 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 about 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 assign some students into a treatment where they were given access to information about their relative position in the distribution of grades. The treatment was effective in informing students about 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 in terms of numbers of exams passed and AGPA. After a short period, however, the treated students catch up in terms of their performance. Moreover, by regularly providing access to this information to the treatment group over the course of their studies, there is no further impact on their performance. As well as an effect on academic performance, we found a positive effect on self-reported student satisfaction with the quality of the courses. This was perhaps a response to the positive surprise about their own ranking. Our results suggest that the impact of relative performance feedback may depend crucially on individuals prior information and their preferences.

Our study highlights a number of important aspects about providing students with feedback, and raises a number of interesting questions that are relevant to policymakers and education authorities. First, the timing of the information is relevant. We showed that the impact of the treatment is confined to the first time 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, which also interacts with the length of the course. If for example the course lasts three or four years, like an undergraduate degree, the optimal timing might be different than the one for an MSc lasting just one year. Second, the reference group might matter. The students in our study compare themselves to the cohort to which they belong, and thus the students' reference group is unchanged over time. This might be one reason for why there is a lack of response to feedback beyond the first time they receive information. If the reference group changed, say because at a certain point in time students specialize, or declare majors, then the information may once again have an impact. Third, the coarseness of the information provided may play a role. We provided very detailed information, in particular, students learnt about the decile to which they belonged. If students were only informed about whether they were above or below an average student, or if they were given the exact percentile, the response might be different. Again, there may an optimal information partition to provide (Diamond 1985). Fourth, the incentives associated with the relative performance could change the response to the information. In our setting there was no explicit reward within the university for ranking high versus low, in other words, there was no immediate competition for a better position. Finally, whether information feedback is provided privately or publicly may have significant impact. In our case, it was provided privately. If the ranking was made public, there may be some consequences because of status seeking, even the absence of explicit rewards within the university.

References

[1] 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.

[2] Azmat, Ghazala and Nagore Iriberri (2014), "The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction," *Journal of Economics and Management Strategy*, forthcoming.

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

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

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

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

[7] Charness, G., Masclet, D., and M.C. Villeval. (2013), "The Dark Side of Competition for Status," *Management Science* 60(1), 38–55.

[8] 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* 80, 272-295.

[9] Diamond, Douglas W. (1985), "Optimal release of information by firms," The Journal of Finance 40, 1071-1094.

[10] Emerson, Tisha LN, and Beck A. Taylor (2004), "Comparing student achievement across experimental and lecture-oriented sections of a principles of microeconomics course," *Southern Economic Journal* (2004), 672-693.

[11] Eriksson, T., Poulsen, A., Villeval, M., (2009), "Feedback and incentives: experimental evidence," *Labour Economics* 16, 679–688.

[12] Gerhards, L. and N. Siemery, (2014), "Private versus public feedback: The incentive effects of symbolic awards," mimeo.

[13] 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* 6 (2011), 3740-3746.

[14] Hannan, R.L., Krishnan, R., Newman, D., (2008), "The effects of disseminating relative performance feedback in tournament versus individual performance compensation plans," *The Accounting Review* pp.83–4.

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

[16] Katreniakova, Dagmara (2014), "Social Comparison, Rewards and Incentives to Learn: A randomized control trial in Uganda," CERGE-EI, mimeo.

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

[18] Kuziemko, I., R. Buell, T. Reich, and M. Norton (2013), "Last-place Aversion: Evidence and Redistributive Implications," *Quarterly Journal of Economics* (forthcoming).

[19] 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* 103, 197-213.

