# RUHR ECONOMIC PAPERS

Friederike Hertweck
Lukas Jonas
Melissa Kistner
Deborah Maffia

# Study Effort in Higher Education: Field Experimental Evidence With Administrative and Tracking Data from Germany

RUHR UNIVERSITÄT BOCHUM RUB rwi

#1138

# Imprint

## Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

## Editors



## Editorial Office



## Ruhr Economic Papers #1138

Responsible Editor: Ronald Bachmann



ISSN 1864-4872 (online) – ISBN 978-3-96973-320-2

Ruhr Economic Papers #1138

Friederike Hertweck, Lukas Jonas, Melissa Kistner, and Deborah Maffia

# Study Effort in Higher Education: Field Experimental Evidence With Administrative and Tracking Data from Germany

**Bibliografische Informationen der Deutschen Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

https://dx.doi.org/10.4419/96973320
ISSN 1864-4872 (online)
ISBN 978-3-96973-320-2

Friederike Hertweck, Lukas Jonas, Melissa Kistner, and Deborah Maffia*

# Study Effort in Higher Education: Field Experimental Evidence With Administrative and Tracking Data from Germany


## Abstract

*This study investigates the impact of a low-cost, color-coded scale intervention designed to inform university students about the expected workload for a course, with the aim of improving students' academic performance and learning behaviors. An initial intervention took place at the beginning of the course, with a follow-up reminder in the middle of the semester. Students who were treated once experienced no significant effect, but those who additionally received the second treatment significantly improved their course grade, scoring 0.51 points (or 21 %) higher on average. Heterogeneity analyses reveal that first-generation, migrant and high-ability students benefited most from the intervention, suggesting that such a treatment may help reduce some forms of educational inequality. To explore the underlying mechanisms, we utilized tracking data from an online learning platform through which the lecturer distributed course materials and provided opportunities for self-paced learning. While we find an overall increase in online activity following the intervention (though imprecisely measured), no specific academic behavior such as online test participation or material downloads can explain the ultimate increase in grades by itself.*



*JEL-Codes: I20, I23, J24, J08*

*Keywords: Student performance; field experiment; higher education; color-coded nudge; RCT*

*December 2024*



** Friederike Hertweck, RWI; Lukas Jonas, RWI and RUB; Melissa Kistner, RUB; Deborah Maffia, RUB.*


*– The authors thank all supporters of the participating university as well as workshop and seminar participants at the German Centre for Higher Education Research and Science Studies (DZHW) and the TH Nuremberg for helpful feedback and comments. The study obtained ethical approval from the Ruhr University Bochum and was pre-registered as AEARCTR-0012459.*


*– All correspondence to: Friederike Hertweck, RWI, Hohenzollernstraße 1–3, 45128 Essen, Germany, e-mail: friederike.hertweck@rwi-essen.de*

# 1 Introduction

Academic performance among college students varies significantly, with many demonstrating suboptimal study habits, struggling to graduate on schedule or even dropping out of university (Bound, Lovenheim, & Turner, 2010). One key factor contributing to this discrepancy is the lack of sufficient investment in study time, often driven by students' limited understanding of the required effort to succeed (Brint & Cantwell, 2010; Babcock & Marks, 2011; Rury & Carrell, 2023). While interventions like mentoring programs have proven effective in providing relevant information especially to underrepresented students or those who lack informal networks (Hardt, Nagler, & Rincke, 2023), their high costs and limited scalability present significant barriers. In contrast, low-cost informational nudges offer a promising alternative, enabling students to better navigate the academic system and potentially reducing achievement disparities by providing the crucial information needed to guide their academic efforts.

In this paper, we report the results of a randomized controlled trial designed to test whether providing information about the required study effort affects students' academic performance and learning behaviors. The study was conducted within a large compulsory first-year undergraduate course at a German university. In the fourth lecture of the course, students participated in a paper-based survey during which they were randomly assigned to receive either information on the required study hours (treatment group) or information on library loans (control group). In the eighth week, students were asked to participate in an additional online survey that included a reminder of the information treatment.

Unlike many related studies, we employ an easily scalable intervention to increase study effort. The treatment contains a color-coded *Study Score* that visually indicates the expected weekly study effort for the course. The colors are based on a traffic light scale ranging from red to green with red indicating insufficient effort, encouraging students to adjust their study habits, while green represents meeting the required study hours. Keeping the treatment that simple avoids the high costs and limited scalability of mentoring programs.

Throughout the course, students' engagement with the course was tracked through their activity on an online learning platform which the lecturer used to distribute lecture notes and problem sets. A key strength of our design is that we track students' actual course-related behaviors, thereby minimizing biases associated with self-reported data. Specifically, we combine three distinct data sources: (1) register data from the university's administrative system which provides information on grades and credits, (2) tracking data from the online learning platform *Moodle*[1] that records students' interactions such as the number of log-ins, time spent on the platform, and the completion of online tests, and (3) survey data collected from two separate surveys that capture students' socio-demographic variables as well as information on personality traits.

The intervention and the subsequent reminder four weeks later increased students' grades by approximately 0.51 points on average, corresponding to a 21 % improvement in academic performance. These findings suggest that even simple, cost-effective interventions such as clarifying academic expectations can have meaningful impacts on student achievement. Yet, providing only the initial treatment (without the reminder) did not lead to significant improvements in grades. Further heterogeneity analyses reveal considerable heterogeneity in the treatment effects across different subgroups of students. First-generation students and students with a migration background exhibited a much stronger response to the intervention, showing an increase of about 41 and 38 % in their course grades. The intervention also had a greater impact on high-ability students, as measured by their high school GPA, whose grades improved more than those of their low-ability peers. This suggests that students with better academic preparation may

[1]Moodle is an open-source learning management system intensively used by universities in Europe, Asia, and the USA to deliver online courses, facilitate collaboration, and track students' learning activities (Klaus Steitz, 2021; Moodle an Hochschulen e.V., 2024).

be more responsive to clarified workload expectations. Consequentially, while the intervention can help reducing inequalities among certain disadvantaged groups such as students with a migration background, it could inadvertently exacerbate disparities for those with lower academic abilities.

The analysis of underlying channels reveals that the intervention primarily benefited students who were initially uninformed about the required study effort, with a 27 % improvement in their grades. However, this effect was not statistically significant, likely due to rather small sample sizes. In contrast, students who were already informed about the workload did not show a meaningful response, suggesting that the intervention had limited impact for those with pre-existing knowledge of the academic demands. This implies that the provision of clearer workload expectations is especially valuable for students lacking such information.

Examining engagement with the online learning platform, we find a modest but non-significant increase in daily effort of 7 to 14 % for students who received both the initial treatment and the reminder. However, no clear differences were found in specific activities, such as completing online tests, downloading materials, or watching video tutorials. The gradual release of course materials throughout the semester may have diluted the impact of the intervention on these behaviors, making it difficult to detect clear effects.

The implementation of interventions aimed at improving students' academic performance and reducing college dropout rates has grown rapidly over the past decade (Hertweck, Kistner, & Maffia, 2025). These interventions are manifold and span peer tutoring (Hardt et al., 2023), coaching or related college support services (Angrist, Lang, & Oreopoulos, 2009; Bettinger & Baker, 2014; Broda et al., 2018; Oreopoulos & Petronijevic, 2019; Evans, Kearney, Perry, & Sullivan, 2020), the opportunity to set academic goals (Dobronyi, Oreopoulos, & Petronijevic, 2019; Clark, Gill, Prowse, & Rush, 2020) or to sign a commitment device (Himmler, Jäckle, & Weinschenk, 2019; Cagala, Glogowsky, & Rincke, 2021), performance feedback (Azmat, Bagues, Cabrales, & Iriberri, 2019), text messages (Castleman & Meyer, 2020; Castleman & Page, 2016, 2015), financial incentives (Angrist et al., 2009), access to specific technology or software tools (Oreopoulos & Petronijevic, 2018), specific pieces of information (Rury & Carrell, 2023; Li, 2018), or reminders about the course schedule and exams (Himmler et al., 2019). Yet, most studies do not target students' allocation of time.

Two exceptions are Oreopoulos and Petronijevic (2019) and Rury and Carrell (2023). Oreopoulos and Petronijevic (2019) implement various interventions at a large Canadian university and find that face-to-face coaching notably improved study habits, increasing overall study time by about 2 hours per week. Although the interventions do not significantly enhance academic outcomes, they improve students' subjective well-being and make them feel more supported. Rury and Carrell (2023) design an information intervention that informed students about the average returns to effort. This intervention was delivered as part of a short survey administered prior to one of several assessments during the course, with students' progress tracked over the following weeks. They find that their treatment leads to an increase in study effort by 7 % which lacks, however, statistical significance. Our results partially align with their finding, showing an increase of 9 to 14 % in daily effort (though also statistically insignificant), while also revealing a notable 21 % average increase in course grades.

Despite the positive outcomes observed in our study, several limitations should be noted. One significant challenge was the relatively small sample size resulting from panel attrition. Out of our initial sample of 509 students, only 173 students (i.e., 34 %) participated in the follow-up online survey, which limited the statistical power of our analyses. This reduced power makes it more difficult to detect smaller treatment effects or subgroup analyses. The structure of the course may also have diluted the impact of the intervention on specific academic behaviors, such as downloading course material or engaging with online assessments, as materials are being released incrementally throughout the semester. These factors likely contributed to the more ambiguous results regarding detailed platform activities, highlighting the complexities involved in measuring student behavior when engagement is spread out over an extended period.

Nonetheless, the intervention itself was remarkably cost-effective. The simple visual cue provided in the form of a color-coded scale designed to clarify the expected study effort resulted in an average improvement of 21 % in students' course grades. In contrast to resource-intensive interventions such as tutoring or mentoring (Bettinger & Baker, 2014; Hardt et al., 2023), which often require significant financial and human resources, our intervention is scalable and requires minimal investment to implement. Overall, our results suggest that even small-scale interventions can lead to meaningful improvements in student performance, when reinforced over time. Given the resource constraints that many universities face, especially in large courses, such simple yet effective strategies could be a powerful tool in supporting student success and enhancing academic achievement across diverse student populations.

Besides contributing to the literature on nudges and information provision in higher education, this study also adds to the growing body of research on the power of color-coded nudges in various settings such as energy efficiency (Waechter, Sütterlin, Borghoff, & Siegrist, 2016; Bengart & Vogt, 2023), resource use (Andor, Goette, Price, Schulze Tilling, & Tomberg, 2023) or food labels (Sothey, 2021). Furthermore, our use of *Moodle* tracking data, which is widely used at universities globally, offers a first insight into their potential on utilizing such data for the evaluation of educational interventions.

Finally, our results also highlight the complex dynamics of educational interventions, suggesting that while they can address some forms of inequality, they may also reinforce existing advantages for students who are already better prepared. First-generation and migrant students but also high-ability students benefited most from the intervention. Thus, this study underscores the need for carefully designed, potentially tailored interventions that consider student characteristics and diverse needs. Future research should explore whether such interventions can be adapted or supplemented with additional support to better address the full spectrum of student backgrounds and abilities.

The remainder of the paper is organized as follows. Section 2 describes the experimental setting including a detailed description of the course, the timeline of the field experiment, and the randomization. Section 3 then describes the different data sources in more detail and also discusses the balancing. Results are presented in Section 4 and discussed in Section 5. Finally, Section 6 concludes.

