Gender Differences in Response to Relative Performance Feedback: A Field Experiment in Education

José María Cabrera and Alejandro Cid *

July 16, 2024

Abstract

Individuals are influenced by both their absolute performance and their performance relative to others. For example, workers' satisfaction is affected not only by their nominal wage but also by how their salaries compare to those of their colleagues. We apply these ideas in the context of education. We analyze the effect of delivering relative performance feedback in a field experiment involving more than a thousand university students. We first find that untreated students tend to misperceive their standing in the grade distribution, with underperforming students often overstating their ranking and high-achieving students, particularly women, understating their performance. We experimentally provided treated students with information about their exact performance relative to peers.

We find asymmetric effects of information feedback on men and women. Treated men reported increased satisfaction with their GPA, while treated women reported reduced satisfaction, regardless of their position in the grade distribution. Additionally, the non-monetary incentive caused a decline in women's academic performance after one and two years. Two potential explanatory channels could account for these findings: women may exhibit a tendency to shy away from competition, and they might face an increasing marginal cost of effort. This paper highlights the nuanced impact of information feedback, showing that more information is not always universally beneficial.

*Cabrera: Universidad de Montevideo (jmcabrera@um.edu.uy); Cid: Universidad de Montevideo, (acid@um.edu.uy). We thank seminar participants at Universidad de Montevideo, Erasmus School of Economics, Universidad de Vigo, 2017 Meeting of the Economics of Education Association and Encuentro Sociedad de Economistas del Uruguay 2016, for helpful comments.

1 Introduction

Researchers in Social Sciences have long been interested in the possibility that individuals care about both their absolute performance and their performance relative to others. This concern surfaces in various contexts such as salary comparisons in the labor market (Card et al., 2012) and grade rankings in education (Carneiro et al., 2023; Denning et al., 2023; Fischer and Wagner, 2023; Brade et al., 2022; Azmat et al., 2019; Bursztyn and Jensen, 2015; Tran and Zeckhauser, 2012).

Our study delves into the impact of introducing academic ranking feedback for undergraduate students in a university setting through a randomized controlled trial. This project was inspired by the work of Card, Mas, Moretti, and Saez (2012), who studied the effect of disclosing information about peers' salaries on workers' job satisfaction at the University of California. They found an asymmetric response to the information, with workers with salaries below the median reporting a lower satisfaction, while high-earners showed no impact. We will look at the effect of privately disclosing one's placement in the grade distribution (an ordinal ranking) on short-term satisfaction with academic performance and medium to longer-term academic results.

The project's design and system setup commenced in mid-2013, with the field experiment launching in July 2014. A total of 1,048 undergraduates at the Universidad de Montevideo (Uruguay) were randomly assigned to a treated or a control group. We examine the impact of information about relative position in the distribution of grades, exploring short-term satisfaction with GPA and, in the medium to longer run (one and two years), objective academic performance. Given gender differences in response to competition, we study the effects by gender.

We first find that, despite students having access to precise information about their GPA through transcripts, there are inaccuracies in their awareness of their *relative* performance. Remarkably, the misplacement of students in the ranking is asymmetrical: those in the lower part of the grade distribution tend to perceive their academic performance as higher than their actual performance (reminiscent of the Dunning-Kruger effect, 1999). Conversely, students in the upper part of the grade distribution, notably high achievers, tend to under place themselves: they say they perform worse than they actually do. The treatment significantly improves the information reducing the difference between actual and perceived positions in the ranking. To illustrate, the probability of a treated student accurately reporting their exact ranking increases fivefold compared to control students.

The misplacement in the ranking may be attributed to a regression to the mean effect, and to students selectively forming groups and then extrapolating the entire distribution of grades based only on the observed performance of their close peers.

We test this hypothesis using information on the network of friends. We find that peer groups are not formed randomly. Good students tend to be friends with other good students. Consequently, when constructing their perceived ranking with the information at hand, they will likely think that the cohort is better performing than it really is. As a result, they will under-place in the ranking given the information they receive from their peers. A same argument applies to underperforming students believing that they are better than they really are.

Subsequently, we proceed to examine satisfaction with GPA through a short-term survey administered two weeks after the treatment. This survey incorporated two measures of satisfaction, one of which utilized anchoring vignettes to enhance intersubject comparisons. Our analysis reveals asymmetric gender responses in satisfaction following the receipt of information on personal ranking positions. Female students in the treatment group report a decline in satisfaction with their GPA compared to their male counterparts. In terms of academic performance, women exhibit negative reactions to the feedback treatment, both after one and two years, particularly relative to males. They undertake fewer exams, achieve lower grades, approve fewer courses, and accumulate a smaller GPA.

We then explore two possible channels that could explain the adverse effects of ranking feedback information on women. Firstly, women might feel less at ease in a competitive environment, which is an inherent aspect of rankings. Male students report higher levels of competitiveness. A second mechanism involves that female students exert greater effort than their male counterparts during the pre-treatment period, as evidenced by their borrowing of books from the library and reporting of longer study hours. If the marginal cost of effort is increasing, it could pose a higher burden for women to maintain or enhance their position relative to men.

In our concluding analysis of the ranking treatment's impact, we assert that additional information may not always yield universal benefits. Policymakers should evaluate the advantages of disseminating more information and explore effective methods of delivery.

The paper most closely aligned with our experiment is the research conducted by Azmat, Bagues, Cabrales, and Iriberri (2019) at the University Carlos III in Madrid, Spain. Over three years, from 2010 to 2013, the authors provided feedback to 354 students in seven lecture groups regarding their position in the grade distribution (with 623 control students in 10 randomly selected lecture groups). They found that, in the absence of treatment, students tended to underestimate their standing in the grade distribution.

Furthermore, they identified negative effects of the treatment on educational performance, with GPAs experiencing a decline of 0.20 standard deviations for treated students in the first year. This adverse effect was short-lived, as students caught up in subsequent periods. By the time of graduation, control students exhibited the same likelihood of graduating and average accumulated GPA. Perhaps surprisingly, despite the negative impact of increased grade transparency on academic performance, students reported higher satisfaction with the quality of the courses.

Our main difference with Azmat et al. (2019) is our focus on specific gender differences recognizing the intrinsic competitiveness inherent in ranking systems. Drawing inspiration from the seminal paper by Niederle and Vesterlund (2007), we anticipated that treated women might underperform relative to men following the introduction of the ranking, a hypothesis substantiated by our findings. Several other distinctions exist between our work and Azmat et al.. For instance, our measure of satisfaction centers around students' GPAs (the foundation for the ranking) rather than the quality of courses. Additionally, our satisfaction measure is constructed using anchoring vignettes, offering a more precise measurement. Another difference is the level of randomization: our study randomized at the individual level (over 1000 students), while Azmat et al. employed randomization at the location+degree level (17 blocks). Furthermore, the information we provide varies; instead of disclosing the decile in the distribution of grades, we provide each student's exact placement in the ranking. Our system's daily updates contrast with Azmat et al.'s six-month intervals, a feature particularly attractive during exam periods when ranking fluctuations might be significant due to new grades entering the GPA (we observe that students significantly increased the use of the system in those periods). Lastly, our exploration encompasses different mechanisms such as competitiveness and the cost of effort. Despite these distinctions, both studies underscore the potential detrimental effects of increased transparency in the form of a ranking, highlighting that more information is not universally beneficial for students.

Our paper intersects with several strands of the literature, primarily within the domain of studies examining the interplay between satisfaction and ranking. We extend upon prior research that has empirically examined the relationship between relative position and satisfaction. Frey and Stutzer (2002) provide a review of this literature. A first reason why information on peers' rewards may affect utility is individuals care directly about their relative rewards. Luttmer (2005) investigates whether individuals feel worse off when others around them earn more. He found that, controlling for an individual's own income, higher earnings of neighbors are associated with lower levels of self-reported happiness. As previously noted, Card et al. (2012) contribute to this discussion by documenting the effects of relative pay comparisons on job satisfaction and job search intentions.

Furthermore, individuals may respond to novel information on peer rewards even when they lack a direct concern for their relative position. Specifically, it is conceivable that students, while not inherently preoccupied with their standing in the grade distribution, rationally use this information to recalibrate their expectations regarding future pay prospects. The relative position in the grade distribution serves as a nuanced signal that informs individuals about potential future wages. This signaling mechanism adds a layer of complexity to the intricate relationship between academic performance, perceptions of one's standing, and the anticipation of future outcomes.

A second strand of research concerns gender dynamics and ranking systems. Advances in psychology and experimental literature have provided a more nuanced understanding of psychological factors that exhibit systematic variations between men and women. While laboratory studies extensively explore these psychological nuances, research on the real-world implications of these factors is comparatively limited. Bertrand (2011) offers a comprehensive review of evidence on gender differences in risk preferences and attitudes toward competition. Findings from studies such as Croson and Gneezy (2009) and Eckel and Grosssman (2008) converge to suggest that women tend to be more risk-averse than men. Moreover, the underrepresentation of women in competitive environments underscores a general preference among women to avoid such settings.

Our work also connects with the literature on overconfidence and misplacement, revealing systematic misjudgements among students regarding their positions in the grade distribution. Underperforming students tend to overstate their ranking, while top-performing students tend to underestimate theirs. In the words of Moore and Healy (2008), "Overconfidence can have serious consequences. Researchers have offered overconfidence as an explanation for wars, strikes, litigation, entrepreneurial failures ...", and, relevant to our context, we could add educational failures. In their analysis of overconfidence types, the overplacement of one's performance relative to others is particularly pertinent to our research. Benoît and Dubra (2011) discuss the rationality underpinning these empirical regularities.

The remainder of the paper is organized as follows. Section 2 outlines the intervention, experimental design, and details our data collection process. In Section 3 we present our main empirical results about the impact of the ranking treatment on both satisfaction and academic performance. Section 4 is dedicated to the exploration of two potential mechanisms contributing to the asymmetric impact on men and women. Finally, we conclude in section 5. Supplementary results are gathered in the Appendix.

2 Experimental Design and Data

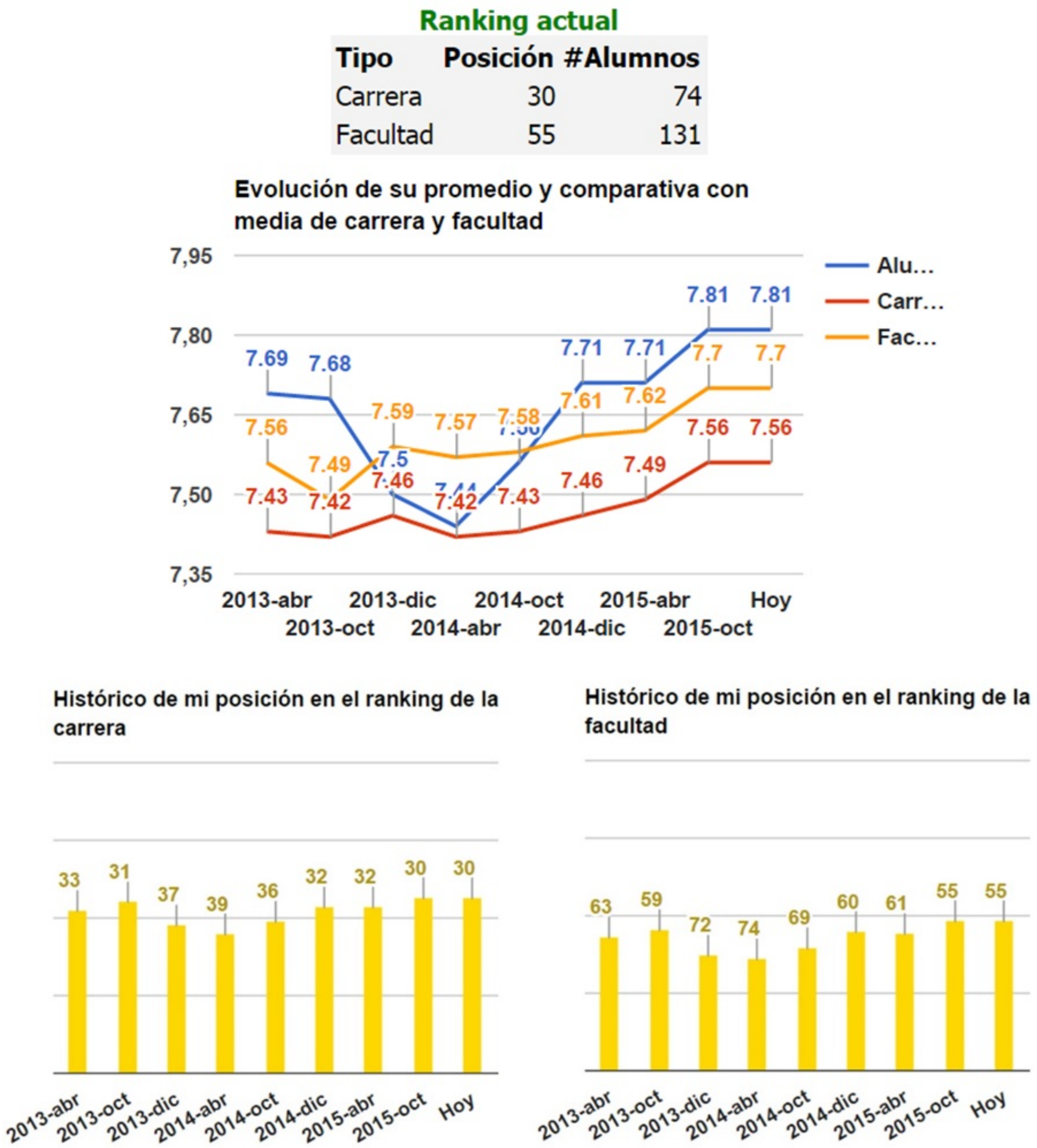
The Experiment

In mid-2013, we decided to conduct an experimental pilot study to gauge how students would respond to having access to information about their position in the grade distribution—an ordinal ranking. This tool was in the developmental stages by the IT team at the University. To assess its impact on students, we proposed conducting a pilot and actively participated in developing the treatment. Our focus was on three Schools at the Universidad de Montevideo: Economics, Engineering, and Law. The evaluation took the form of a randomized control trial.

