Behavioral Interventions and Students' Success at University

Evidence from Randomized Field Experiments

Dissertation
zur Erlangung des Doktorgrades
der Wirtschaftswissenschaftlichen Fakultät
der Georg-August-Universität Göttingen

vorgelegt von
Raphael Brade
aus München

Göttingen, 2021

Erstgutachter: Prof. Dr. Robert Schwager
Zweitgutachter: Prof. Dr. Oliver Himmler
Drittprüfer: Prof. Dr. Thomas Kneib
Tag der mündlichen Prüfung: 22.12.2021

Abstract

Research over the past decades has documented the various individual and social returns to tertiary education. In many countries, however, the realization of those returns is jeopardized by high dropout rates and delayed graduation. Investigating the underlying reasons for the lack of academic success, the recent literature points to the critical role of behavioral biases and the resulting need for interventions that address them. Against this background, this thesis investigates the effectiveness of several low-cost and easy-to-implement measures that aim to increase students' success at university, thereby contributing to the young but growing body of research on behaviorally informed interventions in higher education.

Using randomized field experiments for causal evaluation, the four main chapters of this thesis study i) whether social information about the past behavior of others can raise remedial math course participation and if increased course attendance translates into higher academic performance in the first year of studies (Chapter 2), ii) whether relative performance feedback on accumulated course credits can increase short- and long-term academic achievement (Chapters 3 and 4), and iii) whether a non-binding commitment to study according to the recommended study structure and reminder letters are able to help students succeed in university (Chapter 5).

Taken together, this thesis provides a broad range of evidence for the versatility and potential of behaviorally informed measures. They can improve academic performance directly or indirectly, provide decision-relevant information that is frequently not available to students, or be used to address specific biases – such as procrastination and limited attention. Crucially, Chapters 4 and 5 provide some of the first evidence that behavioral interventions can have substantial positive effects on long-term academic success. Yet, the chapters of this thesis also paint a nuanced picture and show that effects are often heterogeneous. In many situations, a targeted implementation of the proposed measures may therefore be necessary to prevent unwanted effects. At the same time, examples of how econometric methods can be used to identify the beneficiaries of an intervention more precisely are provided.

Keywords: Economics of Education; Behavioral Economics of Education; Higher Education; Randomized Field Experiment; Remedial Course; Social Information; Relative Performance Feedback; Commitment Device; Reminder

Zusammenfassung

In den letzten Jahrzehnten hat die Forschung die verschiedenartigen individuellen und sozialen Erträge der Hochschulbildung dokumentiert. Die Realisierung dieser Erträge wird jedoch in vielen Ländern durch hohe Abbruchquoten und lange Studiendauern beeinträchtigt. Bei der Untersuchung der Ursachen für fehlenden Studienerfolg verweist die neuere Literatur auf die entscheidende Rolle von sogenannten "behavioral biases" und den daraus resultierenden Bedarf an Maßnahmen, die diese ausgleichen. Vor diesem Hintergrund untersucht diese Dissertation die Wirksamkeit mehrerer kostengünstiger und einfach durchzuführender Maßnahmen zur Steigerung des Studienerfolgs und leistet damit einen Beitrag zu der noch jungen, aber wachsenden Forschung über verhaltensökonomisch motivierte Interventionen im Hochschulbereich.

Unter Verwendung von randomisierten Feldexperimenten zur kausalen Evaluation untersuchen die vier zentralen Kapitel dieser Dissertation, i) ob soziale Informationen über das frühere Verhalten anderer die Teilnahme an einem Mathe Vorkurs erhöhen können und ob eine erhöhte Kursteilnahme zu besseren akademischen Leistungen im ersten Studienjahr führt (Kapitel 2), ii) ob relatives Leistungsfeedback bezüglich der akkumulierten Leistungspunkte den kurz- und langfristigen akademischen Erfolg erhöhen kann (Kapitel 3 und 4), und iii) ob eine unverbindliche Selbstverpflichtung, gemäß der empfohlenen Studienstruktur zu studieren, und Erinnerungsschreiben in der Lage sind, den Studienerfolg von Studierenden zu fördern (Kapitel 5).

Insgesamt liefert diese Dissertation ein breites Spektrum an Belegen für die Vielseitigkeit und das Potenzial verhaltensökonomischer Maßnahmen. Sie können die akademischen Leistungen direkt oder indirekt verbessern, entscheidungsrelevante Informationen liefern, die den Studierenden häufig nicht zur Verfügung stehen, oder sie können spezifischen Abweichungen vom rationalen Verhaltensmodell – wie Prokrastination und eingeschränkte Aufmerksamkeit – entgegenwirken. Hervorzuheben ist, dass die Kapitel 4 und 5 einige der ersten Belege dafür liefern, dass verhaltensökonomisch motivierte Maßnahmen auch langfristig erhebliche positive Auswirkungen auf den Studienerfolg haben können. Die Kapitel der Dissertation zeichnen jedoch auch ein differenziertes Bild und zeigen, dass Effekte häufig heterogen sind. In vielen Situationen kann daher eine zielgerichtete Umsetzung der vorgeschlagenen Maßnahmen erforderlich sein, um ungewollte Auswirkungen zu verhindern. Gleichzeitig werden aber auch Beispiele gegeben, wie ökonometrische Methoden eingesetzt werden können, um die Nutznießer einer Intervention genauer zu identifizieren.