# 2 Experimental Setting

## 2.1 Institutional background

This study is situated within the context of university students' academic workload and the challenges they face in managing their time effectively. We conduct the randomized controlled trial (RCT) as part of a compulsory first-year course for undergraduate students in *Business and Economics* at a large German university. The undergraduate program in *Business and Economics* is designed to be completed within three years. It requires students to complete 180 credits to graduate, i.e., 30 credits per semester, equivalent to six courses of five credits each. Despite receiving a structured curriculum and a clear study plan, many students fail to achieve the required 30 credits per term, which suggests that they may be underestimating the amount of time needed for their studies. The gap between expected and actual study hours can be attributed to several factors, such as difficulties with self-organization, ineffective study habits, or a lack of awareness regarding the true scope of the workload.

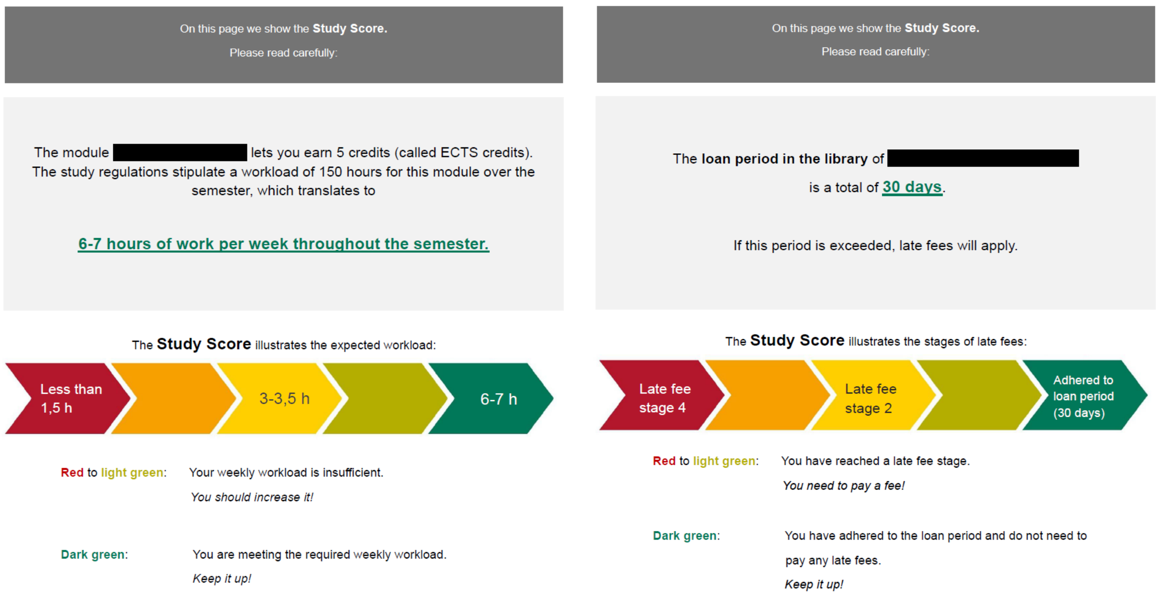
### Actual and expected workload

At the university where the RCT is conducted, the expected workload for a five-credit course is typically between 6 to 7 hours of study per week. A recent survey conducted at the university among *Business and Economics* freshmen during the winter terms of 2022/23 and 2023/24 highlights that students on average spent just 4.8 hours per week on a 5-credits-course, well below the 6 to 7 hours expected. This discrepancy between intended and actual study time suggests a potential information friction that our intervention aims to address.

To reduce this information friction and help students align their study efforts with course requirements, our intervention visualizes the expected study time through a so-called *Study Score* as illustrated in Figure 1. It thereby uses a color-coded scale ranging from red to green to indicate whether students are on track with their expected workload. Along with the color coding the accompanying text provides clear instructions: students in the green zone are meeting their study goals, while those in the red or orange zones are advised to increase their effort. This simple visualization aims to reduce uncertainty and support students in making more informed decisions about their study habits, with the broader goal of reducing educational inequalities stemming from information gaps.

Figure 1: Information provision as part of paper-based survey

A) Treatment group B) Control group




Note: Page with translated treatment (panel A) and control information (panel B). Original German version is displayed in Appendix A.

Overall, the page with the treatment combines several effective features for conveying messages. The treatment includes a concise informational text, a color-coded illustration based on traffic light colors, a brief interpretation and reflection of the scale, and a control question at the end. Color-coding using traffic light colors (red, yellow, and green) has been proven to be an effective method for informing individuals about the quality of their choices. Green commonly reflects a good or optimal choice, encouraging the desired behavior. Yellow indicates a moderate or cautionary choice and suggests that some improvements are needed. Red signals a poor choice, discouraging the behavior and indicating that significant changes are necessary. This intuitive system leverages familiar associations to quickly convey information, helping people make better-informed choices across various contexts, from nutrition labels to environmental impact ratings (Antúnez, Giménez, Maiche, & Ares, 2015; Thøgersen & Nielsen, 2016).[2] Importantly, traffic light labels have been proven to be more effective than text-based labels (Bengart & Vogt, 2023).

## 2.2 Timeline of the intervention

In collaboration with the lecturer of a compulsory first-year course in *Business and Economics*, we conducted the RCT during one of the 5-credits-courses and distributed the *Study Score* to students as part of a paper-based survey in week 4 and a follow-up online reminder in week 8. The course and the surveys are explained in the following.

**The course**

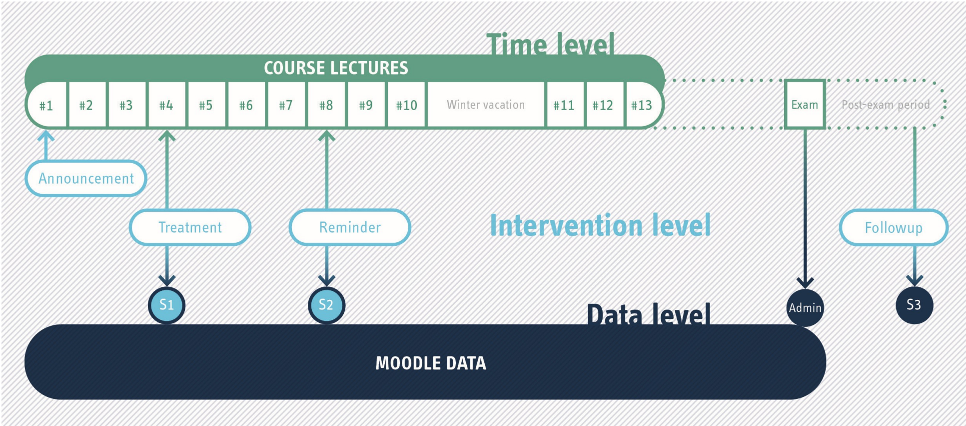
The course is a required component of the undergraduate program in *Business and Economics* and spans 13 weeks of teaching from October to the end of January, with a three-week break over the Christmas period. It culminates in an exam, which is scheduled approximately three weeks after the final lecture to provide students with additional time for preparation. Students attend one 90-minute lecture each week throughout the term, ensuring regular engagement with the course material. However, it is important to note that attendance in these lectures is not tracked, as students are not formally monitored for their presence during class sessions.

Starting in week 4, students furthermore receive problem sets that are discussed during bi-weekly tutorials led by the professor's research assistant. These tutorials provide an interactive space for students to engage with course content, ask questions, and reinforce their understanding through collaborative problem-solving. All course materials, including lecture slides, readings, and additional resources, are uploaded to the university's online learning platform *Moodle* a few days before each lecture or tutorial. This ensures that students have the opportunity to review and prepare for each session in advance.

In addition to the main lectures and tutorials, the lecturer provides supplementary online videos that reinforce key concepts from the course. Students also have access to eight self-assessment quizzes (*online tests*) throughout the term, which are designed to help them assess their understanding of the material and identify areas where further review may be necessary. These resources aim to support students' independent study and improve their overall comprehension of the course content. Figure 2 illustrates the detailed timeline of the course and the interventions.

[2] For instance, the EU Energy label that uses a color-scale from green to red enhances consumers' perceptions of the differences in energy friendliness between the classes and also increases motivation to choose the most energy-efficient products (Waechter et al., 2016).

Figure 2: Timeline of lectures, intervention, and data




Note: The figure illustrates the timeline of the course. S1, S2, S3 indicates the paper-based survey (S1) and two online surveys (S2 and S3).

### The intervention

Students were invited to participate in a paper-based survey during the fourth lecture of the course, which had been announced in class during the first lecture (see Figure 2). Up to this point, there are no differences between the treated and untreated students.

During the day of the intervention, the lecturer delivered his lecture as scheduled and concluded approximately 30 minutes ahead of schedule before leaving the lecture hall. Subsequently, the research team distributed a paper-based survey to students, explaining its purpose to enhance student satisfaction. The team additionally informed students that they could provide their email addresses to be contacted for future surveys and to participate in an online lottery.

The paper-based surveys consisted of different sections. First, students entered their student identification number and their consent to data linkages. They were then asked to respond to questions on their socio-demographic background, personality traits, and learning habits (see Appendix A for details), followed by a page displaying the *Study Score*, questions on their financial situation and finally the possibility to enter their e-mail-adress for future correspondence and the lottery. Two versions of the survey were distributed, with the only variation being the page that included the *Study Score*. Students in the treatment group were provided with information about the required study intensity for the course (Panel A in Figure 1), whereas students in the control group received details about library loans (Panel B in Figure 1) with green indicating that the 30-day loan period per book has been adhered to. The design of the page, including the color scale and font used in both the treatment and control surveys, was designed in a way that ensure the surveys appear identical at a quick glance.

### Reminder and follow-up survey

A first follow-up survey, referred to as the "reminder" due to its inclusion of a short reminder of the treatment, was conducted in the week of lecture 8. It was administered online and participants were invited via the email addresses they provided at the end of the paper-based survey in week 4. The reminder survey included questions on student satisfaction and learning habits, alongside a reminder of the intervention. Thus, students from the treatment group received a reminder on study hours while students in the control group received a reminder on library loans - both in the same form as in the first survey.

The second follow-up survey was conducted about eight months after the first intervention, during the pre-exam period of the following term. This timing was chosen to analyse potential long-term insights while avoiding biases caused by a stressful exam period. Yet, as only 5 % of the baseline survey participated, the results are insufficient for meaningful analysis and not further reported in this paper.

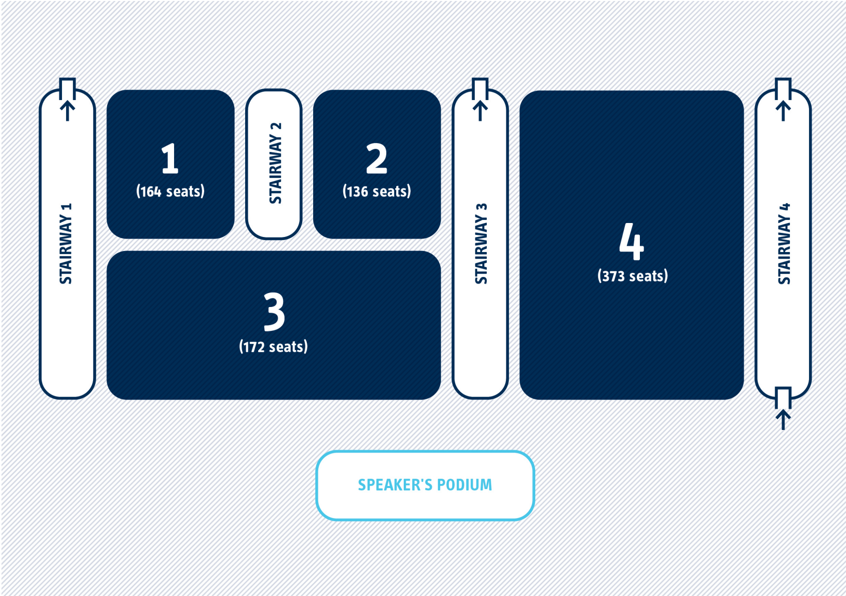
### 2.3 Randomization

In the lecture hall setting, we employed block randomization based on seats in the lecture hall (see Figure 3). By distributing the same surveys within the blocks, we minimized the potential for spillover effects among friends sitting in close proximity. Furthermore, each block included students from rows located at the front, center, and back of the lecture halls. This approach ensures that participation is not biased towards students seated in the front rows, who may be – as anecdotal evidence suggests – more engaged compared to those seated towards the back of the lecture halls.

