RANK INCENTIVES AND SOCIAL LEARNING: EVIDENCE FROM A RANDOMIZED CONTROLLED TRIAL

L.I. DOBRESCU†, M. FARAVELLI‡, R. MEGALOKONOMOU§, AND A. MOTTA⋆

ABSTRACT. In a 1-year randomized controlled trial involving thousands of university students, we provide real-time private feedback on relative performance in a semester-long online assignment. Within this setup, our experimental design cleanly identifies the behavioral response to *rank incentives* (i.e., the incentives stemming from an inherent preference for high rank). We find that rank incentives not only boost performance in the related assignment, but also increase the average grade across all course exams taken over the semester by 0.21 standard deviations. These beneficial effects remain sizeable across all quantiles and extend beyond the time of the intervention. The mechanism behind these findings involves *social learning*: rank incentives make students engage more in peer interactions, which lead them to perform significantly better across the board. Finally, we explore the virtues of real-time feedback by analyzing a number of alternative variations in the way it is provided.

Keywords: relative performance feedback; rank incentives; social learning; academic performance; randomized controlled trial.
JEL codes: J24, J18.

*We are grateful to Zachary Breig, David Byrne, Dimitri Christelis, Luke Chu, Francesco Fallucchi, Miguel Fonseca, Lata Gangadharan, Benjamin Hansen, Boon Han Ko, Umair Khalil, Andreas Leibbrandt, Daniele Nosenzo, Maria Recalde, Tom Wilkening, Haishan Yuan, as well as seminar and conference participants at several institutions, for helpful discussions and suggestions. This work was approved by the Human Ethics Research Committee at University of New South Wales (201700142) and University of Queensland (HREC2017001402).
§Corresponding author. School of Economics, University of New South Wales. Email: dobrescu@unsw.edu.au.
‡School of Economics, University of Queensland. Email: m.faravelli@uq.edu.au.
§School of Economics, University of Queensland. Email: r.megalokonomou@uq.edu.au.
⋆School of Economics, University of New South Wales. Email: motta@unsw.edu.au.
Date: June 19, 2019.

1. Introduction

We live in a world obsessed with rankings. From sport competitions and school ratings to the number of likes on a Facebook post or views of a YouTube video, every day we witness society's fixation with relative performance. Social comparisons are also encouraged within organizations: allowing individuals to compare their performances has the potential to increase their productivity both in educational (Azmat and Iriberri, 2010; Tran and Zeckhauser, 2012; Katreniakova, 2014) and labour settings (Mas and Moretti, 2009; Blanes i Vidal and Nossol, 2011).

Social comparisons, however, are not always a silver bullet. Individuals who strive to improve their relative standing are likely to require both effort and ability to achieve their goal. If they discover that others' ability is lower than initially thought, they might decide to put less effort into the task at hand (Azmat et al., 2019). Interestingly, the same drop in effort can occur due to demoralization effects brought by a lower-than-expected rank (Barankay, 2011).[1] Overall, comparing the existing studies is particularly challenging because of the numerous confounding factors: the behavioral response to incentives stemming from an inherent preference for high rank (*rank incentives* henceforth)[2] is potentially compounded with financial and signaling aspects, learning and experimentation processes, multi-tasking considerations, peer-pressure and changes in beliefs about future compensation schemes and relative ability. This underlying complexity could explain the mixed results in the literature, with rank feedback being productivity-enhancing in certain contexts but not others (Bandiera et al., 2013; Bursztyn and Jensen, 2015; Blader et al., 2016).[3]

In this paper, we report results from a 1-year randomized controlled trial (RCT) in the education sector, involving university students who continuously received private feedback on their relative performance in a semester-long computerized (online) assignment. Our contribution is multifold. First, our experimental design allows to cleanly identify the behavioral response to rank incentives *per se*. Second, we are the first to provide evidence that rank incentives can be effective in higher education.[4] Indeed, we find that feedback has a positive impact on students' performances not just in the online assignment on which

[1] There are also a considerable number of laboratory experiments on the effect of feedback on relative position - see, for instance, Hannan et al. (2008), Eriksson et al. (2009), Khunen and Anieszka (2012), Charness et al. (2013), Gerhards and Siemery (2014), Azmat and Iriberri (2016), Gill et al. (2018), among others.

[2] See Barankay (2012), Tran and Zeckhauser (2012).

[3] From a welfare perspective, the provision of relative performance feedback does not seem to increase stress levels (Katreniakova, 2014), but it can affect satisfaction (happiness and dominance levels) in either direction (Azmat and Iriberri, 2016).

[4] The interventions in Tran and Zeckhauser (2012) and Brade et al. (2018) also involve university students. However, (i) Tran and Zeckhauser (2012) study the impact of rank incentives on the score attained in a standardized, externally administered TOEIC test, while (ii) Brade et al. (2018) provide feedback on accumulated course credits; thus, one cannot disentangle rank incentives from the other incentives associated with the tangible benefit of signaling ability or readiness for the job market.

they were ranked, but also in all (invigilated) course exams and across the entire grades' distribution; this impact is long-lasting, with positive spillovers to other courses beyond the intervention period. Third, we are the first to explore the virtues of real-time feedback; as we will discuss below, this feature may be at the basis of the success of our implementation. In point of fact, the success of our RCT does suggest that results may be very sensitive to even apparently innocuous design features.[5] Our fourth contribution is to examine some of these feedback characteristics by analyzing a number of alternative variations in the way it is provided. Last, but not least, we uncover the mechanism through which rank feedback translates into academic performance. Our findings suggest that feedback disclosure makes students engage more in *social learning* and this vehicle leads them to perform significantly better in the online assignment and, more generally, in the course. From a theoretical perspective, social learning has been proven to emerge when individuals are allowed to observe each other's outcomes. Observing these outcomes, for example, can help one select a more effective technology (Wolitzky, 2018). In our context, students may question whether they are approaching the task (i.e., studying) in the most effective manner: ranking not only conveys information about one's relative ability and effort levels, but it also reveals how efficient one's productive technology is.

Tackling our research questions is challenging because it requires detailed information on how individuals *prepare* for a task. To achieve this, we first develop a semester-long online assignment with a leaderboard system. The assignment is a collection of practice tests (i.e., sets of exercises) tackling a series of simple problems. The leaderboard score is the product of the completion rate (i.e., the number of completed exercises over the total number of exercises available) and the success rate (i.e., the number of successful attempts to solve exercises over the total number of attempts made).[6] This structure ensures that students remain clustered around a similar score as they progress in the assignment (completion rate), while also allowing for enough variation (success rate). All students are free to engage with the assignment at any time during the semester, while those in the treatment group are also shown their score rank for one minute *any time this varies*, either upward or downward.

A consequence of this design is that students can act right after inspecting their relative performance and, therefore, they can affect their assignment rank almost instantly. Hence, any initial boost of effort due to students experimenting with the effort level required to change their rank - and its transitory impact on performance - is unlikely to affect the overall performance by the end of the semester. We also note that the leaderboard score had no

[5] The potential for real-time feedback to address the bias toward what is salient and immediately visible has also been documented by Tiefenbeck et al. (2016) and Jessoe and Rapson (2014) for energy consumption.

[6] Students are provided with random values for each exercise and have to take decisions that lead them to achieve certain economic goals. Once a decision is taken, it is automatically stored on the server. Students receive feedback after every decision: if it is correct, they are allowed to proceed to the next one; if it is suboptimal, they are asked to review their choice.

bearings on the overall course grade, as 20% of it depended exclusively on a student's completion rate at the end of the semester, and not on how well they performed (either in terms of success rate or in terms of leaderboard score). This precludes any possible financial or signaling considerations. Our setup is also characterized by the private nature of feedback. This feature allows us to rule out confounding factors that may otherwise concur to generate the results - such as, for instance, status seeking behavior or attitudes driven by parents' pressure.[7] Finally, to verify whether treated students perform better at the expense of other academic activities. (i.e., address the multi-tasking argument), we collect information on their performance in all the other courses taken in the same semester. All this considered, our treatment effects are likely to cleanly capture *rank incentives* as in Barankay (2012).

To identify the impact of this type of ranking, we exploit the tests required to assess the notification system when the online assignment was first deployed to students. The RCT took place in the first semester of 2016 and involved the students enrolled in a large Principles of Microeconomics course at a major, research intensive, selective university. As part of these tests, students were randomly divided into five groups: one control (receiving no information on their relative performance) and four treatments (featuring the leaderboard system described above and three slight variations of it). Students were not aware of these tests and thus did not know they were being treated. Instructors were not involved in these tests either and did not know the groups their students were assigned to.

We find that treated students perform better in the online assignment, ending up, on average, 62 (out of 1,093) positions higher in the final ranking than students in the control group. Such improvement is robust and of similar magnitude across all quantiles. What really matters, though, is that this direct effect on the assignment also translates into better performance in all course exams taken over the semester. Our results show that providing feedback increases the average grade across all invigilated exams over the semester (i.e., two mid-terms and a final exam) by 0.21 standard deviations (SDs henceforth). These effects remain positive, robust and sizeable across all quantiles, with no further heterogeneity by gender, age, international status or field of study.[8]

These results prompt the question of whether feedback provision in a course assessment can have a spillover impact on the other courses taken in the same semester. And can we identify any long-lasting effects of our intervention on academic performance next semester? To answer these questions, we use the adjusted GPA in Semester 1 2016,[9] as well as GPA in Semester 2 2016, and re-run our analysis. We find that our intervention does not change performance in the other courses taken the same semester, with the corresponding

[7] For an overview on providing private versus public ranking feedback, see also Tran and Zeckhauser (2012) and Gerhards and Siemer (2014, 2016).

[8] Exceptions involve females and 19+ students, with more pronounced beneficial effects in the final exam.

[9] Adjusted GPA is derived from all the courses taken in Semester 1 2016, except for the intervention course.

GPA remaining unaffected by feedback provision across all treatments. This suggests that treated students did not perform better at the expense of other academic activities. Estimates also reveal, however, an increase in academic performance next semester by 0.17 SDs. This might be due to the increased inclination to engage in social learning and its positive long-term spillovers in the following semester. It is also possible that treated students developed healthier study habits, perhaps owing to extra incentives provided by the tighter effort-reward signal embedded in our leaderboards.

These results are noticeable for four reasons. First, providing feedback on a drill that has no bearings on students' grades not only has a positive effect on students' performance in that drill, but also on their grades throughout the semester. Second, this effect translates into a performance boost which is equivalent to being taught by a teacher 2 SDs above the average or, alternatively, in a class 20% smaller - both extremely costly interventions.[10] Third, our findings indicate that relative performance feedback can benefit students across the entire grade distribution. Fourth, the persistence of the effect several months after the intervention is remarkable and points towards ranking being a successful and effectively costless tool to boost performance, not only in the short-term but also over time.

Finally, we turn to our findings pertaining to the underlying mechanism. In this respect, results are consistent with the competitiveness theory (see Azmat and Iriberri, 2010): after receiving feedback, all students - independently of their rank - exert more effort and thus improve their absolute performance, first in the online assignment and, subsequently, in several course exams. Can we actually observe their effort? Fortunately, we possess two such measures. First, we observe the time they spend on the online assignment. In our case, the time spent appears to be the same across all treatment and control groups. However, effort takes different forms and shapes. Our second source of data is related to students' activity as recorded on two separate discussion boards, one internal and one external to the official course website. Both boards are used by students to ask questions related to the course content and materials. Among the students who were engaged in the discussion boards, we find that those exposed to their rank post about 53% more often than their control counterparts. However, we find no effect at the extensive margin, suggesting that our intervention made active students even more active, rather than encouraging more students to contribute to these public forums. Taking the number of posts as an indication of course engagement, we interpret this as evidence that students in the treatment group interacted more with their peers, which resulted in higher performance in all other course exams. This suggests that *social learning* is the likely channel at play here. Treated students learn from each other more and perform better both in the assignment and in the actual course exams.

[10] See Angrist and Lavy (1999); Krueger (1999); Hanushek et al. (2005); Chetty et al. (2014).

A secondary contribution of this paper is to provide the first evidence of the effectiveness of *continuous* relative performance feedback and to show that results on younger, pre-college aged students do carry on to higher education.[11] If the objective of a policy maker is to improve academic performance across the board, our results suggest that students should be allowed to take action right after inspecting their relative performance - effectively giving them the chance to affect their rank almost instantly[12] - rather than updating and revealing their relative performance more infrequently (Azmat et al., 2019; Brade et al., 2018).

Using this new type of continuous feedback raises a number of interesting issues. What is the optimal design? How frequent should the feedback be? What type of feedback works best? In this paper, we attempt to offer a few preliminary answers to these questions by breaking down the main features of the leaderboard system to study what elements are most effective in eliciting effort. Specifically, we exploit the RCT to deconstruct the leaderboard based on two basic components: feedback type and duration. Students in the second treated group were *constantly provided with their rank*, instead of being exposed to it for just one minute when it varied. By comparing this treatment ("Nonstop") with our principal treatment ("Main"), we can study the impact of feedback duration on students' behavior. In the third treatment ("Positive") students were shown their rank for one minute *if and only if this varied upward*. In the fourth treatment ("Negative"), they were shown their rank for one minute *if and only if this varied downward*.[13] We use these two last treatments to examine how performance is affected by the type of news the feedback conveys.

