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

José María Cabrera* and Alejandro Cid

Universidad de Montevideo

September 2026

Abstract

Does giving students precise information about where they stand improve academic outcomes and well-being? We study an individually randomized field experiment with 1,048 undergraduates in Uruguay. Treated students received private, daily-updated information about their position in cumulative-GPA distributions; controls retained access to the same absolute records without the comparison. The feedback reduced errors in perceived rank by about 45 percent. Its downstream effects, however, differed by gender. Women assigned to feedback reported lower satisfaction with their academic performance, were 7.2 percentage points more likely to exit their baseline major after one year, and after two years had passed 2.1 fewer courses and earned 11.9 fewer credits. Twelve years later, they were 10.3 percentage points less likely to graduate from their baseline major. Men did not display a comparable adverse pattern. The long-run gap was concentrated among students exposed upon entering university. Accurate comparative feedback can therefore achieve its informational objective without improving outcomes uniformly. Feedback systems should be assessed for their distributional consequences, not only the precision of the information they provide.

Keywords: relative performance feedback; rank; field experiment; higher education; gender; belief updating.

JEL codes: D83, I23, C93, J16.

1. Introduction

People evaluate outcomes not only in absolute terms but also relative to others. Workers care about how their wages compare with those of their colleagues, and students may care about where their grades place them among their peers. Institutions increasingly make such comparisons easier through rankings, percentiles, progress dashboards, and peer benchmarks. These systems may correct inaccurate beliefs and improve decisions. They may also change satisfaction, confidence, expectations, and the perceived cost of improving one's position. Whether more accurate comparative information helps its recipients is therefore an empirical question.

This paper studies that question in higher education. The project was inspired in part by Card et al. (2012), who showed that disclosing peer salaries reduced job satisfaction among workers paid below

*Acknowledgments. This paper is a revised version of a chapter of José María Cabrera's PhD dissertation at Erasmus University Rotterdam. José María Cabrera is grateful to the members of his dissertation committee—Olivier Marie, Trudie Schils, Hessel Oosterbeek, Jonneke Bolhaar, Bas ter Weel, and Robert Dur—for their comments and suggestions. He is especially grateful to his advisors, Anne Gielen and Dinand Webbink, for their guidance and support. We also thank seminar participants at Universidad de Montevideo, Erasmus School of Economics, Universidad de Vigo, the 2017 Meeting of the Economics of Education Association, and the 2016 Encuentro de la Sociedad de Economistas del Uruguay for helpful comments. Ethical approval for fieldwork was obtained in 2013 from the Board of Universidad de Montevideo (n° FCEE 01/13). The authors declare no conflicts of interest. All remaining errors are our own. Corresponding author: José María Cabrera, jmcabrera@um.edu.uy. Cid: acid@um.edu.uy.

the median but had little effect on higher-paid workers. Academic rank provides a related setting in which individuals observe their own outcome - grades and GPA - but may not know how that outcome compares with peers. Relative position may enter utility directly, or students may use it as a signal about ability, future academic success, or labor-market prospects. Correcting beliefs can therefore affect both immediate satisfaction and subsequent academic investment.

We conducted an individually randomized field experiment at Universidad de Montevideo. The design and information system were developed beginning in 2013, and the intervention was launched in July 2014. The analysis includes 1,048 undergraduates in the Schools of Economics, Engineering, and Law. All students could observe their courses, grades, and cumulative grade point average (GPA) in the university intranet. Treated students additionally received private access to information about their position in the cumulative-GPA distribution within their entry cohort and major and within their cohort and school. The information was updated daily. Control students retained access to the same absolute records but did not receive the comparative ranking.

The experiment follows the full sequence from information to outcomes. Administrative logs show whether students used the new tool. A survey conducted shortly after launch elicits perceived rank and two measures of satisfaction with academic performance, including one constructed with anchoring vignettes. Administrative records then measure exam taking, grades, course completion, credits, cumulative GPA, and exit from the baseline major one and two years later, as well as degree completion over twelve years. This combination lets us ask not only whether the intervention successfully corrected beliefs, but also whether that informational success translated into greater satisfaction, academic progress, persistence, or graduation.

It did correct beliefs. Control students display a patterned misperception of their relative performance: students near the bottom tend to place themselves too high, while students near the top tend to place themselves too low. The intervention reduces absolute errors in perceived position by 3.4 places within a major and 8.8 places within a school, approximately 48 and 45 percent of the corresponding control means. It also raises the probability of reporting the exact position by 29.7 percentage points at the major level and 19.5 points at the school level. Thus, treated students did not merely receive an additional university message; they learned materially more about where they stood.

The consequences differ by gender. On a five-point satisfaction scale, the treatment effect for men is positive but imprecise, while the interaction between treatment and woman is -0.294. The total effect for women is -0.161 points. On the ten-point measure adjusted with anchoring vignettes, the interaction is -0.540 and the total effect for women is -0.382 points. A placebo measure - satisfaction with the university's location - shows no analogous pattern. The two gender interactions remain statistically significant after Holm adjustment within the satisfaction family; the total effects for women are marginal under the same adjustment.

Administrative outcomes show a parallel but concentrated pattern. After one year, treated women are 7.2 percentage points more likely to have exited their baseline major. After two years, treated women have passed 2.1 fewer courses, earned 11.9 fewer credits, and are 7.3 percentage points more likely to have exited their baseline major. Twelve years after the intervention, treated women are 10.3 percentage points less likely to graduate from their baseline major and 9.0 points less likely to receive any new degree from the university. These long-run estimates are concentrated among students exposed upon entering university. The total effects for women on short-run progression and on the two degree outcomes remain significant after Holm adjustment within their respective families. The two-year effect on cumulative GPA is negative but less precise, and the total effects on exam grades and exam passing are not statistically distinguishable from zero. The evidence therefore does not support a claim of deterioration on every academic margin. It identifies a persistent pattern concentrated in

progression, persistence, and eventual degree completion.

Why might the same information produce different responses? Several processes are possible. Unfavorable news may lower confidence or perceived returns to further investment. Even favorable news may make social comparison more salient. Students may differ in their willingness to engage with competitive frames or in the marginal cost of increasing effort. Our data provide suggestive evidence relevant to these explanations: women borrowed more books and reported more study time before treatment, while a post-treatment measure of competitiveness does not display a robust gender difference on its original continuous scale. The survey also asked whether students would care if others knew their exact rank. Exposure to the tool increased this concern among women, suggesting that privacy and audience may be meaningful components of feedback design. None of these exercises separately randomizes a mechanism, so we interpret them as evidence about plausible explanations rather than causal mediation.

The paper contributes to the literature on causal disclosure of relative performance. Early studies show that feedback and rank incentives can improve effort in some settings and reduce performance in others (Azmat and Iriberri, 2010; Tran and Zeckhauser, 2012; Bandiera et al., 2015). The closest earlier university experiment is Azmat et al. (2019), which disclosed GPA deciles every six months and found a short-run performance decline together with greater satisfaction with course quality. Our intervention instead supplies private, daily-updated information about overall academic standing, directly measures rank misperceptions and satisfaction with the performance summarized by the rank, and links short-run academic progress to degree completion twelve years later. Gender heterogeneity is a central object of the design rather than an ancillary subgroup analysis.

Subsequent experiments reinforce the idea that feedback design is part of the treatment. Real-time feedback within a course can improve performance (Dobrescu et al., 2021), while responses to broader benchmarks depend on whether news is favorable and how the comparison is framed (Brade et al., 2022). Personalized feedback can also have effects that vary with course format and student background (Aucejo and Wong, 2025). Timing, reference frame, comparison groups, and distance from degree completion can alter effects as well (Fischer and Wagner, 2023; Chen et al., 2024; Brade et al., 2026). Macías Domínguez et al. (2026), in a later Spanish experiment, find positive effects of one-time percentile feedback, especially for women and initially optimistic students. The contrast in signs is informative: "relative feedback" is not a single treatment, and its consequences depend on content, frequency, prior beliefs, horizon, and recipients.

Our estimand also differs from the effect of occupying a particular ordinal rank. Research using variation in classroom composition or assignment shows that rank can affect confidence, educational investment, attainment, and later outcomes (Elsner and Isphording, 2017; Murphy and Weinhardt, 2020; Denning et al., 2023; Carneiro et al., 2025). Those studies explain why relative position may matter conditional on absolute achievement. Here, absolute academic records are held fixed and the intervention varies whether the institution reveals comparative information that students perceive imperfectly.

Finally, the paper relates to evidence on gender, competition, and responses to feedback (Niederle and Vesterlund, 2007; Wozniak et al., 2014; Jalava et al., 2015; Czibor et al., 2020). The interaction between assignment and a predetermined gender indicator has a causal interpretation because treatment is randomized and gender is measured before treatment. At the same time, the study does not imply that all women respond alike or identify an immutable gender mechanism. It shows that an average treatment effect would conceal an important distributional difference in this setting.

The study first circulated publicly as Cabrera and Cid (2017), before much of the subsequent exper-

imental literature. Brade et al. (2022) and Villeval (2020) cite that version among the earlier field studies of broad relative-performance feedback. Our claim is not that this is the first paper about rank. It is that the experiment provides early evidence on a distinctive combination of private and precise disclosure, direct belief measurement, subjective responses, short-run academic progression, twelve-year degree completion, and gender heterogeneity.

The remainder of the paper is organized as follows. Section 2 describes the setting, intervention, experimental design, and data. Section 3 examines use of the system and belief updating. Section 4 reports effects on satisfaction and academic outcomes. Section 5 develops possible explanations and implications for feedback design. Section 6 concludes.

2. Experimental Design and Data

2.1. Institutional setting and the ranking tool

Universidad de Montevideo is a private university in Uruguay. Before the experiment, students could consult an official transcript through the university intranet. The transcript listed courses, grades, cumulative credits, and GPA, but it did not reveal the student's position in the distribution of grades. Students therefore had precise information about their absolute academic record but no official measure of relative standing. Appendix G reproduces a transcript, two additional treatment trajectories, and the message used to announce the new tool.

In mid-2013, while the university's information-technology team was developing a comparative academic tool, we proposed evaluating it as a randomized pilot and participated in the treatment design. The pilot covered the Schools of Economics, Engineering, and Law. Treated students gained access to a private page in the same intranet. Control students did not receive that page and continued to see only their usual transcript.

Figure 1 reproduces an actual page shown to treated students. For Economics and Law, the first box displayed the current exact position and the size of the comparison group at both the cohort-major and cohort-school levels. The example tells the student that she ranks 30th among 74 students in her major and 55th among 131 students in her school. Engineering used deciles because some of its cohorts and majors were organized differently. The remainder of the page plotted the student's cumulative GPA against the major and school means and showed how her ordinal positions had evolved across semesters. The endpoint labeled "Hoy" ("Today") was recalculated daily as grades entered the system; earlier points were fixed at the relevant academic dates. The tool therefore conveyed both a current comparison and a trajectory that had not previously been available.

The comparison groups were relatively small: an average of 37 students at the cohort-major level and 84 at the cohort-school level. This made relative information potentially valuable but also raised the possibility that students already had good informal knowledge of peer performance. The experiment therefore included direct questions about students' perceived positions and about information learned from classmates. Section 3 returns to those design challenges.