In college settings, two prevalent randomization strategies are commonly employed for interventions, simple and stratified randomization. When employing a simple randomization approach as in Rury and Carrell (2023) or Evans et al. (2020), each participant is randomly assigned to either the treatment or control group with equal probability. In scenarios involving paper-based surveys, this can be executed by shuffling all surveys and distributing them randomly or by alternating treatment and control group surveys. However, simple randomization can increase the risk of spillover effects from treated peers, as there is no control over the proximity or interaction between participants. In stratified randomization as in Hardt et al. (2023) or Bettinger and Baker (2014), randomization is performed separately within each stratum or subgroup based on key variables such as age, gender, or educational background. This method aims to ensure balance across these subgroups. Nevertheless, in our case, stratified randomization was not feasible due to uncertainty regarding which students would attend the lecture, precluding the use of administrative data for stratification. Additionally, a pilot study conducted in a similar course setting a year prior to the intervention revealed considerable attrition among participants from one survey to the other, making it impractical to conduct pre-treatment surveys and apply the intervention in a follow-up.

To address these challenges, block randomization was employed. This method maintains balance across groups while minimizing the risk of confounding factors such as spillovers among peers or systematic attrition that could skew the results. As we show in section 3.3, the treatment and control groups achieved through block randomization are well-balanced across a wide range of characteristics.

Figure 3: Block randomization




Note: The figure illustrates the blocks used for randomization in the lecture hall. Students in blocks 1, 2, and 3 received the same questionnaire as did all students who chose a seat in block 4. To ensure a balanced distribution of students between blocks 1 to 3 and block 4, several rows in block 2 were closed off using barrier tape, creating the appearance that seat replacements or some additional maintenance was pending by technical staff.

# 3 Data and empirical strategy

## 3.1 Data sources and linkages

We use different datasets that cover students' tracking data from an online learning platform, responses from three surveys conducted during the study, and administrative data on each student provided by the university. The datasets were linked using students' names and their student ID numbers, with explicit consent obtained from participants for data use and linkage for research purposes. Due to the sensitive nature of the data, a data trustee handled the linkage process and provided the research team with pseudonymized data. A detailed description of each dataset is provided below.

**Surveys.** Three surveys were conducted throughout the study, with the first administered in class and the second and third distributed online. To encourage participation, incentives were offered through a lottery: respondents of the first and second surveys had a chance to win a 25 € prize, while students who completed both surveys were entered into a separate draw for a prize of 150 €. The surveys collected a wide range of variables, including (1) socio-demographic information such as prior education and employment besides their studies, (2) personality traits like risk aversion, procrastination, and academic motivaton, measured using established scales based on Sirois, Yang, and van Eerde (2019), Dweck, Chiu, and Hong (1995), and Vallerand et al. (1993), (3) satisfaction with the course, instructors, and personal performance, and (4) course-specific metrics such as attendance and goals for the final grade. The surveys were conducted to get valuable insights into both student behavior and their academic goals.

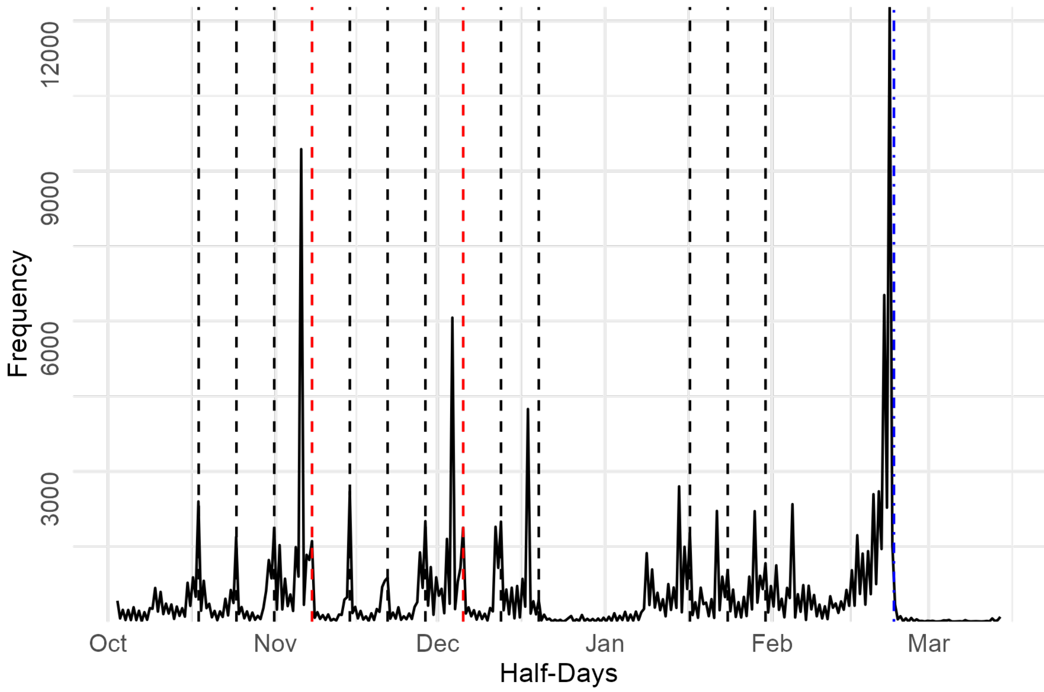
**Administrative data.** Administrative data, provided by the exam office and general administration, includes detailed records related to student exams. These datasets encompass overall academic performance metrics such as end-of-term GPA, total credits earned during the term, and specific details about course-related exams. The latter include results from the final exam and the students' performance in interim (but ungraded) online tests, as well as information on students' intentions to participate in exams. Notably, students at the university are required to self-register for exams, rather than being automatically enrolled, allowing for a precise record of their exam participation decisions. Thus, the data on the course-related exams includes information on whether students intended to participate in the exams but opted out before the exam or were absent due to illness.

**Tracking data.** Tracking data on students' study effort is gathered through the online learning platform *Moodle*, which is extensively used across universities in Europe, Asia, and the USA (Klaus Steitz, 2021; Moodle an Hochschulen e.V., 2024). Moodle provides virtual course rooms where work materials and learning activities for each module are organized. The platform tracks each student's activity with precise timestamps, offering detailed data on every click and download within the course environment. This comprehensive tracking data facilitates an in-depth analysis of student activities throughout the term. Despite its extensive use in course administration, it has not yet been employed to track students' engagement within a course after they have undergone a treatment.

In the specific course where the intervention took place, Moodle's course environment included various work materials and learning activities. *Work materials* comprise lecture notes, problem sets and solutions, and online videos. Lecture notes are uploaded regularly by the instructor, problem sets are made available a few days before tutorials, and solutions are provided shortly afterwards. The *learning activities* feature a discussion forum and voluntary online quizzes designed to review lecture content. Each of the eight quizzes was accessible for several days, allowing students to attempt them multiple times.

The regular updates of work materials and the periodic availability of online quizzes throughout the term were designed in cooperation with lecturer to encourage students to log in to Moodle and remain engaged with the course content. Figure 4 depicts daily Moodle activity aggregated across all students throughout the term. Engagement began with the first lecture, peaking just before or during lectures. Activity was generally low between lectures, particularly in the early weeks, though it increased before the winter break. During the break, Moodle activity almost ceased, likely because students spent their Christmas holidays with their families rather than engaging in coursework. After the break, activity surged as students prepared for exams, with a significant peak the day before the exam. Overall, student engagement was low during the first two-thirds of the semester, increasing only in the final third, particularly as exam dates approached. This pattern suggests that students tended to engage with course materials primarily shortly before exams or when exam dates were imminent. Notable spikes also occurred before lecture 4 (the intervention) and lecture 8 (the reminder), likely in response to announcements regarding the mid-term evaluation.

Figure 4: Total course-related activity on Moodle over the term




Note: The figure presents aggregated Moodle activity (”clicks”) for 498 students who consented to have their survey data linked with their Moodle usage. Activity is consolidated into 12-hour intervals over the course of the term. Dashed vertical lines represent lecture times, while red vertical lines correspond to the lectures associated with the treatment and reminder. The blue dot-dashed line indicates the timing of the exam. Source: Author’s illustration based on course-specific Moodle data.

**Data linkage.** The datasets were linked by a data trustee, as they were based on student ID numbers or names, making them too sensitive for direct access by the research team. The use of a data trustee was mandated by the university’s Data Protection Officer to ensure compliance with data privacy regulations. On the first pages of the paper-based survey in week 4, students were required to provide explicit consent for each specific linkage, including the Moodle data and exam office data. Ten students who did not consent to at least one of these linkages were excluded by the data trustee. In addition, the data trustee was responsible for scanning all paper-based surveys collected in week 4 and subsequently linking them with the online survey data from week 8.

## 3.2 Outcome variables

To assess the impact of the intervention on student performance, we examine four primary outcome variables, each providing valuable insight into different aspects of academic success. These variables allow us to evaluate objective performance measures regarding both the course and the study progress in general, offering a comprehensive understanding of how the intervention may have influenced student outcomes.

- **Course Pass/Fail Indicator**: This variable indicates whether students passed the specific course in which the intervention occurred. It serves as a direct measure of the intervention’s effectiveness in helping students meet course requirements.
- **Course Grade**: This continuous variable reflects students’ final grade in the course measured from 1 (very good) to 4 (worst passing grade). It provides a more nuanced view of academic success, helping us assess whether the intervention improved performance beyond simply passing.

- **Overall GPA at the end of the term**: The overall GPA measures students' academic performance across all courses. It captures potential spillover effects of the intervention on students' performance in courses outside the targeted course, reflecting broader changes in study habits and academic motivation.
- **Total Credits Earned by the End of the Term**: This variable measures the cumulative number of credits successfully earned by students upon the end of the term. It indicates overall academic engagement and completion, helping assess whether the intervention improved students' ability to complete courses and accumulate credits.

These four main outcome variables are sourced from the exam office and are linked to the survey data to differentiate between treatment and control groups. This linkage allows for a clear comparison of academic performance across groups, providing insights into the specific effects of the intervention. In addition, we also consider several secondary variables to further understand the mechanisms at play. Self-reported attendance (collected in week 8, four weeks after the intervention) provides insight into students' engagement with the course in the weeks following the treatment. Attendance is an important proxy for active participation and may help explain how the intervention influenced s tudents' ongoing commitment to their studies.

Furthermore, we analyze tracking data from the course platform, focusing on daily effort and click patterns on specific learning materials. These data points offer a detailed view of students' engagement with the course content, shedding light on how the intervention affected their study behaviors. For example, increased interaction with learning materials or consistent effort could indicate that the intervention helped students better manage their time and focus their efforts on key course components.