Taken separately, all treatments have positive point estimates relative to control, with magnitudes decreasing in the following order in the case of average exam grade: Main, Nonstop, Negative and Positive. A similar trend exists for course engagement, namely: Main, Nonstop, Negative and Positive. That said, only students in our Main treatment significantly outperform control students in terms of both course exam grades and social engagement. This might be due to lack of power or perhaps it is an indication that our continuous-type feedback is very sensitive not only to the *type* of information disclosed, but also to the *way* in which it is disclosed. It is also worth noting that, for instance, there is no significant difference between the Main treatment and Nonstop (either in terms of grades or in terms of social engagement), while we have strong indication that students in the Main treatment outperform Negative students.

[11] See Azmat and Iriberri (2010); Tran and Zeckhauser (2012); Katreniakova (2014).

[12] In our context, This was possible because (i) the technology adopted allowed it, and (ii) the feedback was related to a semester-long assignment, by all means equivalent to regular training/practicing, as opposed to exam grades that are much less frequent by their very nature.

[13] To the best of our knowledge, only a few papers have studied so far how productivity is affected by changing the likelihood of rank feedback, the reference group used, and the informativeness of feedback (Khunen and Tymula, 2012).

All in all, these results suggest that giving only positive or negative news is not as effective as providing full information. It is unclear, however, how feedback effectiveness depends on its duration (and its ability to trigger attention - see Wedel and Pieters (2012)). While these are intriguing ideas, they remain purely speculative and their exploration is beyond the scope of this paper. More research is required to shed light on them.[14]

The paper proceeds as follows. The next section describes the RCT and its context. We then move to present the data and our empirical approach, followed by a discussion of our findings and of the mechanism behind them. Finally, we conclude and discuss future work.

2. Randomized Controlled Trial

2.1. *Environment*

The RCT took place in Semester 1 2016 among the students of a large Principles of Microeconomics course at a research intensive, selective university. The course is taught every semester to more than 1,000 students, over 13 teaching weeks. Each week, students attend a 2-hour live lecture and a 1-hour tutorial; neither classes are compulsory and attendance is not recorded. Lectures are delivered by academic staff, while tutorials are taught by teaching assistants (also known as 'tutors'). All lectures and tutorials take place in the same campus. In our case, all lectures throughout the semester are taught by the same lecturer, who is also the sole course coordinator. As for tutorials, students are randomly assigned to these classes at the beginning of the semester and cannot switch between them during the term. Each tutorial includes, on average, 24 students and effectively consists of solving exercises and discussing course materials, both activities guided by a tutor. All instructors (i.e., lecturer and tutors) use the same teaching material, including textbook, course notes and slides, and tutorial exercises (with standardized solutions provided by the course coordinator). Finally, there are two discussion boards associated with the course, one internal and one external to the official course website. Both discussion boards are accessible by all students and all instructors, and are used to post comments and ask (or answer) any course-related questions.

From 2016 onwards, an educational software was adopted as part of the course material. This software provides students with access to an extensive database of exercises and links to the (electronic) course textbook. The textbook covers all the topics traditionally taught in a standard Principles of Microeconomics course, from the principle of comparative advantage through to externalities and public goods. Exercises, on the other hand, are grouped into several sets, each focusing on a different economic topic and each corresponding to a

[14] Our results are tentatively in line with the demoralization effect reported in Barankay (2012), but are in contrast with Khunen and Tymula (2012) and Azmat et al. (2019), who find those ranking lower (higher) than expected increasing (reducing) effort, and with Gill et al. (2018), who report U-shaped rank response functions.

different textbook chapter. These sets are released progressively throughout the semester, keeping track of the issues discussed in class. Students are required to master them in a certain order, but upon completing them, they can go through them at will, in any order and at any time. Correctly solving all available exercise sets (i.e., fully completing the online assignment) by the end of Week 13 is worth 20% of the overall course grade. In case of partial completion, students receive a proportion (approximated to the first decimal) of the 20% that is equivalent to their completion percentage.

Besides the semester-long online assignment, the course assessment structure also included (i) two invigilated mid-term exams taking place in Week 6 and Week 10 of the semester, containing several essay questions, each worth 20% of the overall course grade, and (ii) one invigilated final exam, taking place at the end of the semester, containing only multiple-choice questions and counting as 40% of the overall course grade. The exam papers for all three invigilated exams are created by the course coordinator, who draws the corresponding questions from a pre-existing database of uniformly difficult questions. Each tutor marks an equal proportion of mid-term exam papers, not necessarily from her own tutorial students; marking is double-blind and follows a strict set of marking guidelines provided by the course coordinator with rigurous consistency checks in place. A machine automatically grades the multiple-choice questions of the final exam.

2.2. *Treatments*

As anticipated in Section 1, there are two basic measures of students' performance in the online assignment. The *completion rate*, at a given point in time, represents the proportion of exercises completed up to that moment. The *success rate*, at a given point in time, represents the percentage of correct decisions taken up to that moment.[15]

Let P denote the performance index, as generated by the product of a student's completion and success rate. As a feature of the online assignment, the software includes a leaderboard ranking all students in the course based on their P index. Every *five* minutes the server updates the ranking and, if an individual's position has varied, notifies the student by displaying on their screen (i) their relative position with respect to all other participants, and (ii) their latest variation in ranking. This information disappears after *one* minute and appears again only once a subsequent server update of the leaderboard picks up another variation. Figure 1 shows an example of a student who dropped 11 positions and is currently ranked 770^{th} out of 1,101. Moreover, by clicking the button "i" on the icon, an information box would appear and explain how the individual rank is constructed.

[15] Thus, if two students have the same completion rate, the one who has made fewer mistakes has a higher success rate. However, a student who has taken only one decision, provided it is correct, has a higher success rate than one who is ahead in the assignment (and possibly already completed it) but has made mistakes.

To test the notification system, the software developer conducted a (platform-wide) RCT by varying the manner in which rank feedback was conveyed online to students.[16] Specifically, students were randomly divided at the start of Semester 1 2016 into five groups, depending on the last digit of their student ID. As mentioned, they were not aware of the nature of these tests and did not know they were being treated.[17] Furthermore, the instructors were not involved in any of these tests and none of them was aware of the treatment group each student was assigned to. This provides us with a unique opportunity to examine the impact of providing relative performance feedback both on students' performance in the online assignment and, more importantly, in the course.

Each of the five groups (one control and four treatments) consists, on average, of 220 students. The first group of students received feedback in the way we described above (i.e., for one minute only, every time the server picked up a ranking variation; we will refer to this as the *Main* treatment). The control group received no information about ranking. The remaining three treatments consist of small variations of the Main treatment. We will call these treatments *Nonstop*, *Positive* and *Negative*. In the Nonstop treatment, ranking would be *constantly* displayed on the screen. In the Positive treatment, feedback would only appear if the student's ranking had *improved* and would be visible for one minute before disappearing. In the Negative treatment, feedback would only appear if the student's ranking had *worsen* and would be visible for one minute before disappearing. Thus, Nonstop conveys the same type of information as Main, but in a more invasive manner. Positive and Negative provide feedback in the same way as Main (i.e., only when rank varies and only for one minute), but their informational content is not as rich. In what follows, we attempt to investigate how these feedback disclosure variations affect performance and identify the corresponding mechanism.

3. Data and Empirical Analysis

This section provides an overview of the data and then discusses our empirical approach. Since we examine the effects of our intervention directly on the course, the mechanism that drives them and whether they extend in any way to the longer-term, we will present these three cases separately.

[16] These tests were conducted (i) to ensure a student's rank was correctly calculated, displayed and updated, and (ii) to identify the rank disclosure format that would elicit the most user activity (which the developer ended up implementing accross the entire platform in its next software iteration). Specifically, while the Main treatment was the default disclosure format, three additional small variation formats were also deployed in the RCT, in order to check which one would attract more user engagement.

[17] During the entire semester, only a handful of students - in 1,101 participants - asked why their assignment did not display their ranking as it did for some of their peers. They were told that this feature was not available to all students, as the developers were testing it. Interestingly, no one asked why they *did have* access to their ranking or why this was provided in a certain way, which suggests that students were not puzzled by the provision of feedback, nor by the way in which it was provided.

3.1. *Data and Descriptive Statistics*

The data we use in our analysis come from university administrative records, as well as from the software developer logs and two course discussion boards. Specifically, our sample consists of 1,101 students, coming from 32 countries and taking the Principles of Microeconomics course in Semester 1 2016. During this period, there were 46 tutorial classes available, taught by 16 different tutors. As mentioned, students are randomly assigned to tutorials at the beginning of the semester and cannot change their class at any point during the term. Below we provide evidence that this randomization worked well.

Table 1 presents the main descriptive statistics for our sample at the student (Panel A.1) and tutorial level (Panel A.2). A quick glance reveals an almost 50:50 split between males and females. Nearly 78% of the students are Australian, while a significant proportion come from Asia (around 20%). Only 18% are Economics students, while most of the others study degrees in Commerce, Business, and Science, Technology, Engineering & Mathematics (STEM henceforth). As for tutors, roughly 48% are males and almost 35% are international.

As discussed in Section 2.2, in Semester 1 2016 the software developer of the online assignment adopted a prototype leaderboard system. This offers a unique opportunity to examine the effects on academic performance of small variations in the way students' relative performance feedback was displayed. Our identification strategy relies on developer's random assignment of students into five groups, based on the last digit of their ID number.

Table 2 reports differences in students' pre-determined characteristics across these groups, both overall and at tutorial level. Such characteristics refer to students' age, gender, degree undertaken, international student status and country of birth, as well as prior academic performance when available. Specifically, while we do have an ex-ante unified measure of prior ability for domestic students, we do not, unfortunately, possess a similar measure for international students. This is because the international students in our sample come from (32) different countries, all with different academic standards. As their high-school graduation and international university admission exams are of different scales and difficulty, the university where our intervention took place uses this information for admission purposes but does not maintain it in its records. As a result, for domestic students (roughly 78% of our sample), we proxy prior academic performance by their comparable high-school score, called the Australian Tertiary Admission Rank (ATAR hereafter).[18] For international students enrolled before Semester 1 2016 (35% of the international sub-sample) we will use their previous semester GPA ("GPA Previous Semester: International" in Table 2).

We compare students' characteristics in each of the feedback groups to those in the control group (see Main, Panel A; Nonstop, Panel B; Positive, Panel C; Negative, Panel D).

[18] ATAR is the primary criterion for entry into Australian undergraduate programs and denotes a student's high-school ranking relative to their peers when completing secondary education.

The first two columns in Table 2 report means and SDs for each treatment group, while the following two display the same descriptive statistics for the control group. The last two columns present the differences in means between the two groups and the related standard errors, respectively. We can quickly confirm that there are no statistically significant differences between treated students and those in the control group in *any* of the pre-determined characteristics at our disposal.[19]

Next, we also check if indeed students are randomly assigned to tutorials. To do so, we compare tutors' characteristics (at tutorial level) for each treated group with the control group. The figures reported in the bottom section of each panel in Table 2 prove this is the case. Finally, an F-test reassures that it is indeed not possible to statistically reject the hypothesis that students' assignment is random, both at the course and at the tutorial level.

In what follows, we will control for all predetermined characteristics, except for prior academic performance. Our reason is threefold. First, doing so would see our sample significantly reduced. For instance, controlling for ATAR would imply using only 78% of the students enrolled in the course, making our estimates much noisier. Second, doing so would also imply dropping a selective, non-random sub-sample of the population - all the international students. International students are, however, likely to differ from domestic students across various unobservable dimensions and may have a different valuation of their university degree.[20] Moreover, while 60% of domestic students are male, this percentage drops to 43% among internationals. Of course, on top of controlling for ATAR we could also control for previous semester GPA for that minority of international students who enrolled at university before the intervention semester. While this would slightly alleviate the small sample issue (i.e., we would only drop 14% of our total sample in this case), we would still be discarding a non-random sub-sample of the population - all of the first year international students. Furthermore, previous semester GPA is a noisier proxy for prior ability than ATAR. Finally, the software developer randomized students *solely* according to their student ID number and, thus, participants are unlikely to differ in their observed and unobserved characteristics across groups. Indeed, Table 2 shows that the randomization worked properly for *all* observable characteristics for which we have information (including ATAR for domestic students and previous semester GPA for internationals). Hence, it is highly improbable that it failed in a single dimension, i.e., (a uniform measure of) prior ability. For all these reasons, we control in our main analysis only for those characteristics which are available to us for the entire sample. (A robustness check that includes an imputed measure of prior ability shows that our results remain unchanged - see Section 6.)

[19] Accounting for prior ability and tutorial fixed effects leaves our balance checks unchanged, except for some country of birth group differences which become significant. All our specifications control for these indicators.
[20] Note, for instance, that international students pay much higher tuition fees, more than three times higher than domestic students.

Besides the administrative data, we also have access to the software developer's logs. For each student, this database provides the following assignment-related data: (i) the final completion rate, (ii) the final success rate, (iii) the final rank, and (iv) the total amount of time spent on the assignment. At any point in time, students knew their completion rate, but they never observed the success rate. (Hence, control students could not infer their rank by observing the rank and completion rate of a treated student.) And, by definition, treated students were aware of their personal rank.

