CESifo WORKING PAPERS

10346
2023

March 2023

Relative Performance Feedback and Long-Term Tasks – Experimental Evidence from Higher Education

Raphael Brade, Oliver Himmler, Robert Jäckle

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10346*

Relative Performance Feedback and Long-Term Tasks – Experimental Evidence from Higher Education

Abstract

We present first experimental evidence that relative performance feedback improves both the speed and quality with which challenging long-term tasks are completed. Providing university students with ongoing relative feedback on accumulated course credits accelerates graduation by 0.12 SD, and also improves grades by 0.063 SD. Treatment effects are concentrated among students with medium pre-treatment graduation probabilities: when these students are informed about an above-average performance, their outcomes improve – otherwise their outcomes deteriorate. Combined with survey evidence, this pattern of results suggests that learning about own ability is a plausible mechanism.

JEL-Codes: C930, D830, D910, I210, I230, I240.

Keywords: relative performance feedback, rank, natural field experiment, higher education, perceived ability, belief updating.

Raphael Brade
ifo Institute – Leibniz Institute for Economic Research
at the University of Munich
Center for the Economics of Education
Poschingerstrasse 5
Germany – 81679 Munich
brade@ifo.de

Oliver Himmler
Faculty of Law, Social Sciences and Economics
University of Erfurt
Nordhäuser Strasse 63
Germany – 99089 Erfurt
oliver.himmler@uni-erfurt.de

Robert Jäckle
Faculty of Business Administration
Nuremberg Institute of Technology
Bahnhofstrasse 87
Germany – 90402 Nuremberg
robert.jaeckle@th-nuernberg.de

March 24, 2023
We would like to thank Cansın Arslan, Robert Schwager, Marie Claire Villeval, Ludger Wosmann, and seminar participants at the University of Gottingen, University of Erfurt, University of Erlangen-Nuremberg, ifo Center for the Economics of Education, and ESA Europe for valuable comments and discussions. Fabian Escher, Peter Frenzel, Burak Akkaymak, Milad Zargartalebi, Anna Bauer, Pauline Wagner, Clara Rattmann, and Maximilian Herrmann provided excellent research assistance. We gratefully acknowledge financial support from the German Federal Ministry of Education and Research under grant 01PX16003A, 01PX16003B, the Staedtler Stiftung, and administrative support from the Nuremberg Institute of Technology. The intervention is listed at the AEA registry (https://www.socialscienceregistry.org/trials/3288), but was not registered until aboutmidway through the intervention.

1 Introduction

Life is full of challenging long-term tasks that require ongoing dedication and effort. Examples include projects such as improving in a sport, maintaining a healthy lifestyle, pursuing a career, and accumulating human capital by attending high school and college as well as via lifelong learning. Many of these long-term endeavors, however, are never completed, take longer than intended, or yield worse outcomes than initially planned.

How can we support individuals facing these issues? Theory suggests that information about ordinal ranks can affect performance if individuals use this information to learn about their ability (assuming that effort and ability are complements), or if they have competitive preferences and therefore strive to outperform others (Azmat and Iriberri, 2010; Dobrescu et al., 2021; Ertac, 2005). It thus seems feasible for policymakers and organizations to support individuals in pursuing long-term tasks by facilitating and accelerating learning about ordinal ranks, for example, by providing relative performance feedback on progress toward completion of the long-term task.

The literature has not yet addressed this. First, although the literature on rank effects in education has shown that high ordinal ranks can have positive effects on performance, even over long periods of time (Denning, Murphy and Weinhardt, forthcoming; Elsner and Isphording, 2017; Murphy and Weinhardt, 2020), it remains unclear whether these findings carry over to the active provision of such rank information. This is because the comparison group as well as individuals' perceived rank are likely to differ depending on whether people learn about their rank through repeated interactions or through external information provision (Delaney and Devereux, 2022; Megalokonomou and Zhang, 2022), and because it is difficult to fully disentangle such rank effects from other peer effects (Delaney and Devereux, 2022; Denning, Murphy and Weinhardt, forthcoming). Second, while there is a large body of literature investigating the effects of relative performance feedback, these studies have mainly covered short time frames and have focused on providing feedback on task quality rather than feedback on progress toward task completion. Much of this literature also stems from lab environments rather than field settings (see Villeval 2020 and Section 2).

We investigate if relative performance feedback can improve the outcomes of challenging long-term tasks by conducting an intervention in higher education and collecting data over a six year period. In two natural field experiments that share the same design, students received ongoing relative feedback on their accumulated course credits, i.e., their progress toward graduation. The context is characterized by several important features: i) obtaining a university degree is a complex and challenging long-term task with high stakes, requiring continuous dedication and effort; ii) as students must obtain a fixed number of course credits to graduate, the higher education system allows to precisely measure individuals' progress toward task completion and to supply them with ongoing relative feedback on it. Additionally, higher education in itself is an important subject of investigation, since iii) many students perform poorly, fail to graduate, or take much longer than the scheduled time, and are thus in need of supportive measures;[1] and iv) improving outcomes in higher education promises

[1] In the U.S. and in other OECD countries, less than 40% of a cohort complete their bachelor's degree within the scheduled study time and three years later around 25% have left tertiary education without obtaining a degree (OECD, 2019). See Bound and Turner (2011) for a discussion of why collegiate attainment rates in the U.S. have stagnated for the last decades.

substantial individual and societal benefits (Hout, 2012; Lovenheim and Smith, 2023; Oreopoulos and Petronijevic, 2013), but there is still a dearth of low-cost and easily implementable interventions effective in improving long-term academic attainment (Azmat et al., 2019; Kim et al., 2022; Oreopoulos and Petronijevic, 2019).

We conducted our field experiments with over 1,600 students from two cohorts pursuing a bachelor's degree at a German university of applied sciences. To graduate in the scheduled study duration of seven semesters, students need to obtain 30 course credits per semester – the standard per-semester course load in European higher education.[2] However, 34.9% of control group students drop out of their program, and those who do graduate take on average 8.62 semesters – typical findings in higher education (see Footnote 1).

In both cohorts, our feedback intervention started in the second semester. In each semester, students in the treatment and the control group received two postal letters informing them about their academic progress by providing the number of credits they had obtained to date. Control group members received no further information, while the individuals in the treatment group were additionally provided with a visual comparison of their accumulated credits to the average and the top 20% in their cohort. Furthermore, we augmented the relative feedback with normative frames that conveyed approval for students with an at least average performance. This design feature is inspired by previously reported "boomerang" effects, which suggest that those who receive information about performing above some reference point may actually decrease their efforts; normative frames can prevent this (Allcott, 2011; Allcott and Rogers, 2014; Schultz et al., 2007). Letters were sent until the eleventh (Cohort I) and tenth (Cohort II) semester, and in both experiments we collected data until the end of the thirteenth semester, at which point over 99% of the students had either dropped out or completed their program.

Overall, the feedback intervention accelerates graduation, thus providing first evidence that relative performance feedback can be considered an effective tool to speed up the completion of challenging long-term tasks. The speed advantage starts to materialize in the first year after the scheduled study duration, i.e., in the eighth and ninth semester. In each of these semesters, the graduation rate of treated students is 4 percentage points (pp) above that of the control group. Over the entire observation period, students in the treatment group obtain their degrees on average about 0.15 semesters earlier than control students (≈ 0.12 standard deviations, SD), implying that roughly one in seven students graduates one semester earlier.

Our results allow, for the first time, to compare the long-term impact of an effective low-touch intervention in higher education with widely used traditional measures, such as grant aid. For example, in a meta-analysis including 43 studies, Nguyen, Kramer and Evans (2019) find that an additional $1,000 in grant aid increases the on-time (i.e., 100% to 125% of the nominal study time) degree completion rate by 1.8 pp. Our relative feedback intervention increases the eight semester graduation rate, i.e., the 114% degree completion rate, by about 4 pp. It thus has roughly the same effect on

[2]Throughout Europe, universities use a standardized point system (European Credit Transfer and Accumulation System, ECTS), in which a full-time academic year consists of 60 credits, with the typical workload for one credit corresponding to 25-30 study hours. See also `https://ec.europa.eu/education/resources-and-tools/european-credit-transfer-and-accumulation-system-ects_en`, retrieved on March 17, 2023.

on-time graduation as \$2,000 of grant aid, but at a much lower cost of about €14.2 per student (see Table A.1).[3]

The overall effect speeds up graduation by one semester for one in seven students. It thus masks heterogeneous responses. We hypothesize that the long-run effects of feedback are driven by students who ranked above average pre-treatment. This is motivated by the heterogeneous effects that we observed after the initial provision of feedback during the second semester. These short-term results are documented in detail in Brade, Himmler and Jäckle (2022) and show that relative feedback increased the obtained second semester credits among students who were informed about an above-average first semester performance. Importantly, a regression discontinuity design around the cutoff for being above average provided evidence that this "effect of being above average" was not driven by these students being able to react better to feedback due to their underlying characteristics (e.g., ability, motivation, learning technology, etc.). Rather, the content of the relative feedback itself generates the effects. We also found no evidence that the approving normative frames affected performance, again pointing to an underlying mechanism of receiving information about being above average. Going beyond our own short-term results, the hypothesis that students who (pre-treatment) ranked above average are the ones driving the effects, can also be derived from the literature on ordinal ranks. It provides ample evidence that receiving feedback about a high rank should be more effective than receiving information about a low rank (see the discussion above, and Section 2) .

Indeed, we find that the positive effects on completion time are driven by individuals who are informed about an above-average pre-treatment performance. In the eighth and ninth semester, their graduation rate improves by 8 pp with treatment, while there are no effects of feedback for students who did not rank above average pre-treatment. Interestingly, in the above-average subgroup, relative feedback not only reduces time to graduation, but it also increases overall degree attainment: after thirteen semesters, these students are about 4 pp more likely to have graduated than the above-average controls.

To explore these heterogeneities further, we use pre-treatment performance measures and background characteristics of control group members to estimate the pre-treatment graduation probability by the end of the eighth semester for all students. For treated students, this can be interpreted as the counterfactual graduation rate in the absence of relative feedback. Our analyses reveal that relative feedback has important distributional implications. Students who are responsive to feedback are concentrated in the middle of the graduation probability distribution. As above, their behavior depends on the content of the feedback: those whom the initial feedback in the second semester informed about an above average performance react positively, which results in about 12 pp higher graduation rates in the eighth and ninth semester and a 7.4 pp increase in their overall degree attainment. Those who are in the middle of the graduation probability distribution, but initially not above average, are affected negatively and are 8.8 pp less likely to have graduated at the end of our observation period. Other students, i.e., those at the bottom and the top of the graduation probability distribution, show little to no response.

[3]Besides positively affecting on-time graduation, grant aid also increases college enrollment (Deming and Dynarski, 2010) as well as persistence in college and delayed graduation, i.e., the 150% or more degree completion (Nguyen, Kramer and Evans, 2019).

From a theoretical perspective, the results for responsive students can be aligned with the predictions of "self-perception theory" (Azmat and Iriberri, 2010; Dobrescu et al., 2021; Ertac, 2005). According to this theory, relative feedback can lead individuals to positively (negatively) update their beliefs about their own abilities. This in turn results in higher (lower) effort and ultimately performance (assuming that ability and effort are complements). Consistent with this pattern, we report survey evidence that after feedback the treated above-average students have significantly higher expectations about their future relative performance than the treated not-above-average students, while there is no difference in the expectations of these two groups among controls. These findings align well with the literature on rank effects, which points to changes in confidence in ability as a channel for the long-lasting impacts of ordinal ranks. For the not-above-average students, the results can additionally be reconciled with Stinebrickner and Stinebrickner (2012, 2014), who show that up to 45% of dropout in the early stages of college can be attributed to students learning about their academic performance.

Under competitive preferences, on the other hand, theory typically predicts performance-enhancing effects of relative feedback across the entire distribution (Azmat and Iriberri, 2010; Dobrescu et al., 2021), which is not consistent with our results. In addition, survey evidence suggests that responsive students in the treatment group do not perceive the study environment as more competitive after receiving feedback compared to controls, and that students at this university are less competitive with respect to their time to degree than their final grades. This lower level of competitiveness in the domain that we give feedback on may thus explain why we find different results than, for example, Azmat and Iriberri (2010), Azmat et al. (2019), and Dobrescu et al. (2021).

For the non-responsive students at the bottom and the top of the distribution, it is more difficult to pin down the exact reasons for their lack of response. We argue that the absence of effects can be attributed to floor and ceiling effects, which leave little room for declines and increases in performance. For those at the top, this is likely also due to the European Credit Transfer and Accumulation System (ECTS) and its recommendation to take exactly 30 credits per semester. First, the degree programs at the university are structured in accordance with the ECTS, and thus provide only limited opportunities for students to accumulate more than the specified number of credits per semester. Therefore, the program structure may effectively prevent students from taking more than 30 credits, and thus stifle potential treatment effects. Second, in addition to the lack of institutional flexibility, the ECTS recommendation to obtain exactly 30 credits per semester may set a reference point which only few students want to exceed.

Finally, we address the concerns that (i) there may be a trade-off between earlier task completion and task-quality, and (ii) providing feedback may have detrimental effects on individuals' mental health (Reiff et al., 2022). Empirical evidence on such negative spillovers in settings with multiple performance dimensions or tasks is provided by Altmann, Grunewald and Radbruch (2022), Eriksson, Poulsen and Villeval (2009), and Hannan et al. (2013). In our experiments, students who receive feedback and consequently manage to graduate earlier, may buy these gains with worse grades or lower well-being. We therefore assess spillovers on these domains and find no evidence supporting these concerns. Quite the contrary, considering task quality, we show that treated students' GPA in the overall sample increases by 0.063 SD, with the positive effects again driven by the responsive above-average students. In terms of well-being, as measured by students' satisfaction with their studies and life as well as study stress, we find little differences between control and treatment.

The paper proceeds by discussing its contributions to the literature, followed by a summary of the institutional background and the design of the intervention in Section 3. Sections 4 to 6 present the average treatment effects and the two heterogeneity analyses. In Section 7, we discuss possible mechanisms behind the results. Section 8 examines spillover effects to other domains and Section 9 concludes.

2 Contributions to the literature

This study contributes to the literature on feedback by providing first evidence that relative performance feedback can improve the speed and quality with which challenging long-term tasks are completed.[4] Unlike our study, most of the related research with adult populations in the field either investigates (i) tasks in workplace contexts that are relatively straightforward or repetitive (Ashraf, 2022; Blader, Gartenberg and Prat, 2020; Blanes i Vidal and Nossol, 2011), or (ii) higher education settings focusing on single courses or subjects, and rather short time spans like one semester (Ashraf, Bandiera and Lee, 2014; Chen et al., 2021; Dobrescu et al., 2021; Kajitani, Morimoto and Suzuki, 2020; Klausmann, Wagner and Zipperle, 2021; Tran and Zeckhauser, 2012).[5] However, spillovers, which can operate both vertically over time, e.g., from one semester to the next, and horizontally across subjects and courses, make it hard to draw conclusions from these settings on whether relative feedback can improve net performance over a long period of time. In fact, the two studies most similar to ours in terms of duration and task complexity have shown that this might not be the case. In a three-year field experiment, Barankay (2012) studies the effects of providing furniture salespeople with their performance rank in terms of year-to-date sales and finds negative effects of relative performance feedback as well as evidence for effort substitution across tasks. In Azmat et al. (2019), university students receive online feedback on their decile rank in the GPA distribution from the second academic year onward. While effects are negative in the short run, after four years, there is no difference between treatment and control students in terms of final grades or degree attainment. But these two studies differ from ours in important ways: i) We provide feedback on a performance dimension (credits) that tracks individual progress toward the completion of one challenging long-term task instead of the quality dimension (grades; Azmat et al. 2019) or repeated task completion (number of furniture sales; Barankay 2012). ii) Our relative performance feedback provides students with coarse instead of precise rank information.

Second, our study contributes to the literature on rank effects in education and links it to the feedback literature by adding experimental evidence that active and repeated provision of rank information over a long period of time can improve academic outcomes in a similar way and through comparable underlying mechanisms as learning about ranks through repeated interaction (see De-

[4] See Villeval (2020) for a comprehensive review of the relative feedback literature that also covers evidence from lab experiments and from settings with tournament and team incentives.

[5] There is also a literature on feedback in primary and secondary education that mostly finds positive effects, but it is unclear whether these results extend to an adult population (Azmat and Iriberri, 2010; Fischer and Wagner, 2023; Goulas and Megalokonomou, 2021; Hermes et al., 2021).

laney and Devereux (2022) for an overview and a discussion of why it is not clear that the implications of the rank literature carry over to active information provision). Denning, Murphy and Weinhardt (forthcoming), Elsner and Isphording (2017), and Murphy and Weinhardt (2020) show that higher ranks in secondary, primary, and elementary school can have positive long-term effects on academic attainment, such as attending and graduating from high school and college. It has been suggested that these effects arise because students use the relative ability information that ranks provide to learn about their true abilities and comparative advantages. Higher-ranked students therefore have greater confidence in their abilities, which then increases performance by reducing the cost of effort (Delaney and Devereux, 2022; Elsner and Isphording, 2017; Elsner, Isphording and Zölitz, 2021; Murphy and Weinhardt, 2020; Pagani, Comi and Origo, 2021). Our approach and results complement and strengthen the findings in this literature, because the active provision of rank information via relative feedback is less susceptible to concerns about how students perceive their rank (Megalokonomou and Zhang, 2022) and whether rank effects can be fully disentangled from other peer effects (Bertoni and Nisticò, 2023; Denning, Murphy and Weinhardt, forthcoming).

Third, our paper provides, to our knowledge, the first evidence that a low-cost and easy-to-implement behavioral intervention can improve long-term achievements in higher education (see Dynarski et al. (2023) for a recent review). While there have been promising attempts to enhance performance in the short-term[6], the very few long-term studies report null results. As noted above, Azmat et al. (2019) find no long-term effects of providing rank feedback on students' GPA. Kim et al. (2022) study the effects of brief online growth mindset and "belonging" interventions over four years and report no overall effects. Oreopoulos and Petronijevic (2019) find no effects for several interventions which ran up to three years and aimed at increasing performance (including goal setting, mindset, and online one- or two-way text coaching interventions).

Fourth, outside of higher education, evidence on the long-term effects of behavioral interventions is similarly scarce – and it is not at all clear that short-term effects will increase or even persist over longer time spans (Ashraf, Karlan and Yin, 2006; Azmat et al., 2019; Blattman, Fiala and Martinez, 2020). Beshears and Kosowsky (2020) analyze 174 studies that estimate the effects of behavioral interventions and find that only 21% collect follow up data at least once or attempt to measure the cumulative effect of a series of nudges. In their meta-analysis, DellaVigna and Linos (2022) find that the average intervention and data collection time among behavioral interventions published in academic journals is 6.7 months ($\sigma = 7.1$), and 4.5 months ($\sigma = 2.0$) for trials that were run by two large U.S. nudge units. The total time span of our intervention and collection of outcome data is over 6 years – within this duration, over 99% of all students who started their bachelors' program either dropped out or completed their degree. The time horizon of our intervention is thus beyond almost all of the studies considered in the above meta-analyses.

[6]This includes measures such as low-touch assistance and information provision (Carrell and Kurlaender, forthcoming; Castleman and Page, 2016; Rury and Carrell, 2020), goal setting (Clark et al., 2020), commitment devices (Himmler, Jäckle and Weinschenk, 2019; Patterson, 2018), and exam sign-ups by default (Behlen, Himmler and Jaeckle, 2022)

3 Institutional background and research design

This section summarizes the institutional background and the design of the relative feedback intervention; see also our previous paper on the short-term effects of relative performance feedback (Brade, Himmler and Jäckle, 2022).

3.1 Institutional background

We conducted the intervention at a large public university of applied sciences (UAS) in Germany. UAS are bachelor's and master's degree granting institutions, and the curriculum at these schools is somewhat more practice-oriented than at more research-oriented German universities. UAS serve a substantial and growing share of the German student population: in the fall semester of 2020, about 39.9% of freshman university students studied at a UAS (Statistisches Bundesamt, 2021). Our subjects were incoming students from two consecutive cohorts who enrolled in one of five bachelor's degree programs at the departments of Business Administration (BuA) and Mechanical Engineering (ME). BuA and ME are among the most popular fields in Germany. In the fall semester of 2020, about 8.4% and 3.1% of freshman students were enrolled in these programs, making them the first and fourth most popular study programs in the country (Statistisches Bundesamt, 2021).

The study programs at the UAS are structured in accordance with the European Credit Transfer and Accumulation System (ECTS, see Footnote 2). The scheduled study duration is seven semesters and students need to accumulate a total of 210 course credits in order to complete their degree. The programs follow a clear structure: In the first and second year, students take a range of compulsory fundamentals courses. Subsequently, all students are required to complete a mandatory 20-week internship, scheduled for the fourth or fifth semester. In the final year, students choose from electives and write their thesis.

Students can at all times track their progress via a web portal of the university, which provides information on current GPA and the number of credits earned so far. Importantly, the web portal only shows information about one's own performance without any comparison to the performance of other students.

Similar to the overall student population in Germany and in other countries, students at this UAS are struggling to complete their degree within the scheduled duration, and many do not graduate at all. By the end of our observation period, 34.9% of the control group have dropped out of their study program, and those who do graduate take on average 8.62 (SD = 1.28) instead of the scheduled seven semesters.

3.2 Relative feedback intervention

Against this background, we designed a relative feedback intervention. Its goal was to decrease time to graduation by providing additional (social) incentives for the accumulation of course credits. We evaluate the effectiveness of the feedback scheme with two natural field experiments of the same design. At the beginning of their second semester, all students from the incoming fall semester cohorts

2014 and 2015 who were still enrolled in their degree programs were assigned to two groups: absolute feedback (= control) and relative performance feedback (= treatment).

Randomization. After the university provided us with information on students' first semester performance (see Figure 1 for a timeline of the intervention), we randomized our subjects into treatment and control using stratification and rerandomization (Morgan and Rubin, 2012). Students in each cohort were allocated to strata based on their study programs and their pre-treatment credit points obtained in the first semester.[7]

Feedback letters. Starting in the second semester and continuing until the eleventh semester (Cohort I) and tenth semester (Cohort II), both control and treatment group students received two unannounced postal letters per semester. The envelopes bore the official seal of the university, and the letters were signed by the dean of the students' department. For both groups, the letters included the current GPA and the credits that the student had accumulated so far – i.e., the same information that was also available via the web portal of the university. The treatment letters additionally contained a graphical illustration that provided students with relative feedback on their credit points (we explain the relative feedback in more detail below).

At the beginning of each semester, the first letter was sent to all students who were still enrolled in their initial study program. About a month before the exam period, we sent the second letter. Its purpose was to keep the relative performance feedback salient at a time when students usually start preparing for their exams. Except for the introductory paragraph, the design of the first and second letter was identical. In some cases, the information on students' academic progress had to be updated, e.g., because grades or credits from some courses were not yet available when the first letter was prepared or because they were updated (students inspect their exams during the first week of the semester, sometimes uncovering grading errors). In later sections of the paper we calculate student rank in the first semester credit point distribution – this is based on the content of the information provided in the second letter.