### 3.3 Descriptive statistics and balancing checks

The sample consists of 509 students who participated in the paper-based survey, with 264 students in the treatment group and 245 in the control group. Of these, 49.3 % are female, and 7.7 % hold citizenship from a country other than Germany. Notably, 60.7 % of the students have a migration background, defined as either the student or at least one of their parents being born abroad. This reflects the diverse demographic composition of the student body.

A significant proportion of the sample, 63.7 %, are first-generation students, which is likely influenced by the relatively young history of higher education in the region, where universities were only established in the 1960s. Additionally, 64.0 % of students are employed while studying, a figure that, while surprising in some international contexts, is typical in Germany and aligns with trends observed in other studies (Demir, Hertweck, Sandner, & Yükselen, 2024). The average high school GPA of the students is 2.54, suggesting a moderate academic background prior to university enrollment.

The vast majority (83.1 %) of students are enrolled in the undergraduate program in *Business and Economics* and is studying in their first or second undergraduate semester. Regarding their study effort, students report aiming to achieve an average of 27.0 credits out of a total of 30 credits expected in their study plan for the semester.

On average, they have attended 3.29 out of the four lectures held thus far. When asked about their study time, students report studying an average of 4.79 hours per week, which is below the 6 to 7 hours per week expected according to the course guidelines. They are rather satisfied with their own performance and the course's content.

In terms of personality traits, students exhibit only mild tendencies toward procrastination: On a scale ranging from -1 to 1, they slightly delay work (0.09) instead of completing work early (-0.27), have rather good self-stated time management skills (0.13) and rather low difficulties in keeping their schedules (-0.04). This suggests that students, on average, delay tasks slightly, yet it does not appear to be a pervasive issue. Interestingly, they believe to have a good use of their learning time (0.25).

Overall, the treatment and control groups are well-balanced, as confirmed by the balancing checks in Table 1, ensuring that any observed differences between the groups can be more confidently attributed to the intervention rather than pre-existing group differences. Most importantly, there is no difference in final high school grade, which is generally considered to be a strong predictor for success in university. The only statistically significant difference between the treatment and control groups pertains to the study program. Specifically, 86 % of students in the treatment group are enrolled in the *Business and Economics* program, compared to 80 % in the control group. The remaining students in both groups are enrolled in different programs or faculties and are taking the course as an elective. To account for this difference, we will include study program fixed effects in our regression models.

Table 1: Comparison of treatment and control group characteristics

| | Treatment | Control | Difference in means | |
|---|---|---|---|---|
| | *Mean* | *Mean* | *Raw diff.* | *p-value* |
| **Socio-demographic characteristics:** | | | | |
| Year of birth | 2002.806 | 2002.726 | 0.080 | 0.577 |
| Female | 0.481 | 0.506 | -0.025 | 0.573 |
| German | 0.928 | 0.917 | 0.011 | 0.645 |
| First-generation | 0.651 | 0.622 | 0.030 | 0.481 |
| Migration background | 0.633 | 0.578 | 0.055 | 0.200 |
| Student job | 0.638 | 0.643 | -0.004 | 0.919 |
| High-school GPA | 2.522 | 2.537 | -0.014 | 0.774 |
| **Study-related information:** | | | | |
| No. of terms at university | 1.630 | 1.616 | 0.014 | 0.858 |
| BA in *Business & Economics* | 0.860 | 0.800 | 0.060 | 0.074 |
| Credits planned for term | 27.152 | 26.845 | 0.306 | 0.514 |
| **Course-specific information:** | | | | |
| No. of lectures attended | 3.343 | 3.223 | 0.119 | 0.151 |
| Weekly effort (in hours) | 4.856 | 4.715 | 0.141 | 0.560 |
| Satisfaction with own performance | 0.055 | 0.081 | -0.027 | 0.596 |
| Satisfaction with content | 0.315 | 0.274 | 0.041 | 0.354 |
| **Personality traits:** | | | | |
| Delay work | 0.126 | 0.051 | 0.076 | 0.178 |
| Complete work early | -0.259 | -0.278 | 0.019 | 0.700 |
| Difficulties in keeping schedule | -0.054 | -0.033 | -0.021 | 0.698 |
| Good use of learning time | 0.238 | 0.256 | -0.018 | 0.712 |
| Good time management | 0.130 | 0.134 | -0.004 | 0.918 |
| Switch goals often | -0.288 | -0.292 | 0.004 | 0.939 |
| Rather risky behaviour | 0.140 | 0.126 | 0.014 | 0.789 |
| Intrinsic motivation | 0.053 | 0.054 | -0.002 | 0.954 |
| Familiar with Nutri Score | 0.865 | 0.853 | 0.012 | 0.689 |

| | Treatment | Control | Difference in means | |
|---|---|---|---|---|
| | *Mean* | *Mean* | *Raw diff.* | *p-value* |
| Number of Observations | 264 | 245 | | |

Notes: Comparison between mean values for treatment group and control group. *Raw diff.* indicates the differences between the means of the treatment with that of the control group. p-value indicates the p-value from a t-test comparing the means of treatment and control group. Personality traits always range from -1 to 1. The exact wording of the questions is in Appendix A. Some students selected "no answer" in the survey instead of choosing a specific option, resulting in missing values for certain variables. However, no student consistently chose "no answer" for every question. Additionally, some students skipped one or more questions. To minimize the loss of observations due to missing values, we imputed the relevant variables using the sample mean. These imputations were rare, with a maximum of fewer than 10 missing values for any given variable.

### 3.4 Processing of treatment information

The intervention was designed using a color-coded scale, ranging from red to green, with green representing the recommended 6-7 hours of weekly study effort (treatment) or a scale on loan periods in the library (control) as illustrated in Figure 1. The use of traffic-light colors was intentional, as they are familiar to most students and provide a clear, intuitive visual cue. To gauge students' familiarity with this type of color-coded scale, we included a question in the survey asking whether they are familiar with the Nutri-Score system.[3] Overall, 86 % of the respondents indicated that they were familiar with the Nutri-Score (see Table 1), suggesting that the traffic-light color scheme would be an effective means of communicating the study effort expectations.

Furthermore, we directly assessed whether students understood the treatment by including a control question in the survey. The treatment group was asked to report the recommended number of weekly study hours (6-7 hours), while the control group was asked about the loan periods for library materials. 93.5 % of the treatment group and 93.6 % of the control group provided correct answers. These high response rates suggest that the majority of students not only understood the information presented through the *Study Score*, but were also able to correctly process the relevant information. Overall, these findings provide strong evidence that students comprehended the treatment and were able to engage with the color-coded study effort scale.

In a follow-up survey conducted in week 8, we assessed how well students remembered the color-coded *Study Score* presented to them. Among the treatment group, 71.4 % recalled the traffic light system, compared to 56.2 % in the control group. This suggests that the combination of a color-coded scale with previously unfamiliar content (study effort requirements) had a more lasting impact on students' memory than the score paired with more commonly known information (library loan periods).

However, when asked to specify what the score represented, only 46.4 % of the treatment group could accurately describe its focus on weekly study effort. In contrast, 43.8 % of the control group were able to recall that their score related to library loan periods, suggesting that while both groups remembered the general concept of a color-coded system, the specific content was less frequently recalled.

Importantly, both groups largely remembered the correct subject of their respective *Study Score*. Only two students from the control group mistakenly associated their score with credit points, suggesting that

[3]The Nutri-Score is a color-coded nutritional label that helps consumers quickly assess the healthiness of food products, with scores ranging from green (A) for the healthiest options to red (E) for less nutritious choices. It takes into account factors like calories, sugar, salt, and fat content, making it an intuitive and widely recognized system (Sothey, 2021).

spillover effects from the treatment group on the control group were minimal. This result indicates that the intervention did not lead to significant cross-group discussion or confusion regarding the content of the *Study Score*. Apparently, the seating-based randomization strategy was effective in preserving the integrity of the experimental groups and ensuring that the content-specific effects of the intervention were not diluted through direct interactions between students in the treatment and control groups.

## 3.5 Data availability and attrition between surveys

In the initial survey, 509 students participated, with only 11 students opting not to allow their data to be linked. However, some students did not respond to all questions, resulting in missing data for certain variables. Additionally, some students could not be linked due to errors such as entering an unknown student ID. As a result, tracking data for the course is available for 98 % of the students. Administrative data, however, was not available for all students from other programs, which limits our ability to link this information for the full sample. Overall, administrative data is only available for 68 % of the students (see Table 6 in Appendix B).

Another issue is the panel attrition between the first survey (conducted in week 4 during the lecture) and the second survey (conducted in week 8 as an online survey including the reminder). Only 34 % of the initial sample participated in the second survey in week 8, representing a substantial drop in response rate - despite the large incentive of entering a lottery of 150 € for those who completed both surveys.[4] As summarized in Table 7, this attrition is statistically significant for several variables: students who participated in both surveys were, on average, less likely to be a first-generation student or have a migrant background, had a better high-school GPA and were slightly more ambitious in terms of the number of credits they intended to achieve by the end of the term, were more satisfied with their own performance and the lecture content, were more likely to complete work early, were more likely to state a good use of learning time and time management, and were generally more risk-averse. These patterns suggest that the reminder survey sample is positively selected, with students who were more engaged and self-assured being more likely to participate.

Panel attrition leads to a reduction in sample size and thereby inevitably diminishes statistical power. However, the balancing check presented in Table 8 in Appendix B indicates that, despite the partly systematic loss of participants, the treatment and control groups remain statistically comparable on key variables. No significant differences were observed when comparing the means of relevant characteristics between the two groups, suggesting that the groups for the strongly reduced sample are still well-matched. The only exception is a small difference in one of the measures for procrastination, with students in the treatment group reporting a slightly higher tendency to delay work compared to those in the control group. Importantly, other measures of procrastination remain balanced across groups. To mitigate the potential impact of this imbalance, we will include personality traits as covariates in the regression models, thereby controlling for individual differences that could influence the outcomes.

## 3.6 Empirical strategy

We provide estimates regarding the average treatment effects on the treated from OLS models. We compare the average outcomes of the control and treatment groups.

$$y_i = \alpha + \beta INFO_i + \gamma X_i + \mu Z_i + \tau_i + \epsilon_i \tag{1}$$

[4] The third survey (follow-up survey in the second semester) was ultimately completed by only 25 students (5 %) of the initial sample. Thus, we could not use the information collected in the follow-up survey in our results.

In Equation 1, $y_i$ are the four major outcome variables (1) course passed, (2) course grade, (3) overall GPA, and (4) overall credits as presented above. $INFO_i$ is an indicator variable that equals 1 if the students received the treatment on expected study effort, and 0 if the students received the information on library loans. $X_i$ are socio-demographic characteristics including age, gender, migration background, high-school GPA, first-generation status, as well as an indicator whether the student has a student job and whether they are satisfied with their financial situation. The vector $Z_i$ captures personality traits including the variables on risk-aversion, intrinsic motivation, and an index combining all variables on procrastination. $\tau_i$ are fixed effects of the study program. $\epsilon_i$ is an error term of which we assume that it is uncorrelated with the treatment variable due to the random assignment of students to treatment and control groups. Thus, the effect of the information treatment on the outcome variable is described by the coefficient $\beta$, our main coefficient of interest. Because all students in the treatment group received the treatment, we measure average treatment effects on the treated (ATT).