Treated students gained access to a new intranet platform, allowing them to view their ranking relative to their peers. Control students did not have access to this new information. Both treated and control groups retained the ability to access their official transcripts of grades, as they could before the experiment's launch. Appendix A.1 displays a sample personal transcript of grades, encompassing all the courses taken and the corresponding grades. Although the transcript provides the student's grade point average (GPA) in the bottom right corner, it doesn't reveal their position in the distribution of grades relative to their peers.

The treatment aimed to contextualize a student's GPA within the framework of their peers. Treated students could utilize the new tool, which provided ranking information on their relative performance. Figure 1 provides a visual representation of an actual example of the treatment.

Figure 1: Information for treated students



Notes. This figure shows an actual intranet page featuring the ranking treatment for the schools of Economics and Law (Engineering had a simplified version in deciles). In the first box, the current ranking is displayed (updated daily), showcasing the exact position within the Major or School. The subsequent section presents the "evolution of your GPA and comparison with the average at the major and school" using a graph with three lines: the student in blue, the major in red, and the school in orange. The 'today' endpoint was updated daily, while the preceding points were fixed at three key moments throughout the year: April, October and December. As time progresses, the lines continue to extend. To illustrate whether a student was improving relative to their peers, the figure also includes the evolution of the position in the ranking in the next two figures. The endpoint of 30 and 55 correspond to the ranking for 'today', as reflected in the first box. The other points indicate the position at different moments in time, offering a comprehensive view of the student's performance over the course of the semesters.

The information treatment was deliberately designed to be visually intuitive and easy to comprehend. It offers students a comparison of their current GPA with peers at both the cohort+major level (with an average size of 37 students) and the cohort+school level (with an average size of 84 students). Additionally, the treatment includes informative figures illustrating the evolution of GPA rankings across semesters. This feature allows students to gauge whether they are making progress or experiencing declines relative to their peers—a dimension not previously accessible to them. In the example depicted in Figure 1, the student discovers that she currently holds the 30th position out of 74 students in their cohort+major and the 55th position out of 131 students in their cohort+school. The treatment doesn't just stop at providing current standings; it also informs about the evolution of a student's GPA over the years relative to their peers, along with the dynamic changes in their ranking. Appendix A.2 showcases two additional real cases of treated students, each with distinct trajectories over semesters. The ranking was recalculated and updated daily. We believe that access to this new information tool empowers students to obtain a highly accurate and insightful picture of their relative performance.

We employed various channels to ensure students were informed about the launch of the new tool and to enhance awareness and accessibility. Banners were strategically placed in the personal intranet, a platform accessed by students nearly every day for course materials and administrative purposes. Additionally, we disseminated information through email (see Appendix 5). It's noteworthy that this email was sent to control group students as well, differing only in the final line, which included a link to the ranking for treated students. The purpose of sending a placebo email to control students was to untangle the impact of receiving an email from the university—potentially leading to increased utilization of existing intranet resources like announcements and exam dates—from the effect of the treatment itself. By employing this placebo email, we can attribute the observed effect to the use of the ranking tool and rule out potential confounding factors related to increased access to the university intranet or other effects triggered by receiving an email from the university staff.

Randomization

The identification strategy employed to address our research question consists of a randomized control trial involving 1,048 university students. We aimed to achieve balance across critical dimensions, including the quality of students, sex, cohort, and all other pre-treatment variables outlined in Table 1. To achieve this balance, we generated 300+ different random assignments of students to treated and control groups. Only assignments that exhibited balance across these dimensions were considered (there were other assignments of participants in which there were differences between treated and control groups, given purely by random chance, which were dropped from the 300 candidate databases). From the pool of balanced assignments, one was selected through a random number generation process. The final selected assignment involved 529 students (50.5%) in the treatment group and 519 students (49.5%) in the control group[1].

Table 1 illustrates the balancing conditions across several pre-treatment characteristics. In the control group, 44% are women. At the beginning of the experiment, control students had taken an average of 25.57 courses, with an average approval of 21.34 courses, resulting in 138.49 credits out of the typically required 330 for a major. The cumulative GPA was 7.54 on a scale of 1 to 12.

A small but statistically significant difference is observed in the number of degrees attended by treated students, who attend slightly more degrees than control students. This discrepancy arose from a loophole in the computer system, a detail brought to our attention by the IT team post-implementation. Specifically, if a student was registered in two majors and placed in the Treated group in one major (e.g., economics), they could access their ranking in the second major (e.g., accountancy). However, the incidence of students registering for two majors is small: 9% in the treated group and 5% in the control group. To address this imbalance in pre-treatment characteristics, all regressions will control for this variable.

We also constructed three variables that we knew beforehand that were correlated to academic performance (and supposedly also to satisfaction): the top three high schools (attended by 29% of students, with 27% in the control group and 30% in the treated group), the proportion of students from Montevideo (the capital city), and the proportion of students with a scholarship exceeding 20% of the tuition fee.

Balancing was also achieved in the cohort, ranging from 2008 to 2014. Students from the 2014 cohort, constituting 21% of the sample, are freshmen without cumulative GPA information (and thus the number of observations in line 3 is smaller). Conversely, students from the 2008 cohort, representing 4% of the sample, have lagged behind and are starting their seventh year at the university, and should have graduated two to three years ago, if they had done their major on track. The exper-

[1]Unfortunately, the pool for randomization inadvertently included students who had dropped out from the university but were not yet accurately recorded as such in the informatic system at the time of database retrieval. Subsequently, after running the randomization process with an initial sample of 1,261 students (631 assigned to T and 630 assigned to C), it was discovered that a subset of students could not participate in the experiment as they were no longer enrolled at the University. Additionally, a group of students pursuing two majors posed a challenge. The randomization was initially assigned at the student*major level. However, due to a loophole in the system, as mentioned in the main text, these students had to be treated as a single observation instead of two as originally planned. Following these adjustments, the final sample consists of 1,048 individual students, with a slightly higher number assigned to the treatment group than the control group

Table 1: Descriptive Statistics

	(1)	(2)	(3)	(4)	(5)		(6)
	Treated	Control	Difference (1-2)	Standard Error	p-value		Obs.
Student characteristics							
1 *Female*	0.47	0.44	0.03	(0.03)	0.40		1,048
2 *Courses*	26.55	25.57	0.98	(1.13)	0.38		1,048
3 *Cumulative GPA*	7.50	7.54	-0.04	(0.10)	0.70		834
4 *Approved courses*	21.83	21.34	0.49	(1.00)	0.62		1,048
5 *Credits earned*	142.16	138.49	3.67	(6.33)	0.56		1,048
6 *Number of degrees*	1.09	1.05	0.04	(0.02)	0.02	**	1,048
7 *School of origin*	0.30	0.27	0.03	(0.03)	0.24		1,048
8 *Montevideo*	0.67	0.65	0.02	(0.03)	0.42		1,048
9 *Large Scholarship*	0.23	0.24	-0.01	(0.03)	0.70		1,048
10 *Cohort 2008*	0.04	0.04	0.00	(0.01)	0.83		1,048
11 *Cohort 2009*	0.05	0.05	0.00	(0.01)	0.94		1,048
12 *Cohort 2010*	0.13	0.12	0.01	(0.02)	0.66		1,048
13 *Cohort 2011*	0.20	0.20	0.00	(0.02)	0.94		1,048
14 *Cohort 2012*	0.20	0.20	0.00	(0.02)	1.00		1,048
15 *Cohort 2013*	0.17	0.16	0.01	(0.02)	0.55		1,048
16 *Cohort 2014*	0.20	0.22	-0.02	(0.03)	0.44		1,048

Notes. This Table presents descriptive statistics for the sample, comprising 529 (50.5%) treated and 519 (49.5%) control students. The difference in the number of observations between the groups is explained in the main text. The t-test for the difference in means is calculated using an OLS regression with robust standard errors. Balance was also performed by schools, degrees and place in the distribution of grades at the major and school level. To maintain brevity in presenting the main results, these additional balancing results are omitted from this table but can be found in Appendix A.1. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

iment excluded previous generations. Balancing was further maintained by major and by decile of the GPA distribution, ensuring an equitable distribution of good students between the control and treatment arms of the experiment (for brevity, detailed descriptives for these variables are omitted but can be found in Appendix A.1).

Data Collection

Our study focuses on two primary outcomes: short-term satisfaction with GPA and long-term academic performance. Short-term satisfaction data was gathered through a web survey administered to both treated and control students, conducted 12 days after the launch of the ranking tool. In the next section we will explain more details of this survey. Academic performance indicators, such as grades and credits achieved, were obtained from administrative records one and two years post-treatment.

The short-term survey achieved a high response rate of 82%, with 861 out of 1,048 participants participating. This success was attributed to robust publicity efforts (by email and other means) and the temporary blocking of access to certain intranet resources for students who hadn't initiated the survey. This blocking practice aligns with the university's standard procedure for various student surveys (for example for the evaluation of courses and lecturers), ensuring widespread participation[2]. Importantly, the survey was not explicitly linked to the ranking experiment; it was presented as a general 'satisfaction survey' originating from the Bedele's office.

There was no statistically significant difference in the response rate between the treated and control groups (83.7% in the Treated group and 80.5% in the Control group, p-value = -0.176). Moreover, response rates also showed no gender-based disparities, a key explanatory variable in our analysis. Appendix A.2 shows that, as expected, the survey was more frequently answered by students with higher cumulative GPAs and those placed in higher deciles of the GPA distribution, indicative of a greater sense of responsibility in better students.

While attrition in the survey was -fortunately- not correlated with treatment status or gender, it is crucial to ensure that balance is maintained among other characteristics for the students who did respond. Appendix A.3 confirms that, even after attrition, balance is preserved, allowing for the analysis of satisfaction outcomes using the random variation in treatment status generated by the experiment.

Importantly, attrition is not a concern in the administrative database, as academic outcomes for all 1,048 students in the experiment are available. This comprehensive dataset includes information on grades, exams, dropouts, and other relevant academic metrics.

Satisfaction module in the short-term survey

The online survey focuses primarily on satisfaction, with two main measures placed as the first and second questions. Satisfaction with the current GPA is initially measured on a 1 to 5 scale, presented within a context of placebo questions that inquire about satisfaction with various aspects of university life. These questions, such as satisfaction with the location of the university campus, serve as placebo outcomes unaffected by the ranking treatment[3].

[2]Our blocking policy for the intranet resources was more lenient than other blockings from administrative staff. We implemented the block mainly as a gentle reminder to students of the ongoing survey, recognizing that some might not regularly check emails. If a student didn´t want to answer this survey, he could opt-out and access his intranet web site with no further delay. Students who preferred not to participate could easily opt-out and access their intranet web site without delay.

[3]The exact wording of the first question is: *'Currently, on a scale from 1 to 5, where 1 is 'Very Unsatisfied' and 5 is 'Very Satisfied', indicate how you feel with: (...) your cumulative GPA'.*

The second satisfaction question is part of a dedicated module, and responses are reported on a 1 to 10 scale[4]. Anchoring vignettes are included in this section to address potential issues with comparability (King et al., 2004). As students may operate on different internal scales, the vignettes serve as anchors by providing fixed and objective situations for evaluation.

When two students report a satisfaction score of, for instance, eight, the comparability of these responses becomes ambiguous, considering that each student may be using a distinct internal scale where an eight corresponds to different satisfaction levels. Notably, a high-ability student might assign less value to a grade of 9 compared to a low-ability student. Additionally, the social context plays a role, as a student with lower academic performance interacts with peers holding similar grades, as we empirically show in a following section. Thus, he might have more information about the left part of the distribution of grades, potentially influencing their scale of evaluation. In contrast, a high-achieving student with high-performing friends might utilize a different scale, and thus evaluate a given GPA in a different social context[5]. The inclusion of anchoring vignettes mitigates these challenges by providing standardized and objective situations for evaluation. Consequently, these vignettes are incorporated to anchor the subjective satisfaction valuations.

The vignettes feature four hypothetical student trajectories, each with different GPAs and expected times for graduation: a top-performing student (*Guille*, with a GPA that placed him in the top 10% of the distribution and who is expected to graduate on time), two students in the middle of the ranking (*Fer* and *Jose*, in percentiles 40 and 60, respectively, with a shorter time to graduation the first one but a lower GPA), and a student with a GPA from the bottom 10% of the distribution who expects to graduate with 6 months delay (*Fran*)[6]. Respondents evaluate these four objective situations using the same 1 to 10 point satisfaction scale used for their own subjective satisfaction[7].

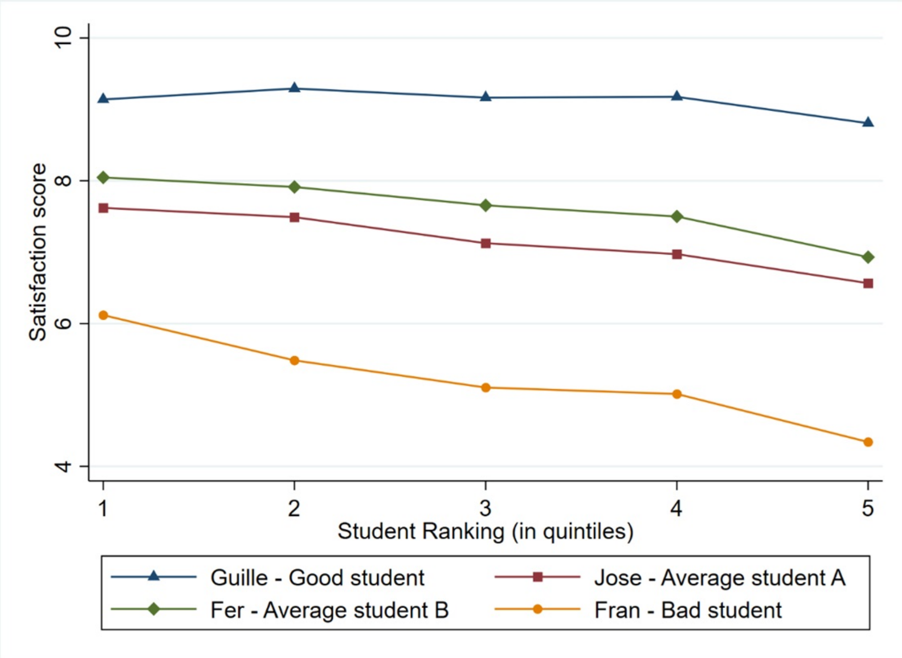
[4]We show the exact wording of the second question in Appendix A.9.