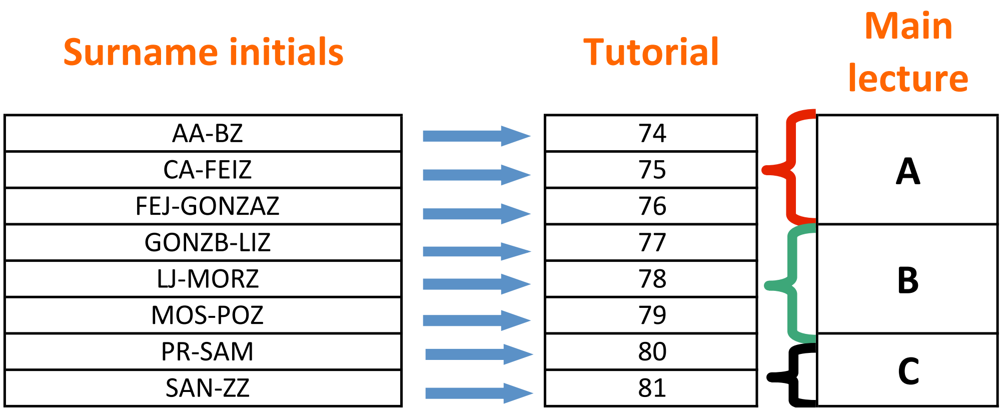
[20] Sacerdote, Bruce (2001), "Peer Effects with Random Assignment: Results for Dartmouth Roommates," *The Quarterly journal of economics*, 116(2), 681-704.

[21] Spence, Michael (1973), "Job market signaling," *The quarterly journal of Economics* 87, 355-374.

[22] 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.