2.2. Assignment, analysis sample, and balance

The experimental sample was constructed from students enrolled in the participating schools and majors. Before the final implementation-eligibility check, this universe contained 1,067 students and 1,131 student-major records. That check excluded 19 assigned students: seven had no academic activity in 2014, with their last recorded activity in 2012 or 2013, and twelve had no calculable GPA-based rank in the July 23 implementation snapshot. These criteria concern eligibility for active comparative

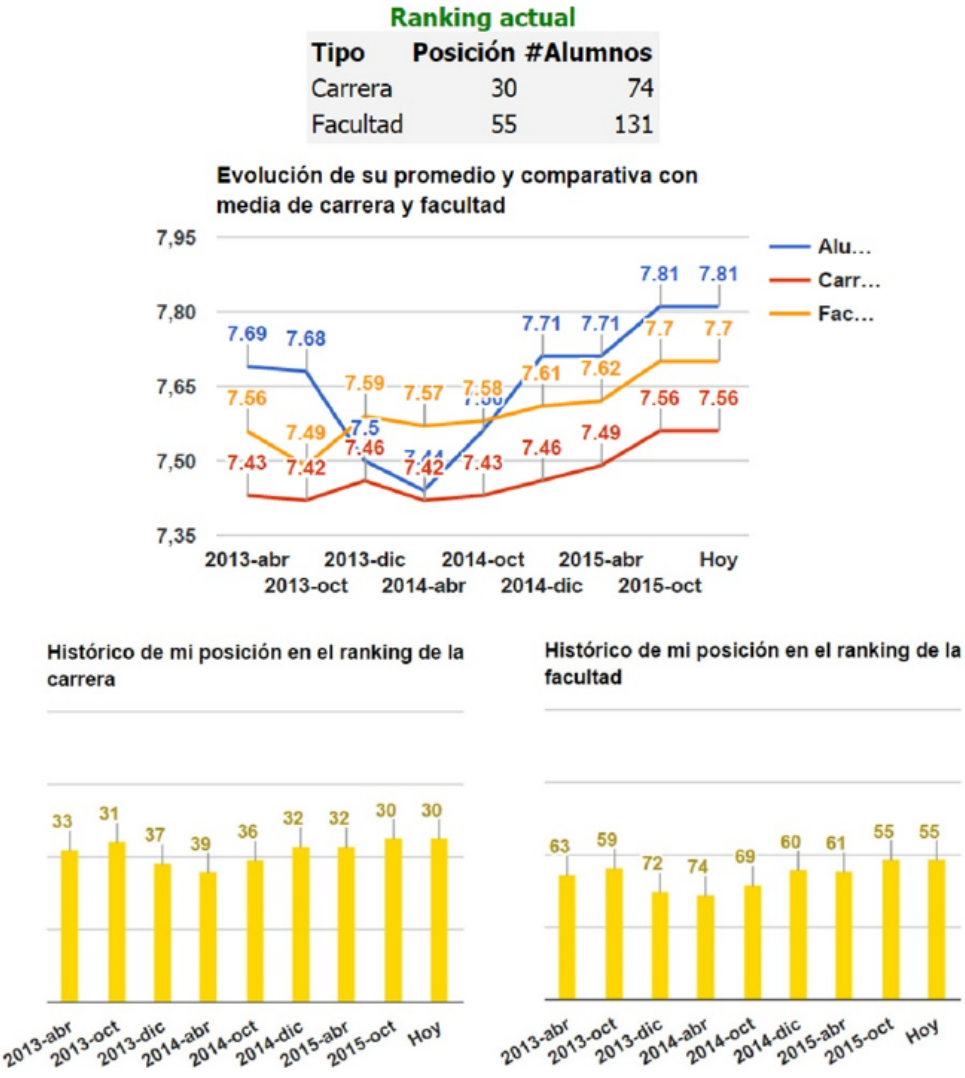


Figure 1. Information shown to treated students.

Notes: This is an actual treatment page for Economics and Law; headings remain in Spanish because they reproduce what participants saw. Ranking actual means current ranking; Carrera and Facultad mean major and school; Evolución de su promedio y comparativa con media de carrera y facultad plots the student's GPA against the major and school means; and the two lower panels show the history of the student's position within each comparison group. The displayed example contains no student name or institutional identifier. Engineering students received deciles rather than exact positions.

feedback, not subsequent academic outcomes. The final analysis sample contains 1,048 distinct students, 529 assigned to treatment and 519 to control. Appendix A documents the sample flow and the balance of these exclusions across assignment arms.

Some students were registered in more than one major. The assignment was initially recorded at the student-major level, but access through any treated major revealed comparative information to that individual. Treatment is therefore defined at the student level: if assignment provided access through at least one major, the student is classified as treated. This rule corresponds to effective eligibility for the tool and is applied consistently across outcomes.

Assignment used restricted randomization in two stages. Students from cohorts already enrolled were assigned first; the entering 2014 cohort was added while the earlier assignment remained fixed. Within each stage, the algorithm formed groups using major, entry cohort, and baseline academic position. Candidate assignments were retained only when a broad set of treatment-control balance tests exceeded a 0.10 p-value threshold, and one admissible assignment was selected at random. The appendix describes the procedure and reports conditional randomization inference that holds earlier assignments fixed and enforces the same balance criterion for the entering cohort.

Table 1 reports core baseline characteristics. Treatment and control students are similar in gender, prior courses, cumulative GPA, approved courses, credits, cohort, school background, geographic origin, and scholarship status. GPA is unavailable for entering students who had not yet completed university exams. Treated students are somewhat more likely to be registered in multiple majors, so all controlled specifications include this and the other pre-treatment variables shown in the table, together with school and cohort indicators and baseline position in the grade distribution.

Table 1. Baseline characteristics and treatment-control balance

Variable	Treated	Control	Difference	Robust SE	p-value	N
Woman	0.46	0.44	0.03	(0.03)	0.402	1048
Courses taken	26.50	25.55	0.95	(1.12)	0.400	1048
Cumulative GPA	7.51	7.54	-0.04	(0.10)	0.695	834
Courses passed	21.79	21.32	0.47	(0.99)	0.633	1048
Credits earned	142.00	138.39	3.61	(6.32)	0.569	1048
Number of registered majors	1.09	1.05	0.04	(0.02)	0.020	1048
Selected prior school	0.30	0.27	0.03	(0.03)	0.243	1048
Lives in Montevideo	0.67	0.65	0.02	(0.03)	0.419	1048
Scholarship of at least 20%	0.23	0.24	-0.01	(0.03)	0.697	1048
Cohort 2008	0.04	0.04	-0.00	(0.01)	0.826	1048
Cohort 2009	0.05	0.05	0.00	(0.01)	0.945	1048
Cohort 2010	0.13	0.12	0.01	(0.02)	0.659	1048

Variable	Treated	Control	Difference	Robust SE	p-value	N
Cohort 2011	0.20	0.20	-0.00	(0.02)	0.937	1048
Cohort 2012	0.20	0.20	0.00	(0.02)	1.000	1048
Cohort 2013	0.18	0.16	0.01	(0.02)	0.550	1048
Cohort 2014	0.20	0.22	-0.02	(0.03)	0.444	1048

Notes: Means are computed in the 1,048-student analysis sample. Woman, selected prior school, residence in Montevideo, scholarship of at least 20 percent, and cohort entries are binary indicators. Number of registered majors is the number of programs in which the student was registered at baseline. GPA is unavailable for entering students without completed university exams. Differences are treatment minus control; p-values use heteroskedasticity-robust standard errors. Additional assignment variables appear in the replication outputs.

2.3. Implementation and the placebo message

The ranking was launched on July 11, 2014, for cohorts already enrolled and on July 16 for the entering 2014 cohort. Students were informed through banners in the personal intranet, which they routinely used for course material and administrative procedures, and through e-mail.

The e-mail was deliberately designed to separate the effect of comparative disclosure from the effect of receiving an institutional message. Both treatment and control students received the same reminder to plan upcoming exams and the next semester using existing intranet resources such as exam dates, the course-modification period, transcripts, and curriculum progress. Only the treated message added a final line and link announcing that the student could access an academic ranking. Thus, both groups received comparable institutional contact, while access to relative-performance information differed by assignment. Appendix G presents the English translation and original Spanish wording.

2.4. Survey data and measuring satisfaction

The university administered an online survey shortly after launch. It was presented as a general satisfaction survey from the academic office rather than as an evaluation of the ranking experiment. The first launch cohort answered between July 23 and August 8, about two weeks after receiving access. Of the 1,048 experimental students, 861 supplied the main satisfaction outcomes: 443 treated and 418 controls.

The response rate is not detectably related to treatment, gender, or their interaction, and the respondent sample remains balanced across treatment arms. Respondents are nevertheless positively selected on prior academic performance: students with higher baseline GPA and rank were more likely to answer. We therefore report intention-to-treat contrasts among respondents, while recognizing that conditioning on survey response can introduce selection that balance on observables does not eliminate. The survey estimates should not be read as mechanically representative of every student in the experimental sample. Administrative outcomes do not depend on survey response and cover essentially the full experimental sample.

Satisfaction with current GPA was measured twice. The first question uses a five-point scale and appears among questions about several aspects of university life. Satisfaction with the university's location, asked in the same format, serves as a placebo outcome that the ranking should not directly affect.

The second measure uses a ten-point ladder running from the worst to the best academic life the respondent can imagine. It is followed by four anchoring vignettes. The reason for the vignettes is simple but important: two students who both report an eight may be using different internal scales. A high-performing student surrounded by other high performers may evaluate a given GPA more severely

than a lower-performing student whose usual comparison group has lower grades. Direct comparisons of subjective scores can then mix true differences in satisfaction with differences in how respondents use the scale.

Anchoring vignettes address this problem by asking all respondents to evaluate the same fixed situations on the same scale they use for themselves (King et al., 2004; Bago d'Uva et al., 2011; Beegle et al., 2012). Guille has a GPA of 9.4, ranks in the top 10 percent, and is expected to graduate on time. Jose has a GPA of 7.8 and expects to graduate six months late. Fer has a GPA of 7.0 but is expected to graduate on time. Fran has a GPA of 5.9, ranks in the bottom 10 percent, and expects to graduate six months late. Respondents evaluate these four profiles and their own satisfaction using the same ten-point scale. The vignettes do not turn satisfaction into a cardinal welfare measure; they provide common reference situations that improve comparability across respondents.

The answers show why the adjustment is informative. Respondents consistently rate Guille highest and Fran lowest, and they tend to prefer graduating on time when comparing the two middle cases. Yet students at different points in the GPA distribution use the scale differently. For example, students in the bottom GPA quintile rate Fran at 6.1 on average, whereas students in the top quintile rate exactly the same profile at 4.4. Figure 2 reports the four profiles by respondent GPA quintile. The second satisfaction specification therefore includes categorical controls for each vignette response.

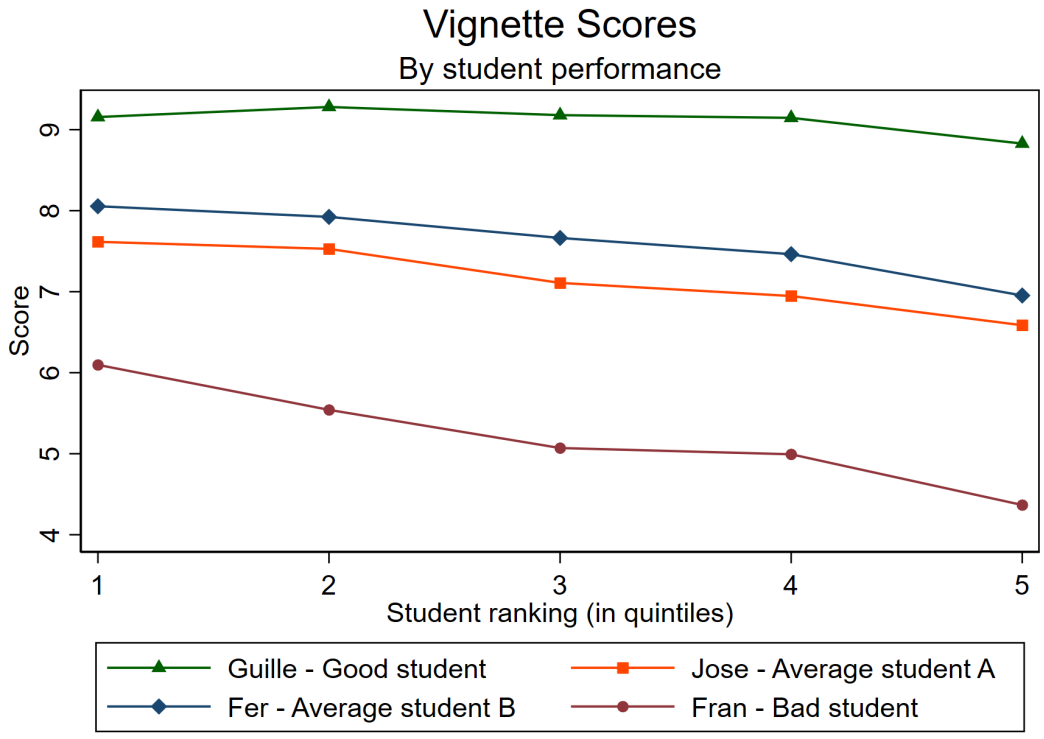


Figure 2. Mean ratings of the four anchoring vignettes.

Notes: The figure reports vignette ratings for the 857 respondents with complete information, separately by quintile of the respondent's actual GPA distribution. Guille has GPA 9.4, top-10-percent standing, and on-time graduation; Jose has GPA 7.8 and a six-month delay; Fer has GPA 7.0 and on-time graduation; Fran has GPA 5.9, bottom-10-percent standing, and a six-month delay. The vignettes enter the second satisfaction regression as categorical controls.

2.5. Administrative outcomes and timing

University records provide exam and student-level outcomes. Exam-level outcomes are the grade, whether an exam was taken, and whether it was passed. Student-level outcomes are cumulative GPA, courses passed, credits earned, and exit from the baseline major. The baseline-major record is selected

deterministically after harmonizing administrative labels across years; this implements the original intention to follow the student's experimental major and avoids order-dependent selection when a student has multiple program rows.

The first post-launch exam year begins at each cohort's actual launch date in July 2014 and ends on August 31, 2015. The second post-launch exam year runs from September 1, 2015, through August 31, 2016, so the periods do not overlap. Student-level outcomes are cumulative snapshots measured in November 2015 and October 2016, which we refer to as the first and second administrative follow-ups (roughly one and two years after launch). A separate sensitivity defines exit from the university only when no program record remains active or completed; because that is a different estimand, the main text retains exit from the baseline major.