Finally, two course-related discussion boards provide further information on peer interactions within the course, via the posts written on these forums by each student in our sample.

3.2. *Empirical Methodology*

Course Effects. Our identification strategy relies on comparing the outcomes of students with similar characteristics, similar classmates and the same tutors, but who are exposed to different feedback treatments. We will analyze two different types of course effects. First, we will start by examining whether the various feedback treatments - to which students were randomly assigned - had any impact on students' performance in the online assignment, as indicated by their final assignment ranking. Next, we will examine the effect of feedback on academic performance as captured by students' grades in the invigilated course exams administered during the semester, namely the two mid-terms and the final exam. To capture the overall effect, we also use the average exam grade computed as the mean of the three aforementioned tests (see Figure 2). Note that this information provides a good measure of students' learning and academic performance. First, we can rule out the possibility that the instructors may, in any conceivable way, artificially drive the effects, as none of them was aware of the treatment group each student was assigned to. Second, all exams are closed-book, administered in-class and invigilated, which makes them an objective measure of individual attainment. Third, the two mid-terms are marked - according to strict guidelines - by tutors, who are unaware of the experiment. As for the final exam, marking is entirely computerized. Finally, exam grades are not adjusted or re-weighted and so, they reflect each student's absolute performance in the course.

Our basic estimating equation takes the following form:

$$Y_{i,d,c} = \alpha + \beta Treatment_{i,d,c} + \gamma X_{i,d,c} + TutorialFE_c + u_{i,d,c} \qquad (1)$$

where $Y_{i,d,c}$ is either (i) the final assignment rank of student i, enrolled in degree d, attending tutorial class c or (ii) the grade achieved by student i, enrolled in degree d, attending tutorial class c in any of the three course exams, either separately or on average. The dummy variable $Treatment_{i,d,c}$ takes the value of one if a student is in one of the treatment groups and zero for control students. $X_{i,d,c}$ refers to students' characteristics (i.e., age, gender, dummies for countries of birth groups, as well as a dummy taking the value of one if a student is enrolled in an Economics degree). To account for any systematic differences between students' learning experiences in tutorials, we include tutorial fixed effects. Standard errors are clustered at tutorial level to further account for the possibility of common shocks at the class level.

Beyond Course Effects. While our main focus is to identify the effects of feedback provision on the related course, it is also interesting to see whether our intervention had any long-lasting performance effects beyond the intervention course. Specifically, we explore whether learning one's ranking in the online assignment affects academic performance (i) in the other courses taken during the intervention semester, as well as (ii) in the courses taken the following semester. To do so, we use two additional outcome variables, namely Semester 1 2016 GPA adjusted to exclude the intervention course grade and Semester 2 2016 GPA. Both measures are reported on a 0-7 scale.

On the one hand, students might end up putting more effort in the intervention course - in order to achieve and/or maintain a higher rank - and this, in turn, may come at the expense of other contemporaneous courses. Hence, feedback may indirectly affect students' performance in other courses taken during the intervention semester. On the other hand, being treated might also change students' behavior more generally, with long-lasting effects on academic outcomes. This might be due to a higher propensity towards social learning, or to healthier study habits developed perhaps because of the extra incentives provided via our leaderboards by the tighter effort-reward signal conveyed. It is also possible that feedback provides information about one's self-perceived ability, while also resolving some of the uncertainty about returns to effort (Goulas and Megalokonomou, 2016). With this in mind, it will be interesting to explore our impacts' heterogeneity by gender, given the recent evidence that feedback provision might affect males and females differently (Mayo et al., 2012; Goulas and Megalokonomou, 2016; Kugler et al., 2017).

The Mechanism. In an attempt to shed light on what drives students' response to feedback provision, we use additional data from the software developer's logs and the two course discussion boards. Both types of data provide us with a proxy for students' effort, either directly via a higher engagement with the online assignment, or indirectly via a broader engagement with with the course through social learning. For instance, the developer's logs provide data on the individual completion rate and time spent in the assignment, both overall and by set of exercises. To see whether feedback provision produces a greater direct engagement with the assignment, we will run a modified version of model (1) with either the completion rate or the total amount of time spent working on the assignment as outcome variables. Furthermore, we use the two discussion boards associated with the course to obtain a measure of each student's total number of posts, while also categorizing them as *relevant* and *irrelevant*.[21] These data allow us to examine whether students' reaction to feedback may be generated by a more active engagement with their peers in the course.

As a result, we employ the following regression model:

$$Posts_{i,d,c} = \alpha + \beta Treatment_{i,d,c} + \gamma X_{i,d,c} + TutorFE_c + u_{i,d,s} \tag{2}$$

where $Posts_{i,d,c}$ refers to the total number of posts that student i, enrolled in degree d, assigned to tutorial class c writes on the discussion boards. We also run separate modified versions of model (2) for specifications in which the outcome variable is either the number of relevant or the number of irrelevant posts, as well as whether one has ever posted on any of the two course discussion boards. In contrast to model (1), note that model (2) includes $TutorFE$ rather than $TutorialFE$. This is because the number of students per treatment who post on the discussion boards is roughly the same as the number of tutorial classes and so, using $TutorialFE$ is not advisable. Including $TutorFE$ is, however, more appropriate since the total number of tutors is roughly one third of the number of tutorial classes.

4. Results

4.1. *Baseline Estimates: Course Effects*

We start our results section by discussing estimates of the impact of feedback on students' performance in the online assignment. Results are reported in the first column in Table 3. The first row documents the effect of the Main feedback on students' final assignment ranking, as estimated via model (1), while the remaining rows report impact estimates for the other three feedback treatments. The estimates show the Main feedback effect to be positive and statistically significant. The corresponding coefficient indicates that, by the end of the semester, ceteris paribus, students in the Main treatment group outrank their peers in the control by 62 (out of 1,093) positions. A different picture emerges, however, when looking at the other feedback groups. Estimates are still positive, but show no significant differences between the control students and those presented with the alternative feedback versions. That said, the impact of Nonstop is descriptively rather similar to the effect of the Main treatment (52 positions), while the effects of Positive and Negative are negligible.

We now turn to the effect of feedback on students' exams performance. Results are presented in Table 3, in (i) columns (2)-(4) and (5)-(7) for the two mid-terms, respectively, (ii) columns (8)-(10) for the final exam, and (iii) columns (11)-(13) for the average exam grade. For each exam, we show: (i) the actual performance - i.e., the grade out of 10 and rounded

[21] Relevant posts relate to economic concepts, course content and materials; irrelevant posts refer, mostly, to course logistics or any other unrelated topic. The assignment of posts to these categories was done by two RAs who independently evaluated each post. When they disagreed on the classification of a post, which happened only a couple of times, a third RA classified that post. None of the RAs was aware of the research question.

to 2 decimal places, (ii) the standardized performance - i.e., the grade transformed into z-scores to facilitate interpretation, and (iii) the ordinal grade rank to allow comparison with the assignment rank.

The first row of Table 3 shows that students in the Main treatment group perform significantly better than control students across all course exams throughout the semester. Specifically, their grades are 3.11%, 4.45% and 2.99% higher than for control students in the first, second and final exam respectively, with an average performance rise of 3.44%. This translates into an improvement of 0.16, 0.21 and 0.18 SDs in Week 6, Week 10 and final exam respectively, for an average increase of 0.21 SDs. Finally, if we rank the grades of all students within each test, we note that the performance rise due to feedback provision translates into a rank increase almost identical to the one observed in the online assignment. Notably, this 62-63 exam positions boost appears constant across all three exams.

Again, a different picture emerges when we look at the alternative ways to provide feedback, as reported in the remaining rows. For all these groups, the estimated coefficients are positive but not statistically different from zero. That said, we find no significant difference between the Main treatment and Nonstop, while Negative, for instance, performs significantly worse than Main (see Table A.1). This implies that while we cannot definitely say that continuously presenting one with her rank information is worse than showing it only when it varies, giving only bad news is clearly detrimental compared to the latter.

The effect of the Main treatment (0.21 SDs) is remarkable. When comparing our findings with results from the education literature, we see, for instance, that our estimates are of comparable magnitude to being taught by a teacher between 1.5 and 2 SDs above the average (Hanushek et al., 2005; Chetty et al., 2014) or to reducing the class size by 20% (Angrist and Lavy, 1999; Krueger, 1999). While these are extremely costly interventions, manipulating the way feedback is disclosed online is virtually costless. Moreover, from a practical perspective, feedback disclosure appears increasingly feasible by the day, due to the fast technological advancements that are quickly becoming an integral part of the education and training sector.

Heterogeneity. The results presented above are quite substantial, but they might vary greatly across different categories of students. To investigate the presence of heterogeneous treatment effects, we split the whole sample along various observable dimensions and re-run our benchmark specification (1) separately for different sub-samples. Results are reported in the appendix, Tables A.2 - A.5. For simplicity, we report only standardized performance, in all course exams and on average.

First, let us analyze the effect of feedback by gender as reported in Table A.2. This dimension is particularly interesting as there exists a growing body of evidence showing that females are more sensitive to feedback interventions than males (Mayo et al., 2012; Goulas and Megalokonomou, 2016; Kugler et al., 2017). In our case, females in the Main treatment group indeed react to feedback provision more than males across all exams. Taken in isolation, female students in the Main treatment group perform, on average, better than females in the control group in all exams except for the first mid-term. This effect is quite substantial, amounting, on average, to a performance rise of 0.36 SDs. While still positive, the same effect drops to 0.10 SDs for males and is no longer statistically significant. Finally, the coefficients of interest for the alternative feedback groups (Nonstop, Positive, Negative) do not reveal any sizeable or significant pattern.

Second, note that another heterogeneous element of our results emerges when we focus on the effect of feedback by age. Table A.3 reports results separately for students below 19 and above (or equal to) 19 years of age.[22] A quick glance reveals a more pronounced feedback effect for older students, statistically significant at the 5% level in all exams except the first mid-term. In terms of magnitude, the coefficient appears almost three times larger for 19+ students than it is for those under 19, suggesting that the Main feedback may be more effective with a more mature sample. This is in line with Barankay (2012) who also reports a more sizable effect of feedback among older salesmen. It contrasts, however, with Blanes i Vidal and Nossol (2011), who find that feedback provision is equally important for workers with different levels of experience. No further feedback alternatives provide any other significant treatment effects.

Third, a considerable proportion of our sample comes from abroad (see Section 3.1). In Table A.4 we report treatment effects by international status, obtained by running regressions for each exam, separately for domestic and international students. The Main feedback impact is more prominent and more precisely estimated within the international sub-sample, albeit statistically significant (at the 10% level) only for the final exam (0.38 SDs). Compared to domestic students, this effect is almost four times larger for internationals. No additional effects are present for any of the other feedback treatments.

Finally, in Table A.5 we look separately at how students majoring in Economics respond to feedback compared to all other students. The effect of the Main feedback seems to be more pronounced, on average, among Economics students, although not statistically different from zero. Indeed, we find no statistically significant effects of feedback in relation to any treatment group.

Non-linearities. Our baseline specification (1) assumes a linear impact of information provision on performance. The effect, however, may very well vary across the grade distribution. It is plausible, for instance, that feedback affects low and high-achieving students

[22] The average student in our sample is 19.5 years old, hence our age split provides us with two rather comparable groups size-wise.

differently. To address this issue, we allow for non-linear effects by running quantile regression models. We estimate the effect of feedback at each decile $\theta \in [0,1]$ of the conditional distribution of grades as follows:

$$Y_{Quant_\theta} = \alpha_\theta + \beta_\theta Treatment_i + \gamma_\theta X_i + TutorialFE + \epsilon_{i,\theta} \qquad (3)$$

Figure 3 plots the $\hat{\beta}_\theta$ coefficients from these quantile regressions (marginal effects) at each decile θ, as well as the associated 95% confidence interval. The outcome variable is the standardized average exam grade.[23] We use bootstrapping (with 500 replications) to compute the standard errors and estimate model (3) for all four treatments.

The top left plot in Figure 3 shows the effect of the Main feedback to be (largely) linear and positive across most of the distribution. Its magnitude is larger between the 20^{th} and the 70^{th} percentile, remaining statistically significant at the 5% level. The impact declines at the highest two deciles, indicating a ceiling effect. As expected, none of the other treatments produces significant effects (see the other plots in Figure 3), except for Nonstop at the third decile. At a descriptive level, we note however that the effect of Nonstop and Negative is positive across the whole distribution. Vice-versa, the bottom left plot in Figure 3 suggests a differential effect of the Positive treatment: the effect is positive up to the 60^{th} percentile, becoming negative beyond that point. Although these results are not significant, such pattern seems to suggest that reporting only positive news may hurt the best students, perhaps because it may induce them to 'rest on their laurels'.