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

Figure 1: Assignment to Tutorial and Lecture Groups



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

Figure 2: Feedback on Relative Performance

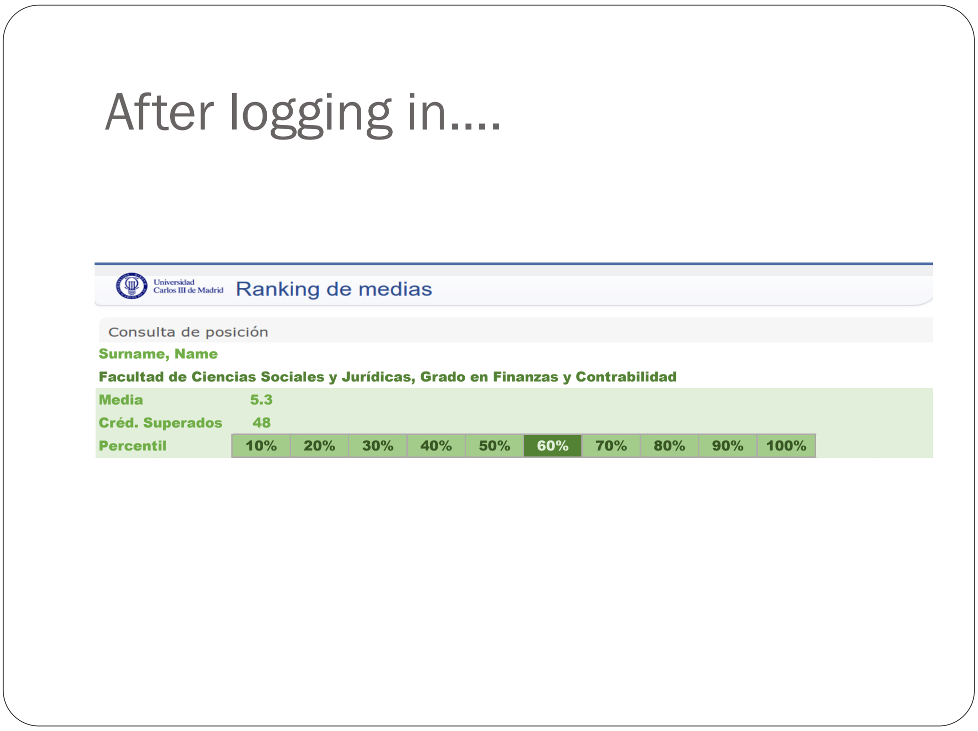


Figure 3: Entry grade and 1st year grades at college

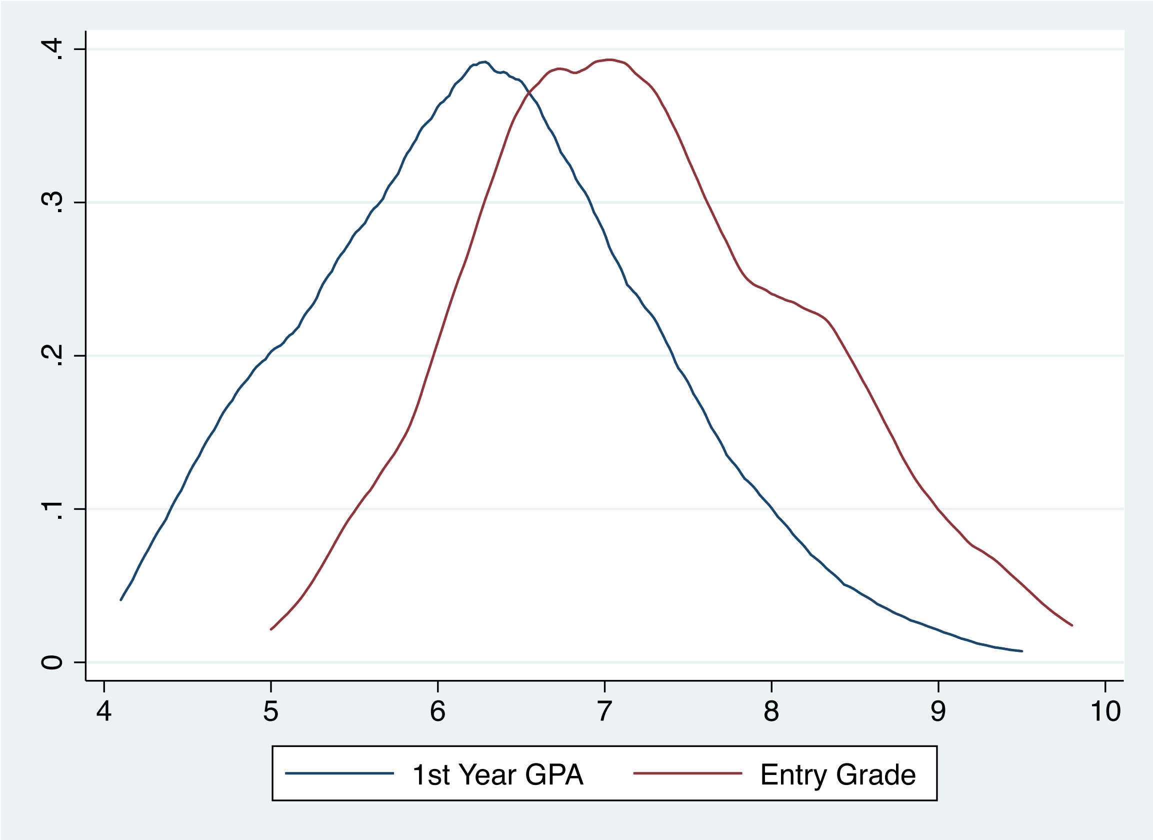
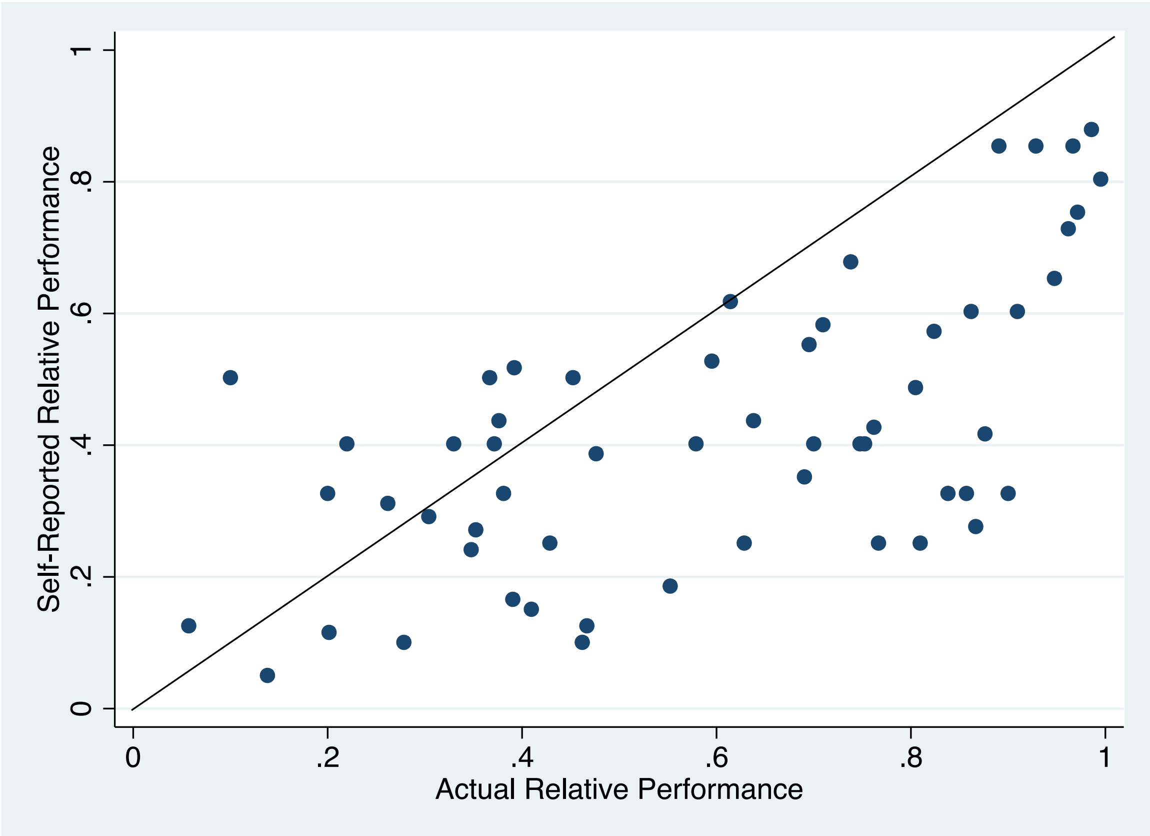