# 4 Results

## 4.1 Main results

This study examines the effect of a color-coded scale intervention on students' academic performance, specifically their probability of passing the course, their final course grade, their overall GPA and total credits achieved by the end of the term. The treatment provided students with a visual representation of the recommended number of study hours required to succeed in the course; the control group received an information regarding library loans. Table 2 presents a summary of the results: Panel A displays the effects on subsequent academic performance for participants who received only the initial intervention in week 4, while Panel B reports the combined effects of the initial intervention in week 4 and a follow-up reminder provided in week 8.

As provided in Panel A of Table 2, the initial treatment administered in week 4 had no statistically significant effects on students' academic performance. Specifically, there were no significant differences between the treatment and control groups in terms of the probability of passing the course (column 1), final course grade (column 2) or overall GPA (column 3). Only the number of total credits earned by the end of the term (column 4 in Panel A) is meaningful in magnitude but lacks statistical significance.

However, when students received the information twice, the treatment had a significant positive effect on students' academic performance as summarized in Panel B of Table 2. Students who received both the treatment *and* the reminder achieved an average course grade that was 0.51 points higher than those who received only the library loan information (column 2 in Panel B).[5] This increase is substantial and represents an improvement in grades of 21 %, given that the grading system allows for only four possible grades ranging from 1 (very good) to 4 (passed).

Furthermore, those who received the treatment twice earned, on average, 4.49 additional credits by the end of the term. Although this increase is economically meaningful – representing nearly the equivalent of completing an additional module – the effect is not statistically significant most likely due to the lack of statistical power.

[5]Please note that the negative coefficient indicates an improvement in grades, as the grading scale ranges from 1 (very good, similar to an A) to 4 (passed, similar to a D).

Table 2: Main results

| | *Dependent variable:* | | | |
|---|---|---|---|---|
| | *Course passed* | *Course grade* | *Overall GPA* | *Overall credits* |
| | (1) | (2) | (3) | (4) |
| **Panel A: Treatment in lecture 4 only** | | | | |
| Treatment | -0.02 | -0.02 | 0.05 | 2.33 |
| | (0.06) | (0.19) | (0.13) | (1.56) |
| Adjusted R-squared | 0.09 | 0.14 | 0.09 | 0.13 |
| Individuals | 325 | 205 | 152 | 210 |
| **Panel B: Treatment in lecture 4 and online reminder** | | | | |
| Treatment | 0.00 | -0.51** | 0.13 | 4.49 |
| | (0.07) | (0.23) | (0.14) | (2.72) |
| Adjusted R-squared | 0.17 | 0.24 | 0.37 | 0.28 |
| Individuals | 173 | 109 | 94 | 109 |
| Mean values | 0.59 | 2.39 | 2.93 | 15.74 |
| Control variables: | | | | |
| Demographic characteristics | yes | yes | yes | yes |
| High school GPA | yes | yes | yes | yes |
| Financial situation | yes | yes | yes | yes |
| Personality traits | yes | yes | yes | yes |
| Study program FE | yes | yes | yes | yes |

**Notes:** The table shows OLS estimation results from the regression specified in Equation 1. Four different outcome variables are presented in columns 1 to 4. First, an indicator whether a student passed the course (column 1), thus including all students except those who did not consent to have their exam data matched to the survey data. In column 2, we present the achieved course grade. These range from 1 (very good) to 4 (pass) with 5 indicating a fail. Students who de-register from the exam do not receive a grade, reducing the number of observations in column 2 compared to column 1 because we only include students if they received a grade. The results for the overall GPA and the overall number of credits achieved by the end of the semester, both excluding the intervention course, are presented in columns 3 and 4. Data on overall GPA and total credits are only available for the subset of students who participated in the intervention course exam. The number of observations differs between these two columns due to the inclusion of students with zero overall credits, who are excluded from the regression in column 3. Robust standard errors are in parentheses. Significance levels: *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

## 4.2 Heterogeneity

To further explore the potential for heterogeneous effects of the intervention, we conducted subgroup analyses based on gender, migration background and first-generation status. Table 3 presents the results. In columns (1a) and (1b), we examined the treatment effects by gender, splitting the sample into male and female students. The results indicate no significant differences between the two groups, as both male and female students exhibited similar responses to the treatment. This suggests that the intervention had a uniform impact on academic outcomes across genders, with no indication of differential effectiveness based on gender.

In contrast, when the sample was split based on migration background[6] in columns (2a) and (2b), we observed substantial heterogeneity in the treatment effects. Specifically, students with a migrant background showed a significantly stronger response to the intervention compared to their non-migrant counterparts. The coefficient for students with a migrant background is -0.92, which corresponds to an approximately 38 % increase in course grades. This result suggests that students from migrant backgrounds benefited much more from the intervention.

Similar results arise, when splitting the sample by their first-generation status[7] in columns (3a) and (3b). We observe that first-generation students exhibit a strong response to the treatment, with a course grade increase of 0.98 points, corresponding to an improvement of roughly 41 %. In contrast, the coefficient for students with at least one parent who has graduated is smaller and does not attain statistical significance.

These findings are noteworthy in the context of educational inequality. Migrant students often face additional challenges in their academic trajectories, such as language barriers, cultural differences, and challenges regarding their social integration (Hertweck & Schneider, 2025). Similarly, first-generation students frequently encounter unfamiliarity with higher education systems and a lack of social or familial academic support. The stronger response to the intervention among these students could indicate that the color-coded scale and its reinforcement through the reminder provided clearer guidance and structure, which might have been particularly beneficial for students facing such challenges. It is possible that the intervention helped reduce disparities in academic performance by offering a more accessible and actionable resource for these students.

### 4.3 Channels

A potential underlying mechanism for the intervention's effects may be the provision of a clearer understanding of the expected workload required to succeed in the course. To understand this potential mechanism, we examine the treatment effects separately for students who were uninformed about the expected workload during the first weeks of the course and those who were already informed. In the baseline survey in week 4, we asked students to proxy the required workload. Students who stated less than six hours per week are considered as being uninformed about the required workload.

Columns 1a and 1b of Table 4 present regression results separately for uninformed and informed students. The coefficient for uninformed students is $-0.64$, indicating an improvement in course grade by roughly 27 %. However, this effect is not statistically significant, which is likely due to the small sample size and the resulting lack of statistical power. On the other hand, we do not observe a meaningful coefficient for informed students, suggesting that their prior understanding of the required workload may have rendered the intervention less effective. These results imply that providing clearer expectations about the workload may have been particularly beneficial for students who initially lack an accurate understanding, but the intervention did not appear to offer additional benefits for those who were already informed.

When we differentiate between high- and low-ability students, as proxied by their high-school GPA, in columns 2a and 2b of Table 4, we observe that high-ability students strongly response to the treatment.

[6] Migrant background in this study is defined as either the student or at least one of their parents being born abroad.
[7] A student is defined to be a first-generation student if neither of their parents has graduated from a higher education institution.

Table 3: Heterogeneity by gender, migration background and first-generation status

| | *Dependent variable: Course grade* | | | | | |
|---|---|---|---|---|---|---|
| | Male students (1a) | Female students (1b) | Migration back-ground (2a) | No-migration back-ground (2b) | First-gen. students (3a) | No first-gen. students (3b) |
| Treatment | -0.47 | -0.32 | -0.92*** | -0.13 | -0.98** | -0.45 |
| | (0.33) | (0.52) | (0.27) | (0.56) | (0.37) | (0.42) |
| Adjusted R-squared | 0.33 | 0.02 | 0.44 | 0.18 | 0.25 | 0.33 |
| Individuals | 60 | 48 | 64 | 44 | 60 | 47 |
| Control variables: | | | | | | |
| Demographic charact. | yes | yes | yes | yes | yes | yes |
| High school GPA | yes | yes | yes | yes | yes | yes |
| Financial situation | yes | yes | yes | yes | yes | yes |
| Personality traits | yes | yes | yes | yes | yes | yes |
| Study program FE | yes | yes | yes | yes | yes | yes |

**Notes:** The table shows OLS estimation results from the regression specified in Equation 1 for students who received the treatment and the reminder (as in Panel B in Table 2. Regressions are run separately for subsamples based on gender in columns (1a) and (1b), migration background in columns (2a) and (2b), and first-generation status in columns (3a) and (3b), respectively. Migration background is defined as the student or at least one of their parents born abroad. A student is defined to be a first-generation student if neither of their parents has graduated from higher education. Robust standard errors are in parentheses. Significance levels: $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

The statistically significant coefficient of $-0.73$ translates into an improvement in their course grades as a result of the intervention by roughly 31 %. In contrast, the coefficient for low-ability students, presented in column 2b, also points into the expected direction, but lacks statistical significance. This pattern suggests that while low-ability students may have also benefited from the intervention, the effect is not as pronounced as among high-ability students, potentially due to greater variability.

One possible underlying explanation for these findings is that high-ability students can more easily integrate new information into their existing academic strategies. These students likely possess stronger foundational study habits and time management skills (for which we control), which may allow them to capitalize on the structured guidance provided by the color-coded scale. In contrast, low-ability students, who may struggle with more fundamental academic skills, might find it more difficult to implement the suggested study hours or to adjust their behavior accordingly, thus attenuating the impact of the intervention. Additionally, high-ability students may be more motivated and capable of translating academic recommendations into concrete improvements in performance, whereas low-ability students might face more substantial barriers that hinder their ability to fully benefit from such interventions.

In addition to examining the effects of the intervention on academic outcomes, we also explore potential changes in student behavior that may explain these outcomes. As shown in Table 5, Panel A and Panel B present the results for students who received the treatment in week 4 and those who additionally received the follow-up reminder in week 8. First, we analyze lecture attendance as a proxy for student engagement. The coefficient for attendance in Panel A is positive, suggesting that the treatment had the expected effect of encouraging students to attend more lectures. However, this effect is not statistically significant, likely due to the limited sample size and the survey-based nature of the attendance data. Similarly, as illustrated in Figure 6 in Appendix B, the self-reported weekly workload of students in the

Table 4: Channel I: Updating of information and ability

| | *Dependent variable: Course grade* | | | |
|---|---|---|---|---|
| | Uninformed students (1a) | Informed students (1b) | High-ability students (2a) | Low-ability students (2b) |
| Treatment | -0.64 | 0.04 | -0.73* | -0.41 |
| | (0.37) | (0.47) | (0.28) | (0.40) |
| Adjusted R-squared | 0.37 | 0.05 | 0.30 | 0.13 |
| Individuals | 46 | 58 | 52 | 56 |
| Control variables: | | | | |
| Demographic characteristics | yes | yes | yes | yes |
| High school GPA | yes | yes | yes | yes |
| Financial situation | yes | yes | yes | yes |
| Personality traits | yes | yes | yes | yes |
| Study program FE | yes | yes | yes | yes |

**Notes:** The table shows OLS estimation results from the regression specified in Equation 1 for students who received the treatment and the reminder (as in Panel B in Table 2. In columns (1a) and (1b), the sample is split according to the number of weekly hours students stated to be needed to complete the module - before the first treatment, Uninformed students are those that stated that less than six hours are required while informed students stated 6 or more hours. In columns (1a) and (1b), regressions are run separately for subsamples based on high-school GPA as a proxy for academic ability. High-ability students are those with an above-median high-school GPA, while low-ability students have a high-school GPA below median. Robust standard errors are in parentheses. Significance levels: *$p$ < 0.10, **$p$ < 0.05, ***$p$ < 0.01.

treatment group is higher than those of the students in the control group. These findings point into the direction that students in fact changed their behavior towards more self-paced studying for the course.