4.2. *Beyond Course Effects*

The administrative records allow us to track students' performance in other courses taken in Semester 1 2016, as well as in the subsequent semester. Table 4 reports evidence of the effect of feedback on these educational outcomes that go beyond the intervention course: column (2) refers to students' GPA in Semester 1 2016, adjusted to exclude the intervention course grade, while columns (3)-(5) focus on whether feedback has any long-lasting effects, by examining next semester GPA. The raw data for these two outcome variables range from 1 to 7. To facilitate comparison with our feedback effects on students' performance in the intervention course, we transform students' average course grades (across all three exams) into the same 1-7 scale[24] and then convert them into a z-score. Column (1) reports estimates of the feedback effect in the intervention course based on this transformation.

[23] Results are very similar when we use any of the three course exam grades in isolation.

[24] The 1-7 range is commonly used in Australia. To rescale the intervention course grade, we use the official university transformation rules and set: Grade=7 if the raw grade is above 8.50, Grade=6 if the raw grade is between 7.50 and 8.49, Grade=5 if the raw grade is between 6.50 and 7.49, Grade=4 if the raw grade is between 5.00 and 6.49, Grade=3 if the raw grade is between 4.50 and 5.49, Grade=2 if the raw grade is between 2.00 and 4.49, and Grade=1 if the raw grade is below 2.00.

The estimated effects for the Main treatment are reported in the first row, while the remaining rows present the effects for our alternative feedback treatments. As shown in column (1), the Main feedback clearly increases treated students' grades in the intervention course. This boost does not appear to come, however, at the expense of performance in other courses. Indeed, column (2) estimates indicate that a student's average performance in all other courses taken in the same semester is not affected by feedback provision in the intervention course; hence, being treated did not crowd out effort in other courses.

The picture is very different when we look at the effect of the Main treatment on students' performance next semester. As reported in column (3), students in the Main feedback group experience a GPA increase in Semester 2 2016 equal to 0.28 SDs. This is a remarkable, and perhaps surprising, effect, greater than the direct impact observed in the intervention course. That said, by performing this exercise, we are conflating two different channels: the direct long-term effect of feedback provision (e.g. via increased students' engagement) and its indirect impact through the intervention course. We are interested in the former. In order to isolate this direct effect, we also control for students' performance in all other courses taken in Semester 1 2016 excluding the intervention course. Although this may introduce a slight endogeneity problem, it is the only way to net out the indirect effect. Moreover, since feedback does not affect performance in other contemporaneous courses, the potential endogeneity does not appear overly problematic. As reported in column (4), the estimate drops to 0.18 SDs - the same magnitude as the (rescaled) effect on the intervention course. Additionally controlling for intervention course performance - see column (5) - leaves our estimates roughly unchanged (0.17 SDs). No further significant results are present for Nonstop, Positive or Negative.

These findings indicate that disclosing relative performance information can have a long-lasting positive impact on university students' academic performance. But is this behavior driven by a particular sub-sample? Also, are these spillovers general or do they come from other Economics courses? Table A.6 presents estimates from specifications similar to the one in column (3) of Table 4, ran on separate sub-samples split by gender and course type (Economics vs. non-Economics). We find that our long-term treatment effect is driven by male students, for whom we report a striking 0.38 SDs performance rise. We also note that this increase is not coming from subsequent Economics courses. That said, this is most likely because 82% of our sample are students enrolled in degrees other than Economics, for whom the option to take Economics courses is quite limited.

5. The Mechanism

Our findings clearly show that students exposed to the Main feedback outperform control students both in the online assignment and, more importantly, in all course exams across the semester. Furthermore, this effect appears to be independent of a student's position across the grade distribution and is, thus, consistent with a model in which competitive preferences induce everyone to exert more effort (see Azmat and Iriberri, 2010). While we are not the first document a positive impact of feedback provision on performance (e.g. Azmat and Iriberri, 2010; Tran and Zeckhauser, 2012; Katreniakova, 2014), up to our best knowledge, no other study has managed, so far, to provide direct evidence of the mechanism driving these results - an increase in the effort exerted. We are able to do so because of the RCT setup, which presents two advantages. First, the feedback provided relates to a continuous drill, and its continuity allows us to observe students' activity over the entire semester. Second, the technology adopted to perform this drill helps us keep track of such activity.

With this in mind, note that effort can manifest itself in various ways. In our context, the most obvious ones are perhaps related to how one engages with the assignment. But it can also take the form of greater engagement at a higher level, both with the course and with other fellow students. To proxy for the first type of effort we will use two different measures of assignment engagement, namely the proportion of assignment completed and the amount of time spent spent doing it. To proxy for the second kind of effort, we will analyze several social learning indicators captured via the number (and type) of posts written by students on the two course discussion boards.

Table 1 Panel B reports the relevant summary statistics on the assignment related outcomes. Recall that 20% of the overall grade depends on completing 100% of the assignment; an equivalent proportion is awarded for partial completion. As we can see, 86% of students finish the assignment, with an average student completing 95% of it and spending roughly 10 hours doing so. This is a substantial amount of time, totalling about one third of the overall face-to-face instruction provided during the entire semester. Table 5 reports the OLS estimates of the corresponding treatment effects.[25] We find that neither the completion rate nor the time spent working on the assignment are in any way robustly related with our treatments. Also, Positive treatment students do spend on average about 1.42 hours less on the assignment, but this is not consistently related to any of our previous findings. Thus overall, our treatments did not affect the way students engage with the assignment *per se*.

We now turn to the impact of our Main treatment (and other variations) on peer interactions, as captured by the posts students upload on the two course discussion boards. We consider not only the total number of posts, but also their split into relevant and irrelevant. (The former group is related to the course material and assessments, while the latter

[25] One concern might be that our assignment time variable may not be very precise, as the server collects information about the total time students were logged in the software platform. There is no reason, however, to assume that this potential measurement error is different across treatments.

includes logistics or fully unrelated issues.) Table 1 Panel B shows that students post on average 1.65 posts, with the relevant ones clearly representing the vast majority (mean=1.46) and the irrelevant ones appearing only sparsely (mean=0.19). Overall, 17% of students contribute to these forums (i.e., write at least one post in any of the course discussion boards).[26] We are interested in investigating whether any particular treatment group is more likely to post in the first place. The preliminary results in Table A.7 show that students who post are very equally split among treatment groups and also compare well to the initial randomized group proportions. Importantly, note that treated students who post - even in this reduced sample - outperform control students. Indeed, we find a mean standardized average grade of treated students of 0.09 (N=36), while for control this mean is 0.04 (N=35).

Table 6 presents our regression results for (i) the total number of posts, in columns (1)-(2), (ii) the number of relevant posts, in columns (3)-(4), (iii) the number of irrelevant posts, in columns (5)-(6), and (iv) whether a student ever posted on any board during the semester. Specifications (1), (3) and (5) employ the baseline model (2), while (2), (4), (6) and (7) include extra controls for students' performance in the other Semester 1 2016 courses. Panel A indicates that while there seems to be no treatment effect on whether one posts or not, the estimated impact of Main feedback (compared to no feedback) on number of posts is positive, sizeable, significantly different from zero, and robust across specifications. In other words, our intervention made active students even more active. In particular, conditional on posting, a student from the Main treatment group writes, on average, 0.79 posts more than a control student. As the mean number of posts that control students write is close to 1.50, this is equivalent to a treated student writing 53% more often than a control one. Importantly, this result is entirely driven by the relevant posts, with a student in the Main treatment group writing about 0.71 more relevant posts than a control student. In contrast, the effect of feedback on the number of irrelevant posts a treated student writes appears small and insignificant. This provides additional evidence supporting the social learning channel: students in the Main treatment react to feedback by getting more involved in the course and interacting more frequently with their peers. This seems to reflect their decision to exert more effort by engaging more actively in social learning, which in turn boosts their performance not only in the short-term but also over time.

One potential concern in this context is related to high-performing students perhaps being, on average, more 'vocal' on this type of forums. We are, however, not particularly worried about such selection as our treatment groups are balanced as far as prior ability is concerned. To allay any further concerns, we also control for a student's average performance in all other Semester 1 2016 courses excluding the intervention course. (Note that

[26] There are no extreme outliers, with only three students posting more than 6 posts. Results continue to hold after dropping these observations.

Section 4.2 shows that feedback does not affect students' performance in any other contemporaneous courses.) Doing so hardly affects our results, with the corresponding estimates changing only by 2-6% (see columns (2), (4) and (6) in Table 6).

One additional remark. The increased participation in public forums could be problematic for our identification strategy if the additional posts generate positive externalities for the general student population (i.e., by being more engaged, the treated students benefit those in the control group as well). In this case, our impacts would be underestimated - they would capture the effect of being more active on public forums net of the positive externality.

Finally, we also explore if there are significant differences between our Main treatment effects and Nonstop, Positive and Negative in terms of posting behavior, which is our proxy for social learning. Panel B in Table 6 shows that no such significant differences between Main and Nonstop are present, which is also the case for course performance when comparing the standardized average grades (see Table A.1). On the other hand, there are significant differences in posting behaviour between those receiving the Main feedback and those in the Positive or Negative group. In particular, students in the Positive (Negative) feedback group write 0.51 (0.67) posts less than students in the Main treatment group. Turning again to their grades, Table A.1 shows that students in the Main treatment group outperform those in the Negative one in the final and average course exam grade by 0.18 and 0.15 SDs, respectively.

Overall, compared to control students, only students in the Main treatment perform significantly better and are more socially engaged. One possible reason for this pattern is related to the effectiveness of feedback potentially relying on its ability to trigger attention, unwaveringly reminding a student of their current relative position with respect to her peers. Marketing experts, for instance, consider attention to be the key driver of advertising success: for consumers to be affected by an ad message, they first have to be paying attention. A similar thinking could apply in our context, but, while intriguing, this idea remains purely speculative. Alternatively, it could just be that we lack power or perhaps our continuous-type feedback is very sensitive not only to the type of information disclosed, but also to the way in which is disclosed. Finally, the treatment effects would also be underestimated if, by being more active on public forums, the treated students benefit those in the control group and, in turn, encourage them to post more. Further research is required to investigate these additional issues.

6. Robustness

To alleviate potential concerns about any confounding factors which might affect our findings, we first perform a *randomized treatments*-type of robustness test. Specifically, we generate placebo treatments that do not reflect students' real treatment assignment and examine whether these treatments can produce a similar pattern as found in our main results. To do so, we randomly re-assign[27] all students into five groups and re-run our main model using the placebo treatments as variables of interest. If the placebo treatments are found to be significant determinants of performance, this would indicate that students might react to confounding factors (not perfectly coinciding with the real treatments) and get a performance boost.[28] If there is a correlation between these confounding factors and the assignment to actual treatment groups, then the placebo treatments would be picking up some of these effects. The results in Table 7 show no effect of these placebo treatments on performance. Hence, we conclude that our findings are unlikely to be driven by simultaneous effects other than our original treatments (as generated by the software developer).

We also wish to investigate whether our results may be, at least partially, driven by differential drop out rates across various treatments. Indeed, one may be concerned that feedback, instead of stimulating students to do better, may be discouraging the low-achieving ones and causing them to drop out at a higher rate than control students. We propose that students are unlikely to drop out from the course in response to the treatment group they are assigned to. To support this claim, first note that students are eligible to drop out only during the first two weeks of the semester and, therefore, their limited exposure to the assignment in this time makes it unlikely that they quit because of the feedback treatment they are assigned to. Furthermore, we look at the actual number of students who drop out from the course by treatment group. Column (2) in Table A.8 shows a similar proportion of drop out students in all treatments (13-14% for all groups).

Another valid concern is related to whether there is a a significant difference in the proportion of students who complete 100% of the online assignment. Indeed, students who complete the whole assignment are more exposed to the treatment than those who only partially complete it. Column (4) in Table A.8 reports the number of students with 100% completion rate. We note that percentages are very similar across treatments (around 82-89%).

To account for prior ability, we also run a specification that includes an imputed measure of previous academic performance. As mentioned, previous ability (ATAR) scores are available only for domestic students, hence controlling directly for this variable would considerably reduce our sample. The missing prior ability values were thus imputed using the coefficients of (i) Semester 1 2016 average course grade (adjusted to exclude the intervention course grade),[29] and (ii) our full set of controls (i.e., age, gender, countries of birth groups,

[27] We create a random variable for all observations in our data and sort observations based on it. Using it, we then assign students to random treatment groups.

[28] For instance, students assigned to different treatments may create effective study groups and benefit from out-of-class interactions.

[29] Note that our treatments are orthogonal to Adjusted GPA Semester 1 2016 (see Table 4).

and whether enrolled in an Economics degree) from an OLS regression on the existing ATAR data (Dobrescu, 2015).[30] Table A.9 estimates show that accounting for prior achievements leaves our treatment effects unchanged (i.e., robustly significant only for the Main group.)

Finally, we also re-run our assignment outcomes analysis with tutor fixed effects to ensure that the lack of results related to this type of effort is not an artefact of the fixed effects nature. Doing so does not change our assignment results, confirming that only our social learning mechanism is at play.

7. Concluding Remarks

Building a unified body of knowledge around the effectiveness of rank feedback is a difficult task because the behavioral response to rank feedback *per se* (rank incentives) is potentially compounded with many confounding factors. This paper is among the first to cleanly identify the impact of rank incentives and to shed light on its underlying mechanism.

In a higher education setting, we find that providing rank feedback had a sizeable positive impact on student performance (0.21 SDs higher exam grades), did not crowd out effort in the other contemporaneous courses and improved academic performance next semester.