[5]A similar rationale is explained in Beegle, Himelein, and Ravallion (2012) regarding welfare and wealth distributions. Suppose that wealth is normalized to be in the [0,1] interval, with M the modal wealth. If a relatively poor individual answers about his subjective welfare, he is only aware of the levels of wealth ranging from 0 to M. On the other hand, a wealthier individual knows from M to 1. Consequently, when they provide responses, they are not utilizing the same scale. The rich citizen might assert that they are worse off than they truly are, while the poor respondent may report a higher position than their actual standing (owing to their limited knowledge about the wealth of a wealthy person).

[6]As is customary when employing anchoring vignettes, we utilized gender-neutral nicknames. In Spanish, unisex names are extremely rare, and delivering gender-specific questionnaires posed a challenge.

[7]Additional assumptions include vignette equivalence, where each participant should perceive the same image or depiction of the hypothetical situation presented in a vignette, and response consistency, implying that the process used by a student to assess their own subjective GPA aligns with the evaluation of GPA in the vignettes. For a more in-depth exploration of these concepts, we direct readers to Bago d'Uva et al. (2011)

Figure 2: Punctuation of the four satisfaction vignettes, by student performance



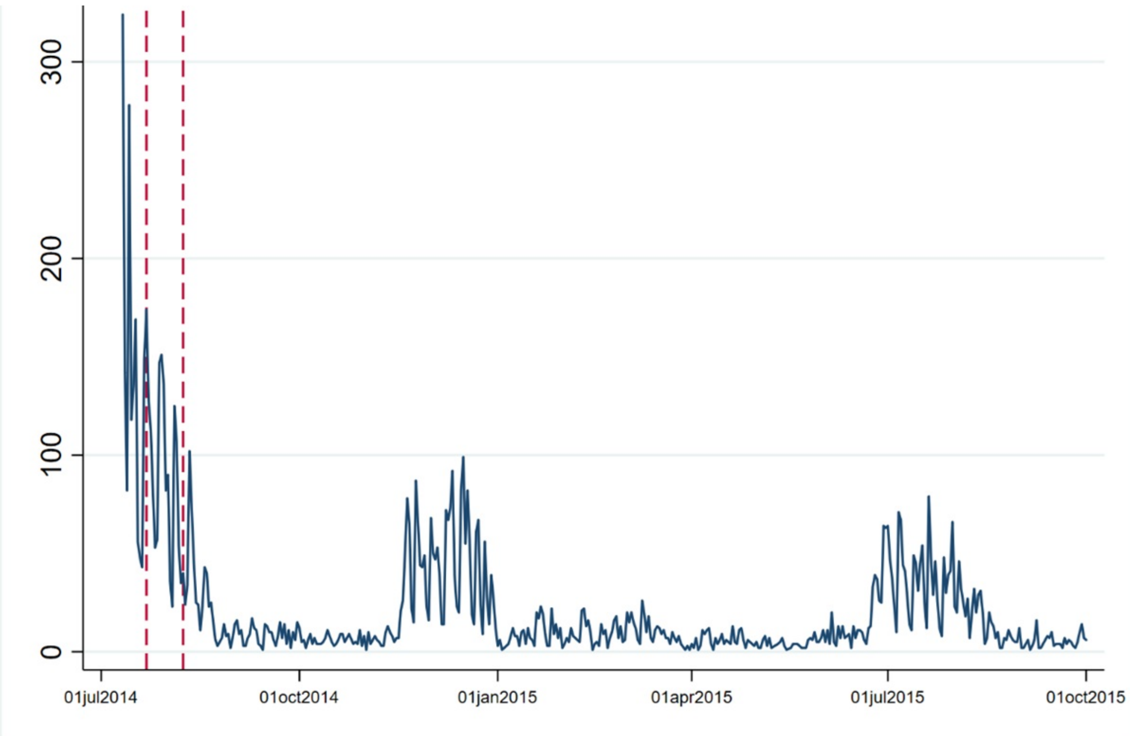
Notes. This figure depicts the satisfaction score assigned by 895 students who answered the short-term satisfaction survey. The assessment of four situations is measured on a 1 to 10 point scale and displayed by quintile of their actual GPA. *Guille*, a hypothetical student with a GPA of 9.4 (top 10%), expects to graduate on time. *Jose*, with a GPA of 7.8, anticipates a 6-months delay in graduation. *Fer*, holding a GPA of 7.0, expects to graduate on time. Notably, students appear to prioritize the graduation date over GPA, at least in the comparison of the two middle situations. Lastly, *Fran*, a hypothetical lower-performing student, with a GPA of 5.9 (bottom 10%), expects a 6-months delay in graduation.

Results from the vignette evaluations indicate consistent patterns (Figure 2). The hypothetical good student *(Guille)* is consistently evaluated as better than any other situation with a lower GPA, while the hypothetical lower-performing student *(Fran)* is consistently evaluated as worse than other situations. Moreover, we find evidence of heterogeneity in reporting satisfaction scales by quintiles of a respondent's GPA. Good and bad students report different satisfaction levels for the four hypothetical situations. Across all situations, there is a downward slope, with good students (quintile 5) assigning fewer points to a given GPA than low-performing students (quintile 1). For instance, satisfaction reported by students placed in the bottom quintile of the ranking with the situation of a bad student *(Fran)* is higher (6.1 points) than the evaluation of good students for the same vignette (4.4 points).

Were students treated?

Figure 3 illustrates the number of accesses to the ranking system. The busiest day was when we launched the system for treated students. On July 11th, 2014, the system was introduced with promotion by email notifications and with customized banners in the individual intranet of treated students, resulting in 326 accesses. Access patterns by weeks, marked by peaks and valleys, are influenced by lower access during weekends and increased activity on Mondays, likely when students utilize university or workplace computers. Access to the system by month mirrors the pattern observed for accessing official transcripts, particularly during main exam periods in December and July. Outside these periods, access intensity decreases since no new grades are added, reducing the incentive to check the ranking.

Figure 3: Treatment. Number of accesses to the ranking by date.



Note: This figure shows the number of accesses to the ranking system by day. The system was launched on the 11th of July, 2014 (324 accesses). A student may access multiple times, as explained in the main text. The two vertical lines show the time window during which the survey on satisfaction with grades was conducted.

The two vertical lines demarcate the time window for the survey on satisfaction with grades, our primary outcome (refer to Appendix A.4 for a detailed breakdown of access patterns during that 17-day period). The system meticulously records each student's use, showing that, from July 11th, 2014, to August 31st, 2015, students collectively viewed their ranking 9,625 times. On average, students accessed the system 18.2 times (min = 0, max = 430). Of the 529 students in the treatment group, 508 accessed the ranking at least once before September 2015, indicating high compliance with the treatment assignment[8]. In the control group, out of 519 students, 518

[8]Additional access statistics, available upon request, reveal no discernible differences between

did not access the ranking (except for one student granted special access by the IT team). Consequently, the main results tables will present intention-to-treat effects from reduced-form models based on the initial assignment. Instrumental variable estimates, using randomization as an instrument for actually receiving the ranking, are similar to ITT due to high take-up rates, and are available upon request.

In summary, the new ranking tool experienced substantial use by treated students. Our next objective is to demonstrate that the ranking system indeed supplied students with novel information.

Did students change their perceptions?

In the preceding sub-section, we established that students actively utilized the ranking system. Now, our focus shifts to demonstrating that treated students acquired a significantly more accurate understanding of their actual placement within the grade distribution. Essentially, the treatment successfully heightened students' awareness of their relative performance. A potential challenge for the experiment could have been that students already possessed accurate perceptions not only of their individual GPAs but also of their relative standings within their cohorts. Given the relatively small size of cohorts at the university, students might naturally have access to good information about the academic performance of their peers, thus having a clear sense of their relative positions.

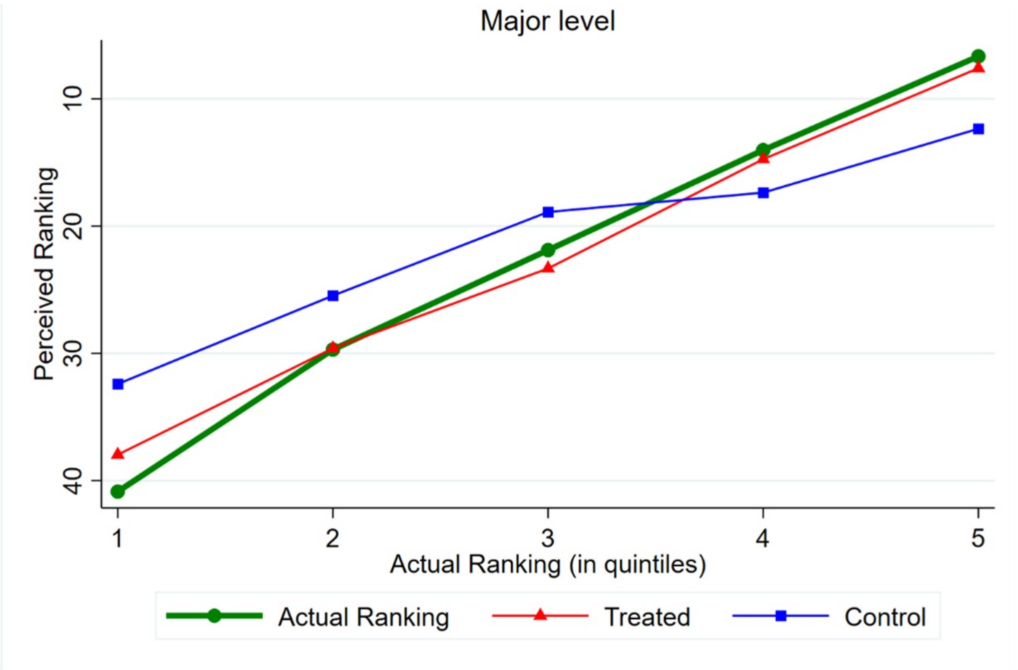
Moreover, there was a risk that control students, once the ranking system was implemented, could obtain information on their relative performance from treated friends. For instance, a control group student may be uncertain about being placed in the top 10% of the cohort. However, if they have a friend with a similar GPA in the treatment arm of the experiment, and that friend knows their ranking, communication of this information can provide the control group student with knowledge about their own standing in the top 10%. So *prior accurate knowledge* and *contamination* could be two major challenges to the experiment.

To tackle these challenges, two questions were included in the Satisfaction Survey. We specifically inquired whether students had learned the ranking of their peers. Out of the 564 students who responded to this question, 38 students in the control group reported learning the exact position of a classmate in the ranking. However, a more accurate measure of contamination involves comparing the differential increase in knowledge caused by the treatment. For this purpose, a second question aimed to assess whether the ranking provided *new* information to students. Both treated

men and women, while students with higher GPAs accessed the system more frequently. Interestingly, students nearing graduation, with a higher credit count, tended to use the system less frequently. Up until September 2016, a total of 14,298 accesses to the system were recorded. Over the two years of the experiment, system usage experienced a slight decline as some students graduated and subsequently left the University.

and control students were asked about their placement in the distribution of grades (ranking), with 1 representing the student with the highest GPA in the cohort (and major or school). Figure 4 illustrates the actual (objective) ranking versus the perceived (subjective) ranking.

Figure 4: Actual vs Perceived relative performance, by treatment status



Note: This figure shows the relation between perceived and actual ranking. The perceived ranking (Y-axis) is reported by the students in the Satisfaction Survey. Each student reports their placement in the distribution of grades (ranking), being 1 the student with the highest GPA. The actual ranking (X-axis) comes from the administrative registries of the University, and is grouped by quintiles. Students from the school of engineering were excluded from this analysis since they did not receive the ranking at the individual placement level but in deciles: thus, the Y-axis has a different unit of measurement for them.

Three lines are depicted on Figure 4. The first is a 45º line, representing the actual ranking (in quintiles) plotted against the actual ranking (in individual positions). The other two lines correspond to control students and treated students (regardless of system access). A student positioned at the right end of the x-axis signifies a high-achieving student. If they accurately place themselves in the ranking, they should report being among the top 10 students on the y-axis. We find that treated students remarkably report their actual ranking with great accuracy, aligning closely with the actual placement. Had they not received the information treatment, their behavior would resemble that of students in the control group. In the case of control students, they misjudge their positions in the distribution of grades. Notably, high-achieving control students (at the right end of the figure) tend to understate their relative performance, reporting a perceived position in the ranking lower than their actual performance. Conversely, underperforming control students (at the left end of the figure) tend to overstate their positions in the distribution of grades, reporting a perceived relative performance higher than the actual one. For a more detailed breakdown by treatment and gender, refer to Appendix A.5.

The disparity between treated and control students indicates that the experiment successfully imparted new information, enhancing the knowledge of relative performance, with no significant spillover effects observed from treated to control students.

Figure 4 reflects the well-known regularity of the Dunning-Kruger effect (1999), at least for unskilled students who may be unaware of their skills. When comparing self-assessment of performance with objective measures, it becomes apparent that lower-performing students tend to overstate their ability. Thus, unskilled individuals, not only have erroneous (overoptimistic) perceptions about their ranking but also, by definition, correlate with being in the lower-performing quintile. A different explanation is required for the misplacement observed among the top-performing students, a phenomenon that cannot be ascribed to a lack of skills.

We propose three hypotheses for control students' misplacement in the ranking:

1. *Statistical inference problem and selection:* Students may selectively form groups of friends and then extrapolate the entire distribution of grades based only on the observed performance of their close peers.