Next, we turn to the course-specific tracking data from the online learning platform, which includes daily login and activity information for almost all students over the duration of the course, spanning from October to February. We examine effort per day in columns 2a and 2b of Table 5, measured by the duration and activity on the platform, using two regression specifications. In column 2a, we include only the treatment indicator and individual fixed effects to control for unobserved heterogeneity at the individual level. In column 2b, we use the same set of individual controls as specified in Equation 1. In both specifications, we observe positive coefficients, indicating an increase in daily effort, but the effects are not statistically significant. Specifically, the coefficient suggests an increase of 1.85 to 3.70 minutes per day (or 13 to 26 minutes per week), corresponding to a 7 % to 14 % increase in daily effort. The increased effort is only pronounced for students who received both the treatment and the reminder (Panel B), which further highlights the importance of the reminder in sustaining or reinforcing the intervention's impact.

To gain deeper insights into how students engaged with the online platform, we explore specific types of activity, as every action (or "click") is tracked. These activities include participation in online tests (which students used to assess their familiarity with the course content, though these tests were not graded), performance on these tests, the time spent on them, downloading course materials (such as lecture notes, problem sets, and solutions), and watching video tutorials. Table 9 in Appendix B presents t-test results comparing the treatment and control groups for students who received both the intervention in week 4 and the reminder in week 8, focusing on selected activities on the platform.

The t-test results indicate that there are no statistically significant differences between the treatment and control groups in terms of their engagement with the various platform activities. However, there is a small tendency for treated students to spend less time on online tests during the semester and more time on the final online tests, though these differences are not statistically significant. For other activities, such as generally engaging in online tests or downloading course materials, there is no clear pattern of behavior that can be attributed to the treatment. An underlying reason may be that the lecturer regularly uploaded new course materials, such as lecture notes and problem sets, throughout the semester to keep students engaged with the platform. Because it was not possible for students to download all course materials at once at the beginning of the term, engagement with these resources was spread out over time. This may have contributed to the statistical insignificance of the results, as the timing and availability of materials could have led to more gradual and less consistent engagement, making it harder to detect a clear treatment effect on material downloads in the data.

Overall, while we find no significant differences in specific activities, there is a small tendency towards increased time spent on the platform as already pointed out in Table 5. These findings suggest that the intervention may have had a broader impact on students' overall approach to studying, rather than influencing specific actions like test-taking or material downloads. It appears that the intervention does not significantly alter particular academic activities but fostered general engagement with the platform as well as students' overall approach to studying.

# 5 Discussion

The presented results provide insights into the potential of informing students about expected workload, particularly through interventions like a color-coded scale, to improve their academic performance. While students who participated only in the initial intervention conducted in week 4 did not exhibit significant effects, those who also participated in the follow-up reminder in week 8 experienced a positive treatment

Table 5: Channel II: Learning effort

| | *Dependent variable:* | | |
|---|---|---|---|
| | *Attendance of lectures* | *Effort per Day* | *Effort per Day* |
| | (1) | (2a) | (2b) |
| **Panel A: Treatment in lecture 4 only** | | | |
| Treatment | 0.15 | -0.23 | -1.12 |
| | (0.30) | (1.37) | (0.86) |
| Adjusted R-squared | 0.14 | 0.04 | 0.00 |
| Observations | 159 | 43,248 | 43,248 |
| **Panel B: Treatment in lecture 4 and online reminder** | | | |
| Treatment | – | 3.70 | 1.85 |
| | – | (3.14) | (1.42) |
| Adjusted R-squared | – | 0.05 | 0.01 |
| Observations | – | 22,984 | 22,984 |
| Mean values | 3.32 | 26.08 | 26.08 |
| Control variables: | | | |
| All controls as in Table 2 | yes | no | yes |
| Individual fixed effects | no | yes | no |

**Notes:** Attendance per lecture is measured in number of lectures. It is self-stated by the students when surveyed in week 8 and thus not available for panel B. Effort per day is measured in minutes spent on the course-specific environment on the learning platform. Robust standard errors are in parentheses. Significance levels: *$p$ < 0.10, **$p$ < 0.05, ***$p$ < 0.01.

effect on their course grade. Despite the lack of statistically significant changes in specific self-paced learning behaviors, such as participation in online tests or downloading course materials, the data indicate a general trend towards increased engagement with the online learning platform. These findings suggest that interventions repeatedly clarifying workload expectations can have a broader impact on students' academic engagement.

Panel attrition in this study presents a notable issue, as students self-selected into the online survey and thereby into the reminder in week 8. This led to positive selection (see Table 7), with students from stronger academic backgrounds and higher ability more likely to engage with the reminder. Despite this, the results demonstrate a substantial impact of the intervention when combined with the reminder, especially for first-generation students and those with a migration background. If there were no self-selection issue into the reminder, it is reasonable to expect that the results would be similar - if not stronger. This is supported by the fact that our main regressions control for relevant demographic and personal characteristics, and by the heterogeneity analysis, which shows that first-generation and migrant students benefited most from repeated treatment. Thus, students who opted out of the reminder in week 8 would likely have experienced comparable or even greater improvements if they had received the treatment twice.

However, several important limitations should be considered when interpreting these results. First, the sample size used in the analysis was limited due to significant panel attrition. While the initial survey was completed by 509 students, only 34 % of these students participated in the follow-up online survey in week 8, resulting in a reduced sample size for the main analyses. This attrition likely reduced statistical power, particularly in detecting smaller treatment effects, and may explain the statistical insignificance of many of the results.

Due to the close collaboration with the lecturer of the course, all course materials - such as lecture notes, problem sets, and solutions - were regularly uploaded throughout the semester to maintain all students' engagement with the online learning platform. However, this incremental provision of additional learning material may have contributed to more diffuse engagement patterns, making it harder to pinpoint the intervention's impact on any particular activity within a short window of time. Future studies could address this issue by examining the timing and frequency of material releases or by testing interventions in settings with more centralized and immediate access to resources. Furthermore, the complexity of measuring engagement over an extended period with multiple, overlapping academic tasks underscores the challenge of capturing the full spectrum of student behavior and engagement. This limitation highlights the importance of considering both the timing and structure of interventions when designing future studies on student learning behavior.

Moreover, the study was conducted within the context of a single course at one university, limiting the generalizability of the findings. Although the course was one of the largest at the faculty, the specific characteristics of the student population and the course format may not be representative of other academic contexts or disciplines. Future research could benefit from replicating this study in different settings or across multiple institutions to assess the broader applicability of the intervention. Furthermore, the study did not investigate the long-term effects of the intervention on students' academic trajectories or their behavior in subsequent courses. Understanding whether the intervention has sustained effects on students' academic performance is up to future research.

Importantly, the intervention implemented in this study was extremely low-cost, essentially involving the use of a single visual cue (a color-coded scale) to convey information about study expectations. This contrasts with more resource-intensive interventions, such as tutoring or mentoring programs (Hardt et al., 2023; Angrist et al., 2009; Bettinger & Baker, 2014) or signing commitment devices (Himmler et al., 2019; Cagala et al., 2021). The simplicity and low cost of the intervention in our study make it an appealing option for institutions looking to implement scalable, cost-effective interventions to improve student performance. Despite the lack of more substantial intervention elements, such as personalized guidance or academic coaching, the results indicate that even minimal interventions can have a meaningful impact, particularly when combined with follow-up reinforcement. Future research could explore how and when such inexpensive interventions could be incorporated alongside other academic support strategies to maximize their effectiveness.

Finally, we cannot distinguish the specific effects of the text of the intervention from those of the color-coded scale. The intervention combined both elements, written information about the required study hours *and* the visual color-coded scale, which makes it impossible to isolate the individual impact of each component. While the color-coded scale may have provided a clear and visually engaging representation of the workload expectations, the text component can also have played a crucial role in conveying this information. Future studies could explore variations of the intervention by isolating these elements to determine whether the visual or textual components, or perhaps their interaction, are more influential in shaping students' academic outcomes.

# 6 Conclusion

This study investigates the effectiveness of a low-cost, color-coded scale intervention aimed at improving students' academic outcomes by clarifying expected study hours. The results indicate that while the initial intervention in week 4 had no immediate impact, the follow-up reminder provided in week 8 significantly improved participating students' course grades, with a substantial effect of approximately 0.51 points on average. This suggests that interventions that repeatedly reinforce academic expectations can have effects on student performance. The improvement in course grades, which represents a 21 % increase, underscores the potential of simple, cost-effective interventions to enhance academic outcomes.

Heterogeneity analyses reveal notable differences in the treatment effects across subgroups. Specifically, first-generation students and students with a migration background showed a significantly stronger response to the intervention, resulting in an increase in their course grade of around 41 and 38 %. This finding highlights the potential for targeted interventions to reduce educational inequality, particularly for groups that may face additional challenges in navigating academic expectations. Additionally, high-ability students demonstrate stronger gains in grades from the intervention than low-ability students, suggesting that students with better academic preparation may benefit more from clarifying workload expectations.

Thus, while the intervention appears to hold promise for improving outcomes for some marginalized groups, it does not necessarily address all forms of academic inequality. Nevertheless, it emphasizes the importance of considering student characteristics when designing interventions, as tailored strategies may enhance the overall effectiveness of academic support efforts. Further research is needed to assess whether interventions like this could be tailored or supplemented with additional support to better serve students across a broader spectrum of academic abilities and backgrounds.

Despite these positive findings, the study does have some limitations. Due to panel attrition, the small sample size reduced the statistical power of the analyses. Furthermore, the course structure, with course materials released incrementally over the semester, may have diluted the impact of the intervention on specific behaviors, such as downloading materials or participating in online tests. The results related to detailed platform activities were therefore more ambiguous, pointing to the challenges of measuring behavior in contexts where engagement is spread out over time.

Importantly, the intervention itself was extremely cost-effective. Just a simple visual cue in the form of a color-coded scale improved course grades by on average 21 %. Unlike more resource-intensive interventions such as tutoring or mentoring, this intervention is easily scalable and requires minimal resources to implement. The findings suggest that even small-scale, low-cost interventions can have a significant impact on academic outcomes, particularly when reinforced over time. Thus, the study presents a compelling case for universities to consider simple yet effective strategies to support students' academic success, especially in large courses or resource-constrained settings.

Summing up, this study provides solid ground for future research and interventions aimed at improving student performance. Future studies could explore how such interventions work in different academic contexts, investigate the long-term effects on student outcomes, and examine how individual characteristics like motivation and self-regulation influence responses to the intervention. Additionally, further research could investigate how interventions can be personalized to meet the needs of specific student groups, such as those from migration backgrounds or varying academic abilities, to maximize their effectiveness in reducing educational inequalities.

# References


Andor, M. A., Goette, L., Price, M. K., Schulze Tilling, A., & Tomberg, L. (2023). Differences in How and Why Social Comparison and Real-Time Feedback Impact Resource Use: Evidence from a Field Experiment. *National Bureau of Economic Research*, *w31845*. doi: 10.3386/w31845

Angrist, J., Lang, D., & Oreopoulos, P. (2009, January). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, *1*(1), 136-63. doi: 10.1257/app.1.1.136

Antúnez, L., Giménez, A., Maiche, A., & Ares, G. (2015). Influence of interpretation aids on attentional capture, visual processing, and understanding of front-of-package nutrition labels. *Journal of Nutrition Education and Behavior*, *47*(4), 292–299. doi: 10.1016/j.jneb.2015.02.010

Azmat, G., Bagues, M., Cabrales, A., & Iriberri, N. (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), 3714-3736. doi: 10.1287/mnsc.2018.3131

Babcock, P., & Marks, M. (2011). The falling time cost of college: Evidence from half a century of time use data. *Review of Economics and Statistics*, *93*(2), 468–478.