Figure 4: Relative performance at the beginning of the 2nd year



Note: The figure includes information from 57 second year Economics students, class of 2014, who were surveyed in November 2011. 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 5: Share of individuals who checks the ranking, by quartile

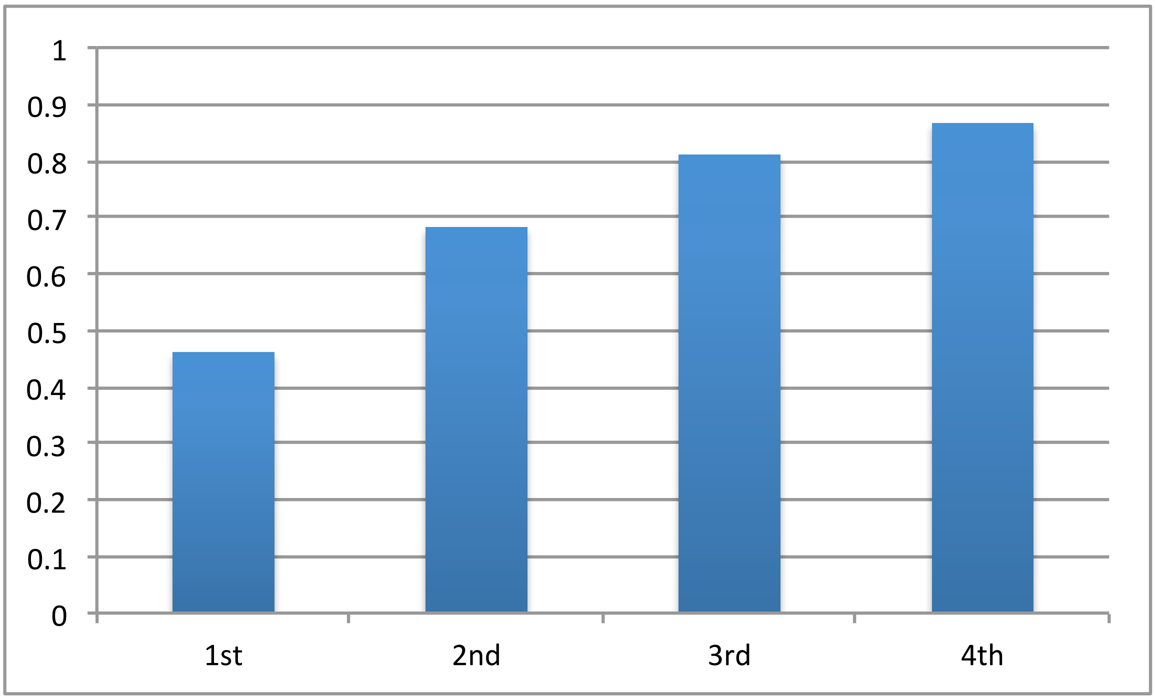
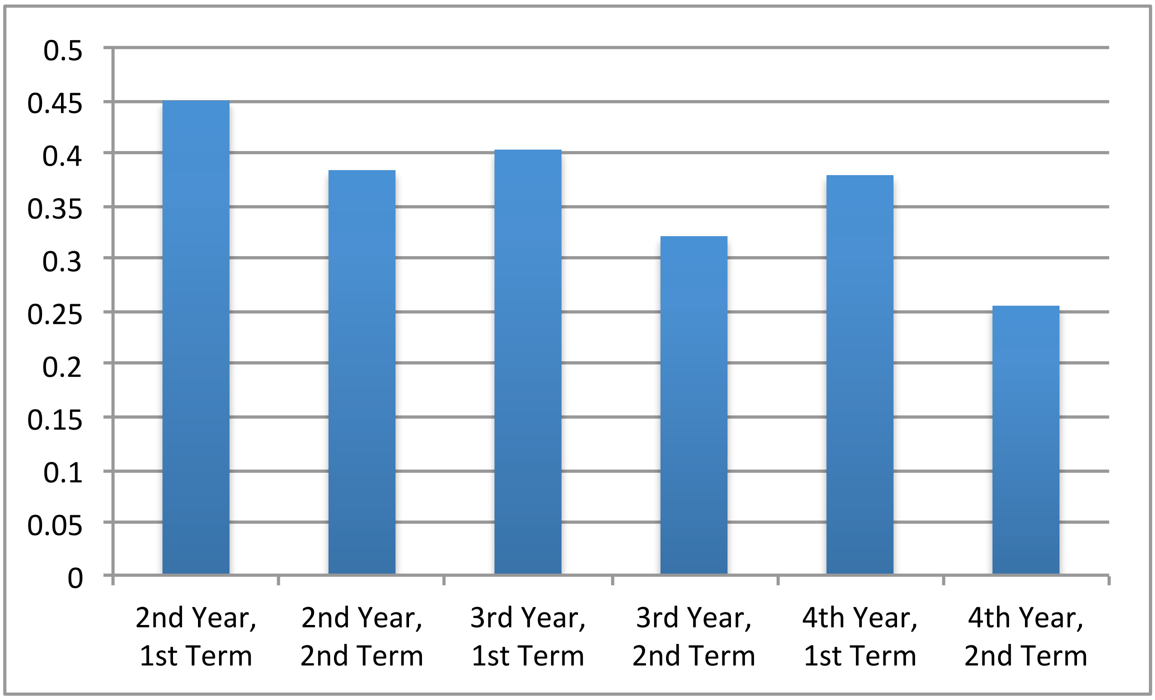
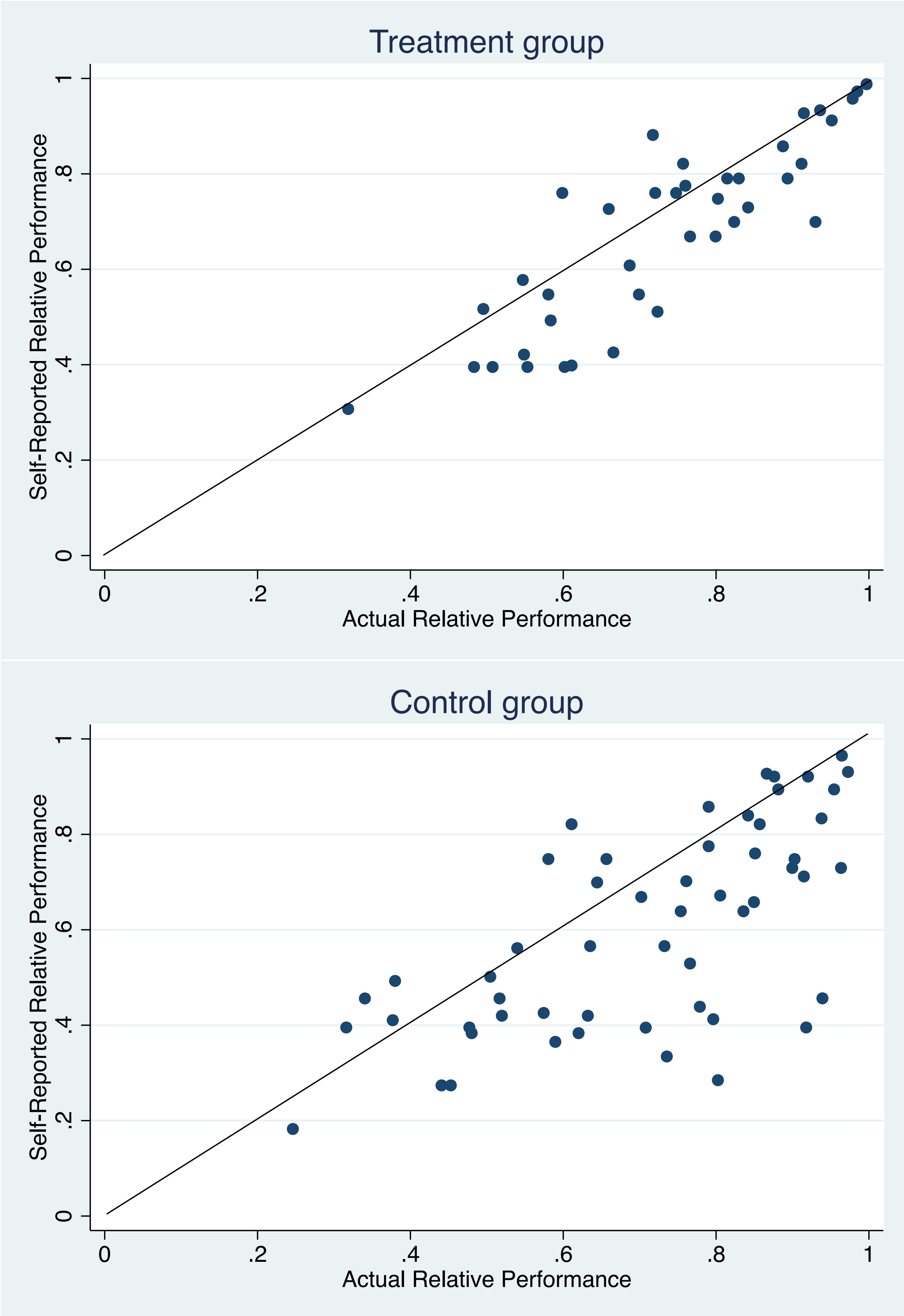