2. *Regression towards the mean:* This effect may explain both the underperforming students placing themselves above the 45º line, and the high-achieving students reporting an expected lower performance relative to their actual ranking (a situation not considered in the original Dunning-Kruger effect, where only unskilled students misplace themselves). The rationale is as follows: A student may recognize a good performance in an exam but remains uncertain whether it is due to their inherent capability (placing them above their peers) or the exam's relative ease (leading to widespread high performance). Consequently, they may adjust their expected placement in the distribution of grades.

3. *Cognitive biases:* Despite having the necessary information, students may possess cognitive biases that impede accurate utilization of the available data.

We have collected data that allows us to test the first hypothesis, drawing inspiration from the work of Cruces, Perez-Truglia, and Tetaz (2013). In the satisfaction survey, we directly asked students to name their best friends in their cohort. Out of 785 students who answered this question, we gathered information on 3,389 friends. This data enables us to construct networks of friends, revealing that peer groups are not formed randomly. Good students tend to be friends with other good students. As shown in Appendix A.6, a student's GPA is correlated with their friend's GPA: an additional point in the friend's grades is associated with an increase of 0.77 in a student's grades (t-value=13.9), indicating a strong selection process. Now, let's examine the inference problem. If a student doesn't know their placement in the grades distribution, they must infer it from the comparison with their closest peers. Given the positive selection of peers, a good student will have high-performing peers. On average, they will be placed lower in the ranking of their peers than if they had a random group of peers representative of the whole distribution of grades. In the data, a student in the 5th quintile of the grades distribution (a good student) has friends with a higher GPA than the median of the whole cohort (the difference is 0.48 σ). Consequently, when constructing their perceived ranking with the information at hand, they will likely think that the cohort is better performing than it really is. As a result, they will under-place in the ranking given the information they receive from their peers.

Figure 4 serves as a concise summary of individual responses, providing a visual representation of the treatment's impact on knowledge. For a more in-depth examination, refer to Appendix A.7, where detailed figures present individual-level data for both control and treated students. Individual survey responses show that underperforming control students tend to overestimate their performance (evidenced by their position above the 45º line), and there is greater dispersion compared to their higher-performing counterparts. In contrast, the figure for treated students reveals a higher concentration along the 45º line, particularly among those excelling academically.

Finally, Table 2 provides additional evidence that the ranking system provided *new information* to treated students, relative to control ones. For each student, we calculate the absolute difference between the reported (perceived) ranking and the actual (objective) one, and conduct the following regression:

$$Y_i = \alpha + \beta_1 Treated_i + \beta_2 CohortSize_i + \beta_3 Grades_i + e_i \tag{1}$$

where Y_i represents a measure of how accurately student i reports her ranking. The first measure is the difference between the actual and the declared ranking, in absolute value, labeled as *absolute difference*. For instance, if a student declares her position in the grade distribution as number 5 (based on survey data), and the actual position (based on administrative data) is number 6 or 4, then the difference will be 1. The second measure is a dummy variable indicating whether the student reported her *exact placement*. In the example, the student will have a 0 if she didn't report her exact position and a 1 if she reported that her position was 5. We expect that the treatment will decrease the error a student makes when reporting her placement and increase the probability that she reports her exact position. The two main controls will be the *cohort size* (larger cohorts may make it more difficult to know the exact position) and the *pre-treatment grades* (better students should report their position more accurately).

Table 2: Impact of the treatment on the perception of relative performance

	(1)	(2)	(3)	(4)
	Absolute Difference		Exact placement	
	Major level	School level	Major level	School level
Treated	-3.360***	-8.583***	0.295***	0.193***
	(0.568)	(1.414)	(0.031)	(0.024)
Cohort Size	0.121***	0.119***	-0.002***	-0.000
	(0.013)	(0.010)	(0.001)	(0.000)
Grades distribution	-0.430***	22.097***	0.028***	-0.097
	(0.114)	(4.973)	(0.005)	(0.076)
Constant	6.854***	22.097***	-0.063	-0.097
	(2.240)	(4.973)	(0.063)	(0.076)
Observations	600	609	600	609

Notes. The outcome *absolute difference* represents the absolute value of the difference between the declared and actual position in the ranking. *Exact placement* is a dummy variable equal to one if the student reports his exact placement in the ranking. *Cohort size* indicates the number of students in the major (columns 1 and 3) or in the school (columns 2 and 4). *Grades distribution* denotes the decile of the student in the distribution of grades, at the major or school level. Students from the School of Engineering are excluded from this analysis because they don't report their ranking in exact positions but in deciles. Controls include cohort dummy variables. Robust standard errors are reported in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

The treatment significantly impacted students' awareness of their relative performance. Column 1 shows that treated students more accurately placed themselves by 3.4 positions, whereas control students misjudged their position by 6.8 positions[9]. The other coefficients indicate that larger cohort sizes contribute to increased reporting errors, while higher GPAs result in more accurate reporting ones ranking (smaller differences). In Column 2, this difference is more pronounced at the school level, with control students misplacing by 22 positions and treated students reducing misplacement by 9 positions. Additionally, 37% of treated students report their exact position, compared to only 8% of control students at the major level (21% vs. 2% at the school level). While regression results capture the treatment's impact on

[9]As previously mentioned, the average cohort size is 32 students at the major level and 131 students at the school level.

mean differences, Appendix A.8 presents a histogram detailing the distribution of accuracy for treated students relative to control ones.

In summary, we find that the randomized treatment successfully altered the endogenous variable—the knowledge students have about their relative performance. To establish the credibility of the experiment's impact on outcomes (satisfaction and academic performance), it was crucial to determine not only the usage of the system (as discussed in a previous sub-section) but, more importantly, whether the experiment effectively influenced the information treated students had compared to untreated students. Fortunately, the treatment provided pertinent information, leading to adjustments in students' perceptions of their relative performance.

3 Results

After detailing the intervention and establishing its effectiveness in improving the information on relative performance for treated students, our analysis now shifts to demonstrate the causal impact of the experiment on satisfaction and academic outcomes. The focus of our investigation will be on the divergent responses between males and females, using the following regression model:

$$Y_i = \alpha + \beta_1 T_i * W_i + \beta_2 T_i + \beta_3 W_i + \boldsymbol{X_i'\theta} + e_i \tag{2}$$

where Y_i represents a measure of the satisfaction or academic performance of student i. T_i takes the value one if the student is assigned to the treatment group. Additionally, W_i takes the value of one if the student is a woman. We also incorporate a set of pre-treatment controls $\boldsymbol{X_i}$ from Table 1, which are balanced by randomization of the treatment. The parameter β_2 represents the treatment effect for men, whereas the treatment effect for women is given by the parameters $\beta_1 + \beta_2$. Hence, the differential treatment effect by gender is represented by β_1.

Satisfaction

The first results, presented in Table 3, are derived from estimating Equation 2 for two distinct satisfaction measures and a placebo. As explained in Section 2, the first satisfaction measure employs a 5-point scale, while the second uses a 10-point scale, but more importantly satisfaction in column 3 is reported using vignettes, providing an anchor for subjective satisfaction valuations. These vignettes are incorporated as dummy variables in the regression. The placebo regression (location of the university campus) is measured on a 5-point scale and is analogous to the first column (both outcomes were reported in the same question).

With the first satisfaction measure, we find that treated men experience an increase in satisfaction by 0.14 points. The interaction term indicates that the

Table 3: Satisfaction with GPA

	(1)	(2)	(3)
	First measure	Placebo	Second mesasure
*Treated*Woman*	-0.310***	0.035	-0.556**
	(0.107)	(0.126)	(0.223)
Treated	0.141**	-0.015	0.164
	(0.07)	(0.079)	(0.146)
Woman	0.148*	-0.204**	0.319*
	(0.078)	(0.095)	(0.168)
Constant	2.592***	4.600***	3.355***
	(0.173)	(0.19)	(0.63)
Observations	860	860	858

Notes. The first measure of satisfaction is constructed on a 1-5 scale. Placebo is a measure of satisfaction not affected by the treatment (satisfaction with the location of the University). The second measure of satisfaction is constructed on a 1-10 scale and includes a full set of dummy variables with information on 4 anchoring vignettes. All models include pre-treatment controls. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

treatment effect is gender-dependent. Specifically, treated*woman has a coefficient of –0.31, meaning that treated female students decrease their satisfaction with their GPA when exposed to the ranking, relative to treated men (equivalent to 0.35 of a standard deviation). The overall treatment effect for women ($\beta_1 + \beta_2 = -0.310 + 0.141$) is negative and significantly different from zero ($p = 0.039$).

The placebo regression in Column 2 reveals no treatment effects on satisfaction with the university location.

Column 3 in Table 3 presents the results of our most accurate satisfaction measure, accounting for vignette evaluations. These findings further underscore that treated women exhibit a noteworthy reduction in reported satisfaction with GPA following exposure to the ranking treatment. The estimated point decrease of -0.556 represents a quarter of a standard deviation in reported satisfaction. The overall negative effect for women ($\beta_1 + \beta_2 = -0.556 + 0.164$) is significantly different from zero ($p = 0.025$). Additionally, the results in Table 3 reveal that, even when controlling for pre-treatment GPA, female students tend to be more satisfied with a given GPA. The decline in satisfaction for treated women surpasses the difference in satisfaction for control women relative to control men. Finally, in Appendix A.10 we visually illustrate the diminished satisfaction for women across the distribution of grades.

Academic outcomes

We obtain the academic outcomes from the university's administrative records, ensuring coverage even for students who did not answer the satisfaction survey. Table 4 presents results for one and two years after the treatment (in panels A and B, respectively). Columns 1 to 3 present results at the exam level, while columns 4 to 7 at the student level. Standard errors in the first three columns are clustered at the student level, with 1,048 clusters and an average of 16.7 exams per student.

Table 4: Academic Outcomes after one and two years of treatment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Exam grade	Took the exam	Passed the exam	GPA	Approved courses	Credits gained	Dropout
Panel A: Medium term academic results							
*Treated*Woman*	-0.262**	-0.032*	-0.024	-0.129	-1.194*	-5.808	0.090***
	(0.127)	(0.017)	(0.017)	(0.119)	(0.723)	(4.371)	(0.033)
Treated	0.151*	0.011	0.013	0.080	0.537	2.201	-0.013
	(0.086)	(0.012)	(0.012)	(0.07)	(0.479)	(2.895)	(0.022)
Woman	0.278***	0.021*	0.011	0.217**	0.413	3.401	-0.019
	(0.096)	(0.012)	(0.013)	(0.098)	(0.569)	(3.42)	(0.022)
Constant	5.968***	0.773***	0.694***	5.683***	10.552***	56.206***	0.139**
	(0.185)	(0.026)	(0.025)	(0.181)	(1.105)	(6.577)	(0.059)
Observations	15,058	17,532	15,058	1,046	1,046	1,044	1,046
Panel B: Longer term academic results							
*Treated*Woman*	-0.150	-0.045**	-0.028	-0.206**	-2.468**	-12.728*	0.061*
	(0.153)	(0.021)	(0.020)	(0.101)	(1.115)	(6.821)	(0.037)
Treated	0.060	0.016	0.023*	0.070	0.428	1.128	-0.000
	(0.101)	(0.013)	(0.014)	(0.064)	(0.728)	(4.405)	(0.024)
Woman	0.366***	0.018	0.037**	0.296***	0.517	4.187	0.006
	(0.112)	(0.015)	(0.015)	(0.079)	(0.853)	(5.198)	(0.026)
Constant	6.545***	0.816***	0.762***	5.816***	18.145***	90.447***	0.197***
	(0.222)	(0.027)	(0.029)	(0.172)	(1.740)	(10.492)	(0.063)
Observations	8,819	9,983	8,819	1,046	1,046	1,045	1,046

Notes. Treatment started in July 2014. Panel A columns (1) to (3) show individual exam-level data between August 2014 and August 2015, while columns (4) to (7) show student-level data measured in Spring 2015 (November in the southern hemisphere). Panel B columns (1) to (3) show individual exam-level data between September 1st, 2015, and August 31st, 2016, while columns (4) to (7) student-level data measured in Spring 2016. Standard errors in columns (1) to (3) are clustered at the student level, while columns (4) to (7) are robust to heteroscedasticity. All regressions include pre-treatment controls as in Table 3. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

We find that treated women decreased their academic performance more than treated men: they took fewer exams and achieved lower grades. These negative results seem to persist even in exams administered two years post-treatment. Examining other academic outcomes, treated women showed a reduction in cumulative GPA, passed fewer courses, and consequently earned fewer credits, while the dropout rate was marginally higher than treated men. The coefficient for *woman* indicates that, on average, female students outperform male students at the university. To contextualize the magnitude of some results, the estimated (negative) impact on exam grades for treated women is nearly equivalent to the gender difference in performance, with control women scoring 0.278 points higher than control men, and a treated woman decreasing their score by 0.262 points, or a tenth of a standard deviation. The estimated impact on GPA (column 4) is approximately two-thirds of the gender GPA gap. Finally, the overall treatment effect for women, considering cumulative GPA ($\beta_1 + \beta_2 = -0.206 + 0.070$) is negative and significantly different from zero ($p = 0.066$).

The different number of observations between column 2 and columns 1 and 3 is due to instances where a student did not *take the exam* ($Y_i = 0$), resulting in no records for *exam grade* (ranging from 1 to 12) or a definition of *passed the exam* (*grade* ≥ 6). The difference between the number of individual exam observations (columns 1 to 3) between Panels A and B is attributed to more exam periods in Panel A than in Panel B, and the longer-term exam results in Panel B excluding those considered in the medium-term results from Panel A. Pooling all individual exams (columns 1 to 3) from Panel A and B in one regression presents the same picture of asymmetric negative effect for treated women, as shown in Appendix A.4 (this exercise is not applied to other outcomes as the long-term effects encompass the short term ones).

Taken together, the short and long-term results suggest that women experienced a decline in various dimensions of academic performance post-treatment, and this effect did not dissipate over time. The feedback from the ranking appears to have had detrimental consequences for them.