Bengart, P., & Vogt, B. (2023). Effects and interactions of labels' color scheme and the individual difference variable lay rationalism on pro-environmental choices. *Journal of Environmental Psychology*, *87*, 101998. doi: 10.1016/j.jenvp.2023.101998

Bettinger, E. P., & Baker, R. B. (2014). The effects of student coaching: An evaluation of a randomized experiment in student advising. *Educational Evaluation and Policy Analysis*, *36*(1), 3-19. doi: 10.3102/0162373713500523

Bound, J., Lovenheim, M. F., & Turner, S. (2010). Why have college completion rates declined? an analysis of changing student preparation and collegiate resources. *American Economic Journal: Applied Economics*, *2*(3), 129–157.

Brint, S., & Cantwell, A. M. (2010). Undergraduate time use and academic outcomes: Results from the University of California Undergraduate Experience Survey 2006. *Teachers College Record*, *112*, 2441–2470.

Broda, M., Yun, J., Schneider, B., Yeager, D. S., Walton, G. M., & Diemer, M. (2018). Reducing inequality in academic success for incoming college students: A randomized trial of growth mindset and belonging interventions. *Journal of Research on Educational Effectiveness*, *11*(3), 317-338. doi: 10.1080/19345747.2018.1429037

Cagala, T., Glogowsky, U., & Rincke, J. (2021). Detecting and preventing cheating in exams: Evidence from a field experiment. *Journal of Human Resources*. doi: 10.3368/jhr.0620-10947R1

Castleman, B. L., & Meyer, K. E. (2020). Can text message nudges improve academic outcomes in college? Evidence from a west virginia initiative. *The Review of Higher Education*, *43*(4), 1125–1165. doi: 10.1353/rhe.2020.0015

Castleman, B. L., & Page, L. C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economic Behavior Organization*, *115*, 144-160. doi: 10.1016/j.jebo.2014.12.008

Castleman, B. L., & Page, L. C. (2016). Freshman year financial aid nudges: An experiment to increase FAFSA renewal and college persistence. *Journal of Human Resources*, *51*(2), 389–415. doi: 10.3368/jhr.51.2.0614-6458R

Clark, D., Gill, D., Prowse, V., & Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, *102*(4), 648–663. doi: 10.1162/rest_a_00864

Demir, G., Hertweck, F., Sandner, M., & Yükselen, I. (2024). Coworker networks from student jobs: A flying start at labor market entry? *Ruhr Economic Papers No. 1127*. doi: 10.4419/96973309

Dobronyi, C. R., Oreopoulos, P., & Petronijevic, U. (2019). Goal setting, academic reminders, and college success: A large-scale field experiment. *Journal of Research on Educational Effectiveness*, *12*(1), 38-66. doi: 10.1080/19345747.2018.1517849

Dweck, C. S., Chiu, C. Y., & Hong, Y. Y. (1995). Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological Inquiry*, *6*(4), 267–285. doi: 10.1207/s15327965pli0604_1

Evans, W. N., Kearney, M. S., Perry, B., & Sullivan, J. X. (2020). Increasing community college completion rates among low-income students: Evidence from a randomized controlled trial evaluation of a case-management intervention. *Journal of Policy Analysis and Management*, *39*(4), 930-965. doi: 10.1002/pam.22256

Hardt, D., Nagler, M., & Rincke, J. (2023). Tutoring in (online) higher education: Experimental evidence. *Economics of Education Review*, *92*, 102350. doi: 10.1016/j.econedurev.2022.102350

Hertweck, F., Kistner, M., & Maffia, D. (2025). Recruitment, consent, and selective attrition in RCTs and surveys in higher education research. *Mimeo*.

Hertweck, F., & Schneider, K. (2025). Hochschule. In M. Küpker & A. Winkler (Eds.), *Bildungsbericht Ruhr 2024: Bildung in der Region gemeinsam gestalten.* Essen: Ruhr Futur & Regionalverband Ruhr.

Himmler, O., Jäckle, R., & Weinschenk, P. (2019). Soft commitments, reminders, and academic performance. *American Economic Journal: Applied Economics*, *11*(2), 114-42. doi: 10.1257/app.20170288

Klaus Steitz. (2021). *Moodle-Instanzen an deutschsprachigen Hochschulen.* `https://blog.e-learning.tu-darmstadt.de/2010/04/09/moodle-instanzen-an-deutschsprachigen-hochschulen/`. (Accessed: 2024-06-30)

Li, H.-H. (2018). Do mentoring, information, and nudge reduce the gender gap in economics majors? *Economics of Education Review*, *64*, 165-183. doi: 10.1016/j.econedurev.2018.04.004

Moodle an Hochschulen e.V. (2024). *Unser Antrieb.* `https://moodle-an-hochschulen.de/`. (Accessed: 2024-06-30)

Oreopoulos, P., & Petronijevic, U. (2018). Student coaching: How far can technology go? *Journal of Human Resources*, *53*(2), 299–329. doi: 10.3368/jhr.53.2.1216-8439R

Oreopoulos, P., & Petronijevic, U. (2019). The remarkable unresponsiveness of college students to nudging and what we can learn from it. *NBER Working Paper No. 26059*.

Rury, D., & Carrell, S. E. (2023). Knowing what it takes: The effect of information about returns to studying on study effort and achievement. *Economics of Education Review*, *94*, 102400. doi: 10.1016/j.econedurev.2023.102400

Sirois, F. M., Yang, S., & van Eerde, W. (2019). Development and validation of the general procrastination scale (gps-9): A short and reliable measure of trait procrastination. *Personality and Individual Differences*, *146*, 26–33. doi: 10.1016/j.paid.2019.03.039

Sothey, F. (2021). 7 European countries team up to propel Nutri-Score rollout. *Food Navigator Europe*. `https://www.foodnavigator.com/Article/2021/02/12/7-European-countries-team-up-to-propel-Nutri-Score-rollout`. (Accessed: 2024-06-30)

Thøgersen, J., & Nielsen, K. S. (2016). A better carbon footprint label. *Journal of Cleaner Production*, *125*, 86–94. doi: 10.1016/j.jclepro.2016.03.098

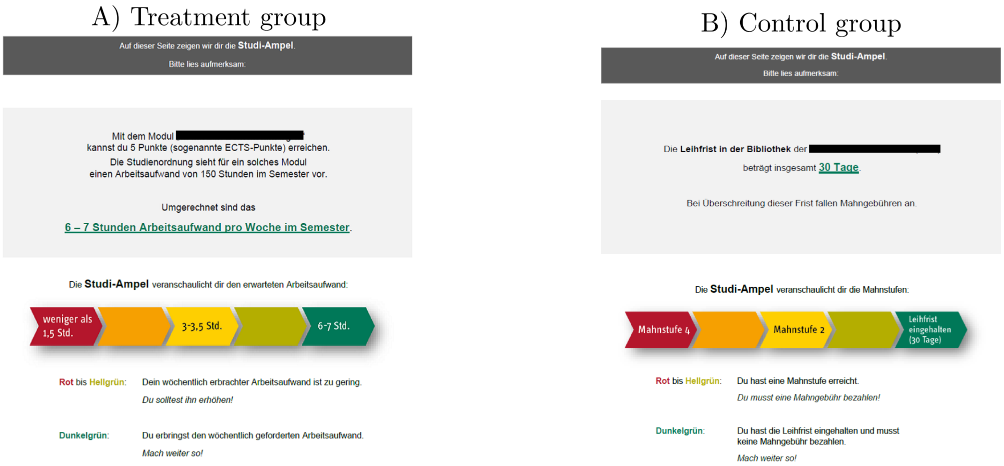
Vallerand, R. J., Pelletier, L. G., Blais, M. R., Brière, N. M., Senécal, C. B., & Vallières, É. F. (1993). Academic motivation scale (ams-c 28) college (cegep) version. *Educational and Psychological Measurement*, *52*, 1992–1993.

Waechter, S., Sütterlin, B., Borghoff, J., & Siegrist, M. (2016). Letters, signs, and colors: How the display of energy-efficiency information influences consumer assessments of products. *Energy Research & Social Science*, *15*, 86–95. doi: 10.1016/j.erss.2016.03.022

# A Details of the survey

## A.1 Treatment in the original German version

Figure 5: Information provision as part of paper-based survey in German




Note: Original page with treatment (panel A) and control information (panel B) in German.

## A.2 Survey questions in paper-based baseline survey (week 4)

Below, we provide an overview of the variables collected and the questions asked in the baseline survey. Please note that the grouping of the variables does not correspond to the order in which the questions were presented. The complete questionnaires (including introductory information on data protection and consent to data linkages) are available on request.

### Socio-Demographic variables

- Month of birth
- Year of birth
- Gender
- German citizenship
- Were you or at least one of your parents born abroad?
- Have at least one of your parents graduated from a university or university of applied sciences?
- Do you work alongside your studies?
  - If yes, how many hours do you work on average per week?
- How satisfied are you with your financial situation?

### Variables related to current studies

- Student ID
- University

- Study program
- Current term
- Is your current course of study your first degree?
- Year of high-school degree
- High-school GPA
- How many ECTS credits do you plan to earn in total during the winter semester 2023/24?

**Student satisfaction**

- How satisfied are you overall with your academic performance so far?
- How satisfied are you with the content of your studies?
- How satisfied are you with your lecturer?
- How satisfied are you overall with your study program?

**Course-related variables**

- How many hours do you invest per week in total for the module (...) including attending lectures and exercises?
- How many lectures of the module (...) have you attended so far this semester?
- What is the minimum final grade you aim to achieve in the (...) module?
- I make good use of my study time for this course.
- Please estimate: How many hours of study effort per week (including attendance at lectures, exercises, etc.) are required throughout the entire semester to successfully complete the (...) module?

**Personality traits**

In total, we asked nine questions regarding the students' personality. The questions are based on (Vallerand et al., 1993; Dweck et al., 1995; Sirois et al., 2019). Students could answer across a 4-point-Likert scale ranging from "fully disagree" to "fully agree" with an additional possibility to not to disclose an answer.

- I study because my course of study allows me to continue learning many things that interest me.
- I study because I want to prove to myself that I can be successful in my studies.
- Please assess yourself: Are you generally a risk-taking person, or do you tend to avoid risks? *(4-point-Likert scale ranging from not risk-loving to very risk-loving).*
- I often don't dedicate much time to this course (e.g., due to other activities).
- I often complete tasks that I had intended to finish days ago.
- I usually hesitate before starting the work I need to do.
- I often complete a task earlier than necessary.
- I often set a goal, but later decide on a different one.
- I find it difficult to stick to a study schedule.

**Nutri Score**

- Are you familiar with the color traffic light system on food products ("Nutri-Score")?

# B Additional tables and figures

## B.1 Data availability

The following table presents the data availability of the different data sources. Due to the small number of responses available in the follow-up survey, the data from the follow-up survey has not been used.