The reason why our intervention was successful, compared to similarly well identified rank incentives studies (Barankay, 2011, 2012) or the other prominent randomized control trials in higher education (e.g. Azmat et al., 2019), might lie in how salient, visible, and immediately actionable our feedback was. In our RCT, students were (randomly) assigned to four treatments that privately presented them with information on their real-time rank as achieved in a semester-long online assignment. Treatments varied in how often the rank feedback was displayed and in what type of information was presented: some students were shown their rank every time it changed either upwards or downwards (Main treatment), others were uninterruptedly exposed to their relative performance position (Nonstop), or they could only see it when their rank position improved (Positive) or worsened (Negative).

A number of instructional design issues emerge as potential avenues for further research. The timing of rank feedback provision seems to be crucial. Indeed, our results suggest that providing it nonstop or not often enough might render the rank information not salient. This raises questions on the optimal frequency of information provision, with an eye on the balance of interest vs. information overload.

How people receive the information also appears to matter greatly. In our case, feedback was provided in real-time and continuously during the semester. Since real-time feedback directly resolves some of the uncertainty about returns to effort (and it does so in a fairly granular, decision-by-decision, manner), restricting it to a specific period would have likely limited its impact. With current technologies making the provision of real-time feedback

[30] Imputations affected 97.12% of the international sample, with the specification attaining an R^2 of 22.72%.

over long periods virtually costless, it will be interesting to see if this type of approach can reduce the demoralization effect (due to a lower-than-expected rank) also in other contexts.

The granularity of the information might have also played a key role. In our RCT, students learnt their exact rank rather than whether they belonged in a specific band. Providing the information partitioned differently would have made the rank changes more difficult to spot and would have likely triggered a different effort response. This direction of research warrants further study and may unveil some potentially interesting lessons.

Finally, an important avenue for future research has to do with the mechanism behind the impact of rank feedback. In our case, results seem to be driven by social learning, i.e., the extent to which one engages with their peers by posting on the two course discussion boards. Main treatment students (and only they) do so 50% more often than control students - a considerable, robust and statistically significant effect. More research is required, however, to shed further light on how the education production function is affected by rank feedback.

Our findings have considerable policy implications. Improving students' attainments is a priority for all policy-makers and practitioners, who tend to focus on a variety of inputs such as (i) reducing class size (Krueger, 1999; Bedard and Kuhn, 2008), (ii) improving quality of teachers (Glewwe et al., 2010; Chetty et al., 2014; Duflo et al., 2015) and schools (Lavy, 2002),[31] (iii) extending term length (Pischke, 2007; McMullen and Rouse, 2012), (iv) improving peer group quality (Zimmerman, 2003; Duflo et al., 2011), (v) providing financial and non-financial incentives (Benhassine et al., 2015; Levitt et al., 2016), or (vi) employing more student-level differentiation (Banerjee et al., 2016), using frequent data to tailor classroom instruction and instilling a culture of high expectations (Abdulkadiroglu et al., 2011; Fryer, 2014). All such interventions are, however, very costly and their effectiveness is uncertain. Technology, on the other hand, is increasingly seen as the leading cost-effective avenue to boost instruction productivity (Gates, 2016; Mead, 2016). There are several channels through which this might occur, from shortening feedback time to creating environments that trigger people's engagement. We shed light on these issues, and show that providing feedback in such contexts is feasible, beneficial and virtually costless.

[31] See also Rockoff (2004); Rivkin et al. (2005); Aaronson et al. (2007); Kane and Staiger (2008).

References

Aaronson, D., L. Barrow, and W. Sander (2007). Teachers and student achievement in the chicago public high schools. *Journal of Labor Economics 25*(1), 95–135.

Abdulkadiroglu, A., J. Angrist, S. Dynarski, T. Kane, and P. Pathak (2011). Accountability in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics 126*(2), 699–748.

Angrist, J. and V. Lavy (1999). Using maimonides' rule to estimate the estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* (114), 533–575.

Azmat, G., M. Bagues, A. Cabrales, and N. Iriberri (2019). What you don't know... can't hurt you: A field experiment on relative performance feedback in higher education. *Management Science*. forthcoming.

Azmat, G. and N. Iriberri (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics 94*(5), 797–811.

Azmat, G. and N. Iriberri (2016). The provision of relative performance feedback: An analysis of performance and satisfaction. *Journal of Economics and Management Strategy 25*(1), 77–110.

Bandiera, O., I. Barankay, and I. Rasul (2013). Team incentives: Evidence from a firm level experiment. *Journal of the European Economic Association 11*(5), 1079–1114.

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shotland, and M. Walton (2016). Mainstreaming an effective intervention: Evidence from randomized evaluations of "teaching at the right level" in india. *NBER Working Paper No.* (22746).

Barankay, I. (2011). Rankings and social tournaments: Evidence from a crowd-sourcing experiment. *mimeo*, University of Pennsylvania.

Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment. *mimeo*, University of Pennsylvania.

Bedard, K. and P. Kuhn (2008). Where class size really matters: Class size and student ratings of instructor effectiveness. *Economics of Education Review 27*(3), 253–265.

Benhassine, N., F. Devoto, E. Duflo, P. Dupas, and V. Pouliquen (2015). Turning a shove into a nudge: A "labeled cash transfer" for education. *American Economic Journal: Economic Policy 3*(7), 86–125.

Blader, S., C. M. Gartenberg, and A. Prat (2016). The contingent effect of management practices.

Blanes i Vidal, J. and M. Nossol (2011). Tournaments without prizes: Evidence from personnel records. *Management Science 57*(10), 1721–1736.

Brade, R., O. Himmler, and R. Jackle (2018). Normatively framed relative performance feedback - field experiment and replication. MPRA Discussion Paper No. 88830.

Bursztyn, L. and R. Jensen (2015). How does peer pressure affect educational investments? *Quarterly Journal of Economics 130*(3), 1329–1367.

Charness, G., D. Masclet, and M. C. Villeval (2013). The dark side of competition for status. *Management Science 60*(1), 38–55.

Chetty, R., J. Friedman, and J. Rockoff (2014). Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American Economic Review 104*(9), 2633–2679.

Dobrescu, L. (2015). To love or to pay: Savings and health care in older age. *Journal of Human Resources 50*(1), 254–299.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review 101*(5), 1739–1774.

Duflo, E., P. Dupas, and M. Kremer (2015). School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools. *Journal of Public Economics 123*, 92–110.

Eriksson, T., A. Poulsen, and M. C. Villeval (2009). Feedback and incentives: Experimental evidence. *Labour Economics 16*, 679–688.

Fryer, R. (2014). Injecting charter school best practices into traditional public schools: Evidence from field experiments. *Quarterly Journal of Economics 129*(3), 1355–1407.

Gates (2016). Finding what works: Results from the leap innovations pilot network. Technical report, Seattle, WA: Bill and Melinda Gates Foundation.

Gerhards, L. and N. Siemer (2014). Private versus public feedback - the incentive effects of symbolic awards. *Mimeo 2014*(1), University of Aarhus.

Gerhards, L. and N. Siemer (2016). The impact of private and public feedback on worker performance: Evidence from the lab. *Economic Inquiry 54*(2), 1188–1201.

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

Gill, D., Z. Kissova, J. Lee, and V. Prowse (2018). First-place loving and last-place loathing: How rank in the distribution of performance affects effort provision. *Management Science*. forthcoming.

Glewwe, P., N. Ilias, and M. Kremer (2010). Teacher incentives. *American Economic Journal: Applied Economics 2*(3), 205–227.

Goulas, S. and R. Megalokonomou (2016). Knowing who you actually are: The effect of feedback on short and long term outcomes. University of Warwick - Warwick Economics Research Papers Series No. 1075.

Hannan, R. L., R. Krishnan, and A. H. Newman (2008). The effects of disseminating relative performance feedback in tournament versus individual performance compensation plans.

The Accounting Review 83(4), 893–913.

Hanushek, E., F. Kain, and G. Rivkin (2005). Teachers, schools and academic achievement. *Econometrica 2*(73), 417–458.

Jessoe, K. and D. Rapson (2014). Knowledge is (less) power: Experimental evidence from residential energy use. *American Economic Review 104*(4), 1417–1438.

Kane, T. and D. Staiger (2008). Estimating teacher impacts on student achievement: An experimental validation. NBER Working Paper 14607.

Katreniakova, D. (2014). Information, aspirations and incentive to learn: A randomized field experiment in uganda. *Mimeo*, CERGE–EI.

Khunen, C. and A. Tymula (2012). Feedback, self-esteem, and performance in organizations. *Management Science 58*(1), 94–113.

Khunen, C. N. and T. Anieszka (2012). Feedback, self-esteem, and performance in organizations. *Management Science 58*(1), 94–113.

Krueger, A. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics 114*(2), 497–532.

Kugler, A., C. Tinsley, and O. Ukhaneva (2017). Choice of majors: Are women really different from men? *NBER Working Paper No.* (23735).

Lavy, V. (2002). Evaluating the effect of teachers' group performance incentives on pupil achievement. *Journal of Political Economy 110*(6), 1286–1317.

Levitt, S., J. List, S. Neckermann, and S. Sadoff (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy 8*(4), 183–219.

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–45.

Mayo, M., M. Kakarika, J. Pastor, and S. Brutus (2012). Aligning or inflating your leadership aelf-image? a longitudinal study of responses to peer feedback in mba teams. *Academy of Management Learning & Education 11*(4), 631–652.

McMullen, S. and K. Rouse (2012). The impact of year-round schooling on academic achievement: Evidence from mandatory school calendar conversions. *American Economic Journal: Economic Policy 4*(4), 230–252.

Mead, R. (2016, March 8). Learn different: Silicon valley disrupts education. *The New Yorker*.

Pischke, J. (2007). The impact of length of the school year on student performance and earnings: Evidence from the German short school years. *Economic Journal 117*(523), 1216–1242.

Rivkin, S., E. Hanushek, and J. Kain (2005). Teachers, schools and academic achievement. *Econometrica 73*(2), 417–458.

Rockoff, J. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review 94*(2), 247–252.

Tiefenbeck, V., L. Goette, K. Degen, V. Tasic, E. Fleisch, R. Lalive, and T. Staake (2016). Overcoming salience bias: How real-time feedback fosters resource conservation. *Management Science 64*(3), 1458–1476.

Tran, A. and R. Zeckhauser (2012). Rank as an inherent incentive: Evidence from a field experiment. *Journal of Public Economics 96*(9), 645–650.

Wedel, M. and R. Pieters (2012). *Visual marketing: From attention to action*. Psychology Press.

Wolitzky, A. (2018). Learning from others' outcomes. *American Economic Review 108*(10), 2763–2801.

Zimmerman, D. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*(1), 9–23.

TABLE 1. DESCRIPTIVE STATISTICS

Variable	Mean	SD	Min.	Max.	N
Panel A.1: Student Level Characteristics					
Age	19.459	2.621	16	47	1101
Male	0.565	0.496	0	1	1101
Undertaking Economics Degree	0.183	0.386	0	1	1101
International Status	0.221	0.415	1	1	1101
COB: Australia	0.777	0.416	0	1	1101
COB: Other Oceania	0.003	0.052	0	1	1101
COB: Europe	0.011	0.104	0	1	1101
COB: Asia	0.203	0.403	1	1	1101
COB: America	0.002	0.043	0	1	1101
COB: Africa and Middle East	0.004	0.060	0	1	1101
ATAR score	90.146	7.555	49	99	861
GPA Previous Semester: International	4.489	1.104	1.667	6	84
Panel A.2: Tutorial Level Characteristics					
Male Tutor	0.478	0.505	0	1	46
Tutor International Status	0.348	0.482	0	1	46
Australian Tutor	0.652	0.482	0	1	46
European Tutor	0.065	0.250	0	1	46
Asian Tutor	0.283	0.455	0	1	46
Panel B: Performance and Effort Indicators					
Week 6 Exam	7.238	1.898	0	10	1,101
Week 10 Exam	6.476	2.142	0	10	1,099
Final Exam	6.203	1.681	0	10	1,101
Adjusted GPA Semester 1 2016	4.867	1.119	1	7	1,080
GPA Semester 2 2016	4.678	1.333	0	7	1,028
Assignment Completion Rate (%)	94.507	17.938	0	100	1,095
Time Spent on Assignment (hours)	10.161	6.322	0	61.351	1,093
Total Number of Posts	1.652	1.241	1	8	184
Number of Relevant Posts	1.467	1.267	0	8	184
Number of Irrelevant Posts	0.185	0.477	0	2	184
Posting (Y/N)	0.167	0.373	0	1	1,101

Notes: The classification of the country of birth (*COB*) follows the Standard Australian Classification of Countries, 2011. The Oceania group includes Oceania countries other than Australia. *ATAR* (Australian Tertiary Admission Rank) score denotes a student's ranking relative to his/her peers when completing secondary education. *GPA Previous Semester: International* is Semester 2 2015 GPA for international students enrolled at the university before the intervention semester (Semester 1 2016). *Assignment Completion Rate (%)* captures how much progress (as a percentage of all the assignment) a student has done in terms of completing all corresponding exercises. *Time Spent on Assignment (hours)* denotes the total number of hours a student spends attempting the assignment during the entire semester. *Total Number of Posts* refers to the posts a student contributes on the two course discussion boards, while *Number of Relevant (Irrelevant) Posts* refers to those posts that are related (unrelated) to the course content, i.e., posts discussing (not discussing) Economics topics. *Posting* is an indicator variable denoting whether a student has ever posted on any of the two discussion boards during the intervention semester.