Design of the relative performance feedback. Figure 2 shows examples of the relative performance feedback included in the treatment letters. The design was inspired by social comparison letters which have been successfully used in the context of energy conservation (Allcott, 2011; Allcott and Rogers, 2014; Schultz et al., 2007). A bar chart compares the student's accumulated credit points to the "Top 20%" and to "All" students who are enrolled in the same degree program and are in the same cohort – i.e., the performance of the student(s) on the 80th percentile and the average performance (students who dropped out of the program or graduated were no longer used for the calculation of the statistics). In the first cohort, the average was defined as the number of credits obtained by the

[7] In Experiment I, we defined five credit strata for each study program (credits ≤ 12, 12 <credits ≤ 18, 18 < credits ≤ 24, 24 < credits ≤ 30, credits > 30). In Experiment II, we defined credits strata based on quantiles (Q); four credit strata in the larger study programs BuA and ME (credits < $Q_{0.25}$, $Q_{0.25}$ ≤ credits < $Q_{0.5}$, $Q_{0.5}$ ≤ credits < $Q_{0.75}$, credits ≥ $Q_{0.75}$) and two credits strata in the other three study programs (credits < $Q_{0.5}$, $Q_{0.5}$ ≤ credits). Within these strata, we performed up to 500 random draws, keeping the allocation with the best balancing properties with respect to age, gender, high school GPA, time since high school graduation, first semester GPA, and – in Experiment II – type of high school degree.

median student(s), and in the second cohort, it was defined as the arithmetic mean. As we discuss in detail in Brade, Himmler and Jäckle (2022), this little tweak in the design allowed to increase the share of students who received information about an above-average performance in the second semester from 37.5% to 56.2%. The tweak was prompted by initial results of the first experiment, which showed that feedback about an above-average performance was more effective. Importantly, switching the average from median to arithmetic mean left the general design of the intervention unaltered, and the use of the term "average" in the letters ensured that students across both experiments would interpret the feedback information in the same way.

The performance comparison group in the letters did not comprise all students in a degree program. Rather relative performance comparisons were within subgroups: In the three smaller bachelor's programs, subjects were compared with others who earned their school leaving certificate in the same year. In the two larger programs, students were compared with students who obtained the same type of school leaving certificate in the same year. These smaller comparison groups served two purposes: First, they aimed to increase perceived similarity among students, thus theoretically increasing the impact (Trope and Liberman, 2010). Second, they were intended to reduce spillovers that could potentially arise from sharing the feedback information, because we hypothesized that the more personalized information in the letters may appear of little interest to other students.[8] Indeed, our analysis of post-treatment beliefs about relative performance in Section 7 suggests that knowledge about relative performance differs between treatment and control.

Drawing from the literature on social comparison, we augmented the descriptive information with normative frames to mitigate possible "boomerang" effects, i.e., negative effects for individuals who placed above the reference points, as has been done in the context of energy conservation (see, e.g., Allcott 2011 and Schultz et al. 2007). Thus, if the performance was at least average, the relative feedback included an approving normative frame. It categorized performance as *good* (plus one "smiley" emoticon) for students at or above the average, and *great* (plus two "smiley" emoticons) for students in the top 20%. For students below average, no approving frame was displayed. Instead, they received the statement "currently below average" (and no emoticon).

With our experimental design, we cannot disentangle the effects of the descriptive performance information from the effects of the normative frames. However, in Brade, Himmler and Jäckle (2022), we provide evidence that approving normative frames alone, i.e., without descriptive performance information that fits the normative frame, are not sufficient to elicit performance gains. Overall, our approach is policy-oriented in that it draws on feedback designs that have been successfully employed in the social comparison literature, and thus promised the greatest benefits for students.

3.3 Data and descriptives

For most of the analyses, we employ administrative student-level data on background characteristics that we received before the randomization, as well as information on performance and progress

[8]Students enrolled in the study programs at the time provided anecdotal evidence that sharing of this information was not common, e.g., there did not appear to be any sharing of the feedback graphs in social media.

toward degree, which we received after the end of the full observation period.[9]

Because relative feedback was provided on credits, we expect effects to materialize on dimensions that are tied to credit accumulation: we study effects on graduation rates, dropout, and time to degree (since increased credit accumulation should translate into earlier graduation). We also investigate effects on accumulated credit points. Finally, because students may buy gains on credit-point-related outcomes with losses on the other relevant performance dimension (grades), we also estimate effects on GPA.[10]

Table 1 presents descriptive statistics and balancing properties for students' background characteristics and measures of baseline performance. It shows that our sample of 1,609 individuals is well balanced.[11] We use standardized and reverse scaled GPAs throughout the analyses of the paper, both for the high school as well as the university GPA. Higher values therefore represent better grades (on the original German scale 1.0 is the best and 4.0 the worst grade).

We complement the administrative data with information from four online surveys that were conducted in the second and fifth semester for the first cohort, and in the second and third semester for the second cohort. Data collection took place in the second half of the semesters, and Table A.2 provides evidence that survey participation is not significantly different between the treatment and the control group. In our analyses, we consider effects on performance expectations, perceived competition and well-being. More details are provided in Sections 7 and 8.2.

4 Average treatment effects

This section presents the overall effects of our relative feedback intervention; we show that treated students graduate significantly earlier and accumulate more credits during the scheduled study duration.

4.1 Empirical approach

Throughout the paper, we pool observations from both cohorts ($N = 1,609$), and our results can thus be interpreted as meta-analytic estimates (Camerer et al., 2016; Open Science Collaboration, 2015). We choose this approach because i) the two experiments share the same design[12], ii) they produced

[9] To ensure data consistency, we use these performance data provided at the end of the entire observation period. As a consequence, the performance data in earlier semesters are not exactly the same as in Brade, Himmler and Jäckle (2022), because in some cases the university updated the information in earlier semesters.

[10] To accurately reflect students' progress toward their degrees, we measure their performance (GPA and accumulated credits) on the module-level. Modules usually consist of one or sometimes also multiple exams that can spread across more than one semester. Students only receive credits if they pass the entire module. The university website and the feedback letters both provide information based on the module-level. Brade, Himmler and Jäckle (2022) includes a detailed discussion of the differences between exam- and module-level performance information and why the exam-level data were used for the analyses of the short-term effects. Importantly, there, we also explain why the short-term effects based on exam-level data can be conservative.

[11] To keep all observations in the sample when using the first semester GPA as a covariate, we impute values for 151 students, who obtained no grade in the first semester (see notes in Table 1 for details).

[12] In the taxonomy of Czibor, Jimenez-Gomez and List (2019), Hamermesh (2007), and Hunter (2001), our second experiment was a statistical replication of the first, as it used the identical model (same protocol as the first experiment) and the

the same pattern of results in the short run (Brade, Himmler and Jäckle, 2022), and iii) pooling increases the power of the statistical analyses.

We report intention-to-treat effects throughout the paper, as we have no information on whether students opened and read the feedback letters. Our results are based on the following specifications estimated via OLS:

$$Y_i^k = \alpha_0 + \alpha_1 T_i + \mathbf{s_i}\boldsymbol{\alpha_2} + \mathbf{x_i}\boldsymbol{\alpha_3} + \mathbf{z_i}\boldsymbol{\alpha_4} + \varepsilon_i, \tag{1}$$

where Y_i^k denotes the outcome measure k for individual i, T_i is an indicator for being randomized into the treatment group, and the vector $\mathbf{s_i}$ controls for the stratified randomization by including strata fixed effects. The *randomization strata* are defined by credits obtained in the first semester (cf. Footnote 7), study program dummies, a cohort dummy, and its interaction with the study program dummies. To increase the efficiency of our estimation, in our preferred specification, we additionally include the vectors $\mathbf{x_i}$ and $\mathbf{z_i}$. $\mathbf{x_i}$ includes information on *baseline performance*, i.e., credits earned in the first semester, the standardized first semester GPA (with missing values imputed as described in Section 3.3), a dummy that indicates if the first semester GPA was imputed, and information on first semester dropout.[13] $\mathbf{z_i}$ includes *additional covariates* (standardized high school GPA, age at randomization, an indicator for women, time since high school graduation, and an indicator for the type of high school degree).

4.2 Effects on graduation rates and time to degree

Our primary interest is in evaluating how feedback changes graduation rates over time (see Figure 3 and Table 2). The figure shows that only 9.3% of the students in the control group manage to earn their degree within the scheduled study duration of seven semesters, thus leaving substantial room for improvement. Among treated students the share is slightly higher, but the effect is not estimated precisely (1.2 pp, $p = 0.396$). The majority of students in the control group graduate in the eighth and ninth semester, and in this period our treatment has significant positive effects on the graduation rate. Figure 3 and Columns (1) to (3) in Table 2 show that the share of students in the treatment group who complete their degree until the end of the eighth semester is 3.7 to 4.6 pp ($p = 0.059$ to 0.027) higher than in the control group (36.0%). We find about the same difference in the ninth semester: at this stage, 48.8% of the students in the control group have already earned their degree and completion in the feedback group is up by 3.7 to 4.5 pp ($p = 0.052$ to 0.025, Columns 4 to 6). Subsequently, the control group catches up, reducing the effect on the graduation rate at the end of the observation period of thirteen semesters to 0.9 to 1.2 pp ($p = 0.514$ to 0.631, Columns 7 to 9), compared to a control mean of 63.8%. Relatedly, we do not find any significant differences in dropout rates (see Figure A.1), which are at about 35% for both groups at the end of the thirteenth semester, implying that only about one percent of students is still enrolled in their degree program at this point in time.

---

same underlying population, but was based on a different sample (new cohort).

[13]The first semester dropouts are due to students whose dropout was not recorded in the administrative data until after the university provided us with information on students' first semester performance.

Taken together, this pattern of results suggests that relative feedback helps students graduate earlier but does not increase the number of graduates.

How large is the effect of feedback on time to degree? In Columns (1) to (3) of Table 3, we provide OLS estimates for the treatment effect on the number of semesters needed to graduate. In the control group, students earn their degree on average after 8.62 semesters. Relative performance feedback reduces the time to graduation by about 0.147 to 0.154 semesters ($p = 0.026$ to 0.029), which corresponds to an effect size of about 0.120 control group standard deviations (SD). However, these estimates cannot account for the (few) individuals who are still studying, because time to degree is not yet observed for them. In Columns (4) to (6) of Table 3, we therefore set time to degree to fourteen semesters for all students who are still enrolled at the end of the thirteenth semester, thus assuming no further dropouts. We then estimate a right-censored Tobit model (at fourteen semesters). The coefficients of 0.188 to 0.214 semesters ($p = 0.009$ to 0.007; 0.126 to 0.144 SD) suggest that not yet observing time to degree for all potential graduates leads to a downward bias in the initial estimates. The reported effects can be interpreted as roughly one in five to one in seven students graduating one semester earlier (since time to graduation is measured in full semesters). This suggests that the average treatment effects mask substantial heterogeneity in the response to the treatment.

4.3 Effects on credit point accumulation

Do these effects also materialize in the number of credits that students accumulate over the course of their studies? Figure 4 shows that by the end of the first treatment semester (= second semester), relative feedback increases the number of accumulated course credits by 0.86 points ($p = 0.107$). In the following semesters, the gap between treatment and control increases and peaks at the end of the scheduled study duration of seven semesters. At this stage, students in the treatment group have accumulated on average an additional 2.351 to 2.937 credits ($p = 0.361$ to 0.280, Columns 1 to 3 in Table 4). After the peak in the seventh semester, the gap between treatment and control closes, because students who already completed their degree do not earn any further credits. The parameter estimates of accumulated credits are less precise in later semesters, as the variance increases due to students dropping out of their program and therefore obtaining no further credit points.

Since we found no evidence of differential dropout between treatment and controls, we also analyze feedback effects on accumulated credits at the end of the seventh semester in the subgroup of students who did not drop out of their study program until the end of the sixth semester (Columns 4 to 6 in Table 4). With coefficients ranging from 2.325 to 2.727 ($p = 0.109$ to 0.091), these conditional estimates are similar in size to the unconditional estimates. As expected, they are estimated more precisely, due to the lower variance of the conditional credits (see the smaller control group standard deviations in the bottom row of Table 4).

Taken together, these results suggest that treated students, on average, accumulated an additional 2.3 to 2.9 credits until the end of the scheduled period of study. A possible interpretation of these estimates is that one in thirteen to one in ten treated students obtains 30 credits more, which corresponds to the regular course load of one semester. This indicates again that the effects of the relative feedback intervention are heterogeneous across subgroups. Note that these results do not contradict our finding that about one in seven students in the treatment group graduates a semester earlier: those who graduate a semester later typically do not lag behind by 30 credits but rather by a smaller number (even lagging behind by a single exam will cause a full semester graduation delay).

5 Heterogeneity I: effects of information about an above average rank

The overall effects presented so far provide evidence that relative performance feedback helps students graduate earlier and that their behavioral response is heterogeneous. In the following sections, we therefore investigate the heterogeneity of the treatment effects.

5.1 Empirical approach

The starting point of our heterogeneity analyses is the question whether relative performance feedback is more effective when it provides information about an above-average rank in the pre-treatment credit point distribution.

There are two reasons to expect this. First, in the short run, the relative feedback intervention increased performance only among treated subjects who performed above the first semester average (Brade, Himmler and Jäckle, 2022). A regression discontinuity design based on the sharp change in the type of feedback at the average provided evidence that the heterogeneous behavioral responses were not caused by differences in the unobserved characteristics of those who placed above or not above average. Instead, the results indicated that it was the information about being above average that caused the effects. Furthermore, the short-run findings did not indicate that the normative frames included in the feedback mattered for the effects on performance.

Second, as discussed in Section 2, the literature on ordinal ranks shows that higher ordinal ranks can improve academic achievements over long periods of time. Based on these considerations, we hypothesize that the differential effects from the short-term analysis will persist in the long-run, and that relative feedback which conveys information about a high rank (= placing above average) will be more effective in the long-term than feedback which informs about a low rank (= not placing above average).

To examine heterogeneous effects, we estimate the following specification:

$$Y_i^k = \alpha_0 + \alpha_1 T_i + \alpha_2 A_i + \alpha_{12} T_i A_i + \mathbf{s_i} \boldsymbol{\alpha_3} + \mathbf{x_i} \boldsymbol{\alpha_4} + \mathbf{z_i} \boldsymbol{\alpha_5} + \varepsilon_i, \tag{2}$$

where A_i indicates ranking above the first semester average, and all other parameters are defined as in Equation 1. It is important to note that we base our above- and not-above-average categorization only on the pre-treatment performance distribution. The reason is that since we found positive effects of feedback (see last section), the rank in the credit point distribution (i.e., being above average or not) in later semesters is not orthogonal to the treatment assignment.

5.2 Effects on graduation rates and credit point accumulation

Figure 5 and Table 5 show treatment effects on graduation rates for the two subgroups. For students who did not place above average, we find no evidence of changes in the graduation rate. Instead, the overall effect is driven entirely by the subgroup of students who initially ranked above the average. For them, relative performance feedback increases the eighth and ninth semester graduation rates by 8.1 to 9.0 pp ($p = 0.012$ to 0.009) and 7.7 to 8.7 pp ($p = 0.007$ to 0.005), respectively (Table 5 also shows that treatment effects for the above- and not-above-average subgroups are significantly different from each other). By the end of the thirteenth semester, the difference between treatment and control is still 3.9 to 4.4 pp ($p = 0.079$ to 0.106). Given that over 99% of treated individuals and controls in the above-average subgroup have either graduated or dropped out by this point, this indicates that the relative feedback helps these students not only to graduate faster, but also makes them more likely to obtain a degree.[14]

In line with the effects on degree attainment, Figure 6 provides evidence that credit accumulation is subject to the same heterogeneous response. Treated students who initially ranked above average earn more credits during the scheduled study duration of seven semesters: starting with an additional 2.42 credits ($p = 0.001$) in the first semester after treatment began, the difference in accumulated credits increases to 7.27 by the end of the eighth semester ($p = 0.042$).

These results provide evidence that the provision of information about relative performance has positive long-term effects on academic achievement – primarily for individuals who initially ranked above average. This confirms our hypothesis that the short-term pattern of effects reported in Brade, Himmler and Jäckle (2022) persists, and is consistent with the effects reported in the literature on learning about ordinal ranks for high-ranking individuals.

6 Heterogeneity II: predicted graduation stratification

The above findings suggest that there are no adverse effects for those who received feedback that they did not perform above average. This is somewhat contrary to what the literature on learning about ordinal ranks typically suggests. Given this discrepancy, in this section, we go beyond our initial hypotheses regarding the long-term effects of the feedback intervention and explore potential heterogeneities in more detail.

Extended heterogeneity analysis can yield insightful results (see, e.g., Smith (2022) for a discussion): i) With respect to underlying mechanisms behind differential responses; thereby facilitating more comprehensive feedback designs in future research. ii) In order to guide policy and to allow for a targeted provision of the intervention, it is important to understand the nature of the heterogeneous responses and to clearly identify those who benefit from it. This can free up resources and help support students for whom the feedback is not effective with other programs that are better suited to their needs. iii) Uncovering heterogeneities in the effects of an intervention is of particular concern

[14] This notion is supported by the effects on the dropout rate shown in Figure A.2. At the end of the thirteenth semester dropout is reduced by 3.4 pp ($p = 0.149$). The figure also shows that these effects on dropout gradually emerge over time and not in specific semesters.

in education, because differential responses may introduce inequalities in educational outcomes or amplify existing ones.

6.1 Empirical approach

We explore heterogeneities in the treatment effect along students' pre-treatment probability of graduating by the end of the eighth semester.[15] Given that we have no clear ex-ante hypotheses, our approach has the following advantages: i) It allows us to assess heterogeneities along students' academic attainment, which is ultimately the dimension that we are most interested in from a policy perspective. ii) Compared to "one-variable-at-a-time" subgroup analyses, it reduces the number of possible heterogeneity dimensions and thus concerns such as multiple hypothesis testing and selective presentation of significant results (see, e.g., Kent, Steyerberg and van Klaveren (2018) for a discussion). iii) Moreover, the reduction in dimensionality also facilitates the potential targeting of the intervention in future cohorts.

To deal with the problem that graduation is a binary outcome and that the counterfactual graduation rate of treated students is unobserved, we use control students' baseline performance and background characteristics to predict the probability of graduating by the end of the eighth semester. Specifically, we take the control group of the first cohort and estimate a logit model that includes the first semester credits, the standardized first semester GPA, background characteristics (standardized high school GPA, age, gender, time since high school graduation, and type of high school degree), and study program fixed effects as explanatory variables.[16] We then predict the eighth semester graduation probabilities for all students (treatment and control) in the second cohort, which essentially collapses the explanatory variables into a single performance index. We repeat the same exercise using the control group of the second cohort for predicting the graduation probabilities in the first cohort. By taking advantage of the two-cohort structure in this way, we prevent the "overfitting bias" that can occur when using within-cohort predictions for the control group students (Abadie, Chingos and West, 2018).[17]

In the final step, we then use the predicted probabilities to construct subgroups, which we refer to as "predicted graduation strata". Based on the heterogeneity analysis in Section 5, we first split our sample into above-average and not-above-average students. Within these groups, we then divide the predicted probabilities at the median of each cohort and study program, thereby creating four predicted graduation strata: the bottom and top half of not-above-average students (Strata 1 and 2) and the bottom and top half of above-average students (Strata 3 and 4).

Figure 7 plots the predicted graduation probabilities for students in the control and in the treat-

[15] We select the eighth semester, because most students graduate in this semester, and because it is the first semester in which the intervention has a significant overall effect on degree attainment. In addition, at this point, there is a reasonable level of differentiation between individuals who have already graduated, dropped out, and are still studying.

[16] We do not include first semester dropout and the dummy that indicates whether the first semester GPA is missing, as this would lead to the exclusion of observations from the logit model for which these two variables take the value of one.

[17] Our approach is similar to the heterogeneity analyses in Dynarski et al. (2021) and is related to "risk stratification" in medical research (Kent, Steyerberg and van Klaveren, 2018). Another related method is "endogenous stratification", in which in-sample control group data are used for the predictions (Abadie, Chingos and West, 2018).

ment group separately for the four strata. This shows that i) the predicted graduation probabilities are well-balanced between treatment and controls in all strata, and ii) in terms of graduation probabilities, not-above-average students in Stratum 2 are almost identical to above-average subjects in Stratum 3. Thus, if the response to feedback is determined by our index of background characteristics and baseline performance (the predicted graduation probability), we would expect treated students in the two middle strata to behave similarly.[18] If, instead, the type of feedback (information about an above- or not-above-average rank) matters for the response, we would expect differential responses between Stratum 2 and 3.

We report treatment effects for the four predicted graduation strata based on the following estimation equation:

$$Y_i^k = \alpha_0 + \alpha_1 T_i + \mathbf{g_i}\boldsymbol{\alpha_2} + T_i\mathbf{g_i}\boldsymbol{\alpha_{12}} + \mathbf{s_i}\boldsymbol{\alpha_3} + \mathbf{x_i}\boldsymbol{\alpha_4} + \mathbf{z_i}\boldsymbol{\alpha_5} + \varepsilon_i, \tag{3}$$

where $\mathbf{g_i}$ is a vector including dummies for the predicted graduation strata (Stratum 1 being the reference group), and all other parameters are defined as in Equation 2.

6.2 Effects on graduation rates

Estimates of the treatment effects on the graduation rate are reported in Figure 8 and Table 6. They show that within the not-above-average subgroup, those with the lowest eighth semester graduation probabilities (Stratum 1) see an increase in the probability of having graduated until the end of our observation period by 4.9 to 5.8 pp. However, this effect is not significantly different from zero on any conventional level (p = 0.220 to 0.130). Treated students in Stratum 2, on the other hand, are 8.8 to 9.4 pp (p = 0.015 to 0.022) less likely to have obtained a degree at this point compared to their control group counterparts.

Among those who ranked above average in the first semester, we find that the positive treatment effect is almost entirely due to Stratum 3, i.e., the bottom half. Relative performance feedback increases the eighth and ninth semester graduation rates of these students by 12.0 to 13.5 pp (p = 0.012 to p = 0.007) and 11.4 to 12.6 pp (p = 0.013 to p = 0.008, see Table 6), respectively. After thirteen semesters, the graduation rate is still increased by 7.4 to 8.0 pp (p = 0.061 to 0.048). These results indicate that information about an above-average performance can help students in the middle of the performance distribution catch up with the best performing students (the raw gap in the control group of 17.1 pp shrinks to 11.5 pp in the treatment group).

Of particular interest is the comparison of the treatment effects in Strata 2 and 3. Despite their virtually identical predicted graduation probabilities, the two groups respond very differently to relative performance feedback: while students in Stratum 2 respond negatively, students in Stratum 3 respond positively. This suggests that it is the content of the feedback, and not students' characteristics, that matters for their behavioral response; we consider potential mechanisms behind this in the

[18] Table A.3 provides descriptive statistics for the predicted graduation strata. The main differences between Strata 2 and 3 are that students in Stratum 2 obtained less first semester credits, that they have a better first semester GPA, are more likely to be female, and more likely to have earned their high school degree from the general track. This suggests that the general academic ability of students in Stratum 2 is not strictly worse compared to those in Stratum 3.

next section.