Figure 6: Information over time



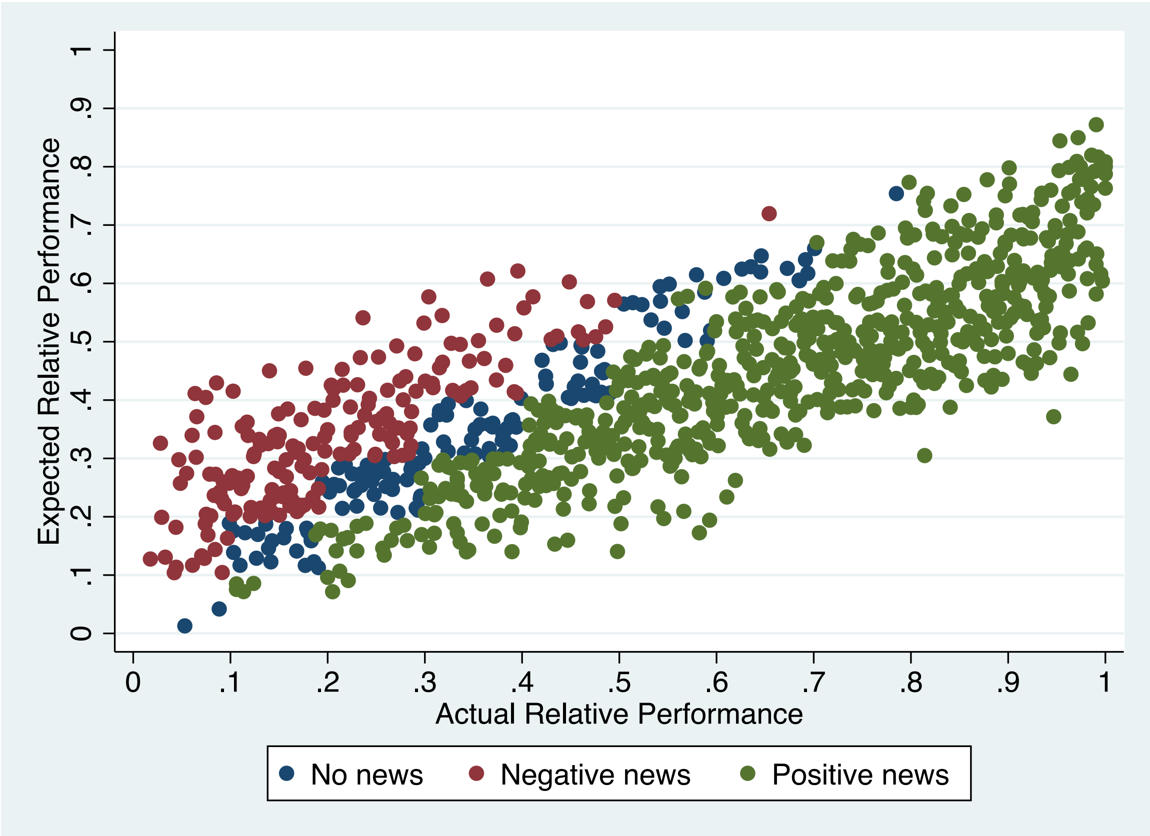
Note: Each bar reflects the proportion of people who experienced mobility from one term to the next in terms of their decile in relative distribution. For instance, approximately 45% of individuals were placed in a different decile at the end of the 1st term of their 2nd year relative to their position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