As a placebo exercise, Appendix A.5 shows the results of regressing the treatment on pre-treatment outcomes. We find that, as expected, there is no impact of the (future) treatment on pre-treatment individual exam grades. Treatment effects for men (β_2) and the differential effect for women (β_1) are of small magnitude and non-significant. These results can also serve as a pre-treatment balance check. Recall from Table 1 (line 3) that we had balanced the cumulative GPA at the student level (n=834, excluding freshmen with no exam grades) and not on pre-treatment *individual exam grades* (n=29,721).

4 Possible channels

Our objective in this section is to answer *why questions*: Not only what has happened with students in the experiment (outcomes) but also why those effects happened (channels). We will explore two potential channels that may explain the observed negative treatment effect on female students. The evidence suggests two potential mechanisms: (i) differences in the willingness to compete between men and women, and (ii) an increasing marginal cost of effort.

Competitiveness. We initially hypothesized that women might be less competitive than men (Niederle and Vesterlund, 2007). Given that the ranking introduces a competitive component, we anticipated that women could underperform relative to men post-introduction of the ranking. To test this hypothesis, we included a question about student's self-declared level of competitiveness[10]. The results in Table 5 indicate that, on average, women report being 10% less competitive than men. In Appendix A.11, we present the distribution of responses on the 10-point scale, revealing a right-skewed density for men, indicating that they are more willing to compete.

Table 5: Competitiveness and effort for men and women

	(1)	(2)	(3)	(4)	(5)	(6)
	Competitiveness		Usage of library		Hours of study	
Woman	-0.072**	-0.101***	6.870***	3.766***	8.396***	8.772***
	(0.033)	(0.035)	(1.295)	(1.019)	(1.239)	(1.171)
Constant	0.382***	0.418***	11.114***	-1.512	52.095***	73.149***
	(0.023)	(0.101)	(0.682)	(9.066)	(0.769)	(3.098)
Controls	NO	NO	NO	NO	NO	NO
Observations	843	842	980	979	25,258	25,226

Notes. Columns 1 and 2 present OLS regression results examining the association between a female student indicator and the competitiveness measure. The dependent variable *competitive* is a binary variable, taking the value of one if a student reports a competitiveness score of 7 or more on the 0-10 scale. Columns 3 and 4 provide insights into library usage, measured by the count of books borrowed from the university library during the pre-treatment period. Columns 5 and 6 focus on study hours per week, measured in minutes, for each subject. The outcomes are derived from three distinct data sources: Competitiveness is obtained from the short-term online survey used for satisfaction assessment, library usage data is extracted from university administrative records, and self-reported study hours come from students' semester-end evaluation of courses. Controls in even-numbered columns are consistent with those used in Table 3. Standard errors are reported in parentheses: In columns 1 to 4, they are robust to heteroscedasticity, while in columns 5 and 6, they are clustered at the student level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Increasing marginal cost of effort. In considering the increasing marginal cost of effort as a potential mechanism, we note that climbing in the rank demands additional study hours. Moreover, if fellow students are improving, maintaining one's position in the ranking requires even more effort. Assuming that the marginal

[10]The exact wording of this question was: *How competitive do you think you are? Please choose a value on the scale below, where a value of 0 means 'I am not competitive at all' and a value of 10 means 'I am very competitive.'*

cost of an additional study hour is increasing, and considering that female students already invest more time, putting in an extra hour might be excessively stressful for them in the pursuit of climbing in the ranking or maintaining their position, thus lowering their overall satisfaction.

We have two measures of study effort (i) an objective measure based on the number of books borrowed from the university library, and (ii) a self-reported measure detailing the number of hours dedicated to each subject.

Examining library records, we find that out of the 1,048 students in the experiment, 980 exhibited activity at the university library, while 68 did not rent any books. On average, a student during the pre-treatment period rented 14.3 books (22.4 considering the entire period). Results from columns 3 and 4 show that female students reported a significantly higher use of library materials, with a raw comparison indicating 6.87 more books (75%) than their male counterparts[11].

Turning to the self-reported number of study hours, we have responses from 1,045 students across an average of 24.17 subjects (n=25,258) during the pre-treatment period. The data, collected through end-of-semester questionnaires, reveals that female students reported an additional 8.4 minutes per week for each subject, accumulating to 45 more minutes of reported study time weekly compared to men, roughly a 15% difference.

In summary, the observed differences between female and male students in competitiveness and effort shed light on potential explanations for the asymmetric gender treatment effects of the ranking intervention. Women may find it less comfortable to engage in competitive academic settings, and the added effort required to maintain or improve the ranking might be disproportionately costly for them, especially considering their already higher investment in study hours relative to their male counterparts.

5 Conclusion

In this paper, we analyzed the effects of a relative performance feedback incentive through a field experiment. Our findings reveal that untreated students often misplace themselves in the grade distribution. Poor-performing students tend to

[11]While we do have information on the *number of days* a student retains borrowed books at home, we regard this as a rather noisy measure of study effort. This is due to its correlation with factors such as forgetting to return the books or unnecessarily prolonged possession. We posit that responsible students are more likely to borrow more books and to return them promptly. Indeed, raw correlations indicate a *positive* relationship between grades and the number of borrowed books ($p < 0.000$), while a *negative* correlation is observed with the number of days a student keeps the books at home ($p = 0.052$). This suggests that better-performing students tend to return the books earlier. Consequently, we assert that the number of days a student retains the books at home may not serve as a reliable proxy for assessing study effort, but rather be correlated with underperformance.

overreport their placement, while high-achieving students, especially women, tend to underplace themselves. The information provided to a randomly selected group of university students, significantly improved their knowledge about their relative academic performance.

We found asymmetric effects of information feedback for men and women. Treated men report higher satisfaction with their GPA after learning their ranking, whereas treated women express relative dissatisfaction with their academic performance reflected in their ranking. Over one or two years, these negative subjective satisfaction effects translate into worse objective academic performance for women compared to men.

We explore two possible explanations for these findings. First, women may be less competitive than men and, as a result, may adversely react to the competitive nature of the treatment, which requires students to outperform their peers to advance in the ranking. Second, women may invest more effort in the production function of academic performance, as reflected in library records and study hours. If the marginal cost of effort is increasing, female students may find it increasingly costly to advance or maintain their position relative to competing men.

At the project's inception, we considered the possibility of making the ranking public. To assess this, the short-term Satisfaction Survey's final question asked students about their comfort level if *others* knew their exact placement. Surprisingly, we found a contrast between men and women in the treated and control groups. While control women reported not caring about what others would think, treated women expressed the contrary, indicating that they would mind (more than men) if the ranking were made public. Thus, it appears that, after treatment, treated women change how they react to the ranking system along this dimension as well.

In light of the results of the pilot program, and before scaling the project to every student, the university administration decided to make several adjustments. They changed the name of 'ranking' to 'academic trajectory' to reduce the competitive element of the new tool. They made access to the system optional (an opt-in design) and excluded freshmen from this information, allowing them to settle into the university before offering the relative performance feedback from year two onwards. Importantly, the adoption of the tool is not implemented university-wide by the President's Office but left to the discretion of school deans.

We emphasize that more information is not always beneficial for everyone, and policymakers should carefully assess the pros and cons. Additionally, the way and context in which information is delivered may matter. The different reactions of males and females, sometimes opposed, are also factors to consider, which may be masked in an average treatment effect.

Bibliography

Azmat, G., M. Bagues, A. Cabrales, and N. Iriberri (2019): "What You Don't Know... Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education," *Management Science*, 65, 3714–3736.

Bago d'Uva, T., M. Lindeboom, O. Donnell, and E. van Doorslaer (2011): "Slipping anchor? Testing the vignettes approach to identification and correction of reporting heterogeneity," *Journal of Human Resources*, 46, 875–906.

Beegle, K., K. Himelein, and M. Ravallion (2012): "Frame-of-reference bias in subjective welfare," *Journal of Economic Behavior Organization*, 81, 556–570.

Benoît, J.-P. and J. Dubra (2011): "Apparent Overconfidence," *Econometrica*, 79, 1591–1625.

Bertrand, M. (2011): "Chapter 17 - New Perspectives on Gender," Elsevier, vol. 4 of *Handbook of Labor Economics*, 1543–1590.

Brade, R., O. Himmler, and R. Jäckle (2022): "Relative performance feedback and the effects of being above average — field experiment and replication," *Economics of Education Review*, 89, 102268.

Bursztyn, L. and R. Jensen (2015): " How Does Peer Pressure Affect Educational Investments? *," *The Quarterly Journal of Economics*, 130, 1329–1367.

Card, D., A. Mas, E. Moretti, and E. Saez (2012): "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction," *American Economic Review*, 102, 2981–3003.

Carneiro, P., Y. Cruz-Aguayo, F. Salvati, and N. Schady (2023): "The Effect of Classroom Rank on Learning Throughout Elementary School: Experimental Evidence from Ecuador," *Journal of Labor Economics*, 0, null.

Croson, R. and U. Gneezy (2009): "Gender Differences in Preferences," *Journal of Economic Literature*, 47, 448–74.

Cruces, G., R. Perez-Truglia, and M. Tetaz (2013): "Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment," *Journal of Public Economics*, 98, 100–112.

Denning, J. T., R. Murphy, and F. Weinhardt (2023): "Class Rank and Long-Run Outcomes," *The Review of Economics and Statistics*, 105, 1426–1441.

Eckel, C. C. and P. J. Grosssman (2008): "Men, women and risk aversion: experimental evidence." Amsterdam: Elsevier, Handbook of Experimental Economics Results.

Fischer, M. and V. Wagner (2023): "Do timing and reference frame of feedback influence high-stakes educational outcomes?" *Economics of Education Review*, 94, 102379.

Frey, B. S. and A. Stutzer (2002): "What Can Economists Learn from Happiness Research?" *Journal of Economic Literature*, 40, 402–435.

King, G., C. J. L. Murray, J. A. Salomon, and A. Tandon (2004): "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research," *American Political Science Review*, 98, 191–207.

Kruger, J. and D. Dunning (1999): "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments." *Journal of Personality and Social Psychology*, 77, 1121–1134.

Luttmer, E. F. P. (2005): "Neighbors as Negatives: Relative Earnings and Well-Being*," *The Quarterly Journal of Economics*, 120, 963–1002.

Moore, D. A. and P. J. Healy (2008): "The trouble with overconfidence." *Psychological Review*, 115, 502–517.

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

Tran, A. and R. Zeckhauser (2012): "Rank as an inherent incentive: Evidence from a field experiment," *Journal of Public Economics*, 96, 645–650.

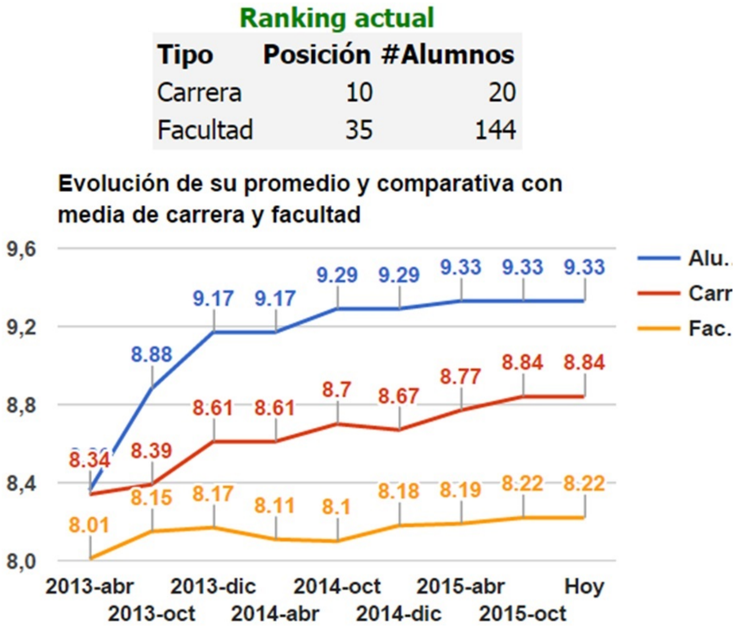
Appendices

Figure A.1: Transcript of grades

Código	Materia	Año cursada	Sem.	Notas		Notas y fechas de exámenes				Creds.
				Final	Curso	Per. 1	Per. 2	Per. 3	Per.Extr.	
01.Obligatorias generales =CP										
00102988	Algebra	2013	I	6	6	NA 3	6			7,5
00102987	Cálculo Básico	2013	I	7	5	NA 5	8			6
00100139	Introducción a la Contabilidad	2013	I	6	7	6				9
00100227	Introducción a la Economía	2013	I	7	7	6				6
00100213	Principios de administración	2013	I	7	7	7				6
00101641	Comunicación Profesional	2013	II	7	6	8				3
00102989	Cálculo	2013	II	7	4	NP	NA 4	10		9
00100216	Derecho privado I	2013	II	6	6	6				6
00100232	Macroeconomía I	2013	II	6	4	8				6
00100050	Antropología	2014	I	8	6	10				6
00100229	Microeconomía I	2014	I	8	10	7				6
00100739	Probabilidad	2014	I	NP	5	NP	NP			7,5
00101642	Etica profesional I	2014	II	9	9	9				4,5
00100735	Matemática financiera	2014	II	10	11	8				6
00100739	Probabilidad	2014	II	8	9	7				7,5
00100322	Finanzas de la empresa I	2015	II	8	9	NP	NA 5	8		8
00103626	Pasantía Social	2015	II	12		12				1,5
02.Obligatorias específicas =CP										
00100145	Contabilidad básica	2013	II	6	5	6				9
00100148	Contabilidad intermedia	2014	I	8	9	7				9
00100217	Derecho privado II	2014	I	7	7	8				6
00100152	Contabilidad de costos	2014	II	10	9	NA 5	11			9
00100215	Derecho laboral	2014	II	6	5	6				6
00100154	Contabilidad avanzada	2015	I	7	8	6				12
00100151	Contabilidad de gestión	2015	I	11	10	11				6
00100203	Costos para la toma de decisiones	2015	I	8	8	8				6
00100220	Legislación tributaria	2015	I	10	9	10				6
00100204	Control interno	2015	II	6	7	6				3
00100159	Teoría contable	2015	II	9	8	10				9
00100223	Técnica tributaria I	2015	II	8	5	NA 4	10			8
00100133	Auditoría I	2016	I	8	8	8				9
00100318	Banca y bolsa	2016	I	8	8	9				6
00100324	Finanzas de la empresa II	2016	I	8	8	8				8
00100225	Técnica tributaria II	2016	I	7	7	6				8
03.Electivas ciencias sociales										
00100751	Historia contemporánea	2013	II	10	10	10				4,5
08.Electivas generales =CP										
00100466	Marketing I	2013	II	9	7	10				6
00101728	Informática Intermedia	2014	I	11	11	11				4,5
00102128	Preparation FCE I	2015	I	9	9	9				6
00102129	Preparation FCE II	2015	II	9	9	9				6
00101732	Contabilidad Sector Agrop. e Industrial	2016	I	10	10		10			4,5
00104627	Procesos de negocio con SAP	2016	I	12	12	12				4,5
19.Prácticas profesionales = FCE										
00100791	Práctica Profesional I	2014	I	8		8				10
										Total de créditos obtenidos: 264
										Promedio de Aprobaciones: 8,2 **Promedio General: 7,7**

Notes: This is an example of a grade transcript for a student who entered the university in 2013 and has taken exams until the second semester of 2016, achieving 264 credits with a GPA of 8.2.