As Figures A.3 and A.4 show, the entire pattern of results that we find for students' graduation rates is also reflected in the effects on dropout and accumulated course credits. In addition, the Figures provide evidence that this pattern starts to emerge immediately after the relative feedback intervention begins in the second semester.

7 Mechanisms behind the pattern of results

The heterogeneous pattern of results in the last sections raises two questions: i) What mechanism is behind the asymmetric behavioral response to above- and not-above-average feedback, particularly among the responsive students in the middle of the predicted graduation probability distribution? We consider whether competitive preferences and self-perception theory – often believed to drive the effects of relative performance feedback – can be aligned with the differential results. In addition, we qualify the theoretical considerations with survey evidence. ii) How can we explain the lack of treatment effects among non-responsive students, i.e., those with the lowest and highest predicted graduation probabilities? We suspect that the low (high) baseline performance in these two groups leaves little room for further deterioration (improvement), and that the top performers may also be constrained by institutional barriers that contribute to the ceiling effects.

7.1 Responsive students

Competitive preferences. Competitive preferences are commonly regarded as one of the mechanisms driving effects of relative performance feedback. The idea is that individuals gain utility not only from absolute but also from relative performance. Under these assumptions, relative feedback can have positive effects because it increases the accuracy of expectations about the performance of others. This in turn increases the weight individuals put on the competitive part of their utility function. The result is higher effort and performance across the entire distribution (Azmat and Iriberri (2010) and Dobrescu et al. (2021) apply this to the context of education and provide empirical results consistent with the theory). However, when applied to our pattern of results, the model cannot explain the negative treatment effects for students in Stratum 2, and it thus seems unlikely that competitive preferences are the central mechanism behind the effects in our study.

In addition, we present survey evidence on perceived competition among students and how it is affected by relative feedback (Figure 9).[19] We asked students *To what extent do you agree with the following statement about your studies: When thinking about my studies, I think of competition among students* (answer categories: completely disagree (= 1) to completely agree (= 7), and no answer). The figure offers the following insights: First, it shows that perceived competition among students is on average rather low in our context (the overall mean is 3.01; SD=1.71). Second, a statistically significant increase in perceived competition can only be observed for the non-responsive students in Stratum 4 (0.45 points, $p = 0.044$); we come back to this result in the next section.

[19] Perceived competition by pre-treatment rank is shown in Figure A.5.

To further investigate whether students in our context generally exhibit low levels of competitiveness and how this depends on the performance dimension, we asked a later cohort of incoming students in 2020 at the same university the following question: *This question refers to the goals you pursue with regard to your degree. To what extent do you agree with the following statements?* a) *It is important to me to graduate sooner than my fellow students* and b) *It is important to me to achieve a better final grade in my studies than my fellow students* (answer categories: completely disagree (= 1) to completely agree (= 7), and no answer). Among students enrolled in the five study programs that were used for the feedback intervention, the average answer to a) was 2.94 and only 23.14% chose category five or higher ($N = 242$), while the average answer to b) was 4.24 and 50% responded with category five or higher ($N = 244$). A t-test on the equality of means confirms that the answers to the two questions are significantly different from each other ($p \leq 0.000$). This suggests that students are not particularly competitive with respect to time to degree, especially when compared to their competitiveness in terms of final grades. We give feedback on credits per semester, which directly translates into time to degree, and so it makes sense that we do not find results consistent with competitive preferences. In addition, the lack of competitiveness plausibly explains the difference in our results from Azmat and Iriberri (2010) and Dobrescu et al. (2021), who study feedback on grades and performance in online assignments and find performance gains across the entire distribution.

Self-perception theory. In contrast to competitive preferences, self-perception theory supposes that utility depends only on one's own performance, and that individuals have incomplete information about their ability and therefore use task performance for inference (Azmat and Iriberri, 2010; Dobrescu et al., 2021; Ertac, 2005). Since performance also depends on the difficulty of the task, its signal about ability is noisy and relative feedback helps to evaluate the difficulty of the task. Intuitively, if one receives the signal that many peers performed well, it increases the likelihood that the task was easy, which in turn decreases the probability that one's own ability is high, and vice versa. Under the assumption that ability and effort are complements, a downward (upward) adjustment in beliefs about one's own ability leads to a reduction (increase) in effort and performance. Applied to our context, it is plausible that treated students adjust beliefs about ability upwards relative to controls when they receive information that they performed above average, and downwards when feedback informs them about a not-above-average performance.[20] The differential updating of beliefs then leads to the observed negative or positive effects on performance that we observe for students in Strata 2 and 3.

To qualify this notion, Figures 10 and A.6 show how our relative performance feedback changed expectations about future performance based on this survey question that we asked students in the fifth (Cohort I) and the second and third semester (Cohort II): *What do you think? What percentage of your fellow students will have obtained more credit points (ECTS) than you at the end of the current*

[20] This simplified interpretation assumes that students process feedback in a rather discrete way, i.e., the exact distance to the average performance is only of minor importance compared to the information about being above or not above average. This is supported by the findings in Brade, Himmler and Jäckle (2022), where an RDD provides evidence that receiving above- instead of not-above-average feedback leads to a large increase in second semester performance for students around the average.

semester. We categorize students as expecting an above average performance if they answer that less than 50 percent of their fellow students will achieve more credits. Two important results emerge: First, and consistent with the literature on ordinal ranks, we find that above-average feedback improves expectations to subsequently perform above average by 12 pp ($p = 0.048$, Figure A.6).[21] This effect is driven by students from both above-average graduation strata (Figure 10).

Second, the pattern of results for Strata 2 and 3 shown in Figure 10 can be well aligned with the theoretical considerations and the heterogeneous effects on academic achievement that we find. For controls, expectations about future relative performance are quite similar, with 65% (Stratum 2) and 68% (Stratum 3) of students expecting that they will perform better than the average. For treated students, on the other hand, there is a big gap in expectations: only 54% of treated not-above-average students in Stratum 2 expect to perform above average, while this share rises by 25 pp ($p = 0.024$) to 79% among treated above-average students in Stratum 3. Assuming that the expectations about relative performance are indicative of students' beliefs about their own ability, the survey results thus suggest that mechanisms that relate perceived abilities to subsequent performance can explain the pattern of effects that we find for students in the middle of the performance distribution.

These results are consistent with findings in the literature on ordinal ranks. As noted in Section 2, it is often argued that rank induced changes in self-confidence, perceived intelligence and abilities, expectations about future grades, and motivation are the mechanism underlying the long-lasting effects. In addition to the effects on expected performance described above, we also find tentative evidence that students who receive information that they ranked above average feel more positive about their performance in terms of accumulated credits (Figure A.8). Treated students who placed above average score 0.38 points (about 0.135 SD) higher on the satisfaction scale ($p = 0.092$).

For students in Stratum 2, our finding of lower graduation and higher dropout rates can also be well aligned with studies that investigate how learning about academic performance influences college dropout. For instance, the results by Stinebrickner and Stinebrickner (2012, 2014) suggest that dropout in the first two years of college would be reduced by up to 45% if no learning about academic performance occurred. Because relative feedback facilitates learning about abilities, it may consequently also induce dropouts, which is exactly what we find for those who are arguably most likely to revise their beliefs about academic abilities downward: the students in Stratum 2 (see Figure A.3). That their counterparts in the control group go on to earn a degree suggests that facilitating learning about academic abilities can lead to adverse outcomes for a considerable fraction of students.

[21] Figure A.7 shows effects on expectations by pre-treatment rank separately across the three semesters in which we conducted the survey (for a split by predicted graduation and semesters the number of observations is insufficient). The results for the second and third semester (Cohort II) replicate the finding in Brade, Himmler and Jäckle (2022) that expectations of control and treated students who ranked above average pre-treatment appear to converge in the third semester (coefficients are slightly different because the results presented here are based on full sample regressions that control for study program FE). However, in the fifth semester (Cohort I), the gap in expectations between treated and controls once again widens.

7.2 Non-responsive students

Although it is difficult to pinpoint the underlying reasons for the absence of treatment effects among the non-responsive students in Strata 1 and 4, we argue below that floor and ceiling effects provide a plausible explanation.

Floor effects leave little room for lower performance in Stratum 1. In accordance with the self-perception theory, we would expect that students in Stratum 1 lower their performance in response to feedback. However, this assumes that there is room to adjust beliefs and performance downward – which may not be the case for these students. The following descriptives illustrate the point: i) About 50% of these students have earned 15 credits or less in their first year (see the top left panel in Figure A.9), which is only a quarter of the credits recommended by the study curriculum. ii) About 60% of them have dropped out of their program by the end of the fourth semester (see Figure A.3). These numbers indicate that the baseline performance in this group is already extremely low, leaving little room for negative treatment effects and making it plausible that many of these students are well aware of their low performance.

Ceiling effects prevent improvements in Stratum 4. Figure 10 showed that treated students at the upper end of the predicted graduation probabilities, i.e., those in Stratum 4, also expect a higher relative performance compared to their control group counterparts, suggesting that feedback changes perceived ability among these students, too. In addition, Figure 9 provided evidence that these students also perceive the study environment as more competitive after receiving feedback. Both results raise the question why we do not observe positive treatment effects on performance for these students.

We argue that the performance of control group students in the top stratum is already so high that treated students have little room to improve it even further. First, this is supported by the results depicted in Figure 8. One year after the scheduled study duration, 84.8% of controls have graduated, and by the end of our observation period this number has increased to 92.4%.

Second, we present additional evidence of ceiling effects in Figure A.10. It plots the cumulative distributions of the accumulated credit points for semesters two through seven for students in the two strata receiving above-average feedback (cumulative distributions for the two not-above-average strata are shown in Figure A.9). Irrespective of their treatment status, a large share of students in Stratum 4 obtains about 30 credits per semester – i.e., the number of accumulated credits that they are expected to earn according to the study curriculum.[22] This suggests two things: i) The study curriculum may set a reference point that students aim to adhere to and are unwilling to exceed. ii) Related to the first point, the fact that study programs, courses, and lectures are structured and scheduled in accordance with the official study duration of seven semesters makes it difficult for students to take courses from later semesters earlier, i.e., they cannot easily take additional courses. As a result, the top performers may have little opportunity to complete their study program faster. A comparison to

[22]The striking jump in the cumulative distribution at approximately 195 credits in the seventh semester is due to the fact that many students do not manage to submit or pass their bachelor's thesis in time to graduate in the seventh semester.

the cumulative distributions of students in Stratum 3 additionally illustrates this: for them the treatment effects are almost exclusively generated in parts of the distribution that lie below the assumed ceiling. From an educational policy perspective, it might therefore be difficult to support the best performing students through relative feedback or other measures without removing these structural barriers.

8 Spillovers to other domains

In the preceding sections of this paper, we have shown that relative performance feedback in terms of accumulated credits can increase performance on this dimension and reduces time to graduation. An important concern is that these gains may be bought with losses in other dimensions. First, it is conceivable that focusing on graduating earlier may result in students achieving poorer grades. This would be suggestive of deteriorating competence levels, and it could also reduce labor market prospects (Piopiunik et al., 2020). Second, relative performance feedback and the faster graduation associated with it may reduce students' satisfaction with their studies, increase stress, or may reduce satisfaction with life in general, because students may, for example, have less time for other activities that they enjoy (Reiff et al., 2022).

8.1 Grade point average

One potential issue regarding the GPA is that it is only observed for students who have earned at least one passing grade. In Columns (1) to (3) of Table 7, we therefore first assess whether our treatment affects the likelihood that the GPA is unobserved by the end of the thirteenth semester (this is the case for students who never obtained a passing grade). We find no statistically significant difference between treatment and control. Figure A.13 shows that this holds over the entire study duration, and we are therefore not concerned that our treatment leads to differential selection.

Results for the treatment effects on students' GPA over the whole study duration and by the end of the thirteenth semester are shown in Figure 11 and in Columns (4) to (6) of Table 7, respectively. We use the standardized and reverse scaled GPA where higher values represent better grades.[23] In contrast to the notion that treated students may buy faster graduation with worse grades, we find that the GPA of treated students is about 0.062 SD better ($p = 0.027$) from the end of the fourth semester onward (Figure 11; note that the last observed GPA of dropouts and graduates is carried forward). Without controlling for baseline performance and additional covariates, the difference is 0.10 SD at the end of our observation period (Table 7). This is mainly due to the omission of the baseline GPA, which is slightly – but not statistically significantly – different between treated and controls (see Table 1).

The heterogeneous effects on GPA generally follow a pattern similar to the heterogeneities that we find for the other performance dimensions. Figure A.11 shows that by the end of the observation

[23] On the original German scale the best passing grade is 1.0 and the worst passing grade is 4.0, but this may be confusing for non-German readers.

period the positive overall effect on GPA is mostly driven by the above-average subgroup. However, we also find positive coefficients for the not-above-average subgroup, in particular in the first semesters after the introduction of feedback. Panel (b) of Figure A.12 shows that the effects of above-average feedback are entirely driven by Stratum 3. Treated students in this Stratum have earned a 0.173 SD (p = 0.001) better GPA than the controls by the end of the thirteenth semester. In the not-above-average subgroup, we find little evidence for heterogeneous effects on GPA (Panel (a) of Figure A.12).[24]

In sum, these findings provide evidence that the relative feedback intervention not only decreases time to graduation, but that it also improves GPA. This suggests that the positive effects of feedback are not driven by a pure maximization of the incentivized outcome dimension, i.e., accumulated credit points. Rather, as argued by the rank literature, it seems plausible that the information about an above-average rank that drives the overall effects, increases self-confidence, motivation, and effort, which raise performance both on the quantitative and qualitative dimension.

8.2 Well-being

We investigate effects on well-being with survey data. We consider the following questions: *During the last weeks, how often did you feel stressed out by your studies?* (answer categories: never (=0), rarely, sometimes, often, very often, always (=5), and no answer), *How satisfied are you currently with your studies, all things considered?*, and *How satisfied are you currently with your life, all things considered?* (answer categories: completely dissatisfied (=0) to completely satisfied (=10), and now answer). In Figure 12, we show treatment effects on a standardized inverse-covariance weighted average of these three outcomes (Anderson 2008, Schwab et al. 2020; feeling stressed out by studies is reversed before, so that for all variables higher values indicate "better" outcomes. Figures A.17 and A.18 depict estimates for the individual outcomes).

We find no strong evidence that relative performance feedback reduces or increases well-being. If anything, the estimates tentatively suggest that relative feedback improves the well-being of students in Stratum 1 and Stratum 4 – i.e., the two strata for which we found the least amount of evidence for effects on academic performance – by 0.25 and 0.26 SD (p = 0.181 and 0.064), respectively.

9 Conclusion

Based on two natural field experiments in higher education, this paper shows that relative performance feedback can be an effective tool for policymakers and organizations to support individuals in completing challenging long-term tasks. Students who received ongoing relative feedback on their accumulated course credits – i.e., their progress toward degree attainment – graduate 0.15 semesters (0.12 SD) earlier compared to controls. Crucially, we show that this does not come at the cost of worse quality or reduced well-being; quite the opposite, treated students' GPA increases by 0.063 standard deviations. These results emphasize the importance of repeated provision of behavioral interven-

[24] Figures A.14 and A.15 show that effects on the likelihood that the GPA is missing for the subgroups. Except for students in Stratum 1, for whom we find some evidence for differences between treated and controls, we do not see any evidence for potential differences in the likelihood to earn at least one passing grade.

tions and the collection of outcome data over long periods of time, as the short-term results in Brade, Himmler and Jäckle (2022) did not indicate that the intervention would have such a substantial long-term impact.

Heterogeneity analyses reveal that relative performance feedback also has important distributional implications that policymakers and organizations should be aware of. We find that students in the middle of the predicted pre-treatment graduation probability distribution are most responsive: relative feedback helps these students catch up with the best performers, when it informs them that they placed above average pre-treatment, but it reduces academic attainment among those who did not place above average. Survey evidence supports the notion that learning about abilities is the underlying mechanism behind this pattern of effects.

Moreover, our results indicate that for the best performing students the European institutional setting may leave little scope for relative feedback to improve performance. This has two important implications: i) It may generally be difficult to improve the performance of these students without removing structural barriers that make graduating before the scheduled study duration difficult. ii) In settings without institutional ceilings, the positive impact of relative performance feedback may be even larger.

References

Abadie, Alberto, Matthew M Chingos, and Martin R West. 2018. "Endogenous Stratification in Randomized Experiments." *Review of Economics and Statistics*, 100(4): 567–580.

Allcott, Hunt. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics*, 95(9): 1082–1095.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review*, 104(10): 3003–3037.

Altmann, Steffen, Andreas Grunewald, and Jonas Radbruch. 2022. "Interventions and Cognitive Spillovers." *Review of Economic Studies*, 89(5): 2293–2328.

Anderson, Michael L. 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects." *Journal of the American Statistical Association*, 103(484): 1481–1495.

Ashraf, Anik. 2022. "Performance Ranks, Conformity, and Cooperation: Evidence from a Sweater Factory." *CESifo Working Paper No. 9591.*

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Household Decision Making and Savings Impacts: Further Evidence from a Commitment Savings Product in the Philippines." *mimeo.*

Ashraf, Nava, Oriana Bandiera, and Scott S. Lee. 2014. "Awards Unbundled: Evidence from a Natural Field Experiment." *Journal of Economic Behavior & Organization*, 100: 44–63.

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

Azmat, Ghazala, Manuel Bagues, Antonio Cabrales, and Nagore Iriberri. 2019. "What You Don't Know… Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education." *Management Science*, 65(8): 3449–3947.

Barankay, Iwan. 2012. "Rank Incentives: Evidence from a Randomized Workplace Experiment." *mimeo.*

Behlen, Lars, Oliver Himmler, and Robert Jaeckle. 2022. "Can Defaults Change Behavior When Post-Intervention Effort is Required?" *MPRA Working Paper No. 113398.*

Bertoni, Marco, and Roberto Nisticò. 2023. "Ordinal Rank and the Structure of Ability Peer Effects." *Journal of Public Economics*, 217: 104797.

Beshears, John, and Harry Kosowsky. 2020. "Nudging: Progress to Date and Future Directions." *Organizational Behavior and Human Decision Processes*, 161: 3–19.

Blader, Steven, Claudine Gartenberg, and Andrea Prat. 2020. "The Contingent Effect of Management Practices." *Review of Economic Studies*, 87(2): 721–749.

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

Blattman, Christopher, Nathan Fiala, and Sebastian Martinez. 2020. "The Long-Term Impacts of Grants on Poverty: Nine-Year Evidence from Uganda's Youth Opportunities Program." *American Economic Review: Insights*, 2(3): 287–304.

Bound, John, and Sarah Turner. 2011. "Dropouts and Diplomas: The Divergence in Collegiate Outcomes." In *Handbook of the Economics of Education.* Vol. 4, , ed. Stephen Machin, Eric A. Hanushek and Ludger Woessmann, 573–613. Elsevier.

Brade, Raphael, Oliver Himmler, and Robert Jäckle. 2022. "Relative Performance Feedback and the Effects of Being Above Average – Field Experiment and Replication." *Economics of Education Review*, 89: 102268.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, et al. 2016. "Evaluating Replicability of Laboratory Experiments in Economics." *Science*, 351(6280): 1433–1436.

Carrell, Scott E., and Michal Kurlaender. forthcoming. "My Professor Cares: Experimental Evidence on the Role of Faculty Engagement." *American Economic Journal: Economic Policy*.

Castleman, Benjamin L., and Lindsay C. Page. 2016. "Freshman Year Financial Aid Nudges: An Experiment to Increase FAFSA Renewal and College Persistence." *Journal of Human Resources*, 51(2): 389–415.

Chen, Jie, Loretti Dobrescu, Gigi Foster, and Alberto Motta. 2021. "Can Rank Feedback Avoid Demoralization? A Randomized Controlled Trial." *mimeo*.

Clark, Damon, David Gill, Victoria Prowse, and Mark Rush. 2020. "Using Goals to Motivate College Students: Theory and Evidence from Field Experiments." *Review of Economics and Statistics*, 102(4): 648–663.

Czibor, Eszter, David Jimenez-Gomez, and John A List. 2019. "The Dozen Things Experimental Economists Should Do (More of)." *Southern Economic Journal*, 86(2): 371–432.

Delaney, Judith, and Paul J. Devereux. 2022. "Rank Effects in Education: What Do We Know So Far?" In *Handbook of Labor, Human Resources and Population Economics.* , ed. Klaus F. Zimmermann. Springer.

DellaVigna, Stefano, and Elizabeth Linos. 2022. "RCTs to Scale: Comprehensive Evidence from Two Nudge Units." *Econometrica*, 90(1): 81–116.

Deming, David, and Susan Dynarski. 2010. "College Aid." In *Targeting Investments in Children: Fighting Poverty When Resources are Limited.* , ed. Phillip B. Levine and David J. Zimmerman, 283–302. University of Chicago Press.

Denning, Jeffrey T., Richard Murphy, and Felix Weinhardt. forthcoming. "Class Rank and Long-Run Outcomes." *Review of Economics and Statistics*.

Dobrescu, Loretti Isabella, Marco Faravelli, Rigissa Megalokonomou, and Alberto Motta. 2021. "Relative Performance Feedback in Education: Evidence from a Randomised Controlled Trial." *Economic Journal*, 131(640): 3145–3181.

Dynarski, Susan, Aizat Nurshatayeva, Lindsay C. Page, and Judith Scott-Clayton. 2023. "Addressing Non-Financial Barriers to College Access and Success: Evidence and Policy Implications." In *Handbook of the Economics of Education.* Vol. 6, , ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 319–403. Elsevier.

Dynarski, Susan, CJ Libassi, Katherine Michelmore, and Stephanie Owen. 2021. "Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students." *American Economic Review*, 111(6): 1721–56.

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

Elsner, Benjamin, Ingo E. Isphording, and Ulf Zölitz. 2021. "Achievement Rank Affects Performance and Major Choices in College." *Economic Journal*, 131(640): 3182–3206.

Eriksson, Tor, Anders Poulsen, and Marie Claire Villeval. 2009. "Feedback and Incentives: Experimental Evidence." *Labour Economics*, 16(6): 679–688.

Ertac, Seda. 2005. "Social Comparisons and Optimal Information Revelation: Theory and Experiments." *Job Market Paper, UCLA*.

Fischer, Mira, and Valentin Wagner. 2023. "Do Timing and Reference Frame of Feedback Influence High-Stakes Educational Outcomes?" *Economics of Education Review*, 94: 102379.

Goulas, Sofoklis, and Rigissa Megalokonomou. 2021. "Knowing Who You Actually Are: The Effect of Feedback on Short- and Longer-Term Outcomes." *Journal of Economic Behavior & Organization*, 183: 589–615.

Hamermesh, Daniel S. 2007. "Replication in Economics." *Canadian Journal of Economics/Revue canadienne d'économique*, 40(3): 715–733.

Hannan, R. Lynn, Gregory P. McPhee, Andrew H. Newman, and Ivo D. Tafkov. 2013. "The Effect of Relative Performance Information on Performance and Effort Allocation in a Multi-Task Environment." *Accounting Review*, 88(2): 553–575.