Figure 7: Relative performance at graduation, treatment group



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 8: Expected & 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 includes individuals who expect a higher ranking than their actual one, the blue group includes individuals with accurate expectations, and the green group includes individuals who are expected to underestimate their relative ranking.

Table 1: Assignment to the 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 2: Predetermined descriptive statistics, individual level

	1	2	3	4
	All		**Treated-Control**	
	Mean	St. Dev.	Difference	p-value
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
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 1st year at university:				
Accumulated GPA	6.02	1.36	-0.05	0.45
Percentile	0.54	0.27	-0.02	0.40
Exams taken	4.89	0.78	-0.06	0.23
Exams passed	3.70	1.49	-0.09	0.64
Retakes taken	2.12	2.37	0.14	0.39
Retakes passed	0.80	0.97	0.01	0.85

Note: The table includes 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 (3) reports the p-value of this difference. *Accumulated GPA* and *Percentile* are measured at the end of the first year. *Exams taken* and *Exams passed* provide information for the second term of the first year.

Table 3: Teaching evaluations

	1	2	3	4
	All		**Treated-Control**	
	Mean	St. Dev.	Difference	p-value
		Before the intervention		
Satisfaction	3.87	0.76	0.01	0.89
Hours of study	2.92	0.45	0.13	0.13
Grading	3.56	0.67	0.00	0.99
		After the intervention		
Satisfaction	3.63	0.85	0.30***	0.01
Hours of study	3.00	0.48	0.15	0.13
Grading	3.15	0.82	0.12	0.34

Note: The upper panel includes information from 182 tutorial groups who completed their teaching evaluations in Fall of academic year 2010-2011, before the intervention took place. The lower panel provides information from 165 tutorial groups who completed their teaching evaluations in Spring of academic year 2010-2011, after the beginning of the intervention. In each panel, the first row 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). The second row reports the average satisfaction with the grading, also coded in a scale from 1 (not at all) to 5 (very satisfied). The third row 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.

Table 4: Prior information on relative rank - 2nd year survey

Dep. var.: Self-reported - Actual Rank	1	2
Female	-0.09*	-0.07
	(0.05)	(0.04)
True rank		-0.66***
		(0.11)
Entry grade		0.11***
		(0.03)
Constant	-0.12***	-0.49**
	(0.03)	(0.21)
Adj. R-squared	0.04	0.50
N	57	52

Table 5: Who checks the information?