Table 6: Data availability after linkage

| | Treatment | Control | Total | Data availability |
|---|---|---|---|---|
| Baseline Survey | 264 | 245 | 509 | 100 % |
| Tracking data | 261 | 237 | 498 | 98 % |
| Admin data | 181 | 167 | 348 | 68 % |
| Reminder | 84 | 89 | 173 | 34 % |
| Follow-up survey | 8 | 17 | 25 | 5 % |

## B.2 Attrition between surveys in week 4 and week 8

Table 7: Comparison of students who completed first or both surveys)

| | Both surveys | First survey | Difference in means | |
|---|---|---|---|---|
| | *Mean* | *Mean* | *Raw diff.* | *p-value* |
| **Socio-demographic characteristics:** | | | | |
| Year of birth | 2002.704 | 2002.800 | -0.096 | 0.524 |
| Female | 0.512 | 0.484 | 0.028 | 0.550 |
| German | 0.913 | 0.928 | -0.015 | 0.558 |
| First-generation | 0.585 | 0.663 | -0.078 | 0.084 |
| Migration background | 0.553 | 0.635 | -0.082 | 0.075 |
| Student job | 0.659 | 0.631 | 0.028 | 0.531 |
| High-school GPA | 2.415 | 2.588 | -0.173 | 0.001 |
| **Study-related information:** | | | | |
| No. of terms at university | 1.616 | 1.627 | -0.012 | 0.892 |
| BA in *Business & Economics* | 0.815 | 0.839 | -0.024 | 0.498 |
| Credits planned for term | 27.619 | 26.688 | 0.931 | 0.050 |
| **Course-specific information:** | | | | |
| No. of lectures attended | 3.319 | 3.268 | 0.051 | 0.562 |
| Weekly effort (in hours) | 4.576 | 4.897 | -0.321 | 0.156 |
| Satisfaction with own performance | 0.135 | 0.033 | 0.102 | 0.047 |
| Satisfaction with content | 0.365 | 0.259 | 0.106 | 0.016 |
| **Personality traits:** | | | | |
| Delay work | 0.029 | 0.121 | -0.093 | 0.117 |
| Complete work early | -0.184 | -0.312 | 0.127 | 0.019 |
| Difficulties in keeping schedule | -0.100 | -0.015 | -0.085 | 0.141 |
| Good use of learning time | 0.339 | 0.199 | 0.141 | 0.006 |
| Good time management | 0.185 | 0.104 | 0.080 | 0.053 |

| | Both surveys | First survey | Difference in means | |
|---|---|---|---|---|
| | *Mean* | *Mean* | *Raw diff.* | *p-value* |
| Switch goals often | -0.279 | -0.295 | 0.017 | 0.745 |
| Rather risky behavior | 0.065 | 0.169 | -0.105 | 0.064 |
| Intrinsic motivation | 0.074 | 0.043 | 0.032 | 0.266 |
| Familiar with Nutri Score | 0.895 | 0.840 | 0.055 | 0.068 |
| Number of Observations | 173 | 336 | | |

Notes: Comparison between mean values for students who participated only in the survey in week 4 and those who also participated in week 8 - irrespective of the treatment allocation. All values are available only for those students for whom their tracking and administrative data is available, thus causing the drop in number of observations compared to Table 1. *Raw diff.* indicates the differences between the means of relevant variables between students who participated in one vs. those who participated in two surveys. *p-value* indicates the p-value from a t-test comparing the means of these groups of students. Personality traits always range from -1 to 1. The exact wording of the questions is in Appendix A.

## B.3 Balancing checks based on participation in reminder

The following table presents the balancing checks of variables asked in the paper-based survey in week 4 with the sample who also responded in the reminder survey in week 8.

Table 8: Comparison of treatment and control group characteristics (based on reminder)

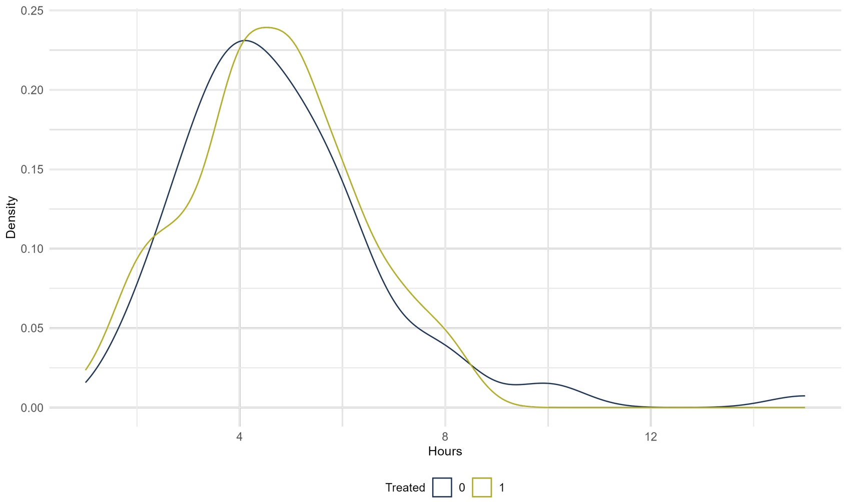
| | Treatment | Control | Difference in means | |
|---|---|---|---|---|
| | *Mean* | *Mean* | *Raw diff.* | *p-value* |
| **Socio-demographic characteristics:** | | | | |
| Year of birth | 2002.786 | 2002.627 | 0.159 | 0.518 |
| Female | 0.571 | 0.455 | 0.116 | 0.126 |
| German | 0.929 | 0.898 | 0.031 | 0.476 |
| First-generation | 0.619 | 0.554 | 0.065 | 0.383 |
| Migration background | 0.595 | 0.512 | 0.083 | 0.275 |
| Student job | 0.655 | 0.663 | -0.008 | 0.911 |
| High-school GPA | 2.480 | 2.353 | 0.128 | 0.162 |
| **Study-related information:** | | | | |
| No. of terms at university | 1.598 | 1.632 | -0.034 | 0.805 |
| BA in *Business & Economics* | 0.798 | 0.831 | -0.034 | 0.570 |
| Credits planned for term | 27.679 | 27.562 | 0.116 | 0.875 |
| **Course-specific information:** | | | | |
| No. of lectures attended | 3.361 | 3.279 | 0.081 | 0.572 |
| Weekly effort (in hours) | 4.765 | 4.397 | 0.368 | 0.248 |
| Satisfaction with own performance | 0.116 | 0.154 | -0.038 | 0.638 |
| Satisfaction with content | 0.387 | 0.345 | 0.042 | 0.533 |

| | Treatment | Control | Difference in means | |
|---|---|---|---|---|
| | *Mean* | *Mean* | *Raw diff.* | *p-value* |
| **Personality traits:** | | | | |
| Delay work | 0.115 | -0.053 | 0.168 | 0.079 |
| Complete work early | -0.197 | -0.172 | -0.025 | 0.780 |
| Difficulties in keeping schedule | -0.126 | -0.075 | -0.052 | 0.578 |
| Good use of learning time | 0.280 | 0.396 | -0.116 | 0.160 |
| Good time management | 0.185 | 0.185 | -0.000 | 0.999 |
| Switch goals often | -0.326 | -0.234 | -0.091 | 0.265 |
| Rather risky behavior | 0.075 | 0.055 | 0.020 | 0.837 |
| Intrinsic motivation | 0.067 | 0.082 | -0.015 | 0.742 |
| Familiar with Nutri Score | 0.876 | 0.913 | -0.038 | 0.412 |
| Number of Observations | 84 | 89 | | |

Notes: Comparison between mean values for treatment group and control group for variables asked in the initial paper-based survey in week 4 but only for those students who also participated in the online reminder survey in week 8. *Raw diff.* indicates the differences between the means of the treatment with that of the control group. p-value indicates the p-value from a t-test comparing the means of treatment and control group. Personality traits always range from -1 to 1. The exact wording of the questions is in Appendix A.

## B.4 Self-reported weekly workload

Figure 6: Density of self-reported weekly workload




Note: The figure illustrates the density of the self-reported weekly workload in hours as stated by the students in the online survey in week 8. The figure distinguished between students in the treatment (yellow line) and in the control group (gray line).

## B.5 Detailed activity on online platform

Table 9: T-tests comparing detailed activity on online platform

| **Detailed activity on online platform** | **Treament group** | **Control group** | **Difference in means: Treatment - Control** | |
|---|---|---|---|---|
| | *Mean* | *Mean* | *Raw difference* | *p-value* |
| **Download of course material:** | | | | |
| Lecture Notes for week 9/10 | 0.878 | 0.852 | 0.026 | 0.625 |
| Lecture Notes for week 10/11 | 0.817 | 0.818 | -0.001 | 0.985 |
| Lecture Notes for week 11/12 | 0.780 | 0.795 | -0.015 | 0.813 |
| Lecture Notes for week 12/13 | 0.780 | 0.807 | -0.026 | 0.674 |
| Problem Set #4 (PS4) | 0.805 | 0.795 | 0.009 | 0.879 |
| Problem Set #5 (PS5) | 0.756 | 0.784 | -0.028 | 0.667 |
| Solution to PS4 | 0.634 | 0.636 | -0.002 | 0.976 |
| Solution to PS5 | 0.659 | 0.648 | 0.011 | 0.883 |
| **Online test activity:** | | | | |
| Completion of test #3 (T3) | 0.220 | 0.307 | -0.087 | 0.198 |
| Completion of test #4 (T4) | 0.207 | 0.239 | -0.031 | 0.626 |
| Completion of test #5 (T5) | 0.171 | 0.136 | 0.034 | 0.538 |
| Completion of test #6 (T6) | 0.171 | 0.250 | -0.079 | 0.206 |
| Completion of test #7 (T7) | 0.134 | 0.216 | -0.082 | 0.161 |
| Completion of test #8 (T8) | 0.378 | 0.352 | 0.026 | 0.729 |
| Time spent on T3 | 41.220 | 53.682 | -12.462 | 0.552 |
| Time spent on T4 | 38.220 | 51.170 | -12.951 | 0.482 |
| Time spent on T5 | 34.488 | 41.670 | -7.183 | 0.683 |
| Time spent on T6 | 70.280 | 54.977 | 15.303 | 0.597 |
| Time spent on T7 | 47.305 | 34.920 | 12.384 | 0.521 |
| Time spent on T8 | 154.073 | 128.966 | 25.107 | 0.592 |
| Correct answers in T3 | 0.626 | 0.692 | -0.066 | 0.538 |
| Correct answers in T4 | 0.458 | 0.454 | 0.003 | 0.969 |
| Correct answers in T5 | 0.440 | 0.534 | -0.095 | 0.475 |
| Correct answers in T6 | 0.290 | 0.290 | -0.001 | 0.994 |
| Correct answers in T7 | 0.350 | 0.325 | 0.025 | 0.773 |
| Correct answers in T8 | 0.656 | 0.659 | -0.003 | 0.972 |
| **Download of video for online tutorials:** | | | | |
| Tutorial video #2 | 0.622 | 0.568 | 0.054 | 0.478 |
| Tutorial video #3 | 0.524 | 0.523 | 0.002 | 0.983 |
| Tutorial video #4 | 0.500 | 0.511 | -0.011 | 0.883 |

**Notes:** Mean values are based on 82 students in the treatment group and 88 students in the control group. All of these students participated in the initial intervention (receiving the treatment or control information in week 4) and a reminder in week 8. All activity is post-reminder, i.e., starting in teaching week 8 at the earliest. Download of course materials is the share of students who downloaded the relevant files. Online test activity is measured as the share of students who completed the respective tests, the time spent on completing the test (in minutes, conditional on completing the test), and the share of correct answers (conditional on completing the test). The download of the videos of the online tutorials is measured as the share of students who downloaded the video.