The university-wide degree register records exact graduation dates through August 28, 2026. Every randomized student has a common twelve-year follow-up from the cohort-specific launch date. We distinguish graduation from the student's baseline major from receiving any new degree at Universidad de Montevideo after the intervention. The first outcome continues the paper's baseline-major estimand; the second allows for changes of major within the university. Sixteen students held another UM degree before the intervention, but none had already graduated from the baseline major. The any-degree outcome therefore counts only degrees awarded after launch. The records do not observe degrees awarded by other institutions.

3. Use of the Ranking and Belief Updating

Before examining satisfaction and academic outcomes, we establish two facts. First, treated students actually used the tool. Second, access materially improved their knowledge of relative performance. This sequence distinguishes an intervention that merely made information available from one that changed what students knew.

3.1. Did students use the ranking system?

Use was voluntary, and the platform recorded each access. Of 529 treated students, 508 entered the ranking at least once before September 2015. Treated students generated 9,622 visits in total. One control student received special access from the IT team and entered six times. The main estimates remain intention-to-treat effects based on initial assignment.

Figure 3 plots daily access among treated students. The highest use occurred at launch, with 324 visits. Subsequent weekly peaks reflect lower weekend use and renewed checking during exam periods, when new grades changed both GPA and relative position. The two vertical lines delimit the satisfaction-survey window. They are shown to connect survey timing to exposure; they do not redefine treatment receipt. The figure shows that interest did not disappear after the initial announcement and returned when the ranking was likely to move.

3.2. Did students change their perceptions?

High use alone does not establish that the intervention supplied new information. We anticipated two threats to the first stage. Students might already know their relative position because cohorts are small and classmates observe one another's performance. In addition, controls might learn about their own standing from treated friends. For example, a control student with a GPA close to that of a treated friend could infer whether she belonged near the top of the distribution after the friend disclosed an exact rank.

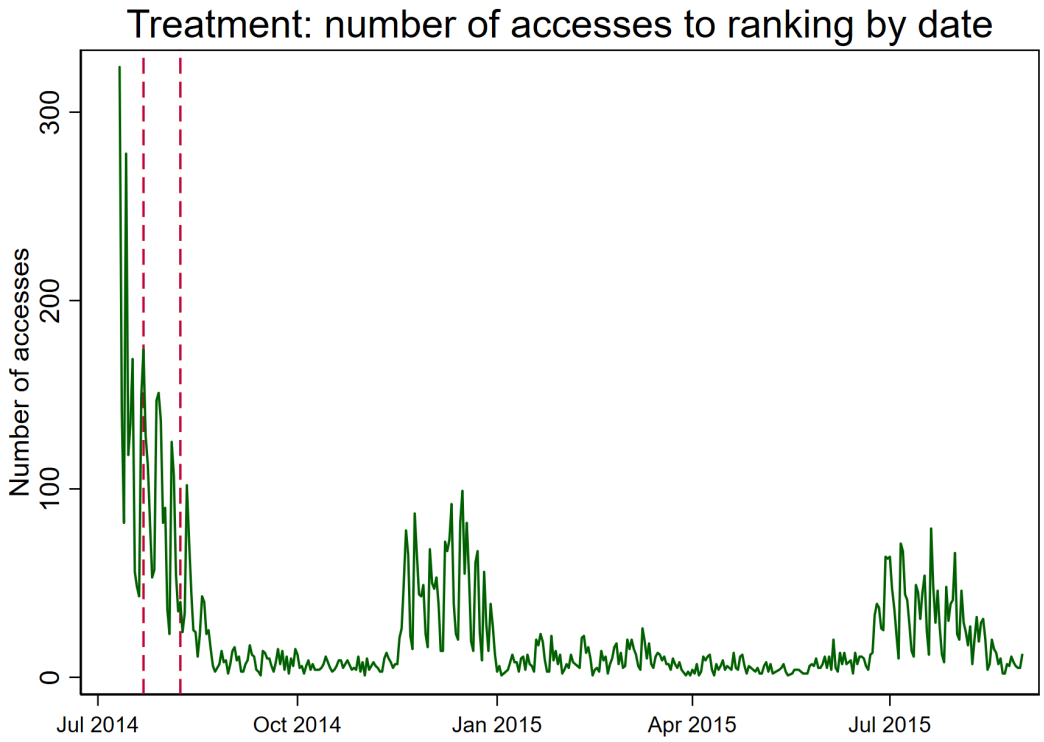


Figure 3. Daily access to the ranking system.

Notes: The figure reports daily visits by students assigned to treatment from July 2014 through August 2015. It records 9,622 visits by 508 treated users. A student can visit more than once. The launch-day maximum is 324 visits. Dashed vertical lines indicate the satisfaction-survey period. The one control student given special access is excluded from the plotted treated series.

To address these challenges, the Satisfaction Survey included two sets of questions. The first asked whether students knew that the tool existed and whether they had learned the exact position of one or several classmates. Among the 169 control students who said they knew about the tool, 38 selected the response indicating that they had learned the exact position of one classmate, 27 selected the response for several classmates, and 10 selected both. In total, 55 distinct controls reported at least one such source of information. This is evidence that some information circulated across students. The questions do not measure every spillover, and we do not claim that contamination was absent or that the intention-to-treat estimate is necessarily a lower bound.

The second set of questions asked both treated and control students to report their own position, with number one denoting the highest GPA. We match those perceptions to administrative positions measured on July 23, 2014. Because Engineering students received and reported deciles rather than exact positions, the exact-position analysis focuses on Economics and Law.

Figure 4 summarizes the relationship between perceived and actual rank at the major level. Actual performance is grouped into quintiles on the horizontal axis; smaller numerical positions on the vertical axis represent better ranks. The thick green series is the mean actual position within each quintile, not a 45-degree line. The red treated series closely follows it. The blue control series is flatter: lower-performing controls tend to place themselves too high, while higher-performing controls tend to place themselves too low. Appendix C reports individual responses and the corresponding school-level evidence.

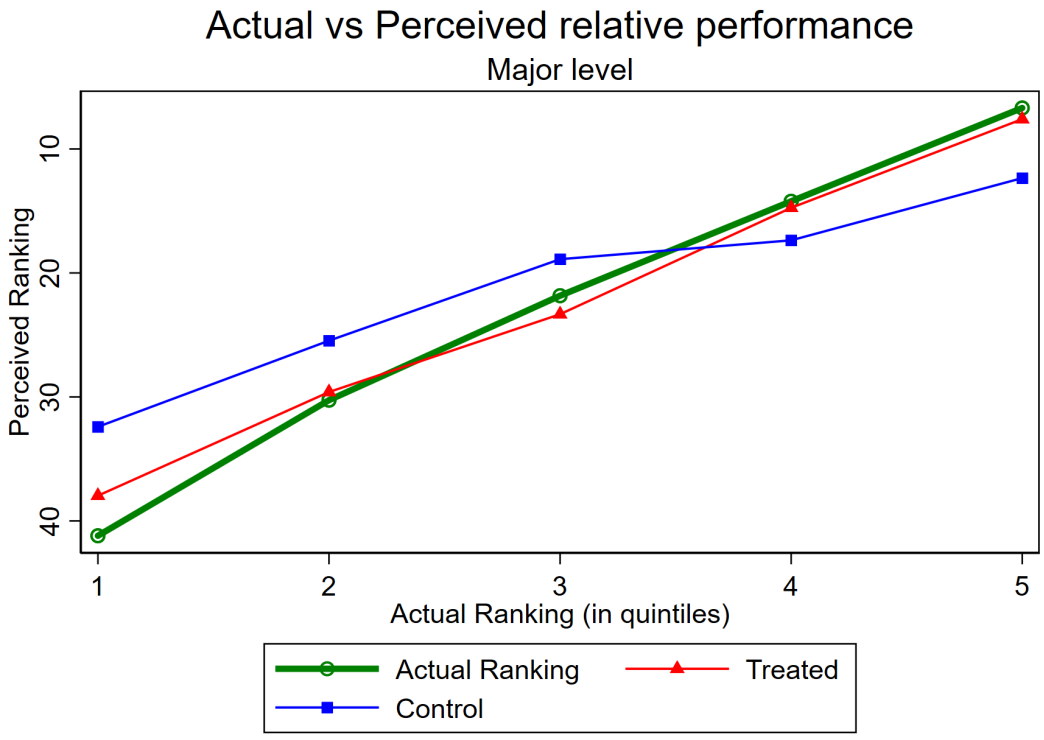


Figure 4. Actual and perceived relative performance.

Notes: The horizontal axis groups students into quintiles of actual GPA rank, ordered from lower to higher performance. The vertical axis reports numerical rank, with smaller numbers indicating a better position. The green line is the mean actual position within each quintile; the red and blue lines report mean perceived position for treated and control students. Engineering is excluded because its treatment and survey response used deciles rather than exact positions.

Several processes may contribute to the control pattern. Regression toward the mean is natural when students are uncertain about whether a good or bad exam reflects ability or exam difficulty. Selective

peer groups can also distort the information on which students base their comparisons. In the survey, 785 respondents named at least one close friend in their cohort, generating 3,413 nominations, 3,389 of which were matched to identified students. A one-point increase in friends' mean GPA is associated with a 0.758-point increase in the respondent's GPA (N=782). High-performing students therefore compare themselves with unusually strong peers, while low-performing students tend to observe peers from the lower part of the distribution. This selection offers a plausible explanation for misplacement, but it is an observational relationship and does not identify a peer mechanism.

Table 2 provides the formal first stage. The first outcome is the absolute difference between perceived and actual position. The second is an indicator for reporting the exact position. Models control for comparison-group size, baseline position in the grade distribution, and cohort indicators. Treatment reduces absolute error by 3.4 places within a major and 8.8 places within a school, from control means of 7.1 and 19.6. It raises exact placement by 29.7 and 19.5 percentage points, from control means of 8.2 and 2.4 percent. All four treatment effects have p-values below 0.001.

Table 2. Effect of treatment on knowledge of relative performance

	Absolute error Major	**Absolute error School**	**Exact placement Major**	**Exact placement School**
Treatment	-3.405***	-8.766***	0.297***	0.195***
	(0.568)	(1.407)	(0.031)	(0.024)
Comparison-	0.122***	0.118***	-0.002***	0.000
group size	(0.013)	(0.010)	(0.001)	(0.000)
Baseline rank	-0.421***	-1.413***	0.028***	0.031***
	(0.114)	(0.301)	(0.005)	(0.005)
Constant	6.806***	22.303***	-0.062	-0.100
	(2.241)	(4.979)	(0.063)	(0.076)
Control mean	7.144	19.559	0.082	0.024
Observations	599	608	599	608

Notes: Engineering is excluded because perceived rank was reported in deciles. Absolute error is |perceived − actual position|. Exact placement equals one when positions coincide. Models include cohort indicators. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.

Together, platform use, perceived-position responses, and the experimental first stage show that treated students acquired substantially more precise information. The circulation documented among some controls counsels against interpreting the experiment as perfect isolation, but it does not erase the large treatment-control difference in knowledge. We next ask how that corrected information affected satisfaction and subsequent academic choices.

4. Effects on Satisfaction and Academic Outcomes

4.1. Estimands and inference

For satisfaction and academic outcomes, we estimate:

$$Y_i = \alpha + \beta_1 Treatment_i + \beta_2 (Treatment_i \times Woman_i) + \beta_3 Woman_i + X_i'\theta + \varepsilon_i.$$

The coefficient on Treatment is the treatment effect for men. The interaction measures how the treatment effect for women differs from that for men. The total treatment effect for women is the

sum of Treatment and Treatment × Woman. We report these three quantities separately because a statistically significant interaction does not by itself establish that the total effect for women differs from zero. Gender is predetermined, so the interaction has a causal subgroup interpretation under random assignment.

The vector X contains the pre-treatment controls described with Table 1. Satisfaction and student-level models use heteroskedasticity-robust standard errors. Exam-level specifications cluster standard errors by student. Conventional p-values are shown in the main tables. Because the analysis contains families of related outcomes, Appendix D reports Holm family-wise and Benjamini-Hochberg false-discovery-rate adjustments for all six outcome families. It also reports conditional randomization-inference p-values for satisfaction, student-level progression, and long-run graduation; the implemented conditional exercise does not cover the repeated exam-level outcomes.

4.2. Satisfaction with academic performance

Table 3 reports the two GPA satisfaction measures and the location placebo. On the five-point measure, the estimated effect for men is 0.133 (p=0.054). The treatment-by-woman interaction is -0.294 (p=0.006), and the total treatment effect for women is -0.161 (p=0.049). The corresponding control means are 3.12 for men and 3.45 for women. The result therefore reflects both a modest positive response among men and a negative response among women, rather than a common fall in satisfaction.