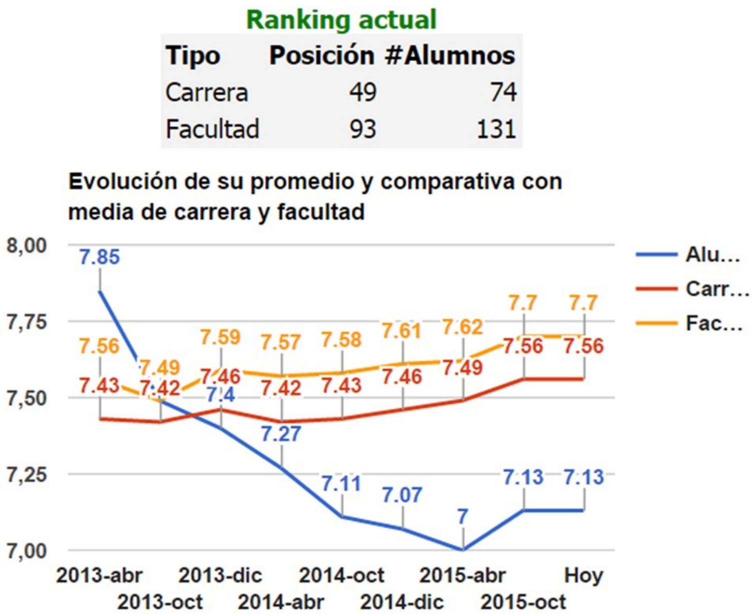
Figure A.2: Ranking treatment: Example 1



Ranking actual

Tipo	Posición	#Alumnos
Carrera	10	20
Facultad	35	144

Figure A.3: Ranking treatment: Example 2



Ranking actual

Tipo	Posición	#Alumnos
Carrera	49	74
Facultad	93	131

Notes. This Appendix shows the first two components of the ranking information system as illustrated in Figure 1, providing information for two actual students. The first student received information not only about his current ranking (10 out of 20), but also that he has been improving since April 2013. Conversely, the second student learned not only that he is placed 49 out of 74 students, but also that he has been decreasing relative to his peers.

Email sent to treatment and control group

Dear students:
We recommend that you carefully plan your upcoming exams and the forthcoming semester. For this, please keep in mind the following details:

- exam dates (available on xxx),
- the modification period (xxx),
- your academic progress and the curriculum grid (xxx),
- ***Starting today, you can access a ranking of your performance at the school.***

Best regards, XXX

Original wording in Spanish:
Estimados alumnos:
Te recomendamos planificar bien tus exámenes y el semestre que viene. Para eso ten en cuenta:

- las fechas de exámenes (disponibles en xxx),
- el período de modificaciones (xxx),
- tu escolaridad y la grilla de avance académico (xxx),
- ***También tienes disponible a partir de hoy un ranking de tu desempeño en la facultad.***

Saludos, XX

Notes:

1. The email was sent to treated and control students, with just a slight variation. The email sent to treated students included the text in italics, letting them know that they had available the ranking of their academic performance at the University.
2. The xxx replace the links to specific intranet web pages.

Table A.1: Descriptive Statistics (continuation of Table 1)

	(1) Treated	(2) Control	(3) Difference (1-2)	(4) Standard Error	(5) p-value	(6) Obs.
Schools and degrees						
17 *School of Business*	0.57	0.56	0.01	(0.03)	0.65	1,048
18 *School of Engineering*	0.26	0.27	-0.01	(0.03)	0.85	1,048
19 *School of Law*	0.16	0.17	-0.01	(0.02)	0.70	1,048
20 *Business Administration*	0.11	0.09	0.02	(0.02)	0.42	1,048
21 *Accountancy*	0.29	0.29	0.00	(0.03)	0.89	1,048
22 *Economics*	0.08	0.08	0.00	(0.02)	0.90	1,048
23 *International Business*	0.08	0.09	-0.01	(0.02)	0.67	1,048
24 *Civil Engineering*	0.12	0.12	0.00	(0.02)	0.99	1,048
25 *Industrial Engineering*	0.11	0.11	0.00	(0.02)	0.92	1,048
26 *Computer Engineering*	0.02	0.02	0.00	(0.01)	0.97	1,048
27 *Telecomunications*	0.03	0.03	0.00	(0.01)	0.80	1,048
28 *Law*	0.12	0.13	-0.01	(0.02)	0.76	1,048
29 *Notary*	0.04	0.04	0.00	(0.01)	0.82	1,048
Place in the distribution of grades						
30 *Decile 1 grades_major*	0.16	0.16	0.00	(0.02)	0.83	1,047
31 *Decile 2 grades_major*	0.13	0.14	-0.01	(0.02)	0.51	1,047
32 *Decile 3 grades_major*	0.11	0.09	0.02	(0.02)	0.21	1,047
33 *Decile 4 grades_major*	0.09	0.10	-0.01	(0.02)	0.53	1,047
34 *Decile 5 grades_major*	0.10	0.08	0.02	(0.02)	0.33	1,047
35 *Decile 6 grades_major*	0.10	0.11	-0.01	(0.02)	0.69	1,047
36 *Decile 7 grades_major*	0.08	0.08	0.00	(0.02)	0.93	1,047
37 *Decile 8 grades_major*	0.08	0.09	-0.01	(0.02)	0.60	1,047
38 *Decile 9 grades_major*	0.08	0.09	-0.01	(0.02)	0.68	1,047
39 *Decile 10 grades_major*	0.06	0.06	0.00	(0.01)	0.74	1,047
40 *Decile 1 grades_school*	0.16	0.15	0.01	(0.02)	0.82	1,047
41 *Decile 2 grades_school*	0.12	0.12	0.00	(0.02)	0.92	1,047
42 *Decile 3 grades_school*	0.09	0.10	-0.01	(0.02)	0.77	1,047
43 *Decile 4 grades_school*	0.11	0.09	0.02	(0.02)	0.40	1,047
44 *Decile 5 grades_school*	0.11	0.10	0.01	(0.02)	0.76	1,047
45 *Decile 6 grades_school*	0.08	0.09	-0.01	(0.02)	0.68	1,047
46 *Decile 7 grades_school*	0.09	0.10	-0.01	(0.02)	0.53	1,047
47 *Decile 8 grades_school*	0.09	0.08	0.01	(0.02)	0.56	1,047
48 *Decile 9 grades_school*	0.08	0.09	-0.01	(0.02)	0.60	1,047
49 *Decile 10 grades_school*	0.07	0.07	0.00	(0.02)	0.94	1,047

Notes. This Table is the continuation of Table 1 in the main text, which contains lines 1-16. Here we show the rest of the variables used in the randomization. The t-test for the difference in means is calculated with an OLS regression with robust standard errors. *** p<0.01, ** p<0.05, * p<0.1.

Table A.2: Characteristics of the students who answered the satisfaction survey

Variable	Completed survey	Did not complete	Difference	
Treated	0.51	0.46	0.05	
	(0.02)	(0.04)	(0.04)	
Female	0.45	0.43	0.03	
	(0.02)	(0.04)	(0.04)	
Courses (number)	26.39	24.60	1.79	
	(0.60)	(1.48)	(1.60)	
Cumulative GPA	7.58	7.23	0.35	
	(0.05)	(0.13)	(0.14)	**
Approved courses	21.89	20.21	1.67	
	(0.53)	(1.34)	(1.44)	
Credits earned	143.14	127.45	15.68	
	(3.41)	(8.16)	(8.83)	*
Number of majors	1.07	1.06	0.01	
	(0.01)	(0.02)	(0.02)	
School of origin	0.29	0.27	0.03	
	(0.02)	(0.03)	(0.04)	
Montevideo	0.65	0.72	-0.07	
	(0.02)	(0.03)	(0.04)	*
Large Scholarship	0.24	0.22	0.02	
	(0.01)	(0.03)	(0.03)	
Cohort	2011.81	2011.86	-0.05	
	(0.06)	(0.14)	(0.15)	
Decile of GPA (Major)	4.95	4.20	0.75	
	(0.10)	(0.21)	(0.23)	***
Decile of GPA (School)	5.09	4.41	0.69	
	(0.10)	(0.22)	(0.24)	***
Observations	861	187		

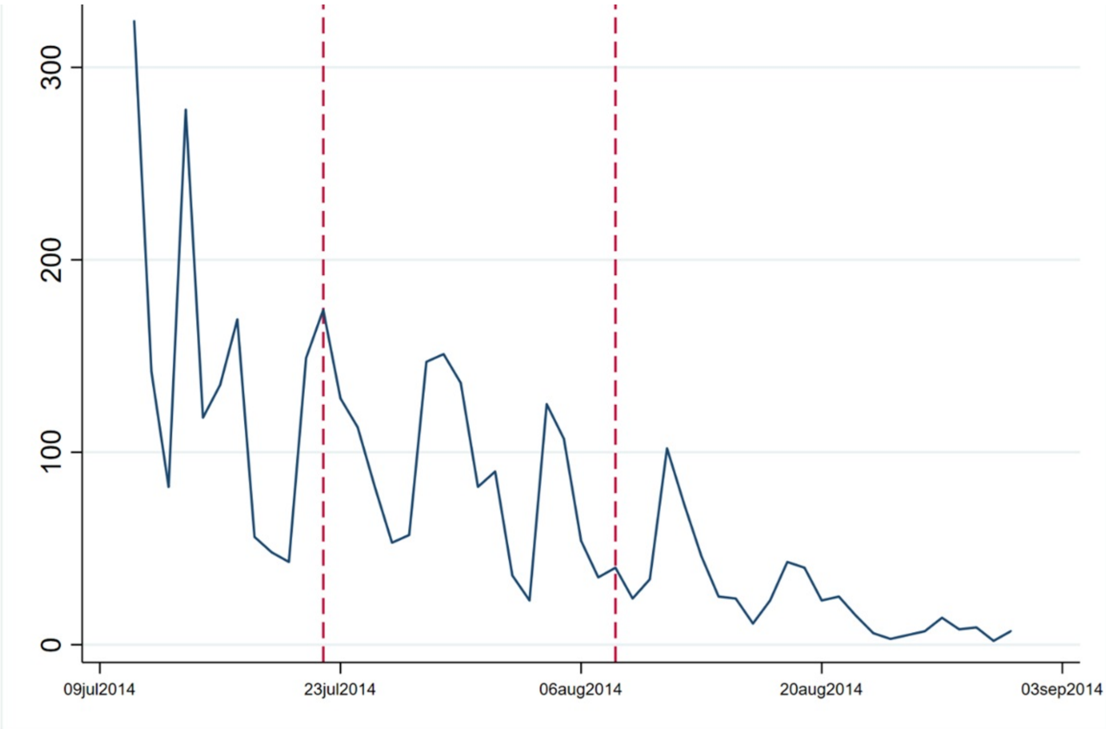
Notes: The figures show point estimates from regression-discontinuity models with a linear specification of the forcing variable and controlling for grade, year and a quadratic of school size. The three discontinuity samples used are +/- 1.5, +/- 1.0 and +/- 0.5 points of the running variable.