Hermes, Henning, Martin Huschens, Franz Rothlauf, and Daniel Schunk. 2021. "Motivating Low-Achievers – Relative Performance Feedback in Primary Schools." *Journal of Economic Behavior & Organization*, 187: 45–59.

Himmler, Oliver, Robert Jäckle, and Philipp Weinschenk. 2019. "Soft Commitments, Reminders, and Academic Performance." *American Economic Journal: Applied Economics*, 11(2): 114–42.

Hout, Michael. 2012. "Social and Economic Returns to College Education in the United States." *Annual Review of Sociology*, 38(1): 379–400.

Hunter, John E. 2001. "The Desperate Need for Replications." *Journal of Consumer Research*, 28(1): 149–158.

Kajitani, Shinya, Keiichi Morimoto, and Shiba Suzuki. 2020. "Information Feedback in Relative Grading: Evidence from a Field Experiment." *PloS one*, 15(4): e0231548.

Kent, David M., Ewout Steyerberg, and David van Klaveren. 2018. "Personalized Evidence Based Medicine: Predictive Approaches to Heterogeneous Treatment Effects." *British Medical Journal*, 363.

Kim, Soobin, John Yun, Barbara Schneider, Michael Broda, Christopher Klager, and I-Chien Chen. 2022. "The Effects of Growth Mindset on College Persistence and Completion." *Journal of Economic Behavior & Organization*, 195: 219–235.

Klausmann, Tim, Valentin Wagner, and Isabell Zipperle. 2021. "Rank Response Functions in an Online Learning Environment." *Economics Letters*, 209: 110096.

Lovenheim, Michael, and Jonathan Smith. 2023. "Returns to Different Postsecondary Investments: Institution Type, Academic Programs, and Credentials." In *Handbook of the Economics of Education*. Vol. 6, , ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, 187–318. Elsevier.

Megalokonomou, Rigissa, and Yi Zhang. 2022. "How Good Am I? Effects and Mechanisms behind Salient Ranks." *IZA Discussion Paper No. 15604*.

Morgan, Kari Lock, and Donald B. Rubin. 2012. "Rerandomization to Improve Covariate Balance in Experiments." *Annals of Statistics*, 40(2): 1263–1282.

Murphy, Richard, and Felix Weinhardt. 2020. "Top of the Class: The Importance of Ordinal Rank." *Review of Economic Studies*, 87(6): 2777–2826.

Nguyen, Tuan D., Jenna W. Kramer, and Brent J. Evans. 2019. "The Effects of Grant Aid on Student Persistence and Degree Attainment: A Systematic Review and Meta-Analysis of the Causal Evidence." *Review of Educational Research*, 89(6): 831–874.

OECD. 2019. *Education at a Glance 2019: OECD Indicators*. Paris:OECD Publishing.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science*, 349(6251): aac4716.

Oreopoulos, Philip, and Uros Petronijevic. 2013. "Making College Worth It: A Review of the Returns to Higher Education." *Future of Children*, 23(1): 41–65.

Oreopoulos, Philip, and Uros Petronijevic. 2019. "The Remarkable Unresponsiveness of College Students to Nudging And What We Can Learn from It." *NBER Working Papers No. 26059*.

Pagani, Laura, Simona Comi, and Federica Origo. 2021. "The Effect of School Rank on Personality Traits." *Journal of Human Resources*, 56(4): 1187–1225.

Patterson, Richard W. 2018. "Can Behavioral Tools Improve Online Student Outcomes? Experimental Evidence from a Massive Open Online Course." *Journal of Economic Behavior & Organization*, 153: 293–321.

Piopiunik, Marc, Guido Schwerdt, Lisa Simon, and Ludger Woessmann. 2020. "Skills, Signals, and Employability: An Experimental Investigation." *European Economic Review*, 123: 103374.

Reiff, Joseph S., Justin C. Zhang, Jana Gallus, Hengchen Dai, Nathaniel M. Pedley, Sitaram Vangala, Richard K. Leuchter, Gregory Goshgarian, Craig R. Fox, Maria Han, and Daniel M. Croymans. 2022. "When Peer Comparison Information Harms Physician Well-Being." *Proceedings of the National Academy of Sciences*, 119(29): e2121730119.

Rury, Derek, and Scott Carrell. 2020. "Knowing What It Takes: The Effect of Information About Returns to Studying on Study Effort and Achievement." *mimeo.*

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science*, 18(5): 429–434.

Schwab, Benjamin, Sarah Janzen, Nicholas P. Magnan, and William M. Thompson. 2020. "Constructing a Summary Index Using the Standardized Inverse-Covariance Weighted Average of Indicators." *Stata Journal*, 20(4): 952–964.

Smith, Jeffrey. 2022. "Treatment Effect Heterogeneity." *Evaluation Review*, 46(5): 652–677.

Statistisches Bundesamt. 2021. "Bildung und Kultur: Studierende an Hochschulen Wintersemester 2020/2021."

Stinebrickner, Ralph, and Todd Stinebrickner. 2014. "Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model." *Journal of Labor Economics*, 32(3): 601–644.

Stinebrickner, Todd, and Ralph Stinebrickner. 2012. "Learning about Academic Ability and the College Dropout Decision." *Journal of Labor Economics*, 30(4): 707–748.

Tran, Anh, and Richard Zeckhauser. 2012. "Rank as an Inherent Incentive: Evidence from a Field Experiment." *Journal of Public Economics*, 96(9): 645–650.

Trope, Yaacov, and Nira Liberman. 2010. "Construal-Level Theory of Psychological Distance." *Psychological Review*, 117(2): 440.

Villeval, Marie Claire. 2020. "Performance Feedback and Peer Effects." In *Handbook of Labor, Human Resources and Population Economics.* , ed. Klaus F. Zimmermann. Springer.

Tables and figures

Table 1: Descriptive statistics and balancing properties

	(1) Control mean (std. dev.)	(2) Treatment coefficient (robust SE)	(3) P-value
Baseline performance			
First semester credits	19.194	0.073	0.803
	(12.059)	(0.291)	
First semester dropout rate	0.024	0.015*	0.071
	(0.152)	(0.008)	
Std. first semester GPA	-0.029	0.060	0.213
	(0.980)	(0.048)	
First semester GPA N/A	0.097	-0.006	0.605
	(0.296)	(0.012)	
Std. imputed first semester GPA[a)]	-0.033	0.062	0.165
	(0.981)	(0.045)	
Additional covariates			
Age	22.466	-0.002	0.991
	(3.230)	(0.152)	
Female	0.370	-0.000	0.999
	(0.483)	(0.021)	
High school degree Abitur	0.415	0.004	0.875
	(0.493)	(0.024)	
Time since high school degree	1.257	-0.050	0.632
	(2.230)	(0.105)	
Std. high school GPA[b)]	-0.028	0.055	0.244
	(1.002)	(0.047)	
N	803	806	

Notes: a) 78 (73) control (treatment) students obtained no passing grade in the first semester. Within each cohort, we impute the values based on a linear regression of the first semester GPA on first semester credits, high school GPA, age, a female dummy, time since high school graduation, a high school degree Abitur dummy as well as study program dummies and their interaction with the other variables. *b*) For 13 (13) control (treatment) students the university had no information on the high school GPA. Within each cohort, we impute those values based on a linear regression of the high school GPA on age, a female dummy, time since high school graduation, a high school degree Abitur dummy as well as study program dummies and their interaction with the other variables. Column (1) presents the raw control group means and standard deviations. Column (2) presents the estimated coefficients of regressing the respective variable on the treatment indicator controlling for credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE. Column (3) tests the null hypothesis of no treatment effect. The (imputed) first semester GPA and the high school GPA are standardized within cohorts and study programs and reverse scaled for easier interpretation. On the original German scale the best passing grade for both is 1.0 and the worst passing grade is 4.0. High school degree Abitur refers to the German general track degree. It is one of the two main secondary school degrees in the tracked school system in Germany that qualifies students to study at a university of applied sciences; the second being the vocational track degree (Fachhochschulreife). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 2: Effect of feedback on graduation rate

	Eighth semester			Ninth semester			Thirteenth semester		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Treatment	0.046**	0.039*	0.037*	0.045**	0.038**	0.037*	0.012	0.009	0.009
	(0.021)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.018)	(0.018)
Randomization strata	yes	yes	yes	yes	yes	yes	yes	yes	yes
Baseline performance	no	yes	yes	no	yes	yes	no	yes	yes
Additional covariates	no	no	yes	no	no	yes	no	no	yes
N	1609	1609	1609	1609	1609	1609	1609	1609	1609
Control mean	0.360	0.360	0.360	0.488	0.488	0.488	0.638	0.638	0.638

Notes: Outcome variable: indicates if a student graduated before or during the respective semester; *randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 3: Effect of feedback on time to degree in semesters

	OLS			Tobit		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	-0.154**	-0.153**	-0.147**	-0.214***	-0.195**	-0.188**
	(0.071)	(0.067)	(0.066)	(0.079)	(0.072)	(0.071)
Randomization strata	yes	yes	yes	yes	yes	yes
Baseline performance	no	yes	yes	no	yes	yes
Additional covariates	no	no	yes	no	no	yes
N	1037	1037	1037	1053	1053	1053
Control mean	8.62	8.62	8.62	8.73	8.73	8.73
(Std. dev.)	(1.28)	(1.28)	(1.28)	(1.49)	(1.49)	(1.49)

Notes: Outcome variable: time to degree in semesters. OLS estimates in Columns (1) to (3) include only individuals who graduated until the end of the thirteenth semester. Tobit estimates in Columns (4) to (6) with right censoring at fourteen semesters also include students who are still enrolled in their study program at the end of the thirteenth semester. Time to degree is set to fourteen semesters for these students. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 4: Effect of feedback on accumulated credits, seventh semester

	Unconditional			Conditional on no dropout		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	2.937	2.378	2.351	2.727*	2.325	2.408*
	(2.719)	(2.584)	(2.573)	(1.611)	(1.449)	(1.416)
Randomization strata	yes	yes	yes	yes	yes	yes
Baseline performance	no	yes	yes	no	yes	yes
Additional covariates	no	no	yes	no	no	yes
N	1609	1609	1609	1093	1093	1093
Control mean	128.72	128.72	128.72	178.48	178.48	178.48
(Std. dev.)	(77.45)	(77.45)	(77.45)	(31.23)	(31.23)	(31.23)

Notes: Outcome variable: number of credits accumulated until the end of the seventh semester. The conditional estimates in Columns (4) to (6) exclude students who dropped out of their study program before or during the sixth semester. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 5: Effect of feedback on graduation rate, by pre-treatment rank

	Eighth semester			Ninth semester			Thirteenth semester		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Not above avg. *(N=856)*	0.009	-0.000	0.001	0.011	0.002	0.005	-0.014	-0.018	-0.016
	(0.024)	(0.024)	(0.024)	(0.026)	(0.026)	(0.026)	(0.028)	(0.027)	(0.027)
Above average *(N=753)*	0.090***	0.085***	0.081**	0.087***	0.081***	0.077***	0.044*	0.041*	0.039
	(0.035)	(0.033)	(0.032)	(0.031)	(0.029)	(0.029)	(0.025)	(0.024)	(0.024)
Abv. avg. - not avb. avg.	0.081*	0.085**	0.081**	0.076*	0.079**	0.073*	0.058	0.058	0.056
	(0.042)	(0.040)	(0.040)	(0.040)	(0.039)	(0.039)	(0.037)	(0.036)	(0.036)
Randomization strata	yes	yes	yes	yes	yes	yes	yes	yes	yes
Baseline performance	no	yes	yes	no	yes	yes	no	yes	yes
Additional covariates	no	no	yes	no	no	yes	no	no	yes
N	1609	1609	1609	1609	1609	1609	1609	1609	1609

Notes: Estimates based on Equation 2. *Outcome variable:* indicates if a student graduated before or during the respective semester; *randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 6: Effect of feedback on graduation rate, by predicted graduation strata

	Eighth semester			Ninth semester			Thirteenth semester		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Not above average									
Stratum 1 *(N=431)*	0.029	0.033	0.028	0.055*	0.061**	0.058*	0.049	0.058	0.055
	(0.022)	(0.023)	(0.024)	(0.030)	(0.031)	(0.031)	(0.040)	(0.038)	(0.039)
Stratum 2 *(N=425)*	-0.023	-0.031	-0.026	-0.048	-0.057	-0.048	-0.088**	-0.094**	-0.088**
	(0.042)	(0.042)	(0.041)	(0.042)	(0.041)	(0.041)	(0.038)	(0.038)	(0.039)
Above average									
Stratum 3 *(N=379)*	0.135***	0.131***	0.120**	0.126***	0.122***	0.114**	0.080**	0.077*	0.074*
	(0.050)	(0.048)	(0.048)	(0.048)	(0.046)	(0.046)	(0.041)	(0.040)	(0.039)
Stratum 4 *(N=374)*	0.037	0.037	0.041	0.039	0.039	0.040	0.003	0.003	0.004
	(0.046)	(0.044)	(0.044)	(0.035)	(0.035)	(0.035)	(0.028)	(0.028)	(0.028)
P-value F-test joint sign.	[0.112]	[0.092]	[0.147]	[0.048]	[0.029]	[0.058]	[0.015]	[0.008]	[0.016]
Randomization strata	yes	yes	yes	yes	yes	yes	yes	yes	yes
Baseline performance	no	yes	yes	no	yes	yes	no	yes	yes
Additional covariates	no	no	yes	no	no	yes	no	no	yes
N	1609	1609	1609	1609	1609	1609	1609	1609	1609

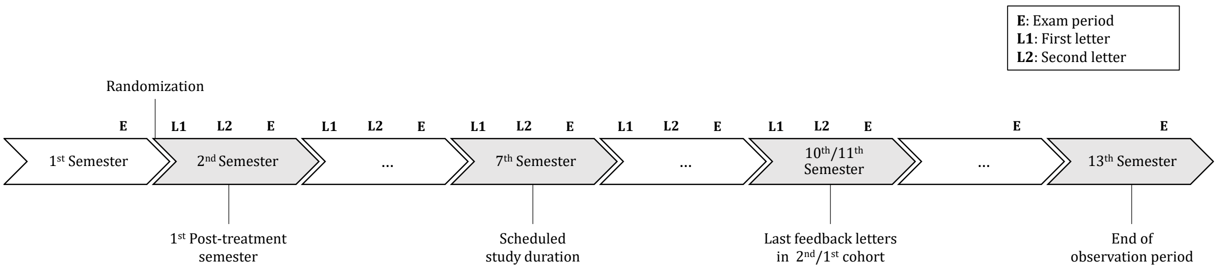
Notes: Estimates based on Equation 3. The depicted p-values are from F-tests on the joint significance of all interaction terms between the treatment indicator and the predicted graduation strata, i.e., $\boldsymbol{\alpha_{12}}$ in Equation 3. *Outcome variable:* indicates if a student graduated before or during the respective semester; *randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 7: Effect of feedback on GPA, thirteenth semester

	GPA N/A			Standardized GPA		
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.011	0.010	0.010	0.102**	0.064**	0.063**
	(0.011)	(0.007)	(0.007)	(0.045)	(0.031)	(0.030)
Randomization strata	yes	yes	yes	yes	yes	yes
Baseline performance	no	yes	yes	no	yes	yes
Additional covariates	no	no	yes	no	no	yes
N	1609	1609	1609	1508	1508	1508
Control mean	0.06	0.06	0.06	-0.06	-0.06	-0.06
(Std. dev.)	(0.23)	(0.23)	(0.23)	(1.00)	(1.00)	(1.00)

Notes: Outcome variables: Outcome variables: GPA N/A: indicates if the GPA at the end of the thirteenth semester is missing. This is the case for students who have never obtained a passing grade. *Standardized GPA:* standardized and reverse scaled GPA (higher values are better) at the end of the thirteenth semester is based on passing grades only. We do not use the original German scale where the best passing grade is 1.0 and the worst passing grade is 4.0. The GPA is unobserved for students who have never obtained a passing grade. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

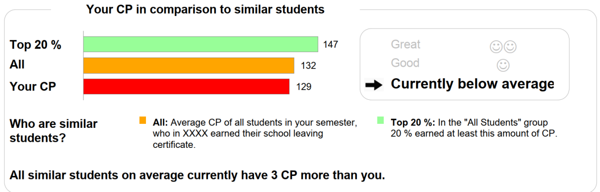
Figure 1: Timeline of intervention



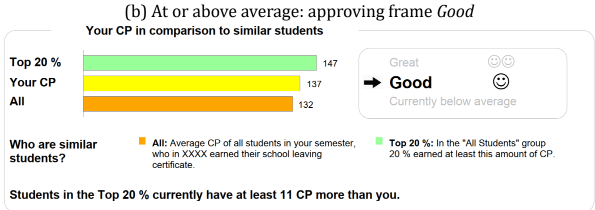
Notes: Both trials began before the COVID-19 pandemic. The last two (four) semesters of our observation period of cohort I (II) took place during the pandemic. During that time, all lectures took place online, and the exams were held partly online and partly face-to-face.

Figure 2: Relative performance feedback – treatment group (examples from sixth semester)

(a) Below average: no approving frame



(b) At or above average: approving frame *Good*



(c) Top 20%: approving frame *Great*

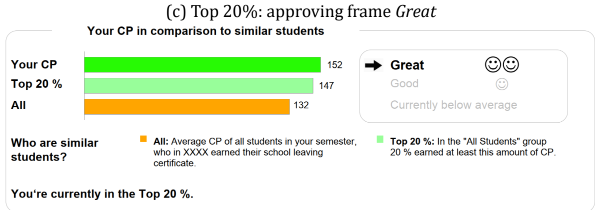
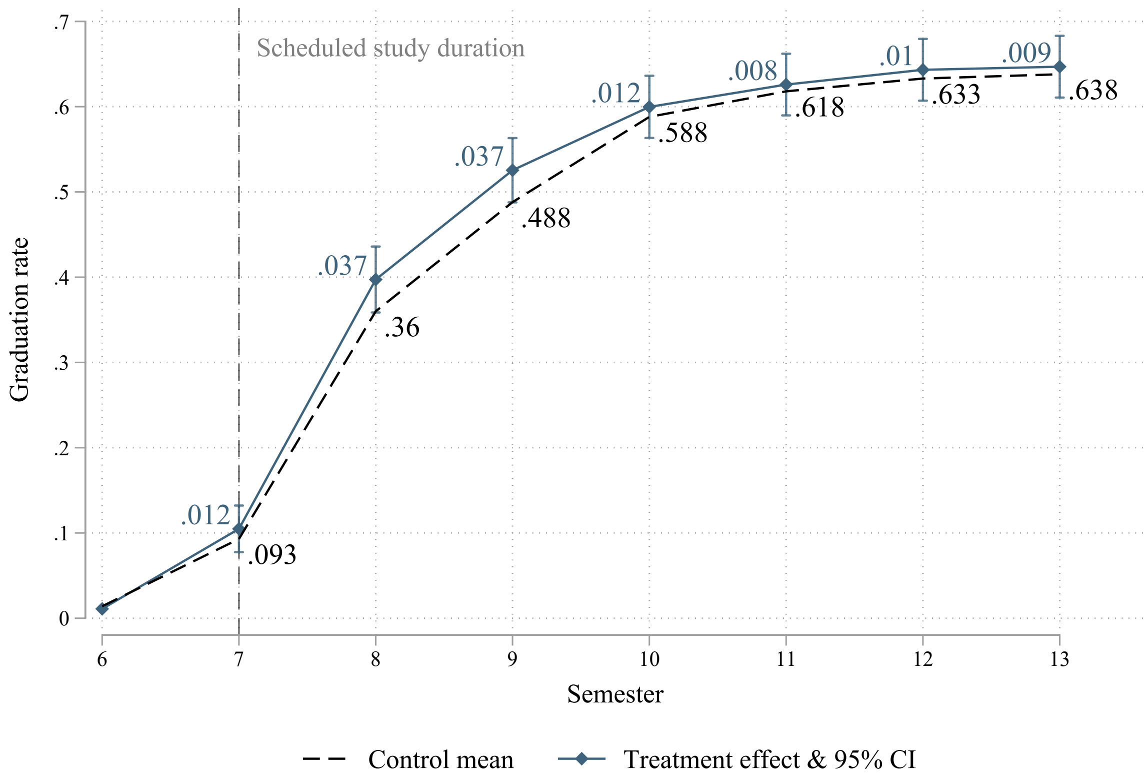
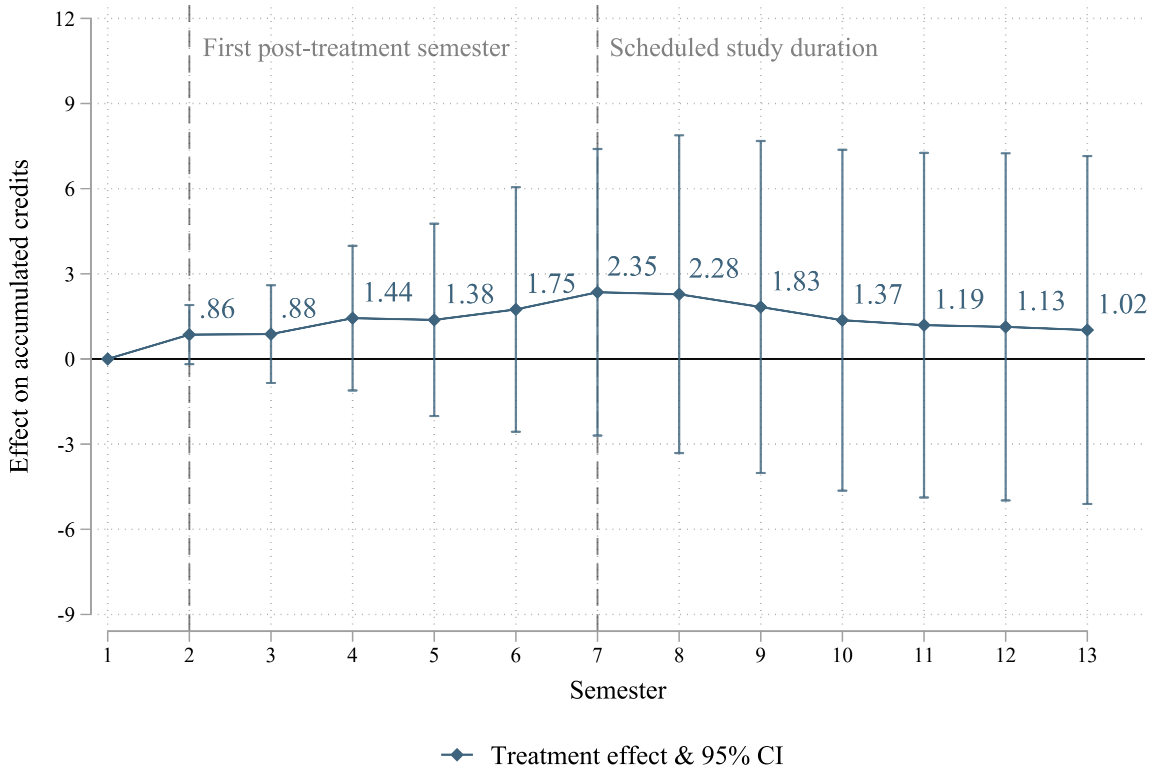