The ten-point measure adjusted by the four vignettes yields the same contrast. The point estimate for men is 0.158 and is not statistically distinguishable from zero. The interaction is -0.540 (p=0.015), and the total effect for women is -0.382 (p=0.028). The location placebo shows neither a treatment effect for men nor a differential effect for women. Holm adjustment across the two satisfaction outcomes leaves both interactions significant; the adjusted p-values for the two total effects for women are 0.056. Conditional randomization inference yields similar or somewhat stronger support, particularly for the vignette-adjusted measure.

Table 3. Treatment effects on satisfaction

	Five-point GPA satisfaction	**Location placebo**	**Ten-point GPA satisfaction**
Treatment effect: men	0.133*	-0.021	0.158
	(0.069)	(0.080)	(0.145)
Treatment × woman	-0.294***	0.041	-0.540**
	(0.106)	(0.126)	(0.222)
Woman	0.144*	-0.205**	0.316*
	(0.078)	(0.095)	(0.167)
Total treatment effect:	-0.161**	0.021	-0.382**
women	(0.082)	(0.096)	(0.174)
Control mean: men	3.125	4.095	6.237
Control mean: women	3.446	3.892	6.925
Observations	859	859	857

Notes: All models include pre-treatment controls. The ten-point specification also includes categorical controls for four anchoring vignettes. Robust standard errors in parentheses. Stars use conventional two-sided p-values; multiplicity and conditional randomization inference are reported in the text and appendix.

Survey nonresponse is unlikely to explain the treatment-control contrast mechanically. Response is not detectably differential by treatment and gender, and inverse-probability weighting based on observables produces estimates close to the main results. These checks do not eliminate selection on

unobservables, and the survey sample remains positively selected on prior performance. We therefore interpret the estimates as intention-to-treat contrasts among observed respondents, stable to observable inverse-probability weighting but subject to this response-selection caveat.

4.3. Academic outcomes after one and two years

Tables 4A and 4B report exam-level outcomes and student-level progression. The exam outcomes provide evidence about intensive academic performance and exam participation. Student outcomes capture accumulated progression and persistence. The distinction matters: fewer courses and credits may partly reflect students leaving their baseline major rather than poorer performance among those who remain. We do not condition the main outcomes on post-treatment enrollment because doing so would select on an outcome affected by treatment.

Descriptively, treatment-control differences among students classified as active or graduated from their baseline major are smaller and imprecise. Because continuation is post-treatment, these comparisons do not identify a causal intensive-margin effect; the unconditional intention-to-treat estimates remain the relevant measure of accumulated progress.

At the first administrative follow-up, treatment-by-woman coefficients are negative for cumulative GPA, courses, and credits, and positive for exit from the baseline major; coefficients are also negative for exam grades and exam taking during the first post-launch exam year. The clearest total effect for women is a 7.2 percentage point increase in exit (p=0.005), compared with a control mean of 5.8 percent among women. That result survives Holm adjustment within the four student-level outcomes. Total effects for women's GPA, courses, credits, and exam outcomes are negative but not precisely estimated after conventional or family-wise adjustment.

At the second administrative follow-up, the persistent pattern is concentrated in progression. Treated women have passed 2.1 fewer courses (p=0.009), earned 11.9 fewer credits (p=0.018), and are 7.3 percentage points more likely to have exited their baseline major (p=0.008). Each remains significant under Holm adjustment within the second-follow-up student-outcome family. The cumulative-GPA effect for women is -0.131 (p=0.093), while total effects on exam grades and passing during the second post-launch exam year are not distinguishable from zero. Exam taking is lower for women in that exam year at the 10 percent level before multiplicity adjustment.

Table 4A. Exam-level academic outcomes

Outcome	Treatment effect: men	Treatment × woman	Total effect: women	Control mean: women
		First post-launch exam year		
Exam grade	0.149*	-0.251**	-0.102	8.156
N = 15,040	(0.086)	(0.127)	(0.093)	
Exam taken	0.010	-0.031*	-0.021*	0.877
N = 17,510	(0.012)	(0.017)	(0.013)	
Exam passed	0.012	-0.022	-0.010	0.871
N = 15,040	(0.012)	(0.017)	(0.012)	
		Second post-launch exam year		
Exam grade	0.062	-0.149	-0.088	8.402
N = 8,819	(0.100)	(0.153)	(0.118)	
Exam taken	0.016	-0.045**	-0.029*	0.896
N = 9,983	(0.013)	(0.021)	(0.016)	

Outcome	Treatment effect: men	Treatment × woman	Total effect: women	Control mean: women
Exam passed	0.024*	-0.028	-0.004	0.900
N = 8,819	(0.014)	(0.020)	(0.015)	

Notes: Each cell reports a coefficient and standard error in parentheses. All specifications include pre-treatment controls. Standard errors are clustered by student. Stars use conventional p-values; family-wise and design-based adjustments appear in the text and appendix.

The interaction coefficients show that treated women generally perform worse relative to treated men on a broader set of outcomes. At two years the interactions are -0.201 for cumulative GPA, -2.892 for courses, -14.931 for credits, and +0.090 for exit from the baseline major. All four conventional p-values are below 0.05 and all four remain below 0.05 after Holm adjustment. These interactions establish treatment-effect heterogeneity. For women, the clearest evidence is concentrated in courses completed, credits earned, and exit from the baseline major, rather than in a uniform decline across every academic margin.

Exploratory estimates further show that the adverse second-follow-up effects on credits and exit among women are concentrated at lower baseline ranks and fade toward zero at higher ranks. Full estimates are reproduced in the replication package and are available from the authors on request.

4.4. Graduation twelve years after the intervention

The final panel of Table 4B extends the administrative follow-up to degree completion. By twelve years, 843 students had graduated from their baseline major and 896 had received at least one new degree from the university after the intervention. Among control women, 87.2 percent graduated from the baseline major. The estimated treatment effect is -10.3 percentage points for women (p=0.001) and +1.5 points for men (p=0.620). The treatment-by-woman interaction is -11.8 points (p=0.007). Using any new UM degree produces a similar pattern: the total effect for women is -9.0 points (p=0.002), the effect for men is close to zero, and the interaction is -9.3 points (p=0.021).

The long-run family contains these two degree outcomes. Both interactions and both total effects for women remain below 0.05 after Holm adjustment. Conditional randomization inference is also supportive: the adjusted randomization p-values are 0.002 and 0.007 for the interactions and 0.004 for both total effects for women. Results at eight years are similar. For the broader any-degree outcome, the pattern also remains when the sixteen students who held another UM degree before treatment are excluded. Appendix D reports the corresponding estimates.

The pooled effect is not uniform across students' stage at launch. For graduation from the baseline major, the adjusted effect for women is -33.6 percentage points among students who entered in 2014 and -3.5 points among earlier cohorts. The difference between the corresponding gender interactions is -36.6 points (p=0.005). For any new UM degree, the effects for women are -26.4 and -3.8 points, while the difference between gender interactions is less precisely estimated (p=0.060). Gender heterogeneity was a central object of the original study, whereas the split by entry cohort is exploratory and speaks to the timing of exposure. We report it because the concentration is material to the scope of the pooled long-run estimate: the evidence is strongest for students who received the feedback upon entering university, when more of their academic trajectory remained ahead.

Table 4B. Student-level academic progression and graduation

Outcome	Treatment effect: men	Treatment × woman	Total effect: women	Control mean: women
		First administrative follow-up		
Cumulative GPA	0.063	-0.183*	-0.119	8.029
N = 1,046	(0.065)	(0.101)	(0.077)	
Courses passed	0.453	-1.261*	-0.808	36.217
N = 1,046	(0.440)	(0.677)	(0.515)	
Credits earned	1.746	-5.886	-4.140	230.854
N = 1,044	(2.613)	(3.988)	(3.034)	
Exit from baseline	-0.019	0.090***	0.072***	0.058
major	(0.021)	(0.033)	(0.025)	
N = 1,046				
		Second administrative follow-up		
Cumulative GPA	0.070	-0.201**	-0.131*	8.100
N = 1,046	(0.064)	(0.100)	(0.078)	
Courses passed	0.769	-2.892***	-2.123***	43.619
N = 1,046	(0.706)	(1.070)	(0.811)	
Credits earned	3.060	-14.931**	-11.871**	277.241
N = 1,045	(4.304)	(6.544)	(4.990)	
Exit from baseline	-0.017	0.090**	0.073***	0.080
major	(0.023)	(0.036)	(0.027)	
N = 1,046				
		Twelve-year graduation outcomes		
Graduated from	0.015	-0.118***	-0.103***	0.872
baseline major	(0.030)	(0.044)	(0.032)	
N = 1,046				
Received any new	0.003	-0.093**	-0.090***	0.903
UM degree	(0.027)	(0.040)	(0.029)	
N = 1,046				

Notes: Each cell reports a coefficient and standard error in parentheses. All specifications include pre-treatment controls. Exit and baseline-major graduation are defined relative to the student's experimental major, selected with a deterministic crosswalk. Any new UM degree counts a degree awarded after the intervention in any major at Universidad de Montevideo. Heteroskedasticity-robust standard errors are reported. Stars use conventional p-values; family-wise and design-based adjustments appear in the text and appendix.

4.5. Robustness, precision, and scope

Three sets of checks support the interpretation that the adverse pattern for women is concentrated in academic progression and is not driven by pre-treatment imbalance or Engineering's use of decile feedback. First, assigning future treatment status to pre-treatment exam outcomes from July 2003 through June 2014 produces no corresponding pattern, consistent with baseline balance; Appendix D reports the estimates. Second, excluding Engineering leaves the main satisfaction and progression pattern qualitatively intact. Third, conditional randomization inference based on 999 admissible assignments generally yields p-values at least as supportive as conventional inference for satisfaction, student-level progression, and long-run graduation. Appendix D reports every outcome and both

estimands for which this conditional exercise was implemented, rather than selecting only results that cross conventional thresholds.

The null and imprecise results also matter. Approximate minimum detectable effects show that several exam-level tests cannot rule out economically meaningful changes. The absence of conventional significance in exam grades or passing should therefore be read as limited precision, not evidence of exactly zero effects. Conversely, the progression results are both sizeable relative to control means and stable across the main inferential approaches.

5. Interpretation and Potential Explanations

5.1. What the experiment identifies

The intervention combined several features of comparative feedback. It was private, precise, updated frequently, and embedded in a particular academic setting. The experiment therefore estimates the effect of offering that feedback system as a whole. The strong first stage establishes that students learned more accurately where they stood, while the subsequent effects may reflect both the content of that news and the way it was delivered. Students may update beliefs about ability or future returns, feel encouraged or discouraged, compare themselves more intensely with peers, or reassess the effort required to improve their position.

The baseline-rank pattern reported in Section 4.3 echoes the belief pattern in Figure 4: lower-performing controls tend to place themselves above their actual position, whereas higher-performing controls tend to place themselves below it. The alignment is consistent with responses depending on the content of comparative feedback.

The remaining evidence asks which of these interpretations is consistent with observed responses. It can illuminate plausible channels, but it cannot attribute the academic effects to a single component because news valence, framing, frequency, privacy, reference group, and accompanying support were not randomized separately.

5.2. Competitiveness and competitive framing

One possible explanation is that women and men respond differently to the competitive framing of rank information. The survey asked students to rate their competitiveness on a zero-to-ten scale, but this question was answered after treatment began. It must therefore be analyzed as a post-treatment response, not as a fixed pre-treatment trait.

On the original continuous scale, treatment has no detectable effect for men, no detectable treatment-by-woman interaction, and no detectable total effect for women. The adjusted gender difference among controls is also close to zero. A gender difference appears only under the stricter cutoff that defines competitive as a score above seven; it disappears when seven is included or when the original scale is retained. We therefore do not treat competitiveness as an established mechanism. The null does not rule out reactions to a competitive frame, but the survey measure does not explain the heterogeneity.

Table 5 summarizes these competitiveness results together with the pre-treatment effort measures discussed next.

Table 5. Measures relevant to potential mechanisms

	Competitiveness 0–10	Competitive score > 7	Library books	Study minutes per week and subject
Woman	0.014	-0.100***	3.793***	8.822***
	(0.186)	(0.035)	(0.981)	(1.172)
Observations	841	841	1046	24960

Notes: Competitiveness is measured in the post-treatment survey; its two columns report pooled adjusted gender associations, with causal treatment specifications in the text and Appendix F. Library use and study time are measured before treatment. All specifications include the baseline controls used in the main analysis. Students without a library loan are coded as zero. Study-time standard errors are clustered by student. These measures do not identify mediation.

5.3. Effort before treatment and the marginal cost hypothesis

Two pre-treatment measures suggest that women were investing more effort before the ranking was introduced. When students with no recorded library loan are coded as zero, women borrowed 3.793 more books conditional on baseline controls ($p<0.001$). In the student-semester-subject measure used here, women reported 8.822 more study minutes per week per subject ($p<0.001$), with standard errors clustered by student.