TABLE 2. BALANCE TESTS FOR TREATMENT AND CONTROL GROUPS IN SEMESTER 1 2016

	Treatment Group		Control Group		Difference	
Variable	Mean	SD	Mean	SD	Diff.	SE
Panel A: Main Feedback and Control Group						
Age	19.384	(2.220)	19.602	(2.787)	0.218	(0.239)
Male	0.558	(0.498)	0.557	(0.498)	-0.002	(0.047)
Undertaking Economics degree	0.170	(0.376)	0.195	(0.397)	0.025	(0.037)
International Status	0.277	(0.448)	0.217	(0.413)	-0.060	(0.041)
COB: Australia	0.723	(0.448)	0.783	(0.413)	0.060	(0.041)
COB: Other Oceania	-	-	-	-	-	-
COB: Europe	0.018	(0.133)	0.018	(0.134)	0.0002	(0.013)
COB: Asia	0.246	(0.431)	0.195	(0.397)	-0.051	(0.039)
COB: America	0.005	(0.067)	0.000	(0.000)	-0.005	(0.005)
COB: Africa and Middle East	0.009	(0.094)	0.005	(0.067)	-0.004	(0.008)
ATAR score	90.798	(7.365)	89.460	(8.304)	-1.338	(0.857)
GPA Previous Semester: International	4.238	(1.361)	4.222	(1.161)	-0.016	(0.409)
Male Tutor	0.496	(0.501)	0.502	(0.501)	0.007	(0.048)
Tutor International Status	0.321	(0.468)	0.358	(0.480)	0.036	(0.045)
Australian Tutor	0.679	(0.468)	0.643	(0.480)	-0.036	(0.045)
European Tutor	0.067	(0.251)	0.086	(0.281)	0.019	(0.025)
Asian Tutor	0.255	(0.437)	0.272	(0.446)	0.017	(0.042)
Panel B: Nonstop Feedback and Control Group						
Age	19.356	(2.375)	19.602	(2.787)	0.246	(0.251)
Male	0.572	(0.496)	0.557	(0.498)	-0.016	(0.048)
Undertaking Economics degree	0.183	(0.387)	0.195	(0.397)	0.012	(0.038)
International Status	0.173	(0.379)	0.217	(0.413)	0.044	(0.038)
COB: Australia	0.817	(0.387)	0.783	(0.413)	-0.035	(0.039)
COB: Other Oceania	0.005	(0.069)	0.000	(0.000)	-0.005	(0.005)
COB: Europe	0.005	(0.069)	0.018	(0.134)	0.013	(0.010)
COB: Asia	0.168	(0.375)	0.195	(0.397)	0.026	(0.037)
COB: America	-	-	-	-	-	-
COB: Africa and Middle East	0.005	(0.069)	0.005	(0.067)	-0.0003	(0.007)
ATAR score	90.567	(7.680)	89.460	(8.304)	-1.108	(0.862)
GPA Previous Semester: International	4.608	(0.946)	4.222	(1.161)	-0.386	(0.422)
Male Tutor	0.452	(0.499)	0.502	(0.501)	0.050	(0.048)
Tutor International Status	0.385	(0.488)	0.358	(0.480)	-0.027	(0.047)
Australian Tutor	0.615	(0.488)	0.643	(0.480)	0.027	(0.047)
European Tutor	0.067	(0.251)	0.086	(0.281)	0.019	(0.026)
Asian Tutor	0.317	(0.467)	0.272	(0.446)	-0.046	(0.044)

Notes: Each panel reports differences in pre-determined characteristics for students in the Main (Panel A) and Nonstop treatment group (Panel B) vs. the control group, respectively. The last two columns report the difference in means and the corresponding standard error of the difference, respectively.

Variable	Treatment Group Mean	Treatment Group SD	Control Group Mean	Control Group SD	Difference Diff.	Difference SE
Panel C: Positive Feedback and Control Group						
Age	19.367	(2.244)	19.602	(2.787)	0.234	(0.243)
Male	0.512	(0.501)	0.557	(0.498)	0.045	(0.048)
Undertaking Economics degree	0.172	(0.378)	0.195	(0.397)	0.023	(0.037)
International Status	0.223	(0.417)	0.217	(0.413)	-0.006	(0.040)
COB: Australia	0.777	(0.417)	0.783	(0.413)	0.006	(0.040)
COB: Other Oceania	0.005	(0.068)	0.000	(0.00)	-0.005	(0.005)
COB: Europe	0.005	(0.068)	0.018	(0.134)	0.013	(0.010)
COB: Asia	0.214	(0.411)	0.195	(0.397)	-0.019	(0.039)
COB: America	-	-	-	-	-	-
COB: Africa and Middle East	0.000	(0.000)	0.005	(0.067)	0.005	(0.005)
ATAR score	89.661	(7.147)	89.460	(8.304)	-0.201	(0.839)
GPA Previous Semester: International	4.736	(0.860)	4.222	(1.161)	-0.514	(0.332)
Male Tutor	0.447	(0.498)	0.502	(0.501)	0.056	(0.048)
Tutor International Status	0.354	(0.479)	0.358	(0.480)	0.004	(0.046)
Australian Tutor	0.647	(0.479)	0.643	(0.480)	-0.004	(0.046)
European Tutor	0.056	(0.230)	0.086	(0.281)	0.030	(0.025)
Asian Tutor	0.298	(0.458)	0.272	(0.446)	-0.026	(0.043)
Panel D: Negative Feedback and Control Group						
Age	19.571	(3.278)	19.602	(2.787)	0.031	(0.286)
Male	0.622	(0.486)	0.557	(0.498)	-0.066	(0.046)
Undertaking Economics degree	0.193	(0.396)	0.195	(0.397)	0.001	(0.037)
International Status	0.210	(0.408)	0.217	(0.413)	0.007	(0.039)
COB: Australia	0.790	(0.408)	0.783	(0.413)	-0.007	(0.039)
COB: Other Oceania	0.004	(0.066)	0.000	(0.000)	-0.004	(0.004)
COB: Europe	0.009	(0.092)	0.018	(0.134)	0.010	(0.011)
COB: Asia	0.193	(0.396)	0.195	(0.397)	0.001	(0.037)
COB: America	0.004	(0.066)	0.000	(0.000)	-0.004	(0.004)
COB: Africa and Middle East	0.000	(0.000)	0.005	(0.067)	0.005	(0.004)
ATAR score	90.270	(7.220)	89.460	(8.304)	-0.811	(0.820)
GPA Previous Semester: International	4.722	(1.064)	4.222	(1.161)	-0.500	(0.359)
Male Tutor	0.476	(0.501)	0.502	(0.501)	0.026	(0.047)
Tutor International Status	0.348	(0.477)	0.358	(0.480)	0.010	(0.045)
Australian Tutor	0.652	(0.477)	0.643	(0.480)	-0.010	(0.045)
European Tutor	0.064	(0.246)	0.086	(0.281)	0.022	(0.025)
Asian Tutor	0.283	(0.452)	0.272	(0.446)	-0.012	(0.042)

Notes: Each panel reports differences in pre-determined characteristics for students in the Positive (Panel C) and Negative treatment group (Panel D) vs. the control group, respectively. The last two columns report the difference in means and the corresponding standard error of the difference, respectively.

Table 3. Treatment Effects of Feedback on Academic Performance

	Assignment	Week 6 Exam			Week 10 Exam			Final Exam			Average Exam		
	Rank	Perf.	Std.	Rank	Perf.	Std.	Rank	Perf.	Std.	Rank	Perf.	Std.	Rank
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Main Feedback	61.715	0.311	0.164	61.948	0.445	0.208	62.902	0.299	0.178	63.582	0.344	0.210	63.233
	(31.858)*	(0.185)*	(0.098)*	(31.272)*	(0.214)**	(0.100)**	(33.063)*	(0.154)*	(0.092)*	(28.168)**	(0.159)**	(0.097)**	(25.825)**
Observations	442	445	445	445	444	444	444	445	445	445	444	444	445
Nonstop Feedback	51.716	0.118	0.062	25.965	0.271	0.127	40.244	0.234	0.139	50.494	0.188	0.115	40.283
	(33.790)	(0.232)	(0.122)	(38.492)	(0.195)	(0.091)	(28.699)	(0.178)	(0.106)	(34.501)	(0.171)	(0.104)	(29.659)
Observations	426	429	429	429	428	428	428	429	429	429	428	428	429
Positive Feedback	6.762	0.006	0.003	16.851	0.135	0.063	21.328	0.087	0.052	18.986	0.084	0.051	18.581
	(30.725)	(0.190)	(0.100)	(35.851)	(0.209)	(0.097)	(32.480)	(0.178)	(0.106)	(36.480)	(0.175)	(0.107)	(32.079)
Observations	434	436	436	436	434	434	434	436	436	436	434	434	436
Negative Feedback	2.676	0.111	0.058	23.872	0.222	0.104	35.882	-0.035	-0.021	10.193	0.090	0.055	23.887
	(36.429)	(0.199)	(0.105)	(34.900)	(0.201)	(0.094)	(29.321)	(0.171)	(0.102)	(32.453)	(0.166)	(0.101)	(28.824)
Observations	451	454	454	454	453	453	453	454	454	454	453	453	454
Tutorial FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in column (1) is the ordinal rank a student achieves in the online assessment (out of 1,093). The dependent variable in columns (2)-(4) is the exam grade (out of 10), standardized exam grade and the grade rank in the first mid-term, while columns (5)-(7) and (8)-(10) refer to the same outcome formulations for the second mid-term and the final exam. Columns (11)-(13) average these series over all three course exams. In all specifications we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered at tutorial level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table 4. Treatment Effects of Feedback Beyond the Intervention Course

	Average Exam (Std.) Intervention Course	Adjusted GPA (Std.) Semester 1 2016	GPA (Std.) Semester 2 2016		
	(1)	(2)	(3)	(4)	(5)
Main Feedback	0.184	0.093	0.285	0.184	0.167
	(0.098)*	(0.095)	(0.114)**	(0.084)**	(0.083)*
Observations	445	437	416	414	414
Nonstop Feedback	0.074	0.035	0.115	0.060	0.053
	(0.106)	(0.113)	(0.129)	(0.085)	(0.088)
Observations	429	426	399	398	398
Positive Feedback	0.033	-0.035	0.089	0.108	0.100
	(0.114)	(0.122)	(0.113)	(0.079)	(0.078)
Observations	436	432	412	410	410
Negative Feedback	0.022	0.029	-0.099	-0.090	-0.093
	(0.104)	(0.102)	(0.120)	(0.095)	(0.097)
Observations	454	445	422	418	418
Tutorial FE	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓
Adjusted GPA Semester 1 2016	x	x	x	✓	✓
Performance in Intervention Course	x	x	x	x	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variables in column (1), (2) and (3)-(5) are the standardized average exam grade achieved in the intervention course (across Week 6, Week 10 and final exam), the standardized GPA of the intervention semester (Semester 1 2016) adjusted to exclude the intervention course, and the standardized GPA next semester (Semester 2 2016), respectively. Adjusted current GPA and next semester GPA are measured on a 0-7 scale. To facilitate comparison, the standardized grade in the intervention course is also transformed to a 0-7 scale, before it is converted into a z-score. In all specifications, we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered at tutorial level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

TABLE 5. TREATMENT EFFECTS OF FEEDBACK ON ASSIGNMENT OUTCOMES

	Assignment Completion Rate (1)	Time Spent on Assignment (2)
Main Feedback	0.919	-0.061
	(1.604)	(0.728)
Observations	443	442
Nonstop Feedback	0.079	-0.108
	(2.226)	(0.585)
Observations	427	426
Positive Feedback	0.570	-1.419
	(1.670)	(0.594)**
Observations	435	434
Negative Feedback	-1.861	-0.266
	(1.802)	(0.707)
Observations	453	451
Tutorial FE	✓	✓
Student Characteristics	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in column (1) is the assignment's completion rate. The dependent variable in column (2) is the total number of hours a student spends completing assignment exercises over the course of the semester. In both specifications, we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered at tutorial level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and, 1% level, respectively.