	1	2
Female	0.106**	0.079*
	[0.047]	(0.045)
True rank		0.585***
		(0.097)
Entry grade		-0.047
		(0.034)
Constant	0.665***	0.708***
	[0.034]	(0.229)
Observations	354	347
R-squared	0.084	0.161

Note: The regression includes information from 354 students who were assigned to the treatment group. The dependent variable is a dummy that takes value one if the students checked at least once the information. Robust standard errors in parenthesis. *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$

Table 6: Available information at graduation

	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 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 7: Impact on academic performance - Intention-to-treat effect

	1	2	3	4	5	6
	All		**Treated-Control**			
			Without controls		**With controls**	
Dependent variable:	Mean	St. Dev.	Difference	St. Error	Difference	St. Error
			Second year			
Exams taken	10.69	3.19	-0.06	(0.17)	-0.05	(0.14)
Exams passed	7.75	3.83	-0.50**	(0.21)	-0.36**	(0.18)
Retakes taken	2.91	2.94	0.47*	(0.27)	0.34	(0.22)
Retakes passed	1.12	1.25	0.23*	(0.12)	0.19*	(0.11)
			Third year			
Exams taken	10.26	4.52	0.12	(0.31)	0.25	(0.27)
Exams passed	8.07	4.06	-0.03	(0.27)	0.13	(0.24)
Retakes taken	2.15	2.65	0.10	(0.16)	0.06	(0.17)
Retakes passed	0.98	1.28	0.07	(0.08)	0.05	(0.09)
			Fourth year			
Exams taken	8.59	4.68	0.06	(0.36)	0.16	(0.32)
Exams passed	6.69	4.41	0.22	(0.33)	0.27	(0.31)
Retakes taken	1.22	2.09	-0.16	(0.17)	-0.17	(0.18)
Retakes passed	0.68	1.11	0.02	(0.07)	0.02	(0.07)
			Overall			
All exams taken	36.46	13.91	0.41	(1.04)	0.49	(0.94)
All exams passed	25.82	11.54	0.01	(0.71)	0.32	(0.70)
Dropout	0.15	0.36	0.00	(0.03)	-0.01	(0.02)
Graduation in 4 years	0.51	0.5	-0.01	(0.03)	0.02	(0.03)
Final AGPA	6.30	1.27	-0.07	(0.10)	-0.03	(0.06)
Grading elective courses	0	1	0.03	(0.03)	0.03	(0.03)

Note: Columns 1 and 2 include information on 977 students that took part in the intervention, except for the variable *graduation rate*, which excludes 200 students enrolled in the six-years degree in Business and Law. The variables *Exams taken* and *Exams passed* refer respectively to the number of exams taken or passed during the regular exam season (January and May). Variables *Retakes taken* and *Retakes passed* refer exams taken and passed during the retake season (June). The lower panel provides information measured at the end of the fourth academic year. *AGPA* refers to the Accumulated Grade Point Average. *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 (9), 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. For instance, the first cell in column 3 indicated that treated students enrolled in 0.06 fewer courses that comparable students in the same degree. In columns 5 regressions also include controls for a set of individual predetermined characteristics. Columns 4 and 6 report standard errors clustered at the tutorial level in parenthesis. *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table 8: Impact on academic performance - IV estimates

	1	2	3	4
	Regular exams		**Retakes**	
	Taken	Passed	Taken	Passed
Second year	-0.074	-0.493**	0.465	0.255*
	(0.186)	(0.244)	(0.301)	(0.150)
Third year	0.347	0.183	0.082	0.066
	(0.374)	(0.331)	(0.235)	(0.120)
Fourth year	0.216	0.373	-0.239	0.027
	(0.433)	(0.423)	(0.243)	(0.089)

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 taken 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 taken. The dependent variable in columns (2) and (4) is the number of exams passes. All regressions include a control for academic performance during the first year and degree fixed effects. Standard errors clustered at the tutorial level in parenthesis. *: $p < 0.10$, ** $p < 0.05$,*** $p < 0.01$.

Table 9: Heterogeneity analysis

	1	2	3	4	5	6	7	8	9	10	
		News			**Gender**		**1st year grades**			**HS grades**	
Sample:	All	Positive	No News	Negative	Female	Male	Low	High		Low	High
Treatment	-0.361**	-0.470**	0.164	0.257	-0.464*	-0.179	-0.331	-0.183		-0.604**	-0.128
	(0.176)	(0.216)	(0.483)	(0.497)	(0.263)	(0.247)	(0.286)	(0.199)		(0.272)	(0.243)
Adj. R-squared	0.634	0.687	0.613	0.588	0.613	0.640	0.541	0.709		0.569	0.640
N	966	644	142	180	521	445	435	531		482	479

Note: The dependent variable is the number of exams passed 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 tutorial level.

Appendix A: Tables

Table A1: Expected rank - 2nd year survey

Dep. var.: Self-reported rank	1
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