These facts motivate an increasing-marginal-cost hypothesis. Advancing in a ranking requires improving relative to peers, and even maintaining a position can require additional effort when peers improve. If marginal effort costs rise, students already exerting more effort may face a higher cost of responding to a newly salient comparison. We view these measures as suggestive correlates: library borrowing and self-reported study time capture effort imperfectly, and neither was experimentally varied. They therefore motivate, but do not establish, the marginal-cost interpretation.

5.4. Privacy, audience, and the design of feedback

The intervention was private, but the survey also asked about possible publicity. At the project's inception, we considered whether rankings might eventually be visible to others. Respondents were therefore asked both how much they wanted to know their own rank and how much they would care if other people knew their exact position. The relatively small comparison groups described in Section 2.1 made the audience for the information especially salient: private access did not eliminate the possibility that students would discuss or infer positions within their cohort. This concern is consistent with evidence that peer observability can change educational investment (Bursztyn and Jensen, 2015).

Interest in knowing one's own rank does not increase for women after treatment. The response to potential publicity is different. On the four-point scale, the treatment-by-woman interaction in concern about others knowing the rank is 0.471 ($p=0.004$), and the total effect for women is 0.305 ($p=0.016$). Using an indicator for the top two concern categories, the total effect for women is 20.3 percentage points ($p<0.001$). The corresponding point estimate for men is negative and imprecise. Because this outcome is post-treatment, exploratory, and one of several survey responses, it should not be treated as a causal mechanism for academic outcomes. It nevertheless points to an important design implication: the audience for comparative information may matter, and privacy should not be viewed as a cosmetic feature of feedback systems.

5.5. External validity and feedback design

The treatment concerns cumulative academic performance at one private university. Comparison groups are institutional, the measure is ordinal, updates are frequent, and choices occur within a particular grading and degree system. The findings do not imply that rankings are universally harmful, that public feedback would have the same effects, or that all women respond in the same way. They show that a system can accomplish its immediate informational objective while producing heterogeneous subjective and academic consequences.

This distinction matters for policy. Evaluation should consider not only whether information is correct and used, but also how the reference group is constructed, when feedback arrives, whether it is opt-in or automatic, whether it is framed competitively, who can observe it, and what support accompanies unfavorable news. Results of different signs in later experiments are consistent with these being substantive treatment dimensions rather than implementation details.

The university's response to the pilot provides a concrete illustration. Rather than simply scaling or abandoning the tool, it renamed it "academic trajectory" to soften its competitive framing, made access opt-in, delayed eligibility until the second year, and left adoption decisions to individual schools. These changes do not identify which design feature would improve outcomes, but they show that the institution treated framing, timing, choice, and implementation scope as substantive policy dimensions. The timing decision is also consistent with the later, exploratory evidence that long-run effects were concentrated among students exposed upon entry.

6. Discussion and Conclusion

The experiment gave students precise and frequently updated information about where their cumulative academic performance placed them among peers. Students used the system, and their knowledge improved markedly. Control students tended to pull their perceived positions toward the middle of the distribution, whereas treated students reported positions close to the administrative ranking. The intervention therefore succeeded as an information treatment.

That success did not produce uniform benefits. Women assigned to feedback reported lower satisfaction with their academic performance on two separately elicited measures. Administrative records then show a persistent decline in progression and persistence: exit from the baseline major rose at the first administrative follow-up, and by the second follow-up treated women had completed fewer courses, accumulated fewer credits, and were more likely to have exited. Twelve years after the intervention, they were also less likely to graduate from the baseline major or to receive any new degree from the university. The long-run difference was concentrated among students exposed upon entering university, an exploratory pattern that narrows the scope of that pooled estimate rather than identifying a mechanism. Men did not display a comparable adverse pattern. Average effects would obscure this contrast.

The pattern is selective rather than universal. Some exam-level and GPA effects are imprecise, and the evidence does not show deterioration on every academic margin. Neither does it establish a single mechanism. Greater pre-treatment effort among women is compatible with rising marginal costs of responding to rank; concern about public visibility suggests that audience matters; and confidence, news valence, expectations, and social comparison remain plausible. Distinguishing these channels would require interventions that separately vary framing, reference groups, publicity, or academic support.

The broader lesson is not that institutions should withhold useful information. It is that accuracy

is not a sufficient criterion for evaluating comparative feedback. A ranking can correct beliefs and still change satisfaction and investment in ways that differ across groups. Feedback systems should therefore be assessed as behavioral interventions, with attention to both average and distributional consequences.

References

Aucejo, E. M., and Wong, K. (2025). "The Effect of Feedback on Student Performance." Journal of Public Economics 241: 105274.

Azmat, G., Bagues, M., Cabrales, A., and Iriberri, N. (2019). "What You Don't Know... Can't Hurt You? A Field Experiment on Relative Performance Feedback in Higher Education." Management Science 65(8): 3714-3736.

Azmat, G., and Iriberri, N. (2010). "The Importance of Relative Performance Feedback Information: Evidence from a Natural Experiment Using High School Students." Journal of Public Economics 94(7-8): 435-452.

Bago d'Uva, T., Lindeboom, M., O'Donnell, O., and van Doorslaer, E. (2011). "Slipping Anchor? Testing the Vignettes Approach to Identification and Correction of Reporting Heterogeneity." Journal of Human Resources 46(4): 875-906.

Bandiera, O., Larcinese, V., and Rasul, I. (2015). "Blissful Ignorance? A Natural Experiment on the Effect of Feedback on Students' Performance." Labour Economics 34: 13-25.

Beegle, K., Himelein, K., and Ravallion, M. (2012). "Frame-of-Reference Bias in Subjective Welfare." Journal of Economic Behavior & Organization 81: 556-570.

Brade, R., Himmler, O., and Jäckle, R. (2022). "Relative Performance Feedback and the Effects of Being Above Average - Field Experiment and Replication." Economics of Education Review 89: 102268.

Brade, R., Himmler, O., and Jäckle, R. (2026). "No Student Left Behind? Relative Feedback and University Completion." Journal of Economic Behavior & Organization 241: 107383.

Bursztyn, L., and Jensen, R. (2015). "How Does Peer Pressure Affect Educational Investments?" Quarterly Journal of Economics 130(3): 1329-1367.

Cabrera, J. M., and Cid, A. (2017). "Gender Differences to Relative Performance Feedback: A Field Experiment in Education." Universidad de Montevideo Working Paper 1704.

Card, D., Mas, A., Moretti, E., and Saez, E. (2012). "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction." American Economic Review 102(6): 2981-3003.

Carneiro, P., Cruz-Aguayo, Y., Salvati, F., and Schady, N. (2025). "The Effect of Classroom Rank on Learning throughout Elementary School: Experimental Evidence from Ecuador." Journal of Labor Economics 43(2): 433-466.

Chen, Y., Dobrescu, L. I., Foster, G., and Motta, A. (2024). "Can Leagues Mitigate the Demoralization Effect of Rank Feedback? A Randomized Controlled Trial." Labour Economics 90: 102602.

Czibor, E., Onderstal, S., Sloof, R., and van Praag, M. (2020). "Does Relative Grading Help Male Students? Evidence from a Field Experiment in the Classroom." Economics of Education Review 75: 101953.

Denning, J. T., Murphy, R., and Weinhardt, F. (2023). "Class Rank and Long-Run Outcomes." Review of Economics and Statistics 105(6): 1426-1441.

Dobrescu, L. I., Faravelli, M., Megalokonomou, R., and Motta, A. (2021). "Relative Performance Feedback in Education: Evidence from a Randomised Controlled Trial." Economic Journal 131(640): 3145-3181.

Elsner, B., and Isphording, I. E. (2017). "A Big Fish in a Small Pond: Ability Rank and Human Capital Investment." Journal of Labor Economics 35(3): 787-828.

Fischer, M., and Wagner, V. (2023). "Do Timing and Reference Frame of Feedback Influence High-Stakes Educational Outcomes?" Economics of Education Review 94: 102379.

Jalava, N., Joensen, J. S., and Pellas, E. (2015). "Grades and Rank: Impacts of Non-Financial Incentives on Test Performance." Journal of Economic Behavior & Organization 115: 161-196.

King, G., Murray, C. J. L., Salomon, J. A., and Tandon, A. (2004). "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." American Political Science Review 98(1): 191-207.

Macías Domínguez, C., Santero Sánchez, R., Sanz Labrador, I., and Tena, J. D. (2026). "Does Relative Performance Feedback Improve Academic Outcomes? Evidence from a Randomised Controlled Trial in a Spanish University." EdWorkingPaper 26-1504. https://doi.org/10.26300/za71-7x46.

Murphy, R., and Weinhardt, F. (2020). "Top of the Class: The Importance of Ordinal Rank." Review of Economic Studies 87(6): 2777-2826.

Niederle, M., and Vesterlund, L. (2007). "Do Women Shy Away from Competition? Do Men Compete Too Much?" Quarterly Journal of Economics 122(3): 1067-1101.

Tran, A., and Zeckhauser, R. (2012). "Rank as an Inherent Incentive: Evidence from a Field Experiment." Journal of Public Economics 96(9-10): 645-650.

Villeval, M. C. (2020). "Performance Feedback and Peer Effects." In Handbook of Labor, Human Resources and Population Economics, edited by K. F. Zimmermann, 1-38. Cham: Springer. https://doi.org/10.1007/978-3-319-57365-6_126-1.

Wozniak, D., Harbaugh, W. T., and Mayr, U. (2014). "The Menstrual Cycle and Performance Feedback Alter Gender Differences in Competitive Choices." Journal of Labor Economics 32(1): 161-198.

Online Appendix — Proposed for Online Publication

A. Assignment, Sample Genealogy, and Baseline Balance

The treatment operated at the student level even when assignment records contained multiple student-major rows. The participating schools contained 1,067 students before the final implementation-eligibility check. Seven assigned students had no academic activity in 2014—their last recorded activity was in the second semester of 2012 or the first semester of 2013—and twelve had no calculable GPA-based position in the July 23 implementation snapshot. These criteria identify students outside the active feedback population or for whom comparative feedback could not operate; they do not use subsequent academic outcomes. The resulting exclusions comprise 13 of 542 assigned to treatment and 6 of 525 assigned to control. The 1.3 percentage point difference in inclusion is not statistically distinguishable from zero (p=0.119). The final file contains one record for each of 1,048 students.

Table A1. Sample genealogy

Stage	Records	Unique students
Students in participating schools before implementation eligibility check	1,131	1,067
Final experimental analysis sample	1,048	1,048

Notes: Assignment begins at the student-major level. The implementation-eligibility check excludes seven students with no academic activity in 2014 and twelve without a calculable GPA-based rank in the July 23 implementation snapshot. Thirteen of 542 assigned to treatment and 6 of 525 assigned to control were excluded at this stage; the difference in inclusion is 1.3 percentage points (robust p=0.119). The final canonical balance file contains one record per student.

Restricted randomization was conducted in two stages. Earlier cohorts were assigned first, and the 2014 cohort was added while holding those assignments fixed. Candidate allocations had to satisfy the same 0.10 balance threshold across the specified subgroups and covariates. For conditional randomization inference, we fix the earlier-cohort assignment and reassign the 2014 cohort under those restrictions. Among 3,604 candidate draws, 999 satisfy the restrictions and form the reference distribution.

The multiple-program rule for student outcomes is deterministic. Labels from each administrative wave are harmonized to the baseline major. A unique crosswalk match resolves 134 of 136 multiple-row students in 2015 and 139 of 142 in 2016. For the five remaining cases, the rule prioritizes an active or graduated row and then accumulated credits, courses, GPA, and the program label. This procedure yields the student-outcome sample used throughout the paper. The main outcome is exit from the baseline major. Exit from the university is reported only as a sensitivity because it answers a different question.

B. Survey Flow, Attrition, and Measurement

The survey has 861 respondents, 443 treated and 418 controls. Each vignette has 859 nonmissing ratings; the complete-case sample used for Figure 2 and the vignette-adjusted regression is 857.

Table B1. Survey response by treatment and gender

Gender	Control response	Treated response	Difference	Cell N (C / T)
Men	79.2%	83.9%	4.7 pp	293 / 285
Women	82.3%	83.6%	1.3 pp	226 / 244

Notes: Response means beginning the survey and providing the main satisfaction outcomes. The treatment-by-woman interaction in a linear probability model is -3.4 percentage points (robust p=0.477).

Table B2. Inverse-probability-weighted satisfaction estimates

Outcome	Estimand	Main estimate	IPW estimate	IPW p-value
Satisfaction measure 1	Treatment × woman	-0.294*** (0.106)	-0.289*** (0.108)	0.008
Satisfaction measure 1	Total effect: women	-0.161** (0.082)	-0.158* (0.084)	0.059
Location placebo	Treatment × woman	0.041 (0.126)	0.048 (0.125)	0.701
Location placebo	Total effect: women	0.021 (0.096)	0.039 (0.095)	0.681
Satisfaction measure 2	Treatment × woman	-0.540** (0.222)	-0.489** (0.230)	0.033
Satisfaction measure 2	Total effect: women	-0.382** (0.174)	-0.374** (0.178)	0.036

Notes: Response propensities are estimated in the experimental sample using treatment, gender, their interaction, and the pre-treatment covariates used in the main analysis. Weighted outcome regressions retain the canonical controls; the ten-point measure also retains the four vignette controls. IPW addresses selection associated with observables and does not identify selection on unobservables.