TABLE 6. TREATMENT EFFECTS OF FEEDBACK ON COURSE ENGAGEMENT

	Total Posts		Relevant Posts		Irrelevant Posts		Posting
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Compared to Control Group							
Main Feedback	0.777	0.788	0.699	0.713	0.077	0.076	-0.009
	(0.325)**	(0.330)**	(0.348)*	(0.351)**	(0.108)	(0.110)	(0.033)
Observations	71	70	71	70	71	70	437
Nonstop Feedback	0.184	0.286	0.081	0.214	0.102	0.072	0.044
	(0.342)	(0.344)	(0.337)	(0.326)	(0.090)	(0.083)	(0.035)
Observations	77	76	77	76	77	76	426
Positive Feedback	-0.068	-0.061	-0.212	-0.223	0.143	0.162	-0.007
	(0.337)	(0.347)	(0.410)	(0.420)	(0.201)	(0.201)	(0.041)
Observations	69	69	69	69	69	69	432
Negative Feedback	-0.024	0.014	-0.082	-0.048	0.058	0.062	0.006
	(0.306)	(0.288)	(0.308)	(0.288)	(0.105)	(0.130)	(0.035)
Observations	72	71	72	71	72	71	445
Panel A: Compared to Main Feedback							
Nostop Feedback	-0.396	-0.369	-0.453	-0.403	0.057	0.034	0.036
	(0.515)	(0.528)	(0.489)	(0.495)	(0.095)	(0.091)	(0.029)
Observations	78	78	78	78	78	76	423
Positive Feedback	-0.508	-0.532	-0.460	-0.494	-0.048	-0.039	-0.009
	(0.318)	(0.298)*	(0.329)	(0.295)	(0.162)	(0.160)	(0.038)
Observations	70	69	70	69	70	69	429
Negative Feedback	-0.673	-0.693	-0.757	-0.800	0.085	0.107	0.002
	(0.340)*	(0.340)**	(0.372)**	(0.375)**	(0.129)	(0.129)	(0.034)
Observations	73	71	73	71	73	71	442
Tutor FE	✓	✓	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓	✓	✓
Adjusted GPA Semester 1 2016	x	✓	x	✓	x	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in columns (1)-(2) is the total number of posts that students contribute to the two course discussion boards. Columns (3)-(4) and (5)-(6) show specifications with the number of relevant and irrelevant posts, respectively as dependent variable. The dependent variable in column (7) is a dummy taking the value one if a student has ever written on any of the discussion boards, and zero otherwise. In all specifications, we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutor fixed effects. Standard errors are clustered at tutorial level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and, 1% level, respectively.

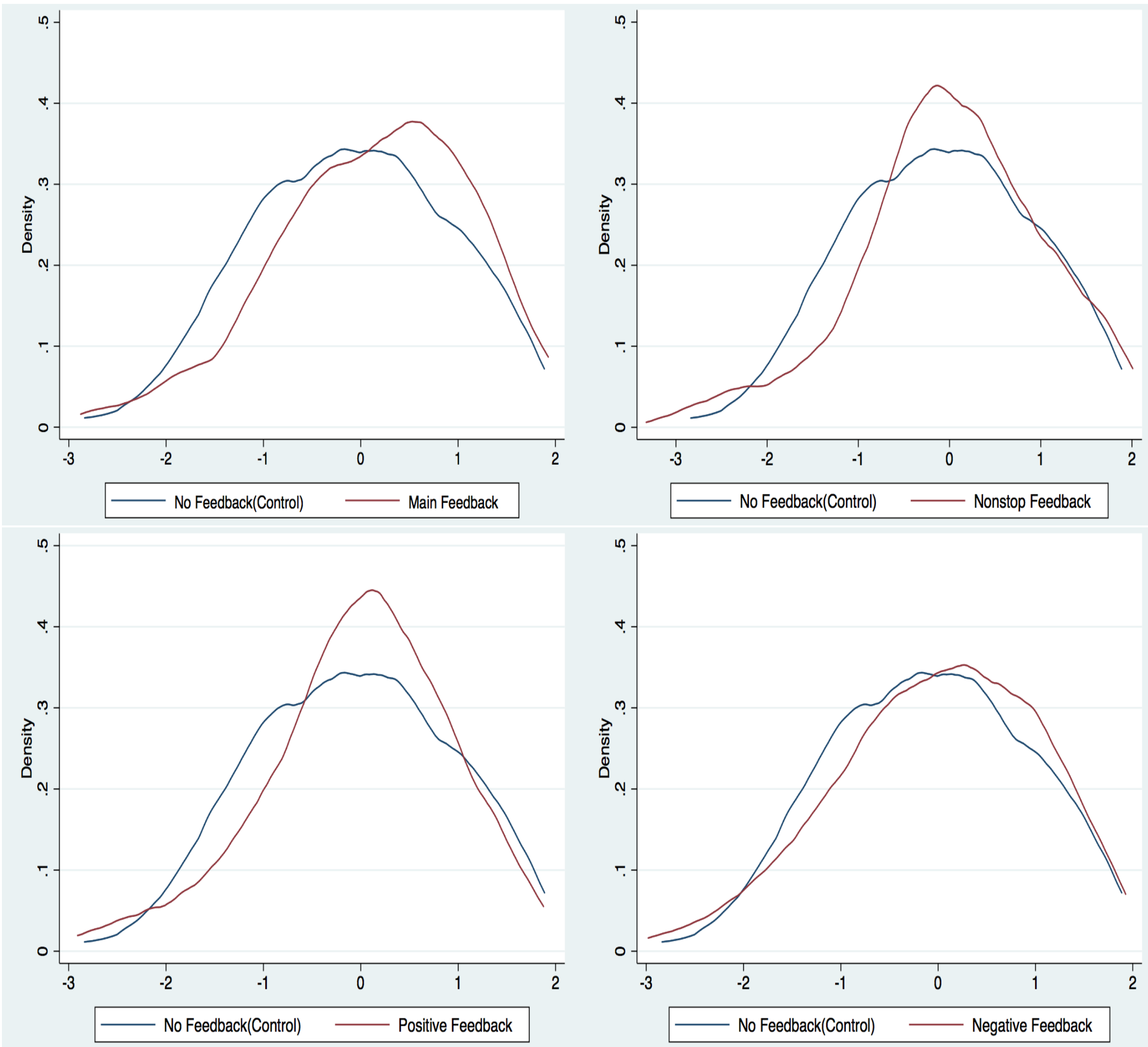
TABLE 7. PLACEBO TREATMENT EFFECTS OF FEEDBACK ON ACADEMIC PERFORMANCE

	Week 6 Exam (1)	Week 10 Exam (2)	Final Exam (3)	Average Exam (4)
Main Feedback	-0.088	0.054	-0.076	-0.030
	(0.106)	(0.097)	(0.118)	(0.109)
Observations	440	439	440	439
Nonstop Feedback	-0.134	-0.072	-0.157	-0.127
	(0.087)	(0.098)	(0.099)	(0.095)
Observations	441	440	441	440
Positive Feedback	0.015	0.155	0.006	0.076
	(0.089)	(0.095)	(0.103)	(0.094)
Observations	440	440	440	440
Negative Feedback	-0.008	-0.030	-0.073	-0.041
	(0.096)	(0.106)	(0.091)	(0.096)
Observations	440	440	440	440
Tutorial FE	✓	✓	✓	✓
Students' Characteristics	✓	✓	✓	✓

Note: Each row presents estimates from separate OLS regressions. Data is generated as follows: First, we create a random variable, next we use it to sort the dataset and then we assign observations to placebo treatment groups. The dependent variables in columns (1)-(4) are the standardized exam grades as achieved in the first and second mid-term, the final exam and on average, respectively. In all specifications, we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered at tutorial level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and, 1% level, respectively.

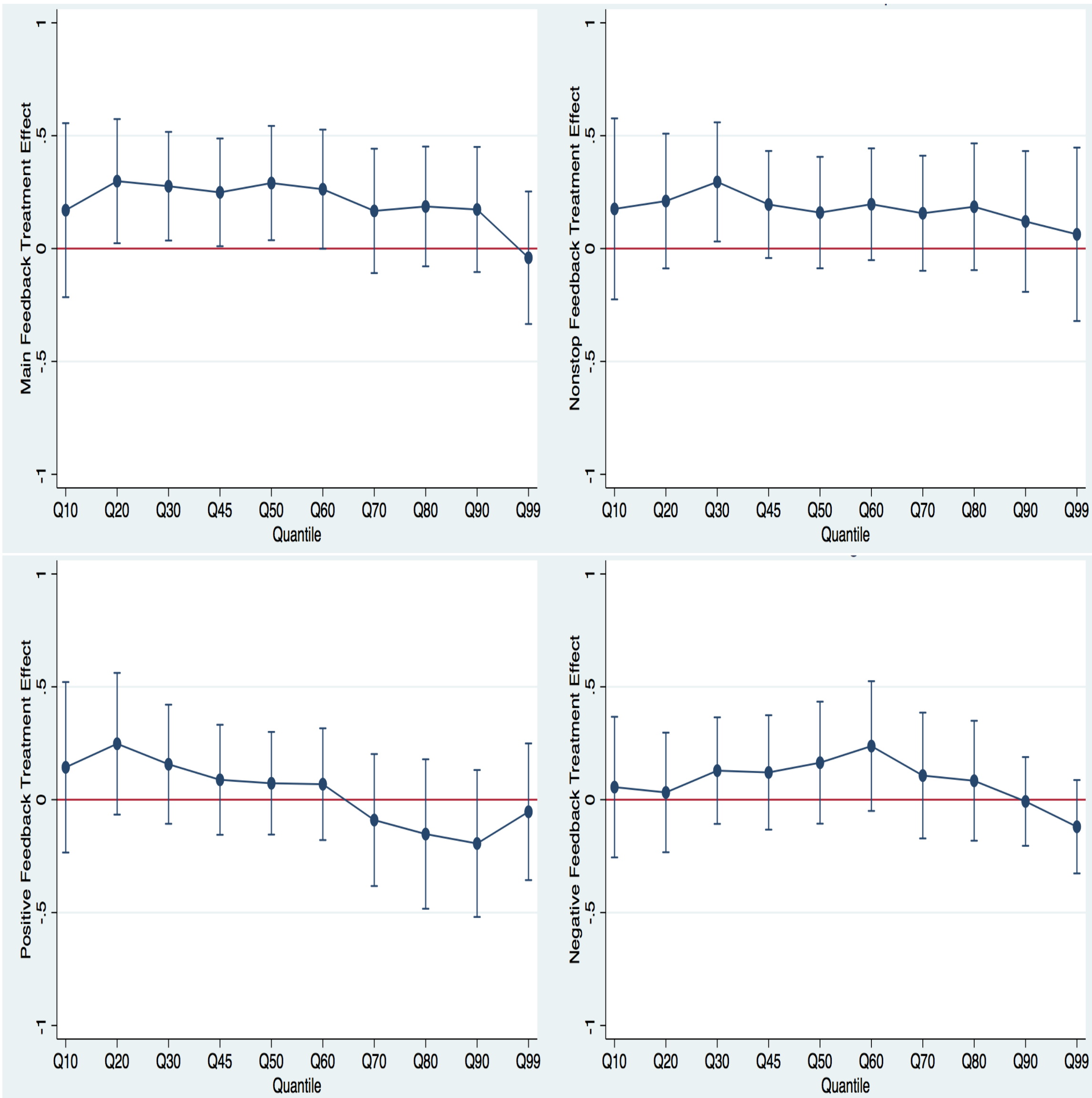
FIGURE 1. EXAMPLE OF RANKING FEEDBACK AS SHOWN IN THE ASSIGNMENT

FIGURE 2. DENSITIES OF AVERAGE EXAM GRADES BY FEEDBACK TREATMENTS



Notes: This figure show the density functions for the standardized average exam grades that treated students achieve compared to their control counterparts. The top left plot shows the densities for the Main treatment group compared to the control group; the top right plot shows the densities for Nonstop vs. control; the bottom left one shows the densities for Positive vs. control; the bottom right plot shows the densities for the Negative vs. control.

FIGURE 3. QUANTILE REGRESSION ESTIMATES OF THE EFFECT OF DIFFERENT FEEDBACK TREATMENTS



Notes: This figure presents the estimated quantile effects (marginal effects) of feedback provision on the standardised grades averaged across all exams at each decile, and the associated 95% confidence interval. The quantile regressions are conditional on students' age, gender, dummies for a student's country of birth and tutorial fixed effects. We use bootstrapped standard errors with 500 repetitions. The top left plot presents the quantile effects for Main treatment; the top right, bottom left and bottom right plots present the quantile effects for the Nonstop, Positive and Negative treatment, respectively.