Figure 3: Effect of feedback on graduation rate



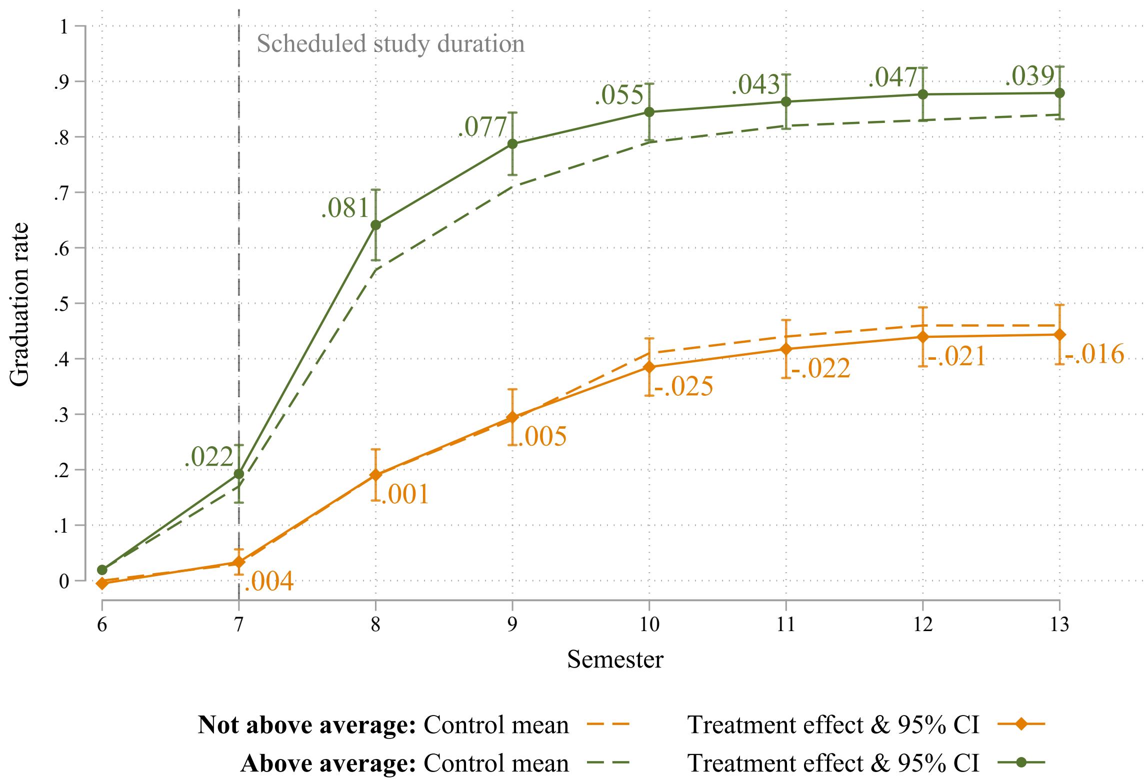
Notes: Outcome variable: indicates if a student graduated before or during the respective semester. The dashed line depicts the raw control group mean. Coefficients are from regressions based on Equation 1 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure 4: Effect of feedback on accumulated credits



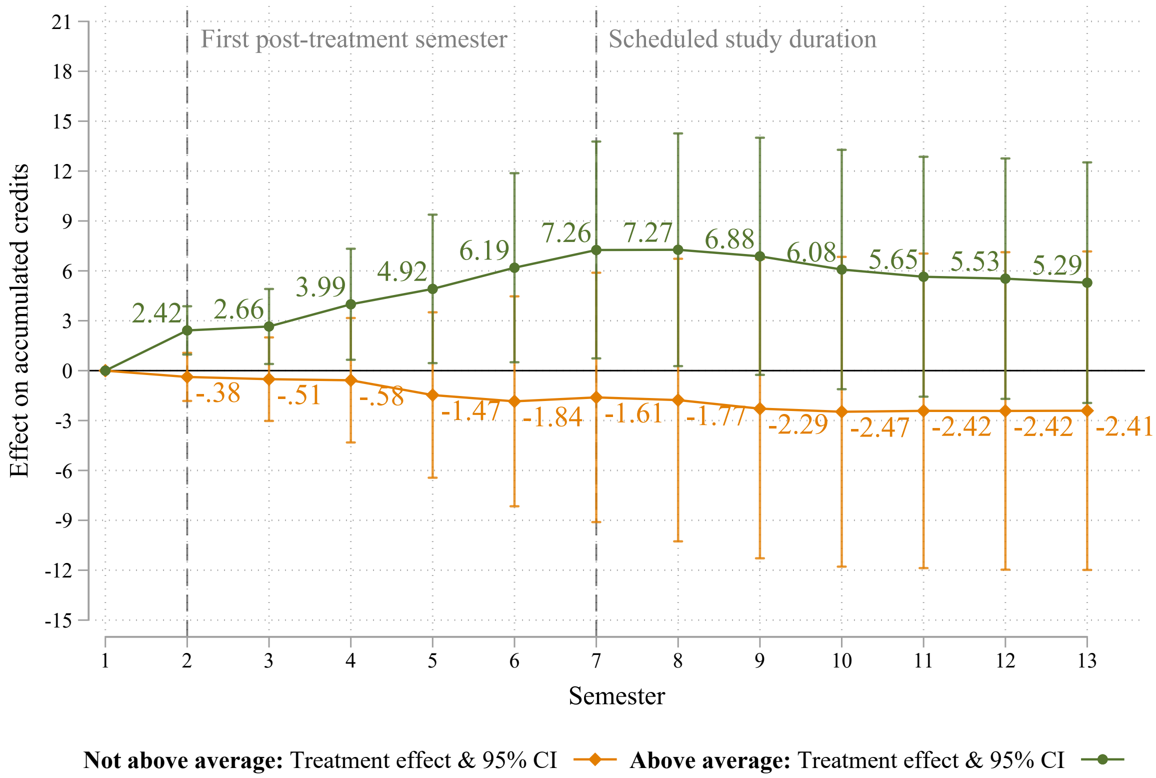
Notes: Outcome variable: number of credits accumulated until the end of the respective semester. Coefficients are from regressions based on Equation 1 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure 5: Effect of feedback on graduation rate, by pre-treatment rank



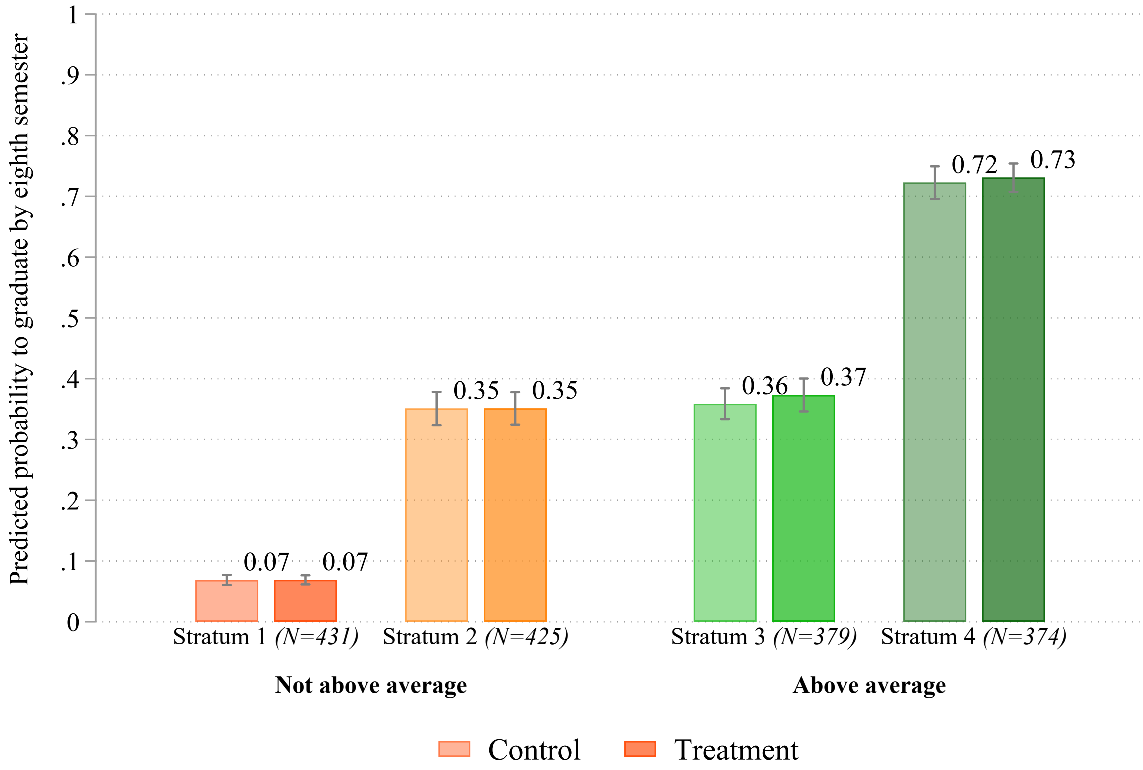
Notes: Not above average (N = 856) and above average (N = 753) refer to the rank in the pre-treatment credit point distribution. *Outcome variable:* indicates if a student graduated before or during the respective semester. The dashed lines depict the raw control group means. Coefficients are from full sample regressions based on Equation 2 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. N = 1609.

Figure 6: Effect of feedback on accumulated credits, by pre-treatment rank



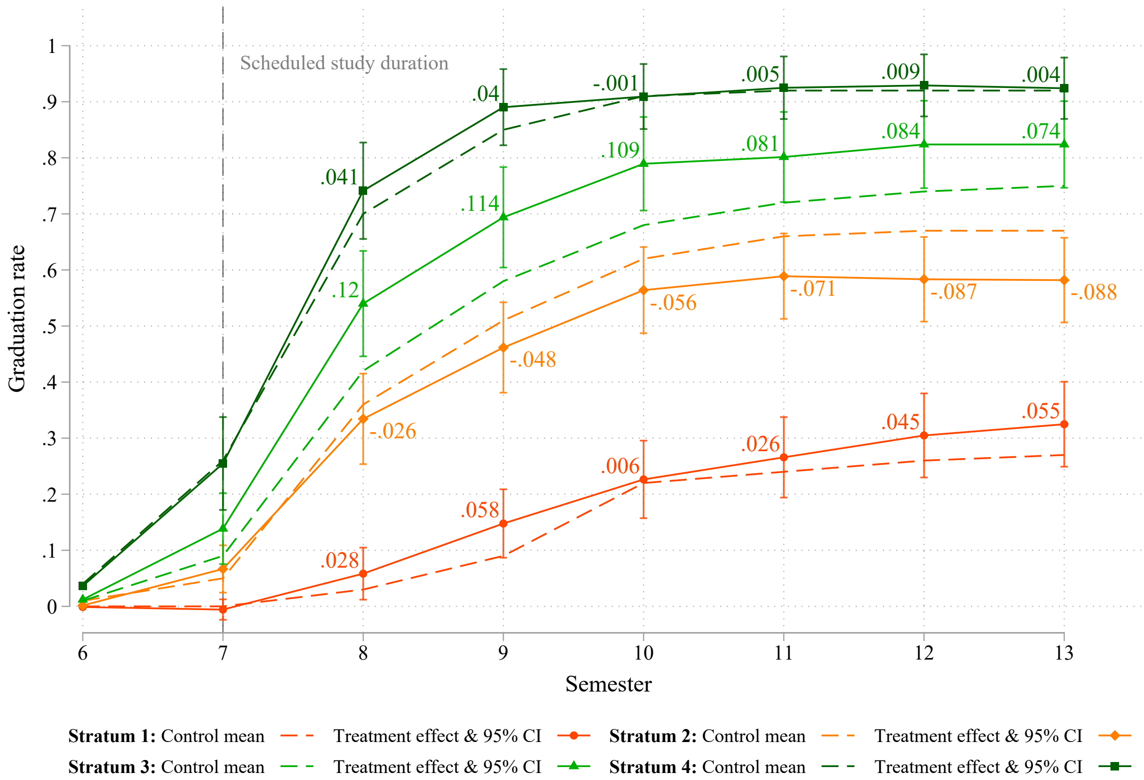
Notes: Not above average ($N = 856$) and above average ($N = 753$) refer to the rank in the pre-treatment credit point distribution. *Outcome variable:* number of credits accumulated until the end of the respective semester. Coefficients are from full sample regressions based on Equation 2 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure 7: Predicted pre-treatment probability to graduate by the end of the eighth semester, by predicted graduation strata



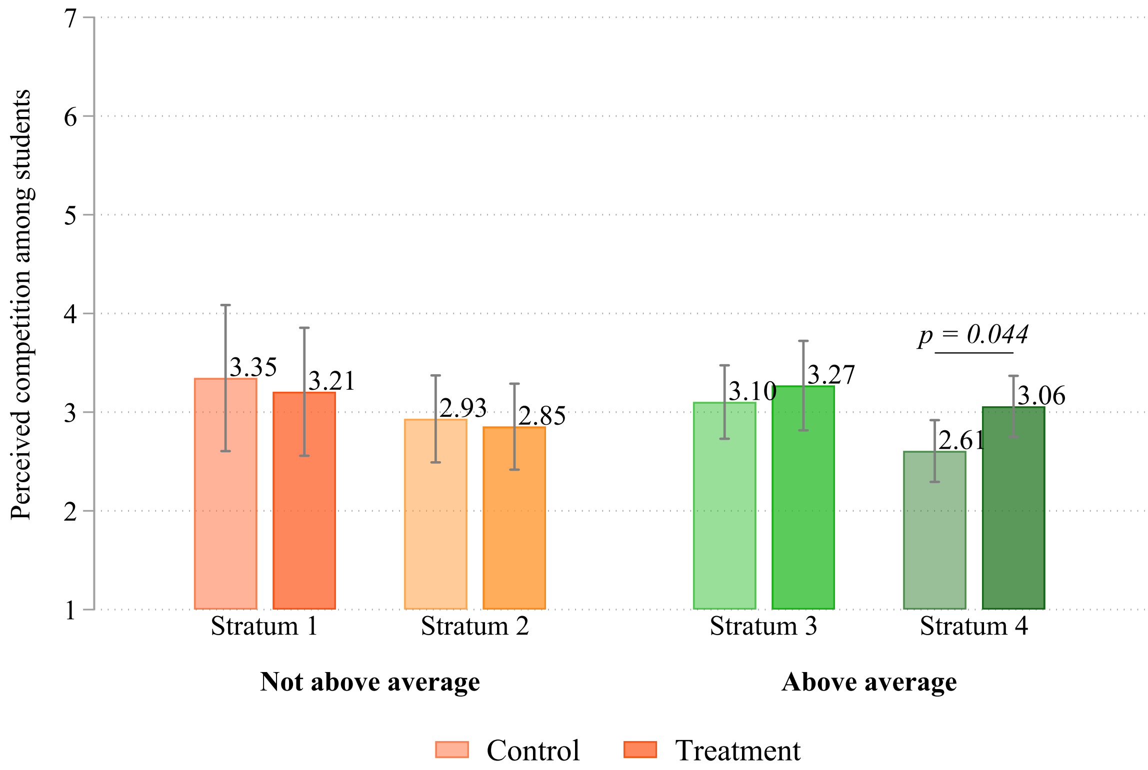
Notes: Not above average ($N = 856$) and above average ($N = 753$) refer to the rank in the pre-treatment credit point distribution.

Figure 8: Effect of feedback on graduation rate, by predicted graduation strata



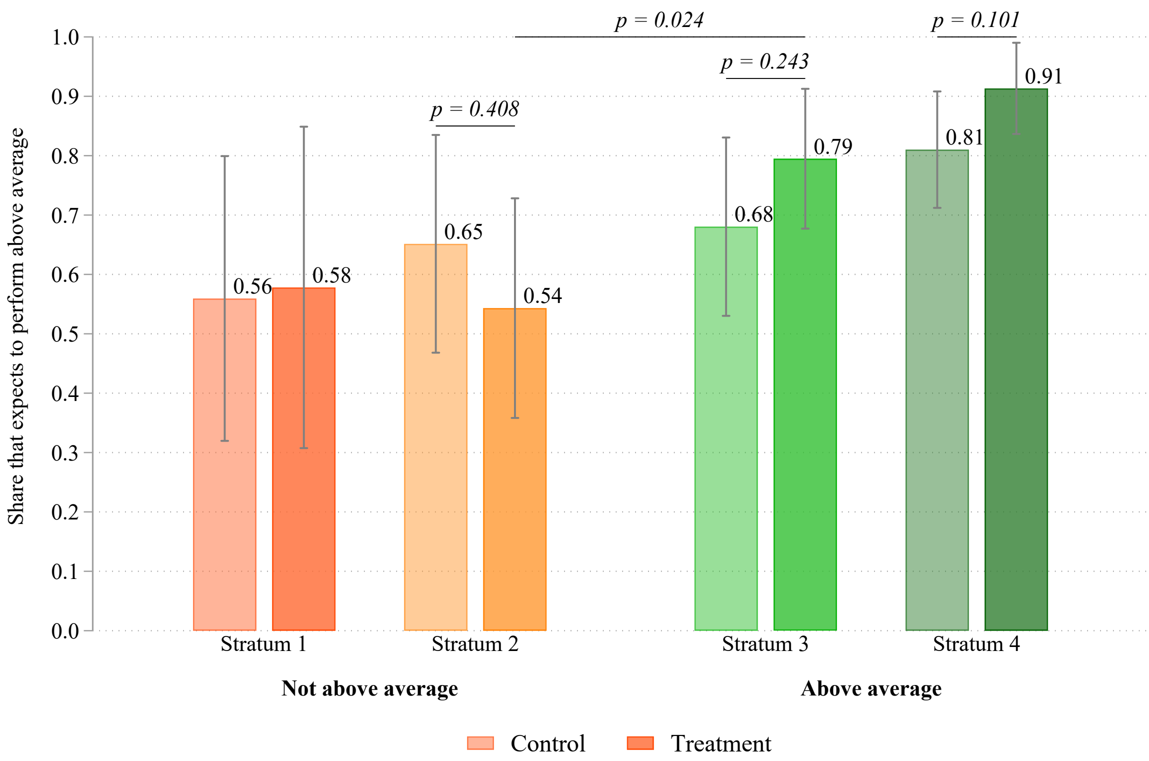
Notes: Outcome variable: indicates if a student graduated before or during the respective semester. The figure plots treatment effects for the predicted graduation Strata 1 (N = 431) and 2 (N = 425) among students who did not rank above average in the pre-treatment credit point distribution and treatment effects for the predicted graduation Strata 3 (N = 379) and 4 (N = 374) among students who ranked above average. Dashed lines depict raw control group means. Coefficients are from full sample regressions based on Equation 3 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. N = 1609.

Figure 9: Effect of feedback on perceived competition, by predicted graduation strata



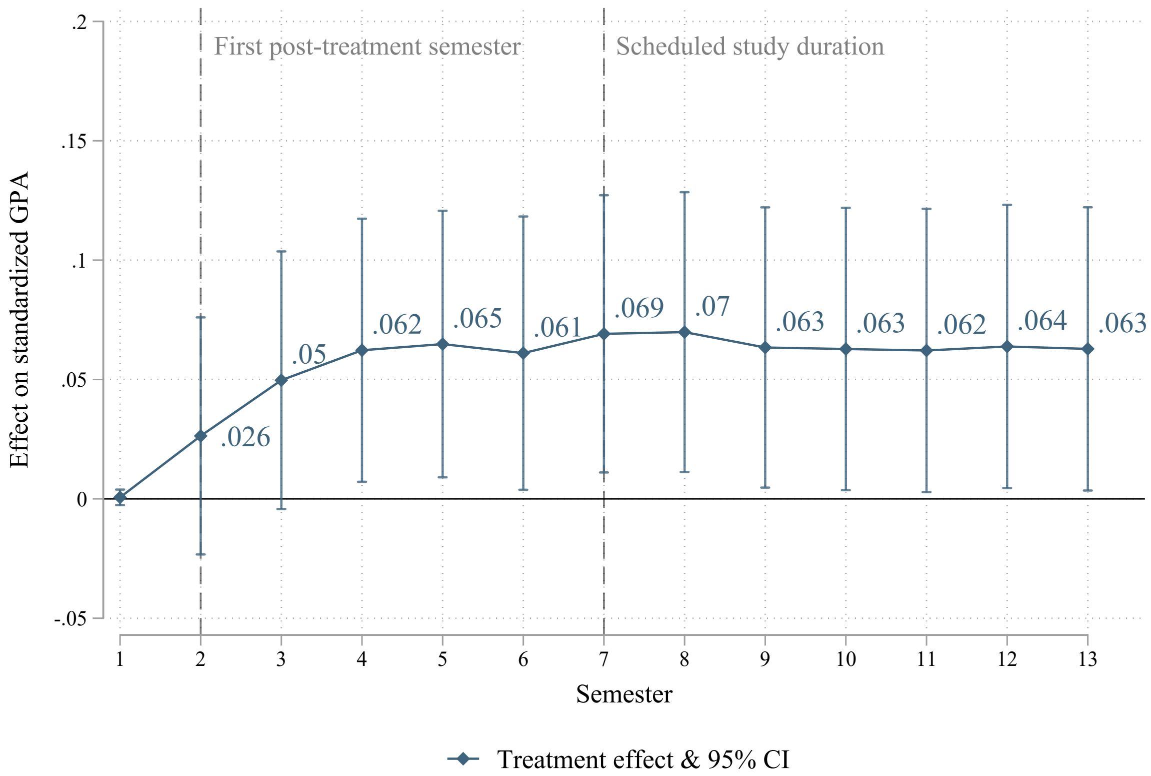
Notes: The figure plots the perceived competition among students for the four predicted graduation strata. The data on perceived competition stem from four surveys that were conducted in the second (both cohorts), the third (Cohort II), and the fifth (Cohort I) semester. In the surveys we asked students *To what extent do you agree with the following statement about your studies: When thinking about my studies, I think of competition among students.*; answer categories; completely disagree (= 1) to completely agree (= 7), and no answer. Estimates are from full sample pooled OLS regressions and control for study program FE and timing of survey FE. 95% confidence intervals are based on robust standard errors clustered at the student level. $N = 598$ from 472 individual students.