Figure 2 presents the same four fixed vignettes described in Section 2.4. Its horizontal axis is the respondent's actual GPA quintile, not treatment status or gender. The downward slopes show systematic heterogeneity in scale use: higher-performing respondents assign lower satisfaction scores to the same objective profiles.

C. Perceived Rank, Spillovers, and Friend Selection

Perceived position comes from the short-run survey and actual position from the July 23, 2014 ranking snapshot. Absolute error is the absolute difference between perceived and actual numerical rank. Exact placement equals one when the two positions coincide. Engineering is excluded from these exact-position outcomes because both treatment and survey used deciles there.

Figure C1 displays the individual observations behind the first-stage comparisons. At both the major and school levels, treated students' answers cluster much more closely around the line of equality. Control responses are more dispersed and exhibit the same compression toward the mean visible in Figure 4.

The survey questions about communication show that 55 of 169 controls who knew the tool existed reported learning the exact position of at least one classmate. This count combines the options for one and for several classmates and avoids interpreting only the singular option as the total. Since

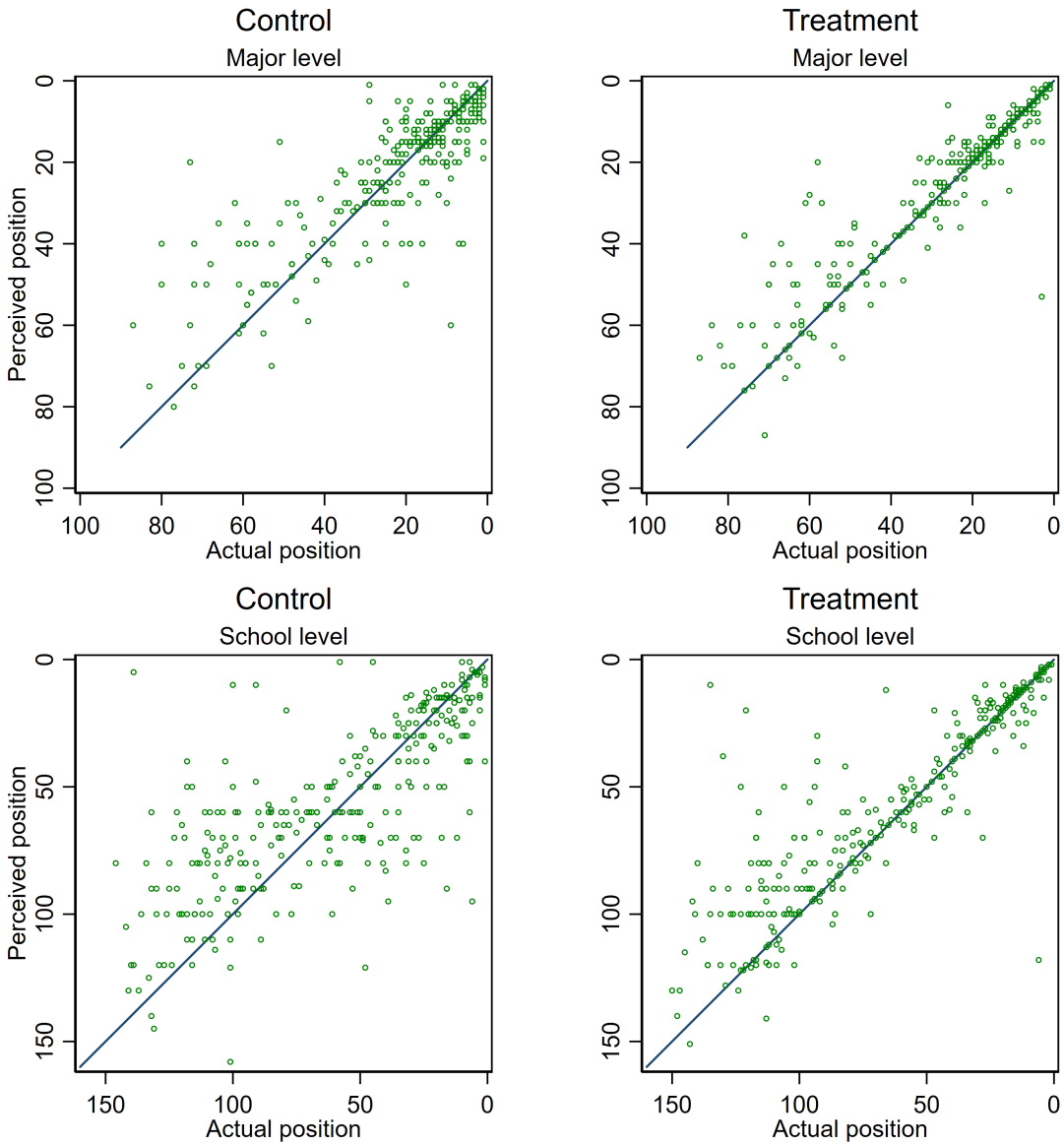


Figure C1. Individual actual and perceived positions.

Notes: Each circle represents one respondent. The top panels report positions within the student's major and entry cohort; the bottom panels report positions within the student's school and entry cohort. The diagonal denotes equality between perceived and actual position. Lower numerical values indicate better positions. Engineering is excluded because its treatment and survey used deciles rather than exact positions. The major-level panels contain 599 respondents and the school-level panels contain 608 respondents.

awareness of the tool is itself post-treatment, conditional treated-control comparisons among aware students are descriptive rather than causal.

Friend nominations produce 3,389 resolved links from 785 respondents. The association between own GPA and friends' mean GPA is 0.758 (robust SE 0.054; N=782). Respondents in the bottom baseline GPA quintile have friends with mean GPA 7.584; respondents in the top quintile have friends with mean GPA 8.361. These facts support peer selection as one explanation for why informal comparison groups may distort perceived position, without identifying peer effects.

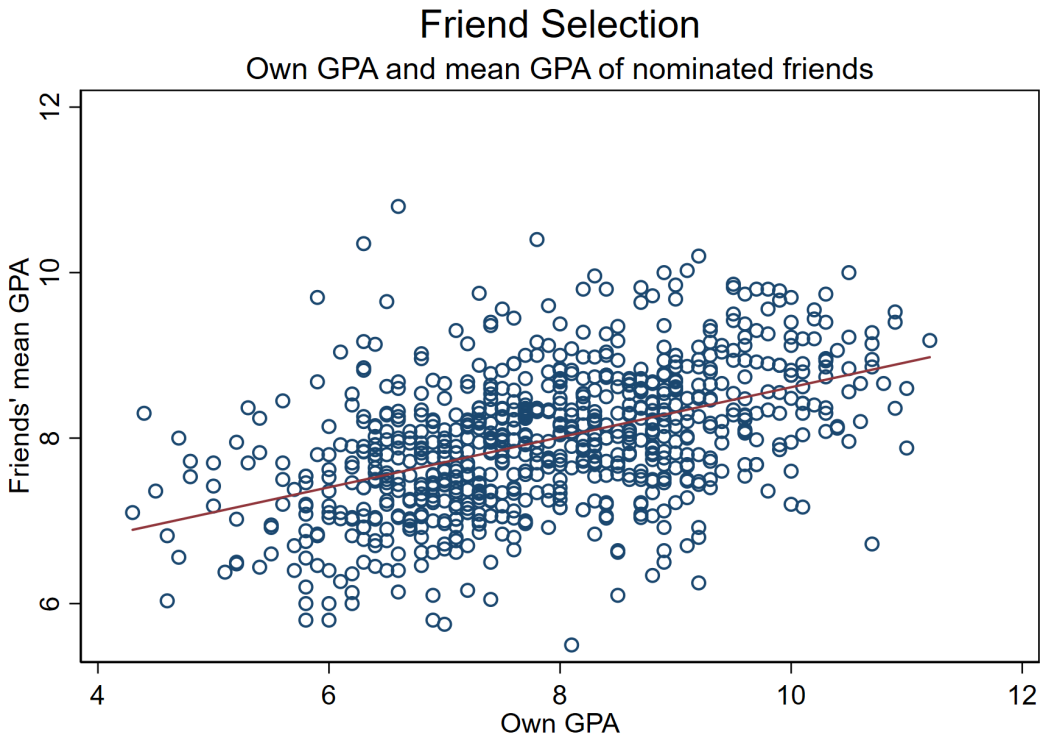


Figure C2. Own GPA and mean GPA of nominated friends.

Notes: The figure plots each respondent's GPA against the mean GPA of up to five nominated friends. The fitted line is descriptive. A one-point increase in friends' mean GPA is associated with a 0.758-point increase in own GPA (robust SE 0.054; N=782). The relationship reflects academic sorting within friendship networks and does not identify peer influence.

D. Inference, Multiplicity, Timing, and Precision

The project did not have a preregistered multiple-testing plan. We group outcomes into six substantively coherent families: satisfaction, one-year exam outcomes, one-year student outcomes, two-year exam outcomes, two-year student outcomes, and long-run graduation. Holm adjustment controls the family-wise error rate within each family and estimand; Benjamini-Hochberg q-values control the false discovery rate. Table D1 reports both adjustments for every outcome and for the treatment-by-woman interaction and total treatment effect for women.

Table D1. Conventional multiplicity adjustments within outcome families

Outcome	Interaction Holm p	Interaction BH q	Women total Holm p	Women total BH q
Five-point GPA satisfaction	0.012	0.012	0.056	0.049
Vignette-adjusted GPA satisfaction	0.015	0.015	0.056	0.049
Exam grade, one year	0.147	0.117	0.553	0.410
Exam taken, one year	0.156	0.117	0.295	0.295
Exam passed, one year	0.200	0.200	0.553	0.410
Cumulative GPA, one year	0.189	0.093	0.351	0.163
Courses passed, one year	0.189	0.093	0.351	0.163
Credits earned, one year	0.189	0.140	0.351	0.173
Exit from baseline major, one year	0.026	0.026	0.020	0.020
Exam grade, two years	0.341	0.331	0.915	0.686
Exam taken, two years	0.095	0.095	0.209	0.209
Exam passed, two years	0.341	0.256	0.915	0.772
Cumulative GPA, two years	0.045	0.045	0.093	0.093
Courses passed, two years	0.028	0.025	0.033	0.018
Credits earned, two years	0.045	0.030	0.035	0.023
Exit from baseline major, two years	0.037	0.025	0.033	0.018
Baseline-major graduation, twelve years	0.015	0.015	0.002	0.002
Any new UM degree, twelve years	0.021	0.021	0.002	0.002

Notes: Holm p-values control the family-wise error rate and Benjamini-Hochberg (BH) q-values control the false discovery rate within each of six outcome families, separately for the treatment-by-woman interaction and the total treatment effect for women. The families are satisfaction, exam and student outcomes at one and two years, and long-run graduation. Families were not preregistered. Conventional unadjusted p-values appear in the main tables.

Table D2. Conditional randomization-inference p-values

Outcome	Interaction RI p	Interaction Holm p	Women total RI p	Women total Holm p
Satisfaction measure 1	0.024	0.024	0.084	0.084
Satisfaction measure 2	0.011	0.022	0.025	0.050
GPA, one year	0.094	0.094	0.170	0.170
Courses passed, one year	0.005	0.020	0.034	0.068
Credits, one year	0.013	0.039	0.016	0.048
Exit, one year	0.015	0.039	0.005	0.020
GPA, two years	0.042	0.042	0.151	0.151
Courses passed, two years	0.003	0.012	0.011	0.030
Credits, two years	0.004	0.012	0.004	0.016
Exit, two years	0.013	0.026	0.010	0.030
Baseline-major graduation, 12 years	0.001	0.002	0.002	0.004
Any new UM degree, 12 years	0.007	0.007	0.002	0.004

Notes: Exit refers to exit from the student's baseline major. Randomization p-values use 999 assignments from the conditional candidate set. Holm adjustment is applied within satisfaction, one-year student outcomes, two-year student outcomes, and long-run graduation outcomes, separately by estimand. The reference distribution holds the earlier-cohort assignment fixed and reassigns the entering 2014 cohort under the recorded balance restrictions. This conditional exercise was not implemented for the repeated exam-level outcomes.

Conditional randomization inference uses studentized statistics for both the treatment-by-woman interaction and the total effect for women. Table D2 reports all twelve outcomes for which the conditional exercise was implemented: two satisfaction measures, four student-level outcomes at each short-run horizon, and two long-run graduation outcomes. It does not report randomization-inference p-values for the repeated exam-level outcomes. Coefficients remain those from the main regressions; randomization inference changes the reference distribution of the tests, not the estimated effect or a second set of standard errors.