Table A.3: Balance in survey response (after attrition)

	(1) Treated	(2) Control	(3) difference (1-2)	(4) Standard Error	(5) p-value	(6) Obs.
Student characteristics						
1 *Female*	0.46	0.44	0.02	(0.03)	0.65	861
2 *Courses*	26.45	26.33	0.12	(1.21)	0.92	861
3 *Cumulative GPA*	7.55	7.61	-0.06	(0.10)	0.57	698
4 *Approved courses*	21.77	22.01	-0.24	(1.06)	0.82	861
5 *Credits earned*	142.70	143.60	-0.90	(6.84)	0.89	861
6 *Number of degrees*	1.09	1.06	0.03	(0.02)	0.10	861
7 *School of origin*	0.31	0.28	0.03	(0.03)	0.34	861
8 *Montevideo*	0.66	0.64	0.02	(0.03)	0.53	861
9 *Large Scholarship*	0.25	0.24	0.01	(0.03)	0.82	861
10 *Cohort 2008*	0.04	0.04	0.00	(0.01)	0.87	861
11 *Cohort 2009*	0.06	0.05	0.01	(0.01)	0.66	861
12 *Cohort 2010*	0.13	0.13	0.00	(0.02)	0.98	861
13 *Cohort 2011*	0.21	0.22	-0.01	(0.03)	0.66	861
14 *Cohort 2012*	0.21	0.21	0.00	(0.03)	0.99	861
15 *Cohort 2013*	0.19	0.16	0.03	(0.03)	0.26	861
16 *Cohort 2014*	0.18	0.20	-0.02	(0.03)	0.45	861

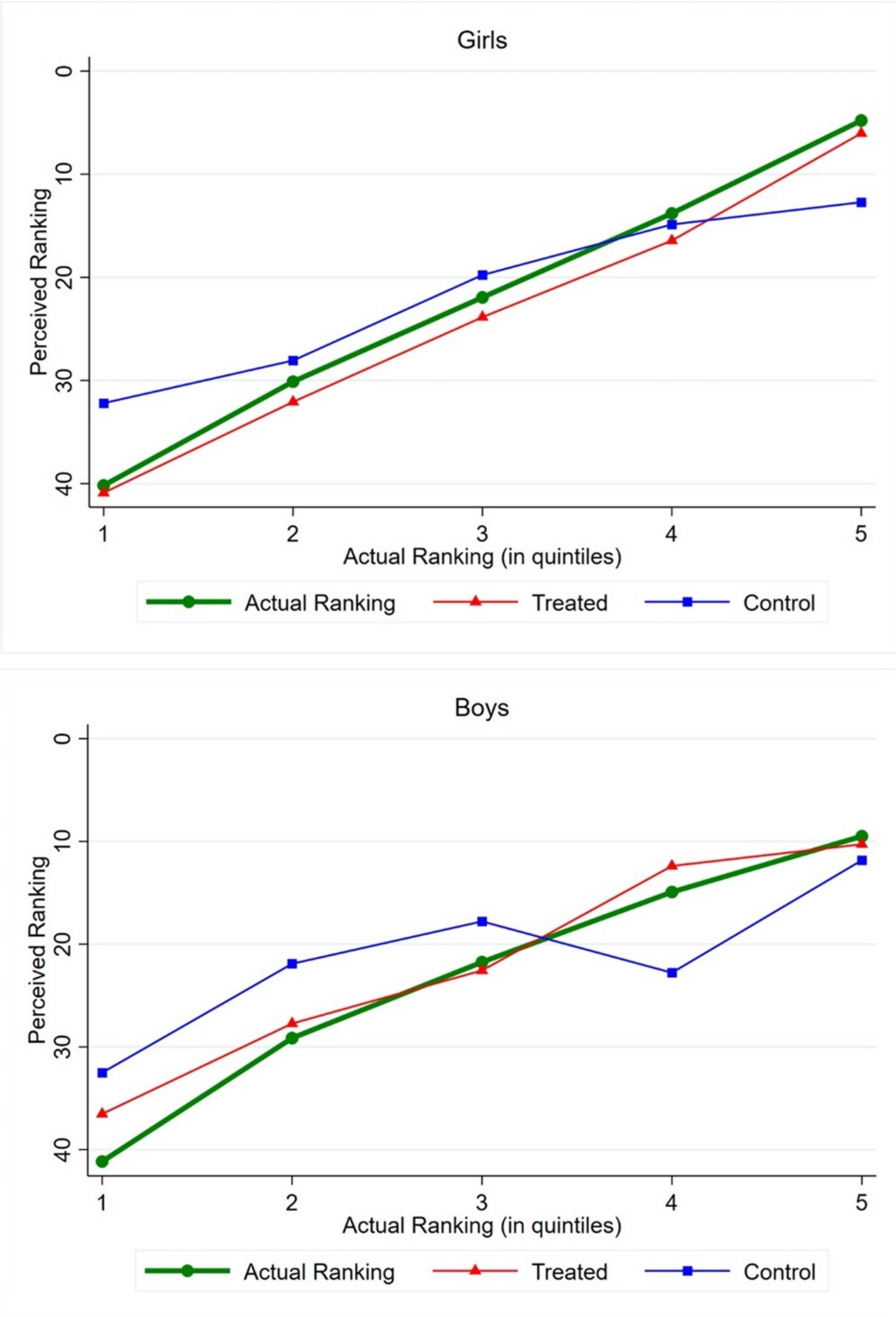
Notes. The difference in means is calculated with an OLS regression with robust standard errors. Of the 861 students, there are 443 (51.5%) in the treated group, and 418 (48.5%) in the control group. Balance was also performed by schools, majors and place in the distribution of grades (at the major and school level). These results are omitted from this table to ease the display of the main results. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure A.4: Access to the Ranking System before and after the time of the satisfaction survey



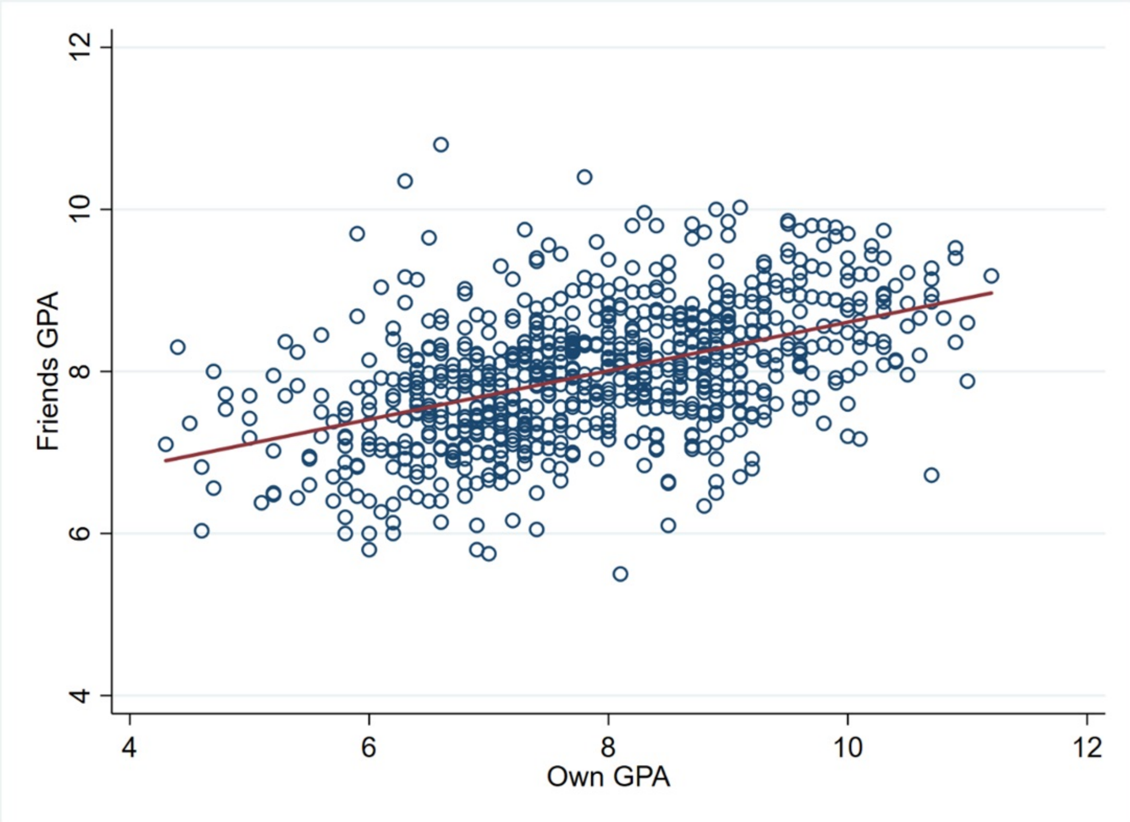
Note: This figure is an enlarged view of Figure 3, concentrating on the period when the student satisfaction survey was administered, delineated by the two vertical lines. The continuous line represents the daily access count to the ranking system. The system was inaugurated on July 11, 2014.

Figure A.5: Actual vs Perceived relative performance by treatment status and gender



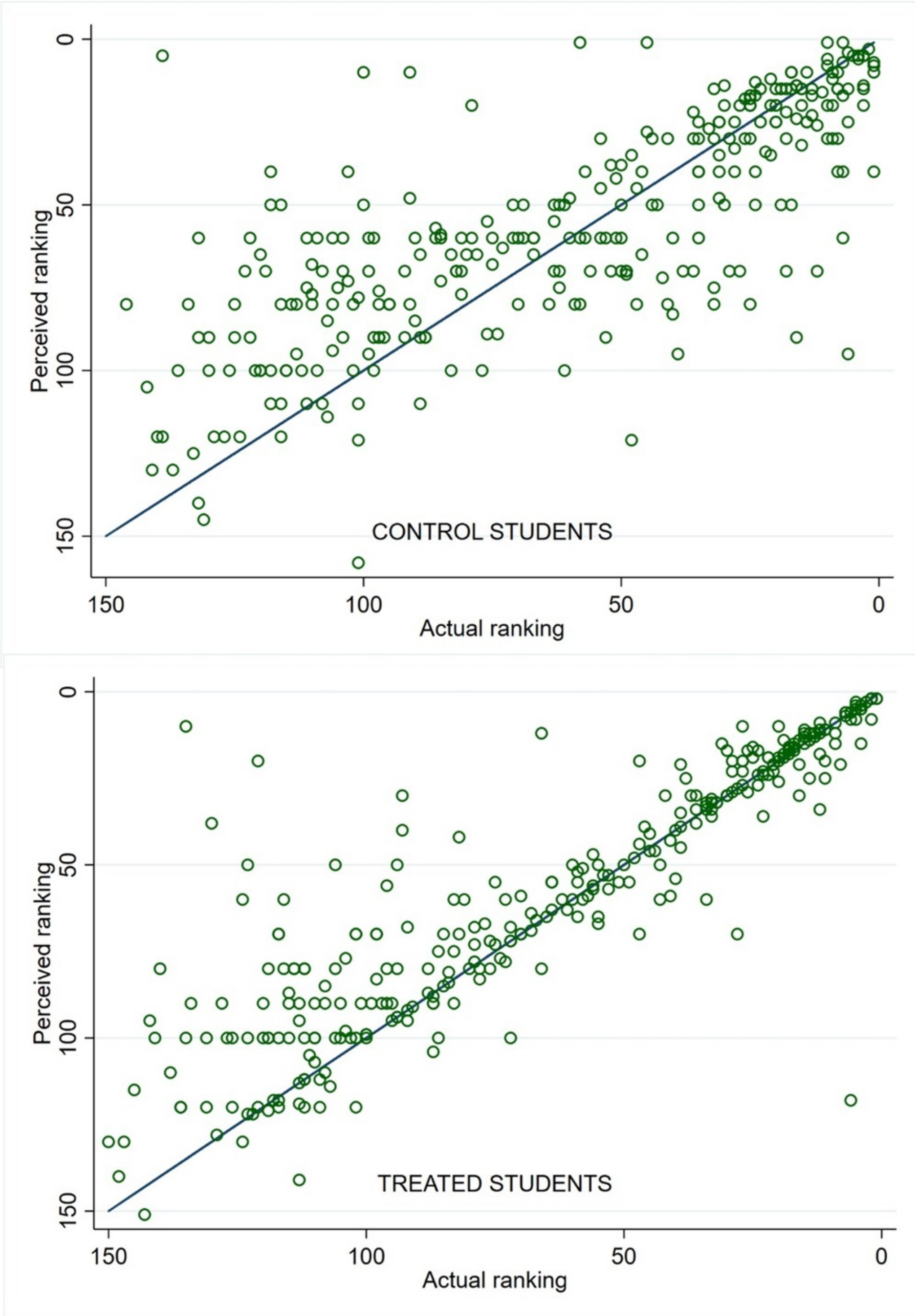
Notes. Treated girls generally exhibit a tendency to modestly underestimate their placement across the entire distribution. Control girls appear slightly more accurate in self-placement compared to control boys. Among the untreated group, the boys who most accurately assess their placement are the high performers. The overall pattern of Figure 4, where lower-performing individuals (quintiles 1 and 2) tend to overestimate their placement and higher-performing individuals (quintiles 4 and 5) tend to underestimate, is broadly consistent across genders.

Figure A.6: Friend Selection: Comparison of Own GPA and Friends' GPA



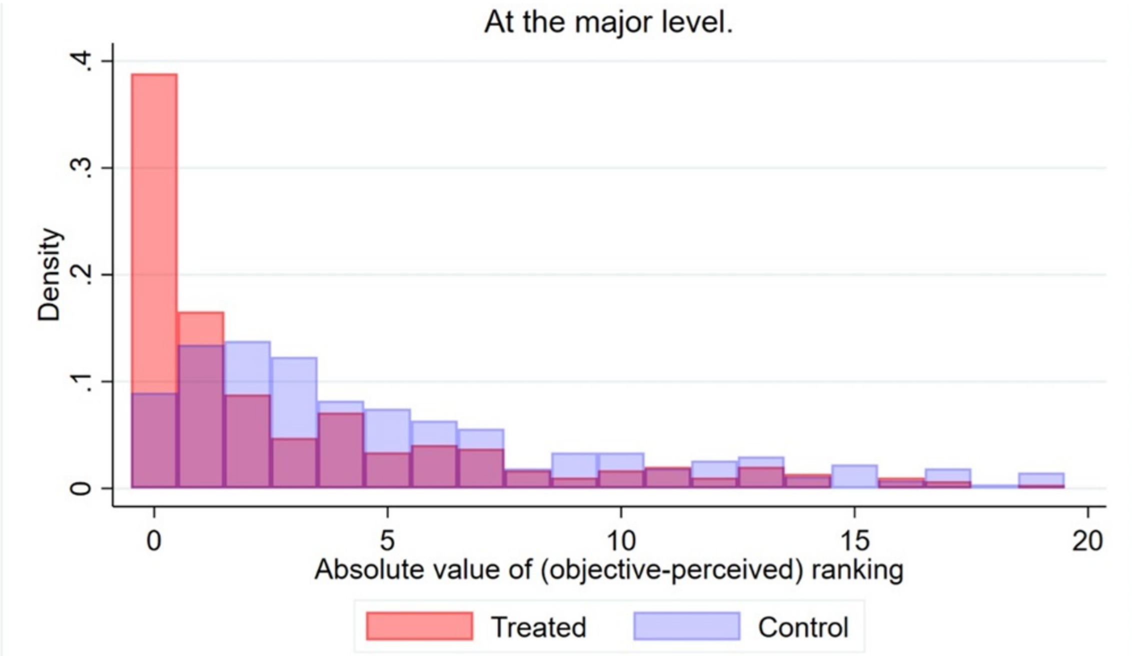
Notes: This figure illustrates the correlation between a student's own GPA and their friends' GPA. An additional point in friends' GPA is associated with an increase of 0.77 points in the student's own GPA. A total of 785 students on the x-axis provided information about their network of friends (n=3,389). The y-axis represents the average GPA of their friends for each reporting student.