Figure 10: Effect of feedback on expectations about relative performance, by predicted graduation strata



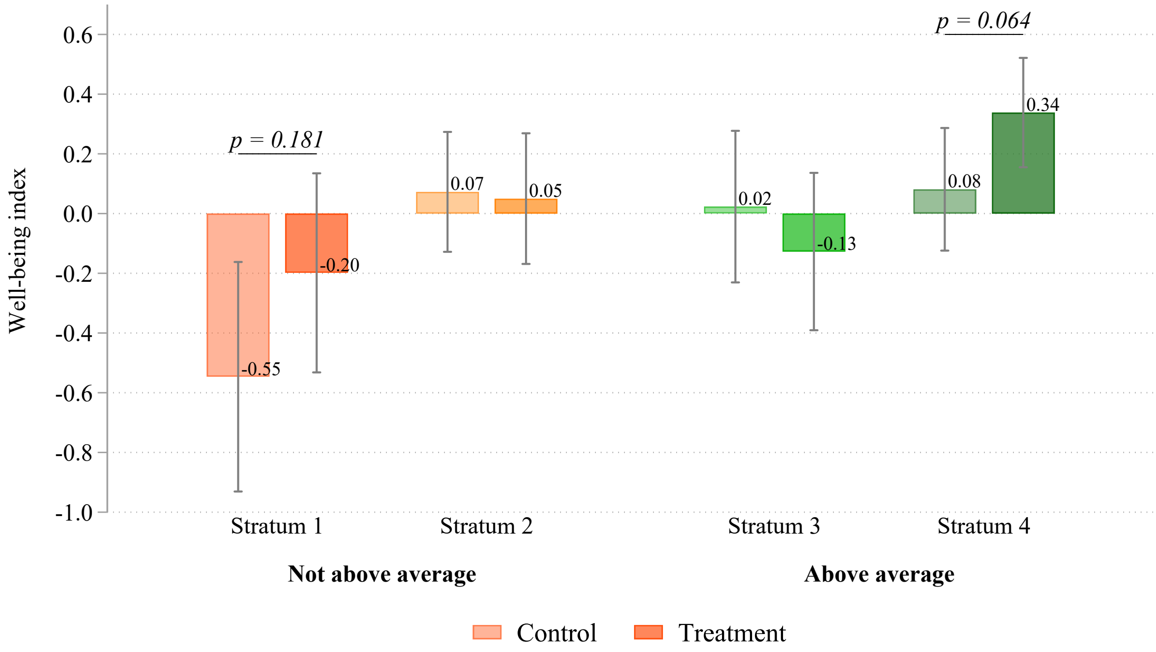
Notes: The figure plots the share of students who expects to rank above the average in the accumulated credit points distribution for the four predicted graduation strata. The data on expectations stem from three surveys that were conducted in the fifth semester (Cohort I) as well as the second and third semester (Cohort II). In the surveys we asked students *What do you think? How many per cent of your fellow students will have achieved more credit points (ECTS) than you at the end of the current semester?*; students could also provide no answer. We categorize students as expecting an above average performance if they expect that less than 50 percent of their fellow students will achieve more credits. Estimates are based on full sample pooled OLS regressions and control for study program FE and timing of survey FE. 95% confidence intervals are based on robust standard errors clustered at the student level. $N = 338$ from 288 individual students.

Figure 11: Effect of feedback on standardized GPA



Notes: Outcome variable: standardized and reverse scaled GPA (higher values are better) is based on passing grades only. We do not use the original German scale where the best passing grade is 1.0 and the worst passing grade is 4.0. The GPA is unobserved for students who have never obtained a passing grade. Coefficients are from regressions based on Equation 1 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure 12: Effect of feedback on well-being, by predicted graduation strata



Notes: The figure plots students' well-being for the four predicted graduation strata. The well-being index is the standardized inverse-covariance weighted average (following Anderson (2008) and using the Stata program by Schwab et al. (2020)) of survey questions that asked students how stressed they feel about their studies, how satisfied they are with their studies, and how satisfied they are with their life in general (see Figure A.18 for separate estimates and the survey questions that we asked). The four surveys were conducted in the second (both cohorts), the third (Cohort II), and the fifth (Cohort I) semester. Estimates are from full sample pooled OLS regressions that control for study program FE and timing of survey FE. 95% confidence intervals are based on robust standard errors clustered at the student level. $N = 612$ from 481 individual students.

Appendix

A Additional tables and figures

Table A.1: Cost calculation for relative performance feedback

Student assistant	(2 cohorts * 60 hours per semester * €12.00)	€1,440
Postage	(2 letters * €0.48 * 1609 students)	€1,544.64
Printing of letters	(2 letters * 2 pages * €0.12 * 1609 students)	€772.32
Printing of letters 2nd language	(2 letters * 2 pages * €0.12 * 246 students)	€118.08
Envelopes	(2 letters * €0.02 * 1609 students)	€64.36
Total cost in second semester		€3,939.40
Cost per student in second semester		€2.45
Total cost		€22,832.32
Total cost per student		€14.19

Notes: For the total cost and the total cost per student, we assume that letters are sent until the eleventh semester and we take into account that the number of students who receives feedback letters decreases, as students drop out or graduate.

Table A.2: Effect of feedback on survey participation

	(1)	(2)	(3)	(4)	(5)	(6)
a) Cohort I		Second semester			Fifth semester	
Treatment	0.008	0.004	0.002	-0.018	-0.020	-0.020
	(0.032)	(0.032)	(0.031)	(0.026)	(0.026)	(0.026)
N	812	812	812	812	812	812
Control mean	0.311	0.311	0.311	0.170	0.170	0.170
b) Cohort II		Second semester			Third semester	
Treatment	-0.001	-0.002	-0.004	-0.034	-0.034	-0.036
	(0.024)	(0.025)	(0.025)	(0.024)	(0.025)	(0.025)
N	797	797	797	797	797	797
Control mean	0.146	0.146	0.146	0.161	0.161	0.161
Randomization strata	yes	yes	yes	yes	yes	yes
Baseline performance	no	yes	yes	no	yes	yes
Additional covariates	no	no	yes	no	no	yes

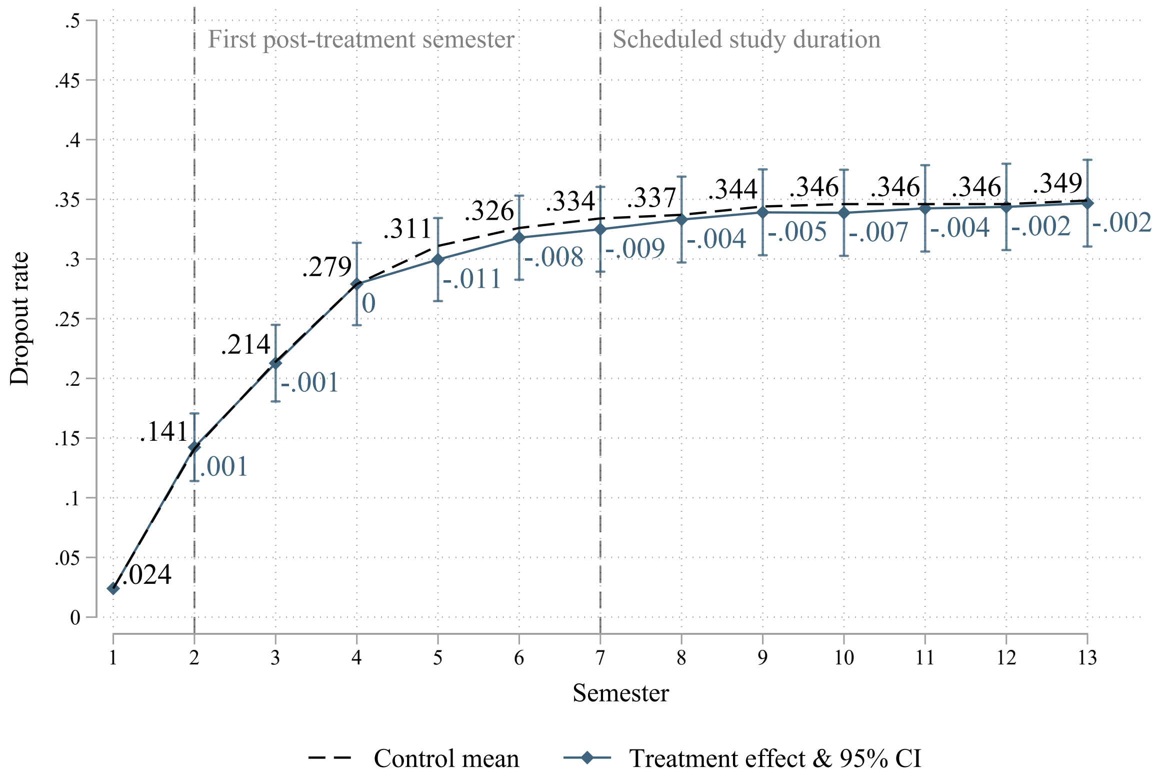
Notes: Outcome variable: indicates if a student answered at least one question in the respective survey. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table A.3: Descriptive statistics, by predicted graduation strata

	Not above average		Above average	
	(1) Stratum 1	(2) Stratum 2	(3) Stratum 3	(4) Stratum 4
Baseline performance				
First semester credits	7.868	17.687	24.290	29.134
First semester dropout rate	0.093	0.019	0.005	0.000
Std. first semester GPA	-1.078	0.171	-0.276	0.931
First semester GPA N/A	0.323	0.024	0.005	0.000
Std. imputed first semester GPA	-0.931	0.242	-0.209	1.009
Additional covariates				
Age	23.480	21.922	22.522	21.858
Female	0.336	0.400	0.277	0.471
High school degree Abitur	0.367	0.421	0.327	0.559
Time since high school degree	1.517	1.052	1.172	1.168
Std. high school GPA	-0.513	0.017	-0.109	0.682
N	431	425	379	374

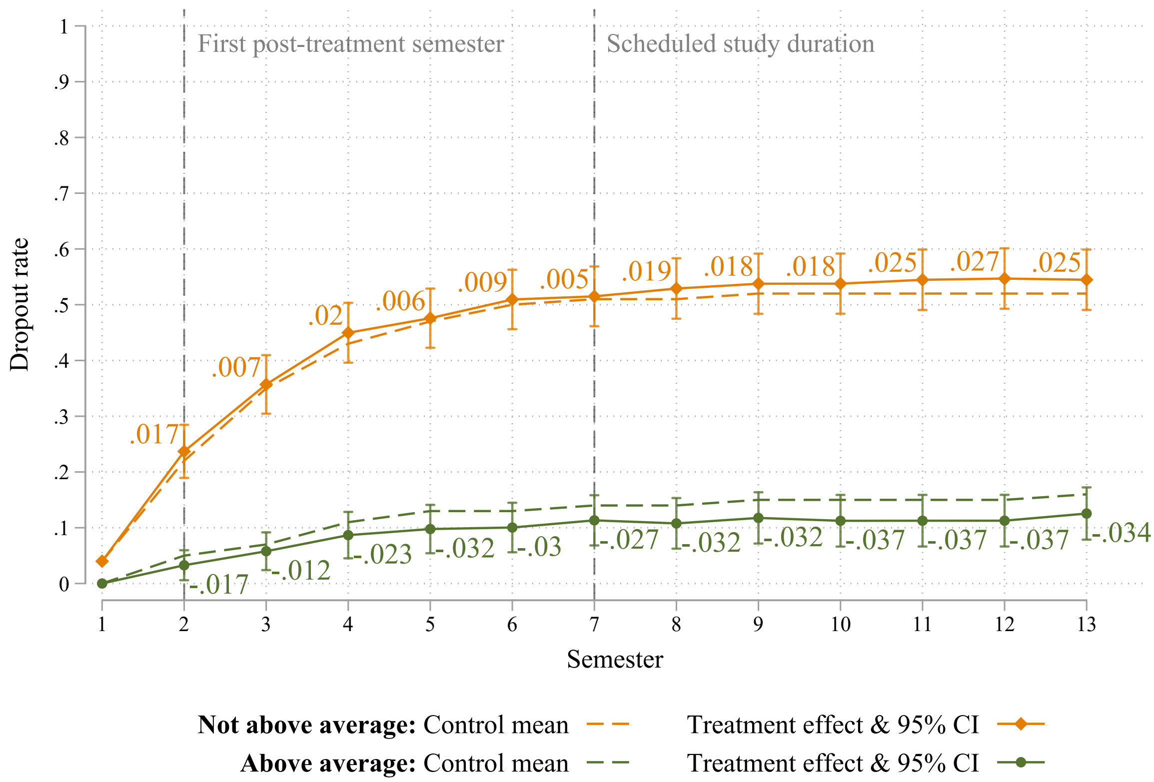
Notes: Above average and not above average refer to the rank in the pre-treatment credit point distribution.

Figure A.1: Effect of feedback on dropout rate



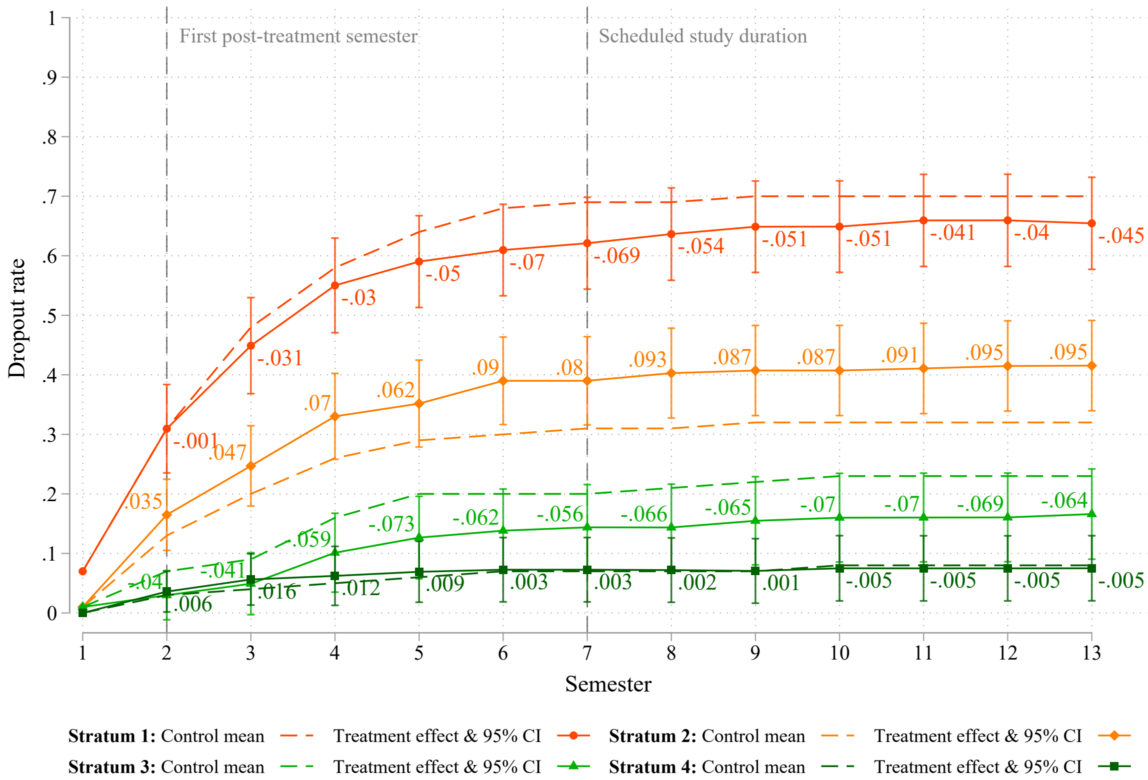
Notes: Outcome variable: indicates if a student dropped out of their study program before or during the respective semester. The dashed line depicts the raw control group mean. Coefficients are from regressions based on Equation 1 that are estimated separately for each semester and control for *randomization strata, baseline performance*, and *additional covariates. Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure A.2: Effect of feedback on dropout rate, by pre-treatment rank



Notes: Not above average (N = 856) and above average (N = 753) refer to the rank in the pre-treatment credit point distribution. *Outcome variable:* indicates if a student dropped out of their study program before or during the respective semester. The dashed lines depict the raw control group means. Coefficients are from full sample regressions based on Equation 2 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. N = 1609.

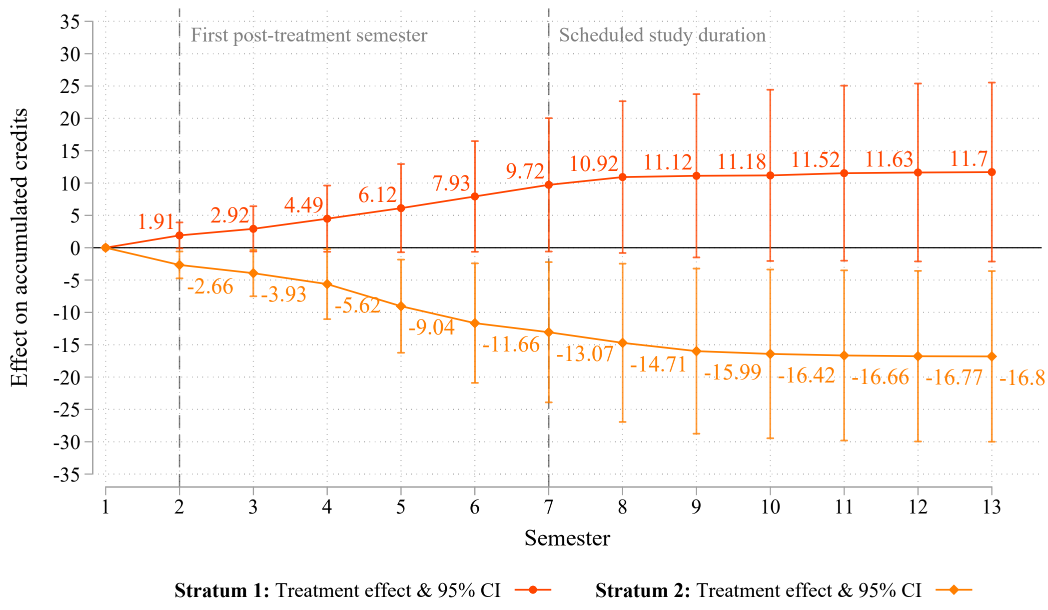
Figure A.3: Effect of feedback on dropout rate, by predicted graduation strata



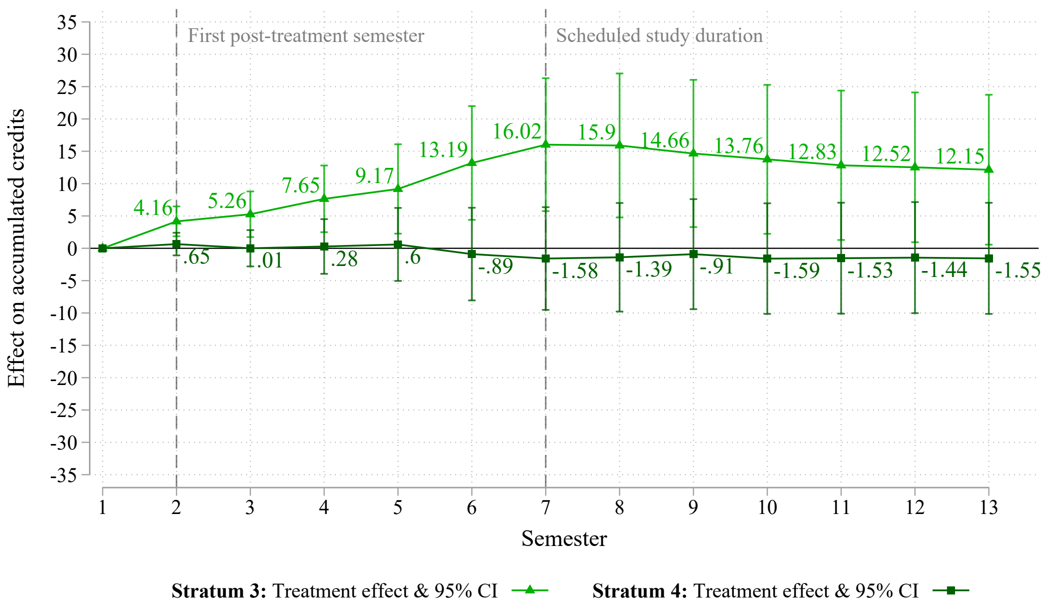
Notes: Outcome variable: indicates if a student dropped out of their study program before or during the respective semester. The figure plots treatment effects for the predicted graduation Strata 1 ($N = 431$) and 2 ($N = 425$) among students who did not rank above average in the pre-treatment credit point distribution and treatment effects for the predicted graduation Strata 3 ($N = 379$) and 4 ($N = 374$) among students who ranked above average. Dashed lines depict raw control group means. Coefficients are from full sample regressions based on Equation 3 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure A.4: Effect of feedback on accumulated credits, by predicted graduation strata

(a) Not above average

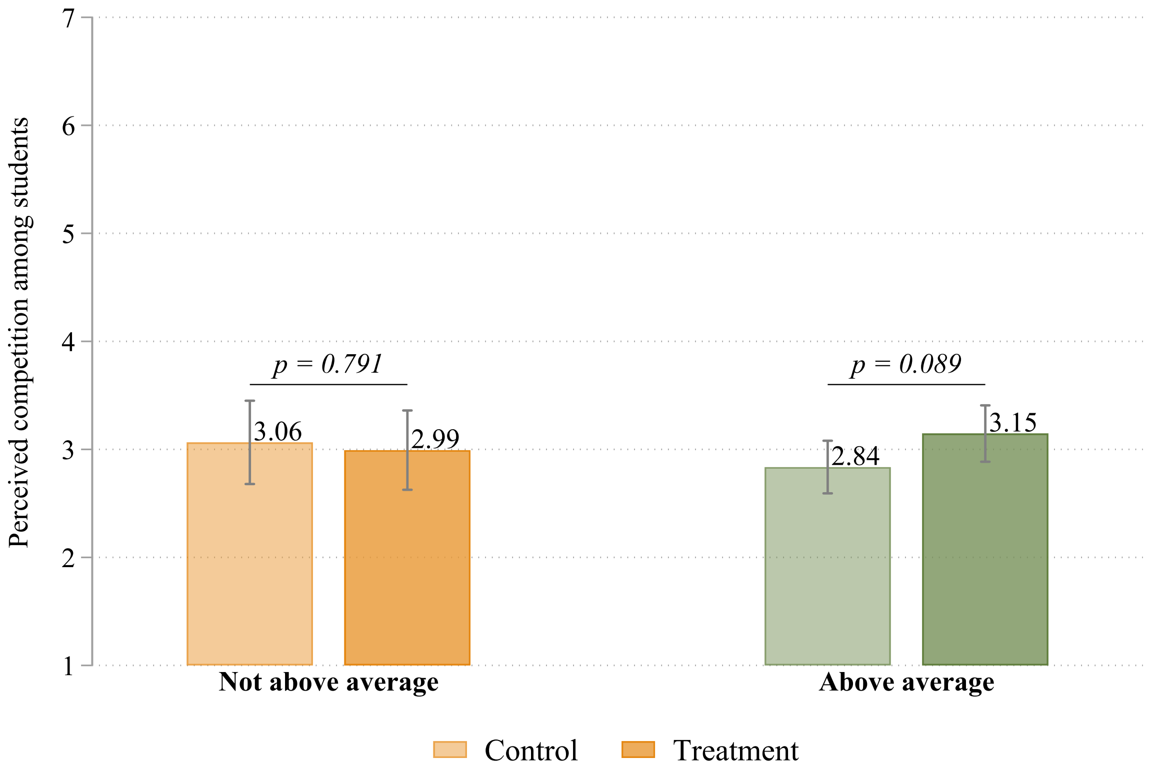


(b) Above average



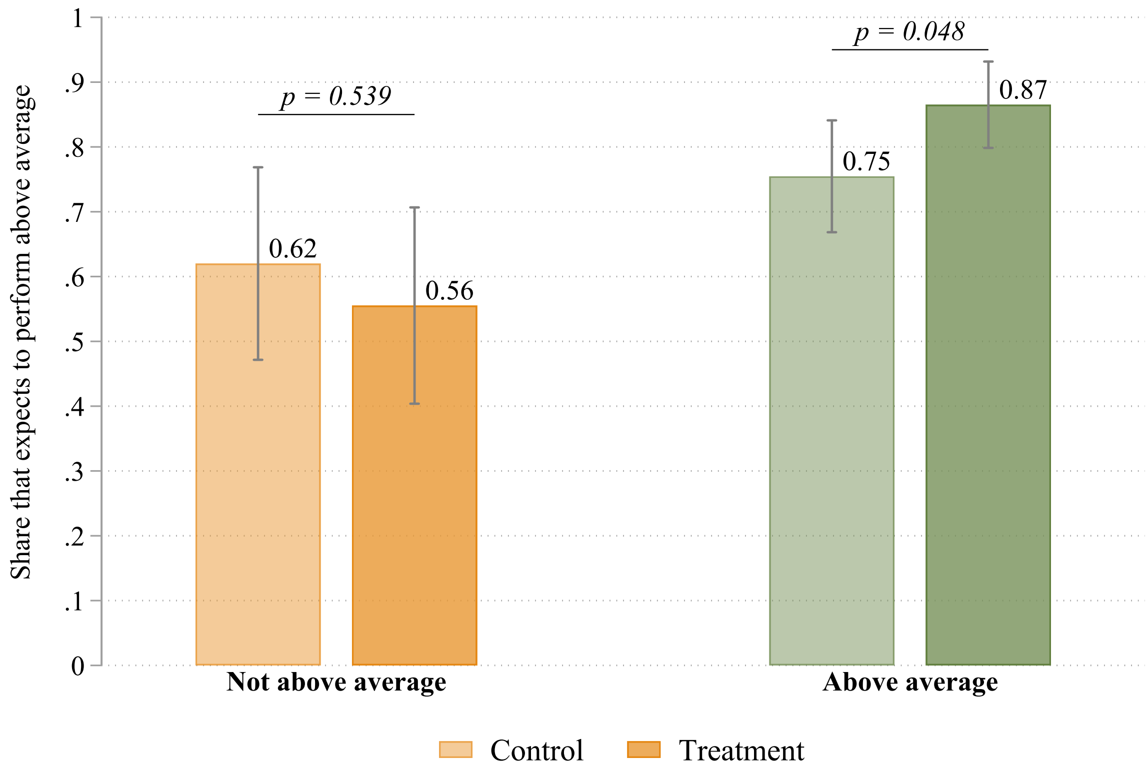
Notes: Outcome variable: number of credits accumulated until the end of the respective semester. Panel a) plots treatment effects for the predicted graduation Strata 1 ($N = 431$) and 2 ($N = 425$) among students who did not rank above average in the pre-treatment credit point distribution. Panel b) plots treatment effects for the predicted graduation Strata 3 ($N = 379$) and 4 ($N = 374$) among students who ranked above average. Coefficients in all panels are from full sample regressions based on Equation 3 that are estimated separately for each semester and control for *randomization strata, baseline performance*, and *additional covariates. Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure A.5: Effect of feedback on perceived competition, by pre-treatment rank