Table D3. Confidence intervals and approximate minimum detectable effects

Horizon / level	Outcome	Estimand	Estimate	95% CI	MDE (80%)
One year, exam	Exam grade	Interaction	-0.251	[-0.500, -0.001]	0.357
One year, exam	Exam grade	Women total	-0.102	[-0.285, 0.082]	0.262
One year, exam	Exam taken	Interaction	-0.031	[-0.065, 0.003]	0.049
One year, exam	Exam taken	Women total	-0.021	[-0.046, 0.004]	0.036

Horizon / level	Outcome	Estimand	Estimate	95% CI	MDE (80%)
One year, exam	Exam passed	Interaction	-0.022	[-0.056, 0.012]	0.049
One year, exam	Exam passed	Women total	-0.010	[-0.035, 0.014]	0.035
Two years, exam	Exam grade	Interaction	-0.149	[-0.450, 0.152]	0.430
Two years, exam	Exam grade	Women total	-0.088	[-0.319, 0.144]	0.330
Two years, exam	Exam taken	Interaction	-0.045	[-0.086, -0.004]	0.059
Two years, exam	Exam taken	Women total	-0.029	[-0.060, 0.002]	0.045
Two years, exam	Exam passed	Interaction	-0.028	[-0.068, 0.012]	0.057
Two years, exam	Exam passed	Women total	-0.004	[-0.034, 0.025]	0.042
One year, student	Cumulative GPA	Interaction	-0.183	[-0.380, 0.015]	0.282
One year, student	Cumulative GPA	Women total	-0.119	[-0.271, 0.032]	0.217
One year, student	Courses passed	Interaction	-1.261	[-2.590, 0.068]	1.900
One year, student	Courses passed	Women total	-0.808	[-1.818, 0.203]	1.443
One year, student	Credits earned	Interaction	-5.886	[-13.712, 1.939]	11.182
One year, student	Credits earned	Women total	-4.140	[-10.094, 1.814]	8.508
One year, student	Exit from baseline major	Interaction	0.090	[0.025, 0.155]	0.093
One year, student	Exit from baseline major	Women total	0.072	[0.022, 0.121]	0.071
Two years, student	Cumulative GPA	Interaction	-0.201	[-0.399, -0.004]	0.282
Two years, student	Cumulative GPA	Women total	-0.131	[-0.284, 0.022]	0.219
Two years, student	Courses passed	Interaction	-2.892	[-4.991, -0.793]	3.000
Two years, student	Courses passed	Women total	-2.123	[-3.715, -0.532]	2.274
Two years, student	Credits earned	Interaction	-14.931	[-27.773, -2.089]	18.350
Two years, student	Credits earned	Women total	-11.871	[-21.663, -2.079]	13.991

Horizon / level	Outcome	Estimand	Estimate	95% CI	MDE (80%)
Two years, student	Exit from baseline major	Interaction	0.090	[0.019, 0.160]	0.101
Two years, student	Exit from baseline major	Women total	0.073	[0.019, 0.126]	0.077

Notes: Confidence intervals use the conventional heteroskedasticity-robust or student-clustered standard errors from the main specifications. The MDE is the approximate two-sided effect detectable with 80 percent power at a 5 percent significance level, using the realized standard error and residual degrees of freedom; it is a precision diagnostic, not an ex ante power calculation, and is not adjusted for multiplicity.

The MDE entries are ex post precision diagnostics calculated from realized standard errors. They are not ex ante power calculations and are not adjusted for multiplicity. They help distinguish an imprecise null from evidence of a substantively negligible effect.

The pre-treatment placebo assigns future treatment status to exam outcomes observed from July 2003 through June 2014. Across exam taking, passing, and grades, interaction p-values range from 0.155 to 0.925 and p-values for the total effects on women range from 0.231 to 0.491. At eight years, the adjusted treatment effects for women are -9.8 percentage points for graduation from the baseline major (p=0.002) and -8.5 points for any new UM degree (p=0.004). Excluding the sixteen students with another UM degree before launch yields a -9.4 point effect on any new UM degree for women (p=0.001) and a -10.0 point treatment-by-woman interaction (p=0.012).

Figure D1 plots unadjusted treatment-control differences in cumulative graduation at each common horizon. The women's gap emerges after the first several years and then remains between roughly eight and eleven percentage points; the men's series remains close to zero. The figure is descriptive and is not used to select the twelve-year endpoint, which is the maximum common follow-up available for every randomized student.

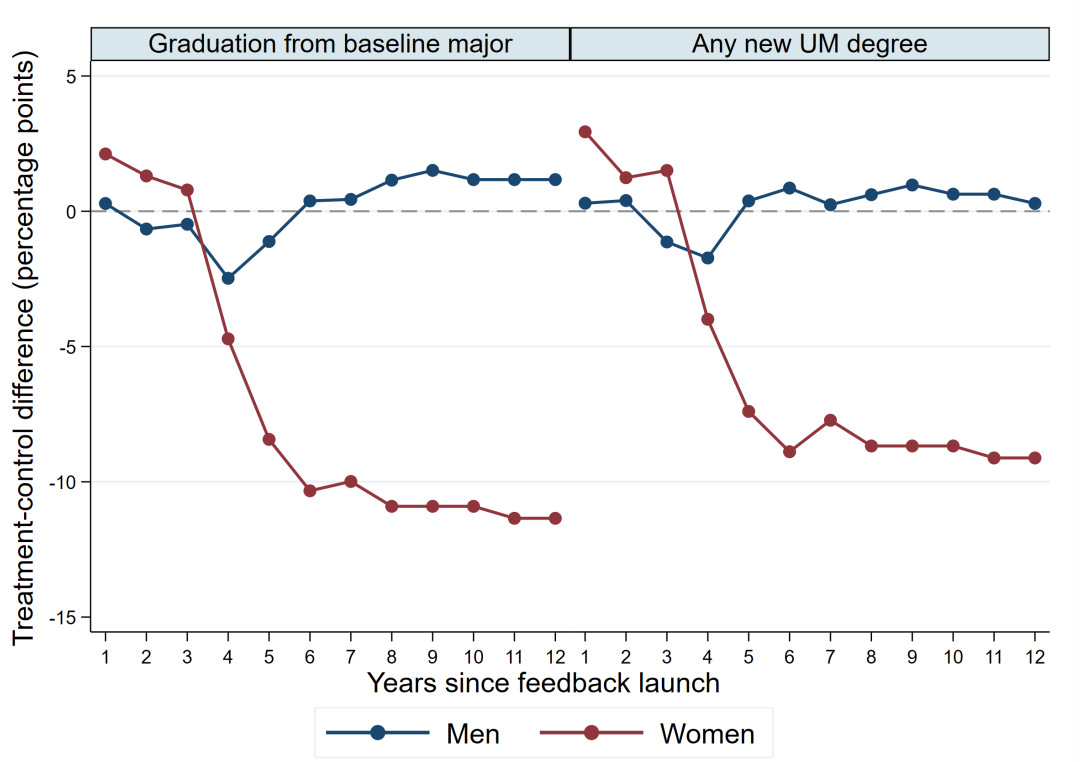


Figure D1. Treatment-control differences in cumulative graduation.

Notes: Each panel plots the unadjusted difference between treated and control graduation rates at horizons from one to twelve years after the cohort-specific feedback launch. The left panel uses graduation from the baseline major; the right panel uses any new degree awarded by Universidad de Montevideo after launch. The series are descriptive; conventional inference for the fixed twelve-year endpoints appears in Table 4B and Table D1, and conditional randomization inference appears in Table D2.

Table D4 reports the assignment-stage decomposition discussed in Section 4.4. It keeps the main treatment effect for men, the treatment-by-woman interaction, and the total effect for women visible within each stage, and reports the cross-stage difference in the interaction.

Table D4. Long-run graduation by assignment stage

Outcome / assignment stage	**Treatment effect: men**	**Treatment × woman**	**Total effect: women**	**Difference in interaction**
Baseline-major	0.003	-0.037	-0.035	
degree	(0.031)	(0.045)	(0.032)	
Earlier cohorts				
Baseline-major	0.068	-0.404***	-0.336***	-0.366***
degree	(0.088)	(0.121)	(0.083)	(0.129)
Entering cohort 2014				
Any new UM	0.005	-0.044	-0.038	
degree	(0.029)	(0.042)	(0.030)	
Earlier cohorts				

Outcome / assignment stage	Treatment effect: men	Treatment × woman	Total effect: women	Difference in interaction
Any new UM degree	-0.002	-0.262**	-0.264***	-0.218*
	(0.075)	(0.108)	(0.078)	(0.116)
Entering cohort 2014				

Notes: The table reports adjusted estimates from a model fully interacting treatment, woman, and an indicator for the entering 2014 cohort. The final column is the difference between the treatment-by-woman interactions for the entering cohort and earlier cohorts. The stage decomposition is exploratory, motivated by the experiment's two-stage assignment, and reported to clarify the scope of the pooled long-run estimate. Heteroskedasticity-robust standard errors are in parentheses.

Taken together, these exercises sharpen rather than overturn the main results. Conditional randomization inference and Holm adjustment continue to support the treatment-by-woman differences and the total effects for women on courses, credits, exit from the baseline major, and graduation. At the same time, the confidence intervals and MDEs show that several exam-level nulls remain too imprecise to rule out economically meaningful effects. The evidence is therefore strongest for progression, persistence, and graduation, not for a uniform change in every academic outcome.

E. Sensitivity Excluding Engineering

Engineering received rank deciles rather than exact numerical positions. Table E1 therefore reproduces every main satisfaction and academic outcome after excluding Engineering. The exercise is a treatment-format sensitivity, not a randomized comparison of exact rank and decile feedback.

Table E1. Sensitivity excluding the School of Engineering

Domain / horizon	Outcome	N	Interaction	p-value	Women total (p)
Satisfaction	Satisfaction measure 1	641	-0.252	0.043	-0.082 (0.370)
Satisfaction	Location placebo	641	0.154	0.290	0.037 (0.724)
Satisfaction	Satisfaction measure 2	639	-0.486	0.058	-0.273 (0.158)
One year	Exam grade	10,631	-0.277	0.055	-0.072 (0.467)
One year	Exam taken	12,589	-0.027	0.209	-0.016 (0.251)
One year	Exam passed	10,631	-0.047	0.021	-0.012 (0.356)
Two years	Exam grade	6,069	-0.154	0.400	-0.011 (0.929)
Two years	Exam taken	6,955	-0.069	0.006	-0.035 (0.051)
Two years	Exam passed	6,069	-0.028	0.277	-0.002 (0.886)
One year	Cumulative GPA	771	-0.188	0.099	-0.093 (0.259)
One year	Courses passed	771	-1.075	0.151	-0.608 (0.226)
One year	Credits earned	769	-5.803	0.193	-3.120 (0.302)
One year	Exit from baseline major	771	0.103	0.009	0.070 (0.010)

Domain / horizon	Outcome	N	Interaction	p-value	Women total (p)
Two years	Cumulative GPA	771	-0.203	0.075	-0.100 (0.230)
Two years	Courses passed	771	-2.690	0.020	-1.738 (0.029)
Two years	Credits earned	770	-13.075	0.063	-9.403 (0.053)
Two years	Exit from baseline major	771	0.094	0.029	0.064 (0.029)

Notes: The table reports every outcome from the main satisfaction and academic specifications after excluding Engineering. p-values are conventional robust or student-clustered values. The location placebo is included.

F. Potential Explanations and Feedback Design

Table F1 reports treatment effects for self-reported competitiveness on the zero-to-ten scale and for indicators based on two alternative cutoffs. Across all three definitions, the treatment effect for men, the treatment-by-woman interaction, and the total effect for women are imprecisely estimated and not statistically different from zero. The estimated gender difference among controls changes with the cutoff, whereas it is near zero on the continuous scale. We therefore do not interpret competitiveness as an explanation for the heterogeneous academic effects.

Table F1. Post-treatment self-reported competitiveness

Outcome definition	Treatment effect: men	Treatment × woman	Total effect: women	Woman among controls
Competitiveness,	-0.037	0.191	0.154	0.069
0–10	(0.257)	(0.363)	(0.249)	(0.290)
Score ≥ 7	-0.009	-0.016	-0.024	-0.008
	(0.047)	(0.070)	(0.051)	(0.054)
Score > 7	0.003	-0.037	-0.034	-0.090*
	(0.046)	(0.066)	(0.048)	(0.052)

Notes: Competitiveness was elicited in the survey after ranking access began. The first three estimands come from regressions including treatment, woman, their interaction, and the baseline controls used in the main analysis. The final column estimates the adjusted gender difference among controls. Heteroskedasticity-robust standard errors are in parentheses. N=841 for the treatment specifications and N=414 for the control-only comparisons.