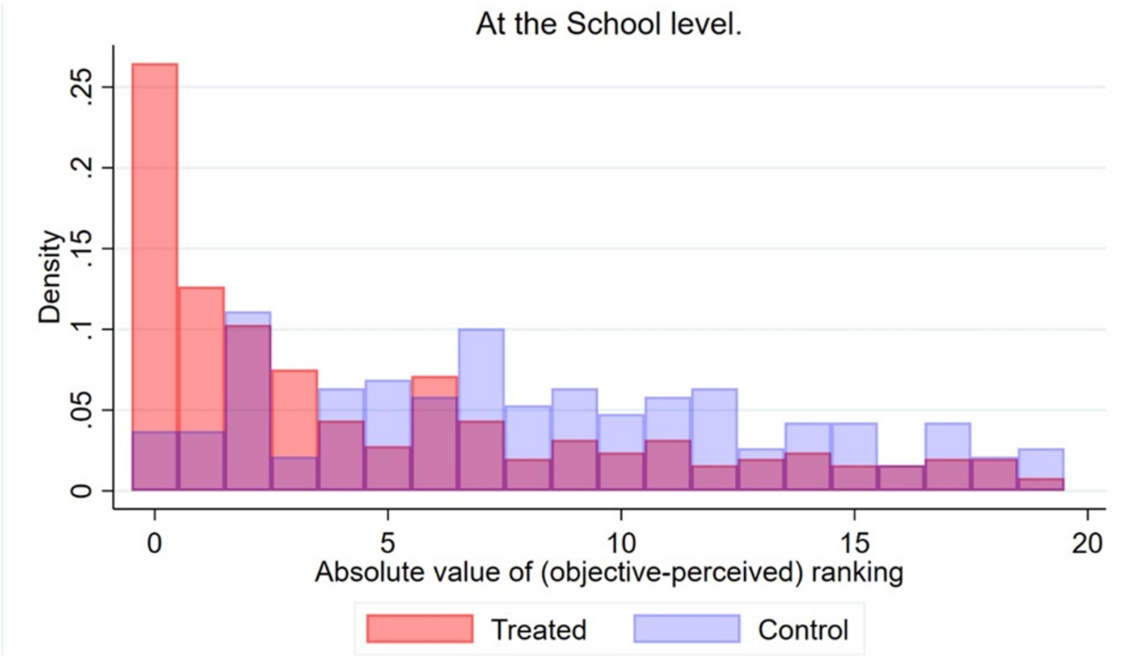
Figure A.7: Actual vs perceived relative performance. Individual answers.



Notes. These figures illustrate the relationship between perceived and actual ranking. The perceived ranking (Y-axis) is reported by the students in the Satisfaction Survey, where each student indicates their placement in the distribution of grades (ranking), with 1 representing the student with the highest GPA. The actual ranking (X-axis) is derived from the administrative registries of the University. Students from the school of engineering were excluded from this analysis, as explained in the main text. These two figures constitute the raw data used for plotting Figure 4 in the main text.

Figure A.8: Accuracy of students report of their placement in the grade distribution



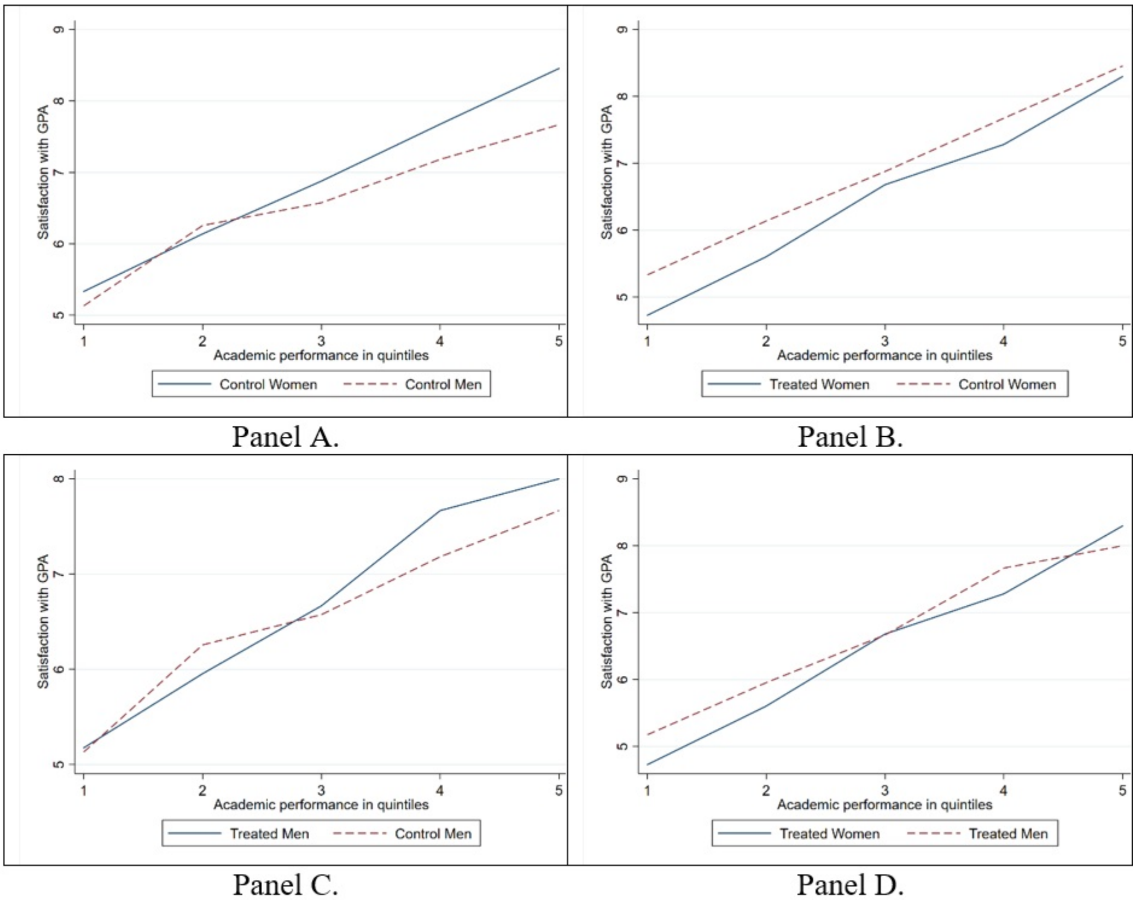


Note. These figures display the histogram of the (absolute) difference between perceived and actual ranking, by treatment status, at both the major and school levels. The absolute value of the difference is limited to < 20 for improved visualization.

Figure A.9: Second question about satisfaction

Notes. This picture shows the beginning of the second question about satisfaction. The translation of the Spanish text is as follows: *'Imagine a staircase with steps numbered from 1 at the bottom to 10 at the top. The top of the ladder represents the best academic life for you and the bottom the worst academic life for you.'* At the top of the ladder it says: 'Completely satisfied', while at the bottom it says 'Completely dissatisfied'. This introduction was followed by the question about satisfaction with academic performance on a 10-point scale, and then by the four hypothetical situations for the anchoring vignettes.

Figure A.10: Visual representation of results from Table 3, col 3.



Notes. These figures illustrate the underlying satisfaction comparisons outlined in Table 3, column 3, now displayed across the GPA distribution. The four figures show that satisfaction with GPA is increasing in actual GPA. Panel A depicts the difference between satisfaction lines for women and men in the control group (an average of β_3 from Equation 1, estimated to be 0.319). In Panel B, the average difference between treated women and control women is the sum of parameters $\beta_1 + \beta_2$, estimated to be -0.392 (significant at the 5% level), meaning that treatment decreased satisfaction for women, irrespective of their position in the distribution of grades. Panel C suggests that the information treatment increased satisfaction for men in the upper quintiles of the grade distribution (good news for them). The estimated average difference for men β_2 is not statistically different from zero (Table 3, column 3, coefficient for treated). Finally, panel D shows the difference between treated women and treated men (with an average of $\beta_1 + \beta_3$, which is not statistically different from zero).

The average satisfaction reported by treated women was 6.5, while the control group reported 6.9. For males, treated students reported an average satisfaction of 6.3, while control students reported 6.2. The coefficient β_1 is in essence an average difference in difference parameter, between males increasing their average satisfaction after treatment while women decreasing.

In summary, untreated women report higher satisfaction with their GPA than men, treated men (especially higher performers) experience increased satisfaction, while treated women report decreased satisfaction. The combined overall result reveals a relative decrease in satisfaction for treated women compared to treated men by $\beta_1 = -0.556$ (Table 3, col 3).

Table A.4: Treatment effect from pooling medium and long-term individual exam results from Table 4.

	(1)	(2)	(3)
	Exam grade	Took the exam	Passed the exam
*Treated*Woman*	-0.224*	-0.037**	-0.026*
	(0.122)	(0.016)	(0.016)
Treated	0.116	0.013	0.017
	(0.081)	(0.010)	(0.011)
Woman	0.313***	0.021*	0.021*
	(0.091)	(0.011)	(0.012)
Constant	6.421***	0.805***	0.751***
	(0.177)	(0.022)	(0.023)
Observations	23,877	27,515	23,877

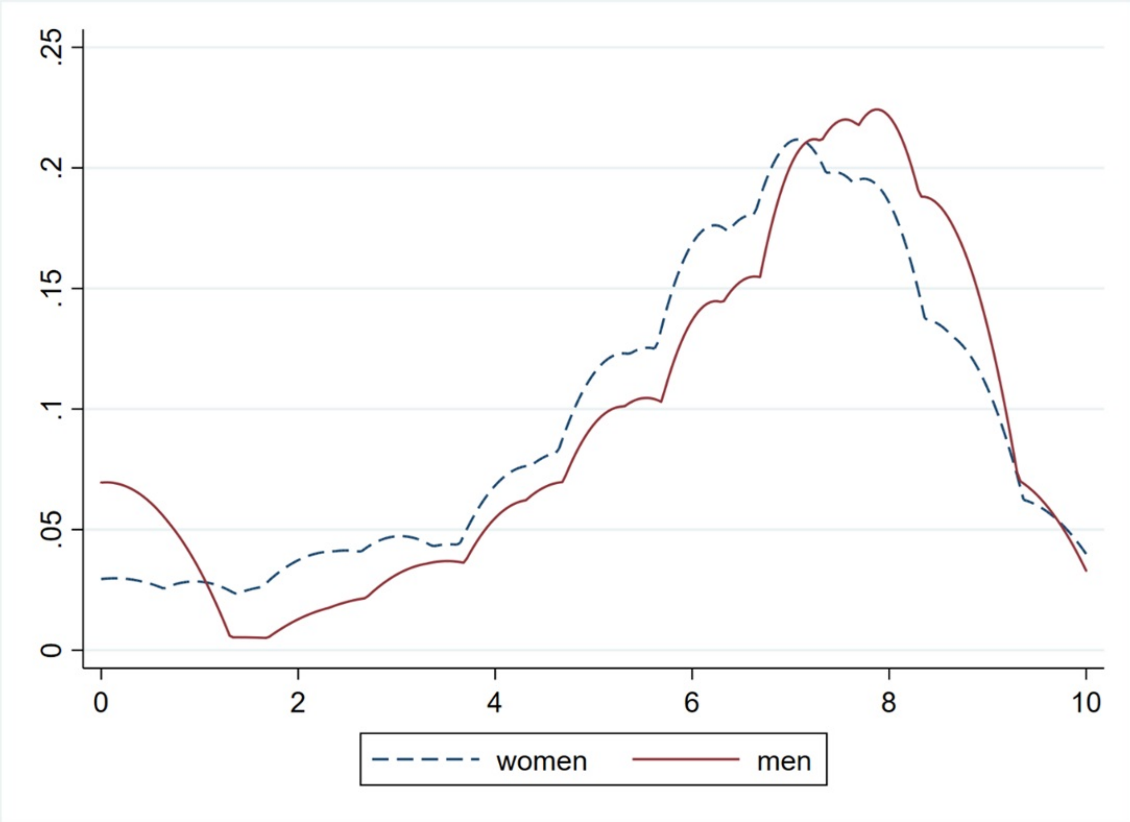
Notes. This table shows the results from pooling medium and long-term observations from Table 4. All regressions include pre-treatment controls as in Table 3 plus a dummy indicating if the exam belongs to the medium-term period. Standard errors are clustered at the student level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table A.5: Pre-treatment (placebo) impact

	(1)	(2)	(3)	(4)	(5)	(6)
	Exam grade		Took the exam		Passed the exam	
*Treated*Woman*	-0.136	-0.016	-0.017	-0.012	-0.010	0.012
	(0.178)	(0.065)	(0.022)	(0.015)	(0.024)	(0.010)
Treated	-0.038	-0.031	-0.007	-0.002	-0.004	-0.008
	(0.114)	(0.041)	(0.014)	(0.010)	(0.017)	(0.007)
Woman	0.781***	0.037	0.021	0.001	0.106***	-0.002
	(0.127)	(0.049)	(0.015)	(0.011)	(0.017)	(0.007)
Constant	7.210***	4.888***	0.843***	0.687***	0.754***	0.590***
	(0.081)	(0.356)	(0.010)	(0.043)	(0.012)	(0.052)
Observations	29,721	29,695	35,201	35,175	29,721	29,695
Controls	NO	YES	NO	YES	NO	YES

Notes. This table shows the results from the (future) treatment on pre-treatment academic outcomes at the exam level for the period July 2003 to June 2014. Treatment started in July 2014. All regressions include pre-treatment controls as in Table 3. Standard errors are clustered at the student level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Figure A.11: Self-reported competitiveness for men and women



Notes. This figure displays a kernel density estimation for the competitiveness measure, for men and women. Surprisingly, a considerable proportion of men reported a competitiveness score of 0. We lack a clear explanation for this anomaly. The results in the main text would have been more robust if we had excluded these peculiar answers.