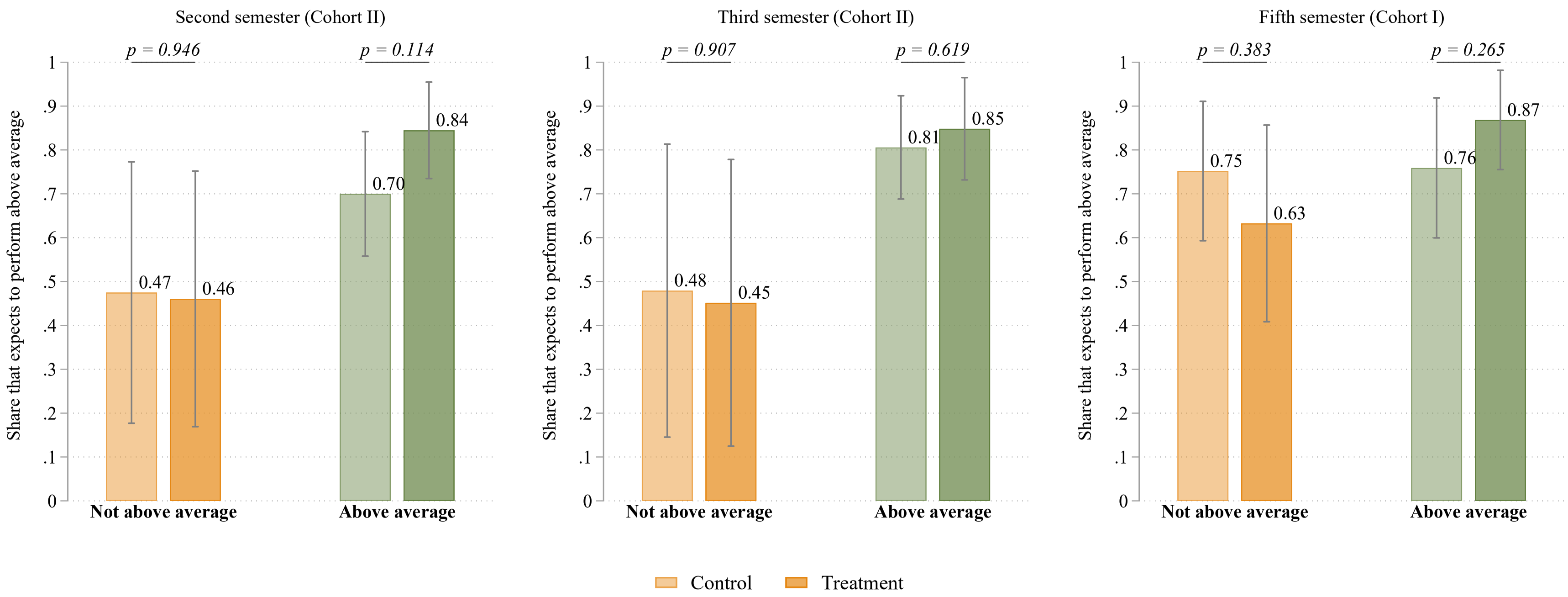
Notes: The figure plots the perceived competition among students by pre-treatment rank. The data on perceived competition stem from four surveys that were conducted in the second (both cohorts), the third (Cohort II), and the fifth (Cohort I) semester. In the surveys we asked students *To what extent do you agree with the following statement about your studies: When thinking about my studies, I think of competition among students.*; answer categories; completely disagree (= 1) to completely agree (= 7), and no answer. Estimates are from full sample pooled OLS regressions and control for study program FE and timing of survey FE. 95% confidence intervals are based on robust standard errors clustered at the student level. $N = 598$ from 472 individual students.

Figure A.6: Effect of feedback on expectations about relative performance, by pre-treatment rank



Notes: The figure plots the share of students who expects to rank above the average in the accumulated credit points distribution by pre-treatment rank. The data on expectations stem from three surveys that were conducted in the fifth semester (Cohort I) as well as the second and third semester (Cohort II). In the surveys we asked students *What do you think? How many per cent of your fellow students will have achieved more credit points (ECTS) than you at the end of the current semester?*; students could also provide no answer. We categorize students as expecting an above average performance if they expect that less than 50 percent of their fellow students will achieve more credits. Estimates are based on full sample pooled OLS regressions and control for study program FE and timing of survey FE. 95% confidence intervals are based on robust standard errors clustered at the student level. $N = 338$ from 288 individual students.

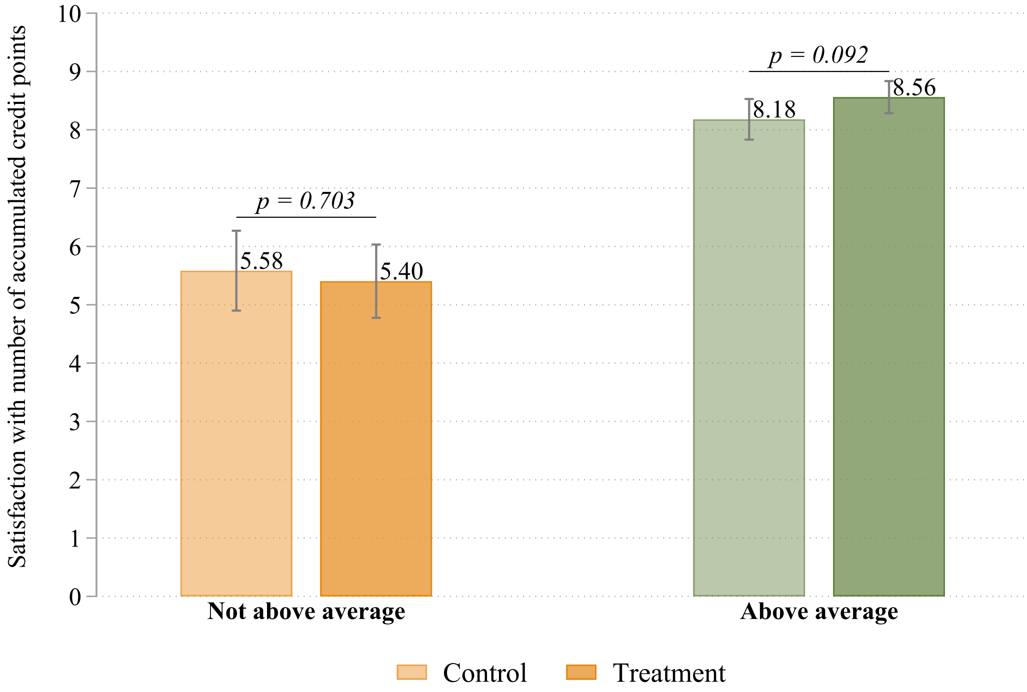
Figure A.7: Effect of feedback on expectations about relative performance, by pre-treatment rank and semesters



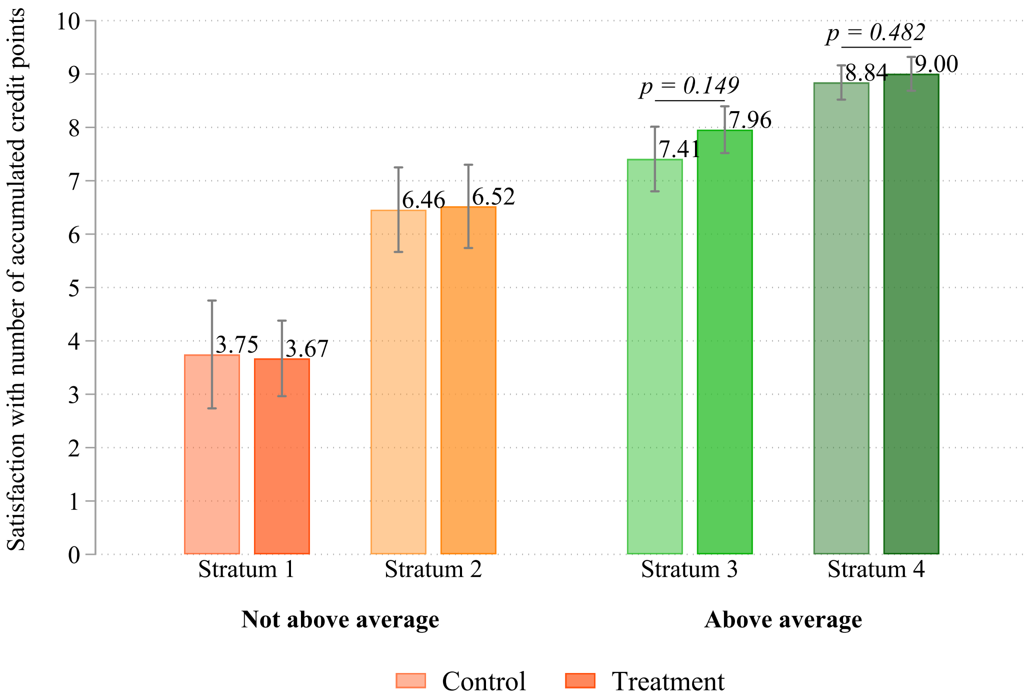
Notes: The figure plots the share of students who expects to rank above the average in the accumulated credit points distribution by pre-treatment rank and semesters. The data on expectations stem from three surveys that were conducted in the fifth semester (Cohort I) as well as the second and third semester (Cohort II). In the surveys we asked students *What do you think? How many per cent of your fellow students will have achieved more credit points (ECTS) than you at the end of the current semester?*; students could also provide no answer. We categorize students as expecting an above average performance if they expect that less than 50 percent of their fellow students will achieve more credits. Estimates are based on full sample pooled OLS regressions and control for study program FE. 95% confidence intervals are based on robust standard errors. $N = 108$ in the left, $N = 108$ in the middle, and $N = 122$ in the right panel.

Figure A.8: Effect of feedback on satisfaction with number of accumulated credits

(a) By pre-treatment rank

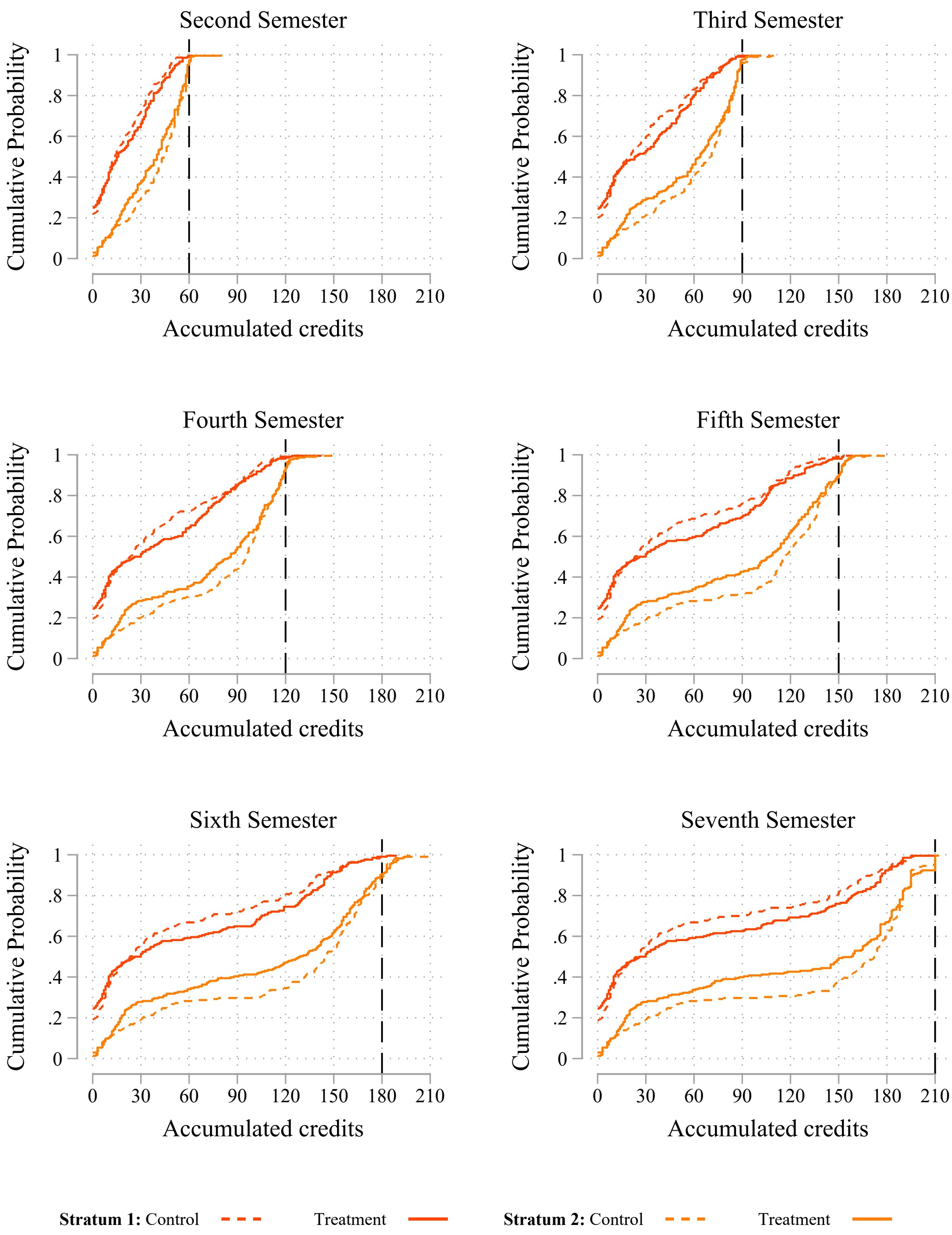


(b) By predicted graduation strata



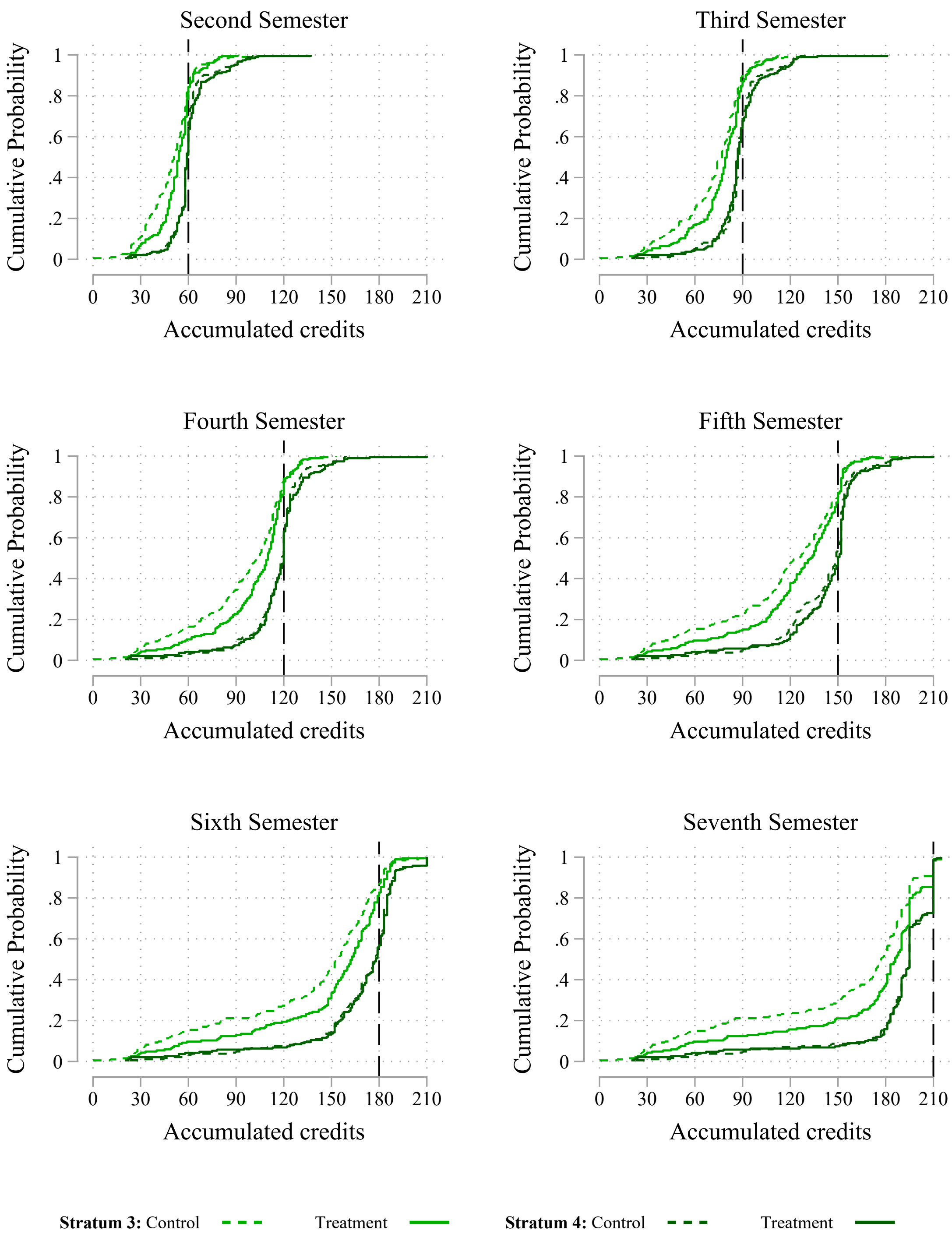
Notes: The figures plot the satisfaction with the number of accumulated credits by pre-treatment rank (Panel a) and for the four predicted graduation strata (Panel b). The data on satisfaction stem from four surveys that were conducted in the second (both cohorts), the third (Cohort II), and the fifth (Cohort I) semester. In the surveys we asked students *How satisfied are you with your performance in your studies so far? With my attained credit points (ECTS), I am ...* ; answer categories; completely dissatisfied (= 0) to completely satisfied (= 10), and no answer. Estimates are from full sample pooled OLS regressions and control for study program FE and timing of survey FE. 95% confidence intervals are based on robust standard errors clustered at the student level. $N = 597$ from 471 individual students.

Figure A.9: Cumulative distributions of accumulated credits by predicted graduation strata – not-above-average students



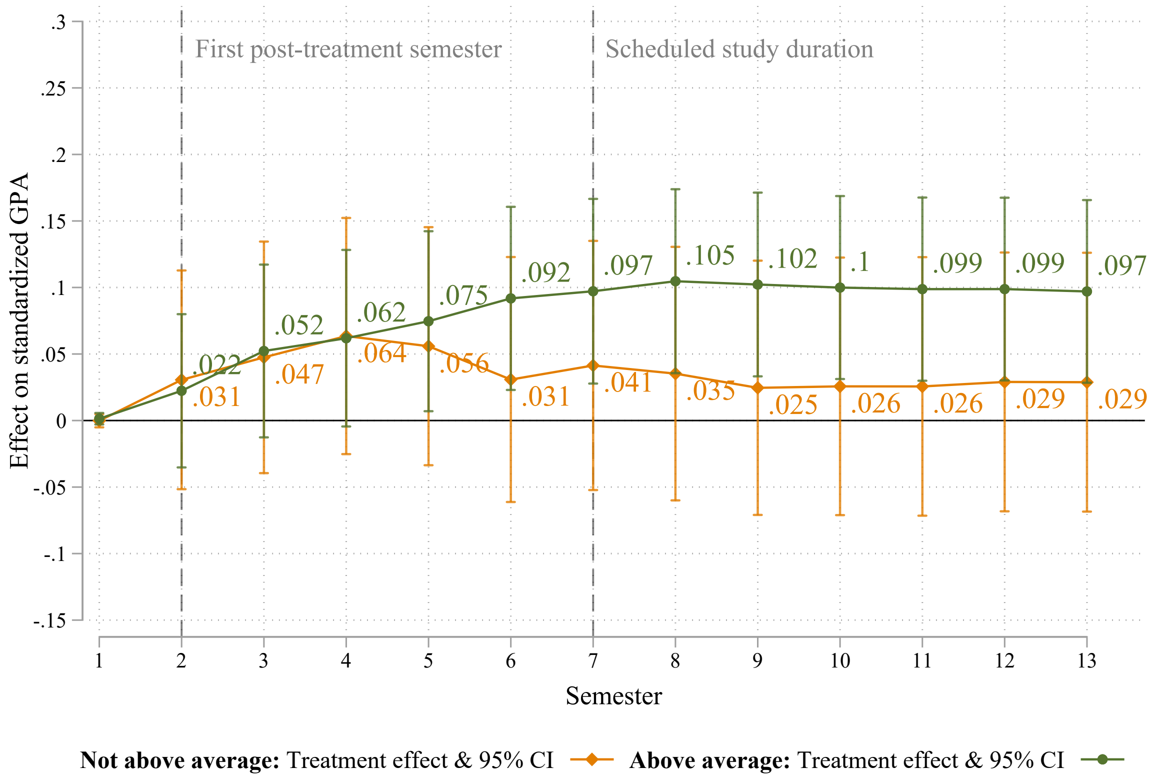
Notes: The six panels plot cumulative distributions of the accumulated credits at the end of the respective semester by treatment status for students in the predictive performance Strata 1 ($N = 431$) and 2 ($N = 425$). The black dashed lines indicate the number of accumulated credits that students should have obtained at the end of each semester in order to finish within the scheduled study duration of seven semesters.