Table F2 reports survey evidence on students' preferences for learning and potentially publicizing rank. The interaction for concern about others knowing one's exact position is visible on both the four-point scale and the binary high-concern measure. These outcomes are exploratory and post-treatment. They inform the design of future feedback systems but cannot establish mediation of the academic effects.

Table F2. Preferences for learning and publicizing rank

Outcome	Treatment effect: men	Treatment × woman	Total effect: women	Control mean: women
Interest in knowing own rank (1-4)	-0.227**	0.160	-0.066	2.584
	(0.112)	(0.164)	(0.124)	
Interest in own rank: high	-0.092	0.118	0.026	0.545
	(0.062)	(0.091)	(0.068)	
Concern if others knew rank (1-4)	-0.167	0.471***	0.305**	1.662
	(0.106)	(0.164)	(0.126)	
Concern if others knew rank: high	-0.064	0.266***	0.203***	0.130
	(0.052)	(0.075)	(0.055)	

Notes: Outcomes were elicited after treatment in the satisfaction survey (N=562 in the controlled specifications). High indicators equal one for the top two categories. Models include treatment, woman, their interaction, and the baseline controls used in the main analysis. Robust standard errors are in parentheses. These exploratory post-treatment outcomes describe responses to feedback design; they do not identify a mechanism.

Library and study-time measures are pre-treatment correlates. Students with no recorded loan are assigned zero rather than omitted. Study-time observations are aggregated to the student-semester-subject level before estimation, and standard errors are clustered by student.

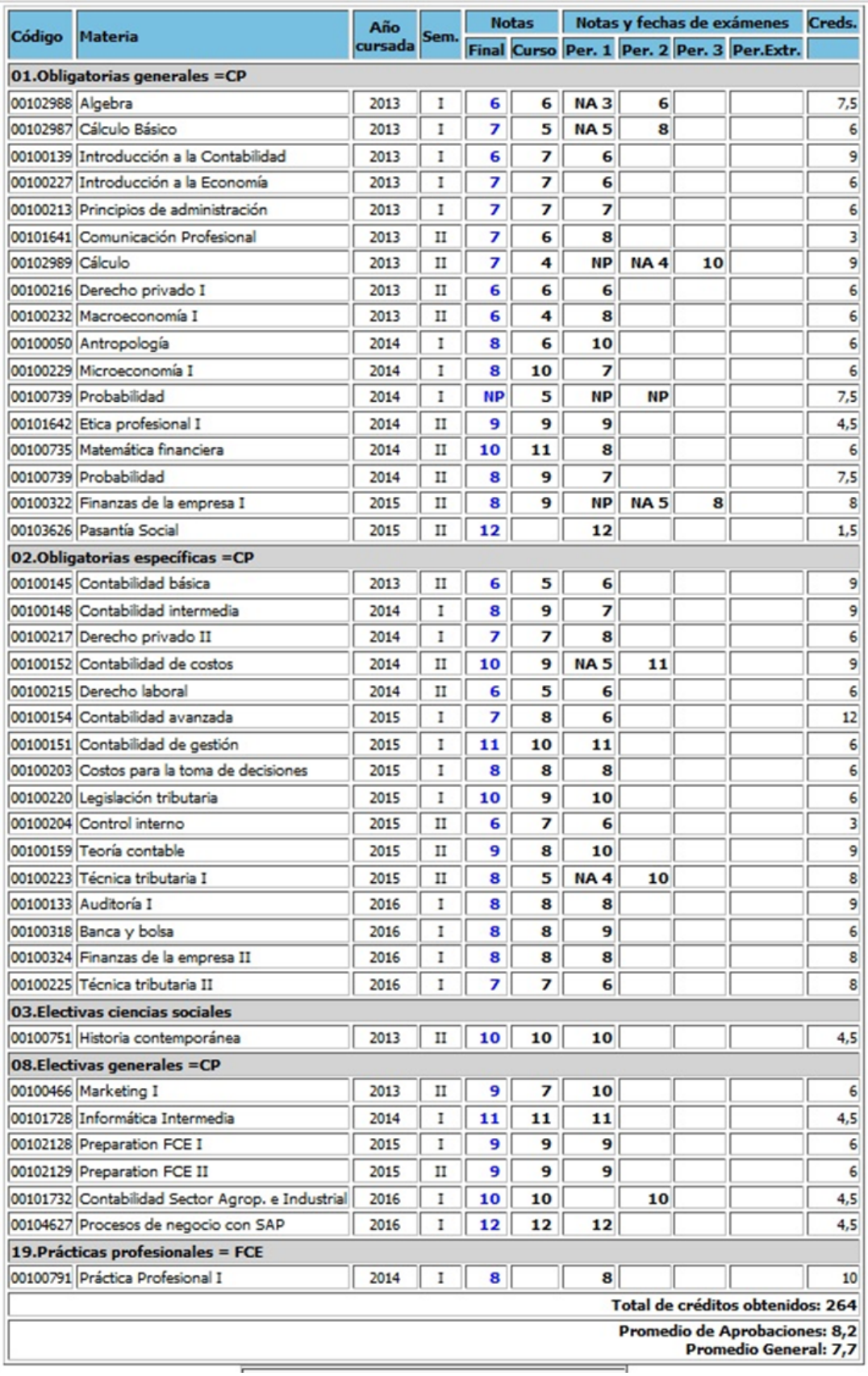
G. Treatment Materials

The treatment and control e-mails were identical except for one final sentence announcing access to the ranking. The original Spanish common text was: "Estimados alumnos: Te recomendamos planificar bien tus exámenes y el semestre que viene. Para eso ten en cuenta: las fechas de exámenes (disponibles en [enlace]); el período de modificaciones ([enlace]); y tu escolaridad y la grilla de avance académico ([enlace]). Saludos, [firma institucional]." The treated message added before the closing: "También tienes disponible a partir de hoy un ranking de tu desempeño en la facultad."

The corresponding English translation is: "Dear students: We recommend that you plan your upcoming exams and the next semester carefully. To do so, please keep in mind: the exam dates (available at [link]); the course-modification period ([link]); and your transcript and academic-progress grid ([link]). Best regards, [institutional signatory]." The treated message added before the closing: "Starting today, you can also access a ranking of your performance within your school." Bracketed text replaces identifying intranet links and the institutional signatory. There was no experimental task, payment protocol, or separate set of behavioral instructions.

Before the intervention, this transcript was the principal source of academic information available in the intranet. It reported courses, grades, credits, and cumulative GPA, but not a student's position relative to peers.

Figures G2 and G3 reproduce two additional real treatment examples. The first student learned that he ranked 10th of 20 in his major and that his GPA had improved relative to both comparison means. The second learned that she ranked 49th of 74 and that her GPA had declined relative to peers. These examples show why the treatment was more informative than a one-time rank: it placed the current position alongside an academic trajectory.



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**										

Figure G1. Example of the academic transcript available before the intervention.

Notes: The transcript lists courses, grades, examination attempts, credits, and cumulative GPA. Direct student identifiers are not displayed. It does not report a position in the distribution of grades.

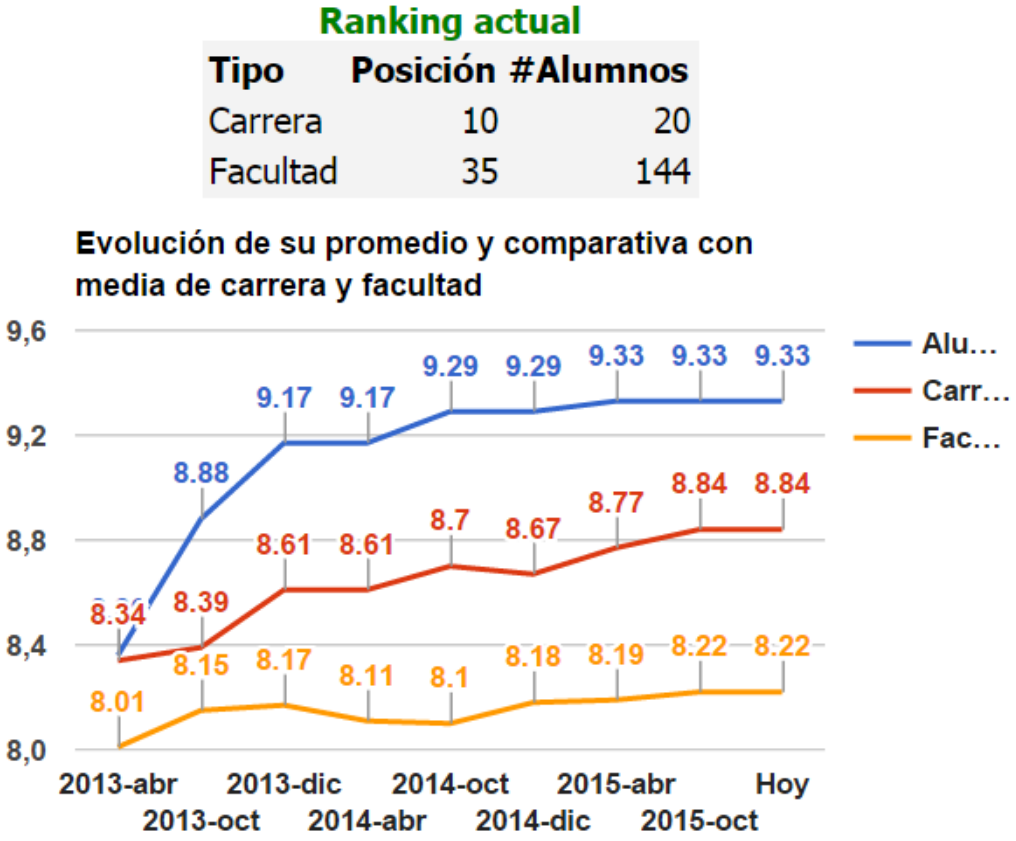


Figure G2. Real treatment example: improving relative performance.

Notes: The example shows the student's current position and the evolution of the student's GPA (blue) relative to the major mean (red) and school mean (orange). The screenshot has been cropped to remove all direct personal identifiers.

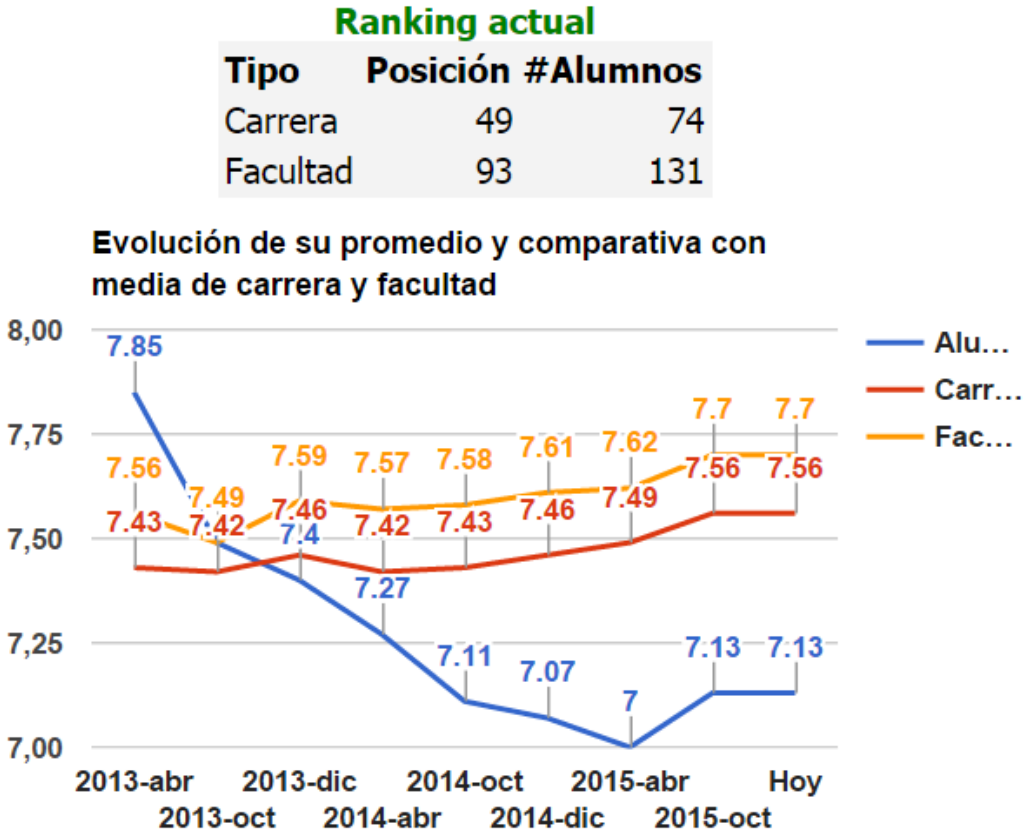


Figure G3. Real treatment example: declining relative performance.

Notes: The example shows the student's current position and the evolution of the student's GPA (blue) relative to the major mean (red) and school mean (orange). The screenshot has been cropped to remove all direct personal identifiers.