Appendix

Table A.1. Treatment Effects of Feedback on Academic Performance: Main vs. Alternative Disclosures

	Assignment	Week 6 Exam			Week 10 Exam			Final Exam			Average Exam		
	Rank (1)	Perf. (2)	Std. (3)	Rank (4)	Perf. (5)	Std. (6)	Rank (7)	Perf. (8)	Std. (9)	Rank (10)	Perf. (11)	Std. (12)	Rank (13)
Nonstop Feedback	-6.269	-0.075	-16.121	-0.040	-0.089	-0.042	-13.484	-0.046	-0.027	-7.171	-0.070	-0.043	-12.259
	(36.954)	(0.187)	(29.555)	(0.098)	(0.208)	(0.097)	(30.743)	(0.171)	(0.102)	(31.507)	(0.161)	(0.098)	(26.632)
Observations	428	432	432	432	432	432	432	432	432	432	432	432	432
Positive Feedback	-68.997	-0.160	-20.498	-0.084	-0.320	-0.149	-48.710	-0.172	-0.102	-34.411	-0.207	-0.127	-35.094
	(28.958)**	(0.216)	(34.948)	(0.114)	(0.178)*	(0.083)*	(25.228)*	(0.186)	(0.110)	(34.650)	(0.162)	(0.099)	(27.247)
Observations	436	439	439	439	438	438	438	439	439	439	438	438	439
Negative Feedback	-15.917	-0.194	-39.706	-0.102	-0.251	-0.117	-28.869	-0.296	-0.176	-49.179	-0.247	-0.151	-39.251
	(32.409)	(0.170)	(29.023)	(0.090)	(0.182)	(0.085)	(28.912)	(0.152)*	(0.091)*	(26.349)*	(0.132)*	(0.081)*	(22.519)*
Observations	453	457	457	457	457	457	457	457	457	457	457	457	457
Tutorial FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in column (1) is the ordinal rank a student achieves in the online assessment (out of 1,093). The dependent variable in columns (2)-(4) is the exam grade (out of 10), standardized exam grade and the grade rank in the first mid-term, while columns (5)-(7) and (8)-(10) refer to the same outcome formulations for the second mid-term and the final exam. Columns (11)-(13) average these series over all three course exams. In all specifications we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered at tutorial level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

TABLE A.2. HETEROGENEOUS TREATMENT EFFECTS OF FEEDBACK ON ACADEMIC PERFORMANCE: GENDER

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam	
	Males (1)	Females (2)	Males (3)	Females (4)	Males (5)	Females (6)	Males (7)	Females (8)
Main Feedback	0.176	0.292	0.020	0.277	0.134	0.381	0.100	0.364
	(0.146)	(0.210)	(0.135)	(0.190)	(0.129)	(0.179)**	(0.134)	(0.193)*
Observations	248	197	247	197	248	197	247	197
Nonstop Feedback	0.002	0.094	0.155	0.057	0.121	0.126	0.071	0.104
	(0.199)	(0.197)	(0.143)	(0.143)	(0.163)	(0.172)	(0.162)	(0.169)
Observations	242	187	241	187	242	187	241	187
Positive Feedback	0.043	0.063	0.052	0.106	0.107	0.111	0.065	0.126
	(0.146)	(0.167)	(0.138)	(0.155)	(0.144)	(0.180)	(0.140)	(0.166)
Observations	233	203	232	202	233	203	232	202
Negative Feedback	0.060	0.194	0.044	0.260	-0.005	0.070	0.019	0.212
	(0.138)	(0.186)	(0.140)	(0.180)	(0.151)	(0.202)	(0.133)	(0.189)
Observations	268	186	267	186	268	186	267	186
Tutorial FE	✓	✓	✓	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variables in columns (1)-(2), (3)-(4) and (5)-(6) are the standardized exam grade in the first, second and final course exams, respectively. The dependent variable in columns (7)-(8) is the standardized exam grade averaged across all three exams. In all specifications, we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered at tutorial level and reported in parentheses.*, **, and *** denote significance at the 10%, 5% and, 1% level, respectively.

TABLE A.3. HETEROGENEOUS TREATMENT EFFECTS OF FEEDBACK ON ACADEMIC PERFORMANCE: AGE

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam	
	Age< 19	Age>= 19	Age< 19	Age>= 19	Age< 19	Age>= 19	Age< 19	Age>= 19
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Main Feedback	0.065	0.232	0.125	0.294	0.045	0.301	0.095	0.308
	(0.182)	(0.141)	(0.151)	(0.141)**	(0.206)	(0.127)**	(0.173)	(0.139)**
Observations	186	259	186	258	186	259	186	258
Nonstop Feedback	0.002	0.079	0.210	0.132	0.128	0.132	0.136	0.108
	(0.152)	(0.200)	(0.155)	(0.153)	(0.166)	(0.166)	(0.147)	(0.167)
Observations	184	245	184	244	184	245	184	244
Positive Feedback	0.027	0.081	0.052	0.191	-0.054	0.171	0.015	0.176
	(0.153)	(0.138)	(0.209)	(0.132)	(0.178)	(0.149)	(0.186)	(0.141)
Observations	177	259	177	257	177	259	177	257
Negative Feedback	0.083	0.088	0.148	0.077	-0.184	0.094	0.034	0.089
	(0.140)	(0.143)	(0.144)	(0.142)	(0.150)	(0.143)	(0.148)	(0.138)
Observations	188	266	188	265	188	266	188	265
Tutorial FE	✓	✓	✓	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variables in columns (1)-(2), (3)-(4) and (5)-(6) are the standardized exam grade in the first, second and final course exams, respectively. The dependent variable in columns (7)-(8) is the standardized exam grade averaged across all three exams. In all specifications, we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered by student and reported in parentheses.*, **, and *** denote significance at the 10%, 5% and, 1% level, respectively.

Table A.4. Heterogeneous Treatment Effects of Feedback on Academic Performance: International Status

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam	
	Domestic (1)	International (2)	Domestic (3)	International (4)	Domestic (5)	International (6)	Domestic (7)	International (8)
Main Feedback	0.067	0.373	0.148	0.239	0.084	0.377	0.108	0.377
	(0.115)	(0.255)	(0.098)	(0.267)	(0.104)	(0.216)*	(0.103)	(0.255)
Observations	335	110	334	110	335	110	334	110
Nonstop Feedback	0.059	0.159	0.158	-0.051	0.095	0.213	0.110	0.112
	(0.129)	(0.375)	(0.095)	(0.288)	(0.110)	(0.369)	(0.110)	(0.363)
Observations	345	84	344	84	345	84	344	84
Positive Feedback	-0.007	-0.092	0.112	0.062	0.032	-0.083	0.064	0.014
	(0.101)	(0.431)	(0.096)	(0.282)	(0.108)	(0.374)	(0.103)	(0.347)
Observations	340	96	339	95	340	96	340	96
Negative Feedback	0.012	0.253	0.130	0.138	-0.084	0.333	0.024	0.272
	(0.114)	(0.275)	(0.108)	(0.291)	(0.121)	(0.290)	(0.115)	(0.296)
Observations	357	97	356	97	357	97	356	97
Tutorial FE	✓	✓	✓	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variables in columns (1)-(2), (3)-(4) and (5)-(6) are the standardized exam grade in the first, second and final course exams, respectively. The dependent variable in columns (7)-(8) is the standardized exam grade averaged across all three exams. In all specifications, we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered by student and reported in parentheses.*, **, and *** denote significance at the 10%, 5% and, 1% level, respectively.

TABLE A.5. HETEROGENEOUS TREATMENT EFFECTS OF FEEDBACK ON ACADEMIC PERFORMANCE: FIELD OF STUDY

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam	
	Econ. (1)	Non-Econ. (2)	Econ. (3)	Non-Econ. (4)	Econ. (5)	Non-Econ. (6)	Econ. (7)	Non-Econ. (8)
Main Feedback	-0.090	0.192	0.373	0.176	0.367	0.120	0.253	0.189
	(0.233)	(0.121)	(0.426)	(0.121)	(0.334)	(0.102)	(0.360)	(0.119)
Observations	81	364	81	363	81	364	81	363
Nonstop Feedback	0.063	0.083	-0.172	0.133	0.234	0.122	0.029	0.120
	(0.358)	(0.142)	(0.455)	(0.104)	(0.305)	(0.118)	(0.357)	(0.121)
Observations	81	348	81	347	81	348	81	347
Positive Feedback	0.054	0.008	0.268	0.046	0.121	0.032	0.179	0.041
	(0.391)	(0.112)	(0.313)	(0.126)	(0.361)	(0.120)	(0.349)	(0.127)
Observations	80	356	80	354	80	356	80	354
Negative Feedback	0.226	0.053	0.408	0.045	0.222	-0.068	0.341	0.009
	(0.298)	(0.125)	(0.268)	(0.107)	(0.322)	(0.122)	(0.292)	(0.115)
Observations	88	366	88	365	88	366	88	365
Tutorial FE	✓	✓	✓	✓	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variables in columns (1)-(2), (3)-(4) and (5)-(6) are the standardized exam grade in the first, second and final course exams, respectively. The dependent variable in columns (7)-(8) is the standardized exam grade averaged across all three exams. In all specifications, we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered by student and reported in parentheses.*, **, and *** denote significance at the 10%, 5% and, 1% level, respectively.

TABLE A.6. HETEROGENEOUS TREATMENT EFFECTS OF FEEDBACK ON ACADEMIC PERFORMANCE NEXT SEMESTER: GENDER AND TYPE OF COURSES

	GPA (Std.) Semester 2 2016			
	Males (1)	Females (2)	Non-Econ. (3)	Econ. (4)
Main Feedback	0.381	0.194	0.264	0.022
	(0.207)*	(0.160)	(0.125)**	(0.314)
Observations	234	182	270	135
Nonstop Feedback	0.159	0.022	0.132	-0.239
	(0.211)	(0.170)	(0.168)	(0.216)
Observations	223	176	240	147
Positive Feedback	0.283	-0.047	0.119	-0.032
	(0.176)	(0.147)	(0.154)	(0.236)
Observations	219	193	258	142
Negative Feedback	-0.145	-0.016	0.007	-0.126
	(0.176)	(0.200)	(0.150)	(0.206)
Observations	252	170	252	152
Tutorial FE	✓	✓	✓	✓
Student Characteristics	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS regressions. In columns (1) and (2) we look into the effect of feedback on males' and females' standardized GPA in the following semester, respectively. In columns (3) and (4) we look into the effect of feedback on students' next semester standardized GPA in Economics and non-Economics courses, respectively. In all specifications, we include controls for age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and tutorial fixed effects. Standard errors are clustered at tutorial level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and, 1% level, respectively.

TABLE A.7. DIFFERENCES IN PROPORTIONS OF STUDENTS ACROSS SAMPLES AND TREATMENT GROUPS

	Full Sample	Discussion Board Sample	Difference	
Variable	Mean	Mean	Diff.	SE
Main Feedback	0.203	0.196	0.008	(0.032)
Nonstop Feedback	0.189	0.228	-0.039	(0.033)
Positive Feedback	0.195	0.185	0.010	(0.031)
Negative Feedback	0.212	0.201	0.011	(0.032)

Notes: This table displays the differences in the proportion of students in each treatment group across the full sample (N=1,101) and the sub-sample that posts in the two discussion boards associated with the course (N=184).

TABLE A.8. DROP OUT AND FULL ASSIGNMENT COMPLETION RATES

	N (1)	Drop Outs (#) (2)	Continuing (#) (3)	100% Completions (#) (4)
Main Feedback	261	37	224	199
Nonstop Feedback	240	32	208	171
Positive Feedback	247	32	215	186
Negative Feedback	268	35	233	195

Notes: Each column shows the number of students belonging to each treatment group, as they enrol in the course, decide to drop out or to remain enrolled and complete 100% of the online assignment by the end of the semester.

TABLE A.9. TREATMENT EFFECTS OF FEEDBACK ON ACADEMIC PERFORMANCE: ROBUSTNESS CHECKS

	Week 6 Exam			Week 10 Exam			Final Exam			Average Exam		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Main Feedback	0.367	0.353	0.251	0.478	0.496	0.381	0.311	0.310	0.230	0.371	0.377	0.285
	(0.171)**	(0.182)*	(0.159)	(0.206)**	(0.213)**	(0.153)**	(0.153)**	(0.163)*	(0.111)**	(0.158)**	(0.168)**	(0.103)***
Observations	445	445	437	444	444	436	445	445	437	444	444	436
Nonstop Feedback	0.085	0.108	0.086	0.346	0.360	0.260	0.123	0.159	0.202	0.170	0.198	0.173
	(0.216)	(0.207)	(0.189)	(0.198)*	(0.194)	(0.144)*	(0.177)	(0.171)	(0.125)	(0.169)	(0.163)	(0.109)
Observations	429	429	426	428	428	425	429	429	426	428	428	425
Positive Feedback	0.092	0.116	0.052	0.171	0.205	0.166	0.067	0.121	0.132	0.109	0.152	0.116
	(0.171)	(0.167)	(0.139)	(0.189)	(0.200)	(0.127)	(0.160)	(0.158)	(0.108)	(0.152)	(0.157)	(0.096)
Observations	436	436	432	434	434	431	436	436	432	434	434	431
Negative Feedback	0.117	0.137	0.139	0.308	0.294	0.231	-0.004	-0.019	0.037	0.126	0.122	0.132
	(0.182)	(0.185)	(0..147)	(0.186)*	(0.194)	(0.173)	(0.159)	(0.159)	(0.133)	(0.154)	(0.157)	(0.117)
Observations	454	454	445	453	453	444	454	454	445	453	453	444
Prior Ability	X	X	✓	X	X	✓	X	X	✓	X	X	✓
Tutorial FE	X	X	✓	X	X	✓	X	X	✓	X	X	✓
Student Characteristics	X	✓	✓	X	✓	✓	X	✓	✓	X	✓	✓

Notes: Each row presents estimates from separate OLS regressions. The dependent variable in column (1)-(3), (4)-(6) and (7)-(9) is the exam grade (out of 10) in the first mid-term, the second mid-term and the final exam. Columns (10)-(11) average these series over all three course exams. Prior Ability is the imputed measure of previous academic performance. Standard errors are clustered at tutorial level and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.