Figure A.10: Cumulative distributions of accumulated credits by predicted graduation strata – above-average students



Notes: The six panels plot cumulative distributions of the accumulated credits at the end of the respective semester by treatment status for students in the predictive performance Strata 3 (N = 379) and 4 (N = 374). The black dashed lines indicate the number of accumulated credits that students should have obtained at the end of each semester in order to finish within the scheduled study duration of seven semesters.

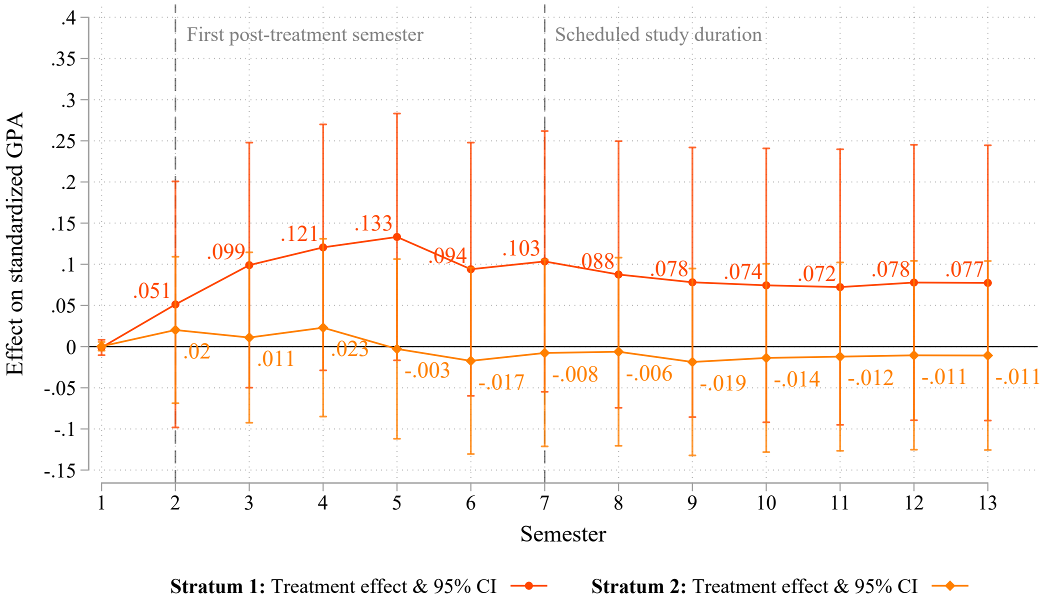
Figure A.11: Effect of feedback on standardized GPA, by pre-treatment rank



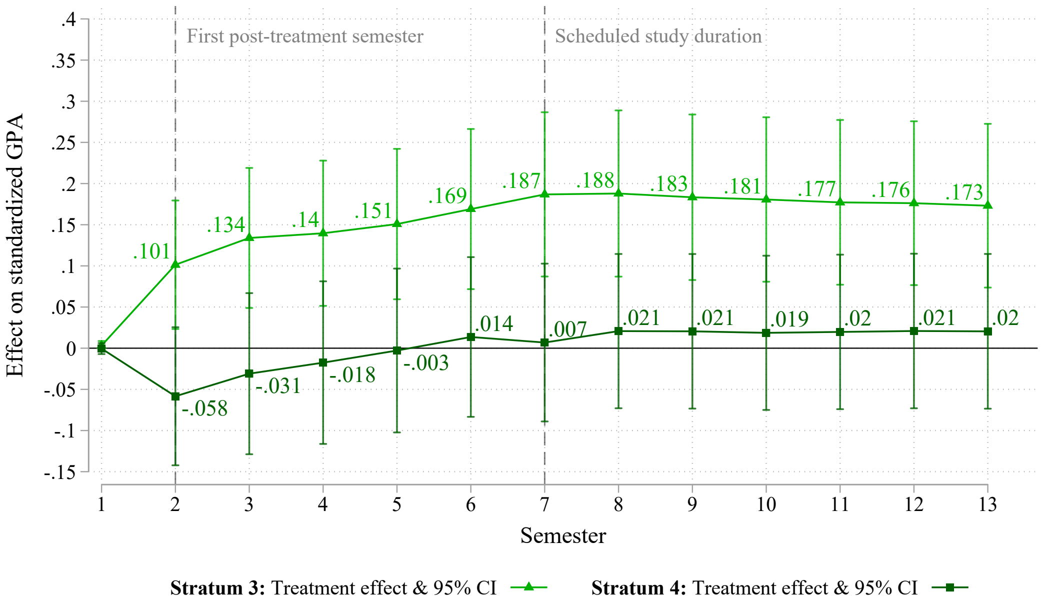
Notes: Not above average ($N = 856$) and above average ($N = 753$) refer to the rank in the pre-treatment credit point distribution. *Outcome variable:* standardized and reverse scaled GPA (higher values are better) is based on passing grades only. We do not use the original German scale where the best passing grade is 1.0 and the worst passing grade is 4.0. The GPA is unobserved for students who have never obtained a passing grade. Coefficients are from full sample regressions based on Equation 2 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure A.12: Effect of feedback on standardized GPA, by predicted graduation strata

(a) Not above average

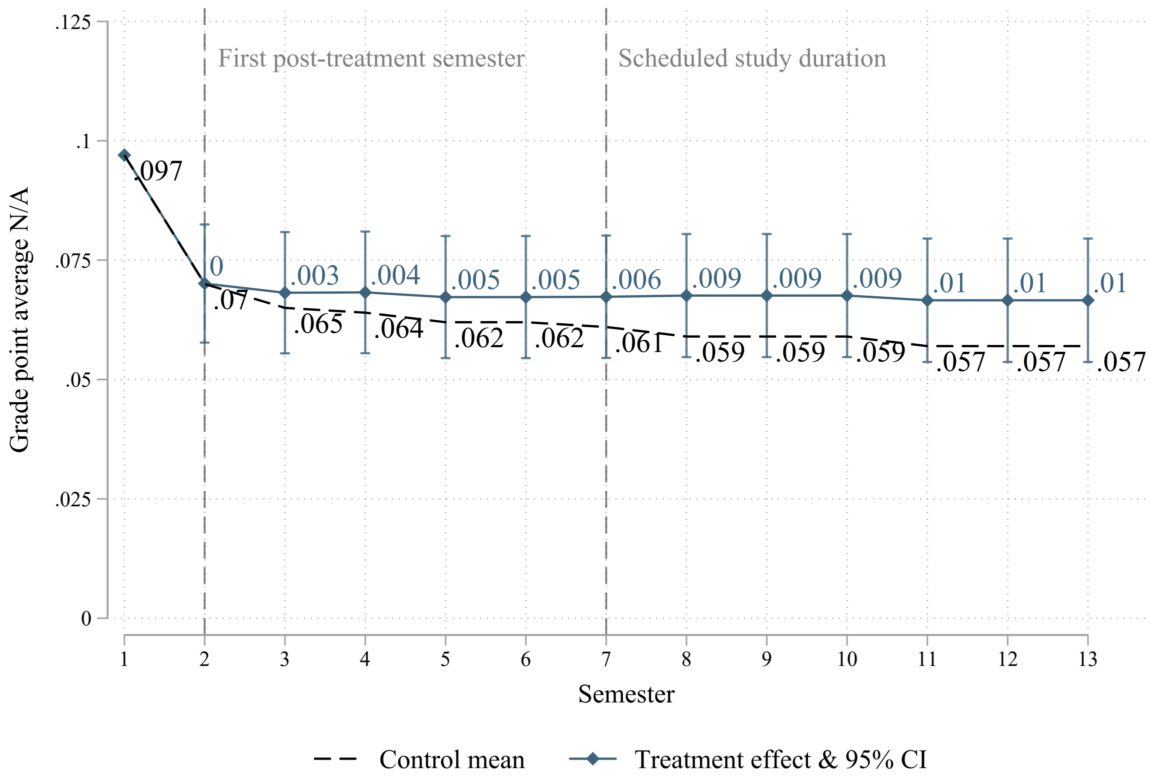


(b) Above average



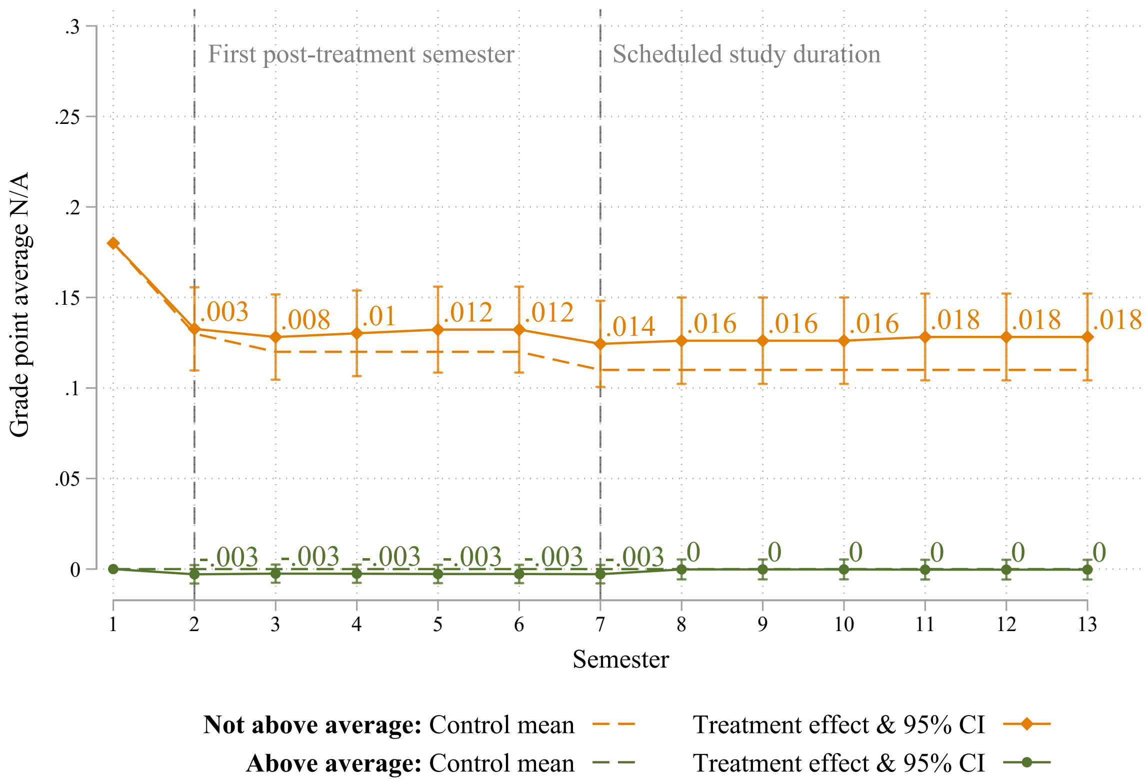
Notes: Outcome variable: standardized and reverse scaled GPA (higher values are better) is based on passing grades only. We do not use the original German scale where the best passing grade is 1.0 and the worst passing grade is 4.0. The GPA is unobserved for students who have never obtained a passing grade. Panel a) plots treatment effects for the predicted graduation Strata 1 ($N = 431$) and 2 ($N = 425$) among students who did not rank above average in the pre-treatment credit point distribution. Panel b) plots treatment effects for the predicted graduation Strata 3 ($N = 379$) and 4 ($N = 374$) among students who ranked above average. Coefficients in all panels are from full sample regressions based on Equation 3 that are estimated separately for each semester and control control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure A.13: Effect of feedback on GPA N/A



Notes: Outcome variable: indicates if the GPA at the end of the respective semester is missing. The dashed line depicts the raw control group mean. Coefficients are from regressions based on Equation 1 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

Figure A.14: Effect of feedback on GPA N/A, by pre-treatment rank

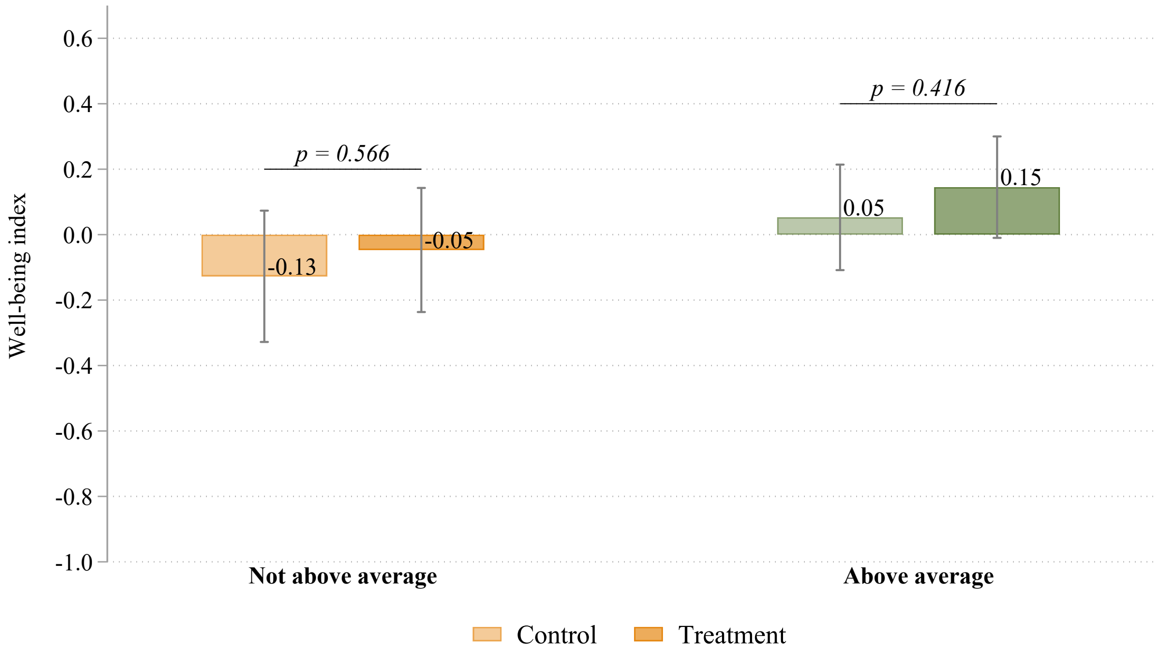


Notes: Not above average (N = 856) and above average (N = 753) refer to the rank in the pre-treatment credit point distribution. *Outcome variable:* indicates if the GPA at the end of the respective semester is missing. The dashed lines depict the raw control group means. Coefficients are from full sample regressions based on Equation 2 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates. Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. N = 1609.

Figure A.15: Effect of feedback on GPA N/A, by predicted graduation strata

Notes: Outcome variable: indicates if the GPA at the end of the respective semester is missing. The figure plots treatment effects for the predicted graduation Strata 1 ($N = 431$) and 2 ($N = 425$) among students who did not rank above average in the pre-treatment credit point distribution and treatment effects for the predicted graduation Strata 3 ($N = 379$) and 4 ($N = 374$) among students who ranked above average. Dashed lines depict raw control group means. Coefficients are from full sample regressions based on Equation 3 that are estimated separately for each semester and control for *randomization strata*, *baseline performance*, and *additional covariates*. *Randomization strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *baseline performance:* first semester credits, standardized first semester GPA (missing values imputed), first semester GPA imputation dummy, and first semester dropout; *additional covariates:* standardized high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Confidence intervals are based on robust standard errors. $N = 1609$.

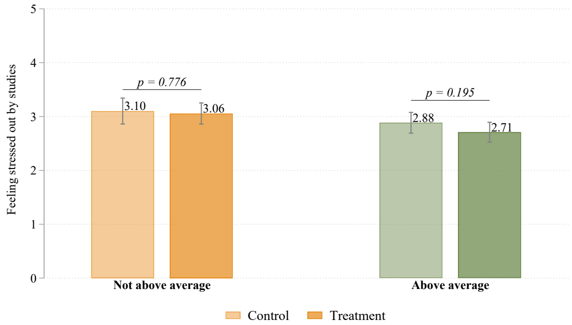
Figure A.16: Effect of feedback on well-being, by pre-treatment rank



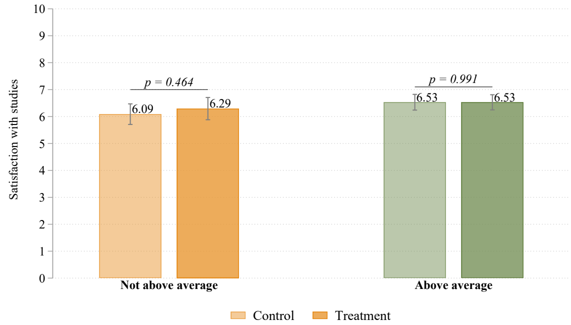
Notes: The figure plots students' well-being by pre-treatment rank. The well-being index is the standardized inverse-covariance weighted average (following Anderson (2008) and using the Stata program by Schwab et al. (2020)) of survey questions that asked students how stressed they feel about their studies, how satisfied they are with their studies, and how satisfied they are with their life in general (see Figure A.17 for separate estimates and the survey questions that we asked). The four surveys were conducted in the second (both cohorts), the third (Cohort II), and the fifth (Cohort I) semester. Estimates are from full sample pooled OLS regressions that control for study program FE and timing of survey FE. 95% confidence intervals are based on robust standard errors clustered at the student level. $N = 612$ from 481 individual students.

Figure A.17: Effect of feedback on well-being, by pre-treatment rank

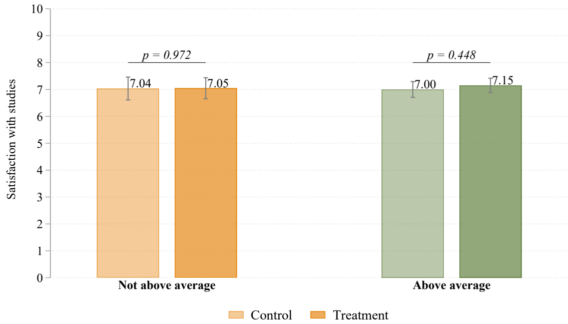
(a) Stressed out by studies



(b) Satisfaction with studies



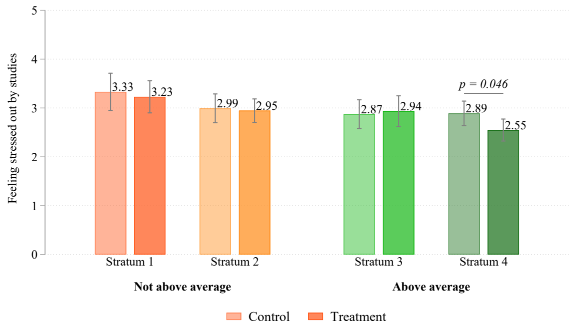
(c) Satisfaction with life in general



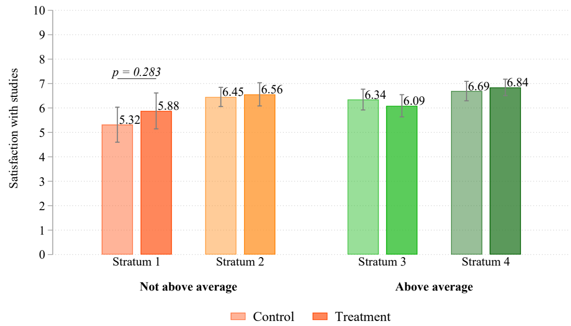
Notes: The figures plot students' well-being by pre-treatment rank. The data stem from four surveys that were conducted in the second (both cohorts), third (Cohort II), and fifth (Cohort I) semester. For the outcome in Panel (a), we asked students *During the last weeks, how often did you feel stressed out by your studies?*; answer categories: never (= 0), rarely, sometimes, often, very often, always (= 5), and no answer. For the outcome in Panel (b), we asked students *How satisfied are you currently with your studies, all things considered?*; answer categories: completely dissatisfied (= 0) to completely satisfied (= 10), and no answer. For the outcome in Panel (c), we asked students *How satisfied are you currently with your life, all things considered?*; answer categories: completely dissatisfied (= 0) to completely satisfied (= 10), and no answer. In the surveys, the last question was asked before the other two. Estimates are from full sample pooled OLS regressions that control for study program FE and timing of survey FE. 95% confidence intervals are based on robust standard errors clustered at the student level. $N = 605$ from 475 individual students in panel (a), $N = 603$ from 473 individual students in panel (b), and $N = 602$ from 474 individual students in panel (c).

Figure A.18: Effect of feedback on well-being, by predicted graduation strata

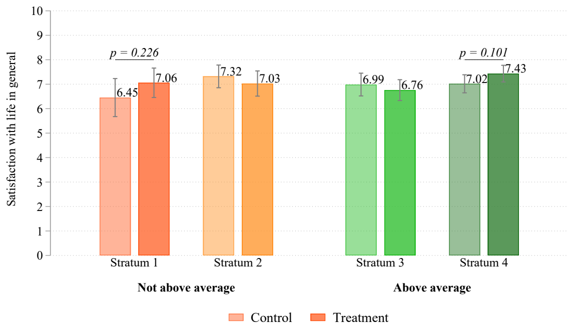
(a) Stressed out by studies



(b) Satisfaction with studies



(c) Satisfaction with life in general



Notes: The figures plot students' well-being separately for the four predicted graduation strata. The data stem from four surveys that were conducted in the second (both cohorts), third (Cohort II), and fifth (Cohort I) semester. For the outcome in Panel (a), we asked students *During the last weeks, how often did you feel stressed out by your studies?*; answer categories: never (= 0), rarely, sometimes, often, very often, always (= 5), and no answer. For the outcome in Panel (b), we asked students *How satisfied are you currently with your studies, all things considered?*; answer categories: completely dissatisfied (= 0) to completely satisfied (= 10), and no answer. For the outcome in Panel (c), we asked students *How satisfied are you currently with your life, all things considered?*; answer categories: completely dissatisfied (= 0) to completely satisfied (= 10), and no answer. In the surveys, the last question was asked before the other two. Estimates are from full sample pooled OLS regressions that control for study program FE and timing of survey FE. 95% confidence intervals are based on robust standard errors clustered at the student level. $N = 605$ from 475 individual students in panel (a), $N = 603$ from 473 individual students in panel (b), and $N = 602$ from 474 individual students in panel (c).