Acknowledgements

First of all, I am very grateful for the continuous guidance and support that my supervisors Robert Schwager and Oliver Himmler provided throughout the preparation of this thesis. I would also like to thank Thomas Kneib for being the third member of my thesis and examination committee. Special thanks also go to Robert Jäckle for his excellent supervision and collaboration in the research project.

I would also like to thank my co-authors Oliver Himmler, Robert Jäckle, and Philipp Weinschenk as well as my colleague Lars Behlen in the research project "Verhaltensökonomisch motivierte Maßnahmen zur Sicherung des Studienerfolgs" for insightful discussions and our productive and constructive collaboration. I gratefully acknowledge that large parts of the research in this thesis were carried out as part of the research project just mentioned, which was funded by the German Federal Ministry of Education and Research under grant numbers 01PX16003A and 01PX16003B. I would also like to thank all the student assistants for their contributions to the research in this dissertation as well as the staff at the universities for their administrative support in conducting the field experiments.

Furthermore, my thanks go to Robert Schwager, Monika Jackmann, Florian Rottner, Ann-Marie Sommerfeld, Ansgar Quint, and Kamila Danilowicz-Gösele for the wonderful atmosphere and working environment at the Chair of Public Economics and the insightful conversations over lunch and coffee. I am equally grateful for all the other people I met during my time in Göttingen, especially Vivi and Ann-Kathrin for great conversations about research and life in general.

Finally, I am deeply indebted to the people who provided the necessary balance to my research and who made the time off much more enjoyable. In particular Lea, for enriching my life tremendously over the past few years, my parents – Siegi and Wolfgang, for their lifelong support and encouragement, and my friends in Munich, who always gave me a very warm welcome when returning home.

Contents

List of Tables viii

List of Figures xi

List of Abbreviations xiv

1 General Introduction 1

1.1 Behavioral economics of higher education 1
1.2 Randomized field experiments 3
1.3 Outline of this thesis 4

2 Social Information and Educational Investment – Nudging Remedial Math Course Participation 7

2.1 Introduction 7
2.2 Institutional background and research design 11
2.2.1 Institutional background 11
2.2.2 Experimental design 13
2.2.3 Data and estimation 16
2.3 Effects of the invitation letter intervention 20
2.3.1 Main effects 20
2.3.2 Heterogeneity 21
2.3.3 Effects on academic achievement 27
2.4 Effects of the reminder letter intervention 31
2.5 Discussion and conclusions 33
Appendix 36
A Additional tables and figures 36
B Randomization and balancing properties 45
C Statistical power 51
D Pre-registered analyses by cohort 53

3 Relative Performance Feedback and the Effects of Being Above Average – Field Experiment and Replication 60

3.1 Introduction . . . 60
3.2 Institutional background and research design . . . 65
3.2.1 Institutional background . . . 65
3.2.2 Field Experiment I . . . 65
3.2.3 Field Experiment II: Replication . . . 69
3.2.4 Data and estimation . . . 70
3.3 Main results . . . 72
3.3.1 Field Experiment I . . . 72
3.3.2 Replication experiment and pooled results . . . 76
3.4 Mechanisms (i): regression discontinuity designs . . . 76
3.4.1 Characteristics of above-average students do not explain the response to above-average feedback . . . 77
3.4.2 Why are the positive effects smaller for students in the top 20 percent? 80
3.5 Repeated treatment . . . 82
3.6 Mechanisms (ii): beliefs about relative performance and theoretical considerations . . . 84
3.6.1 Expectations and treatment effects . . . 84
3.6.2 Theoretical considerations . . . 87
3.7 Negative spillovers on other domains? . . . 89
3.8 Conclusion . . . 90
Appendix . . . 93
A Additional tables and figures . . . 93
B Robustness of regression discontinuity designs . . . 100
C Data and methods appendix . . . 106

4 Relative Performance Feedback and University Completion 110

4.1 Introduction . . . 110
4.2 Institutional background and research design . . . 115
4.2.1 Institutional background . . . 115
4.2.2 Intervention . . . 116
4.2.3 Data and descriptives . . . 119
4.3 Main results . . . 121
4.3.1 Empirical approach . . . 121
4.3.2 Main effects . . . 122

4.3.3 Heterogeneity: above- versus not-above-average feedback 129
4.4 Further heterogeneity and mechanisms 131
4.4.1 Subgroup responses and persistence of treatment effects 133
4.4.2 Predicted performance stratification 135
4.4.3 Explaining the pattern of results 141
4.5 Conclusion 144
Appendix 146
A Additional tables and figures 146
B Choice and performance of the prediction model 151

5 Helping Students to Succeed – The Long-Term Effects of Soft Commitments and Reminders 152
5.1 Introduction 152
5.2 Key design challenges 156
5.3 Institutional background and experimental design 158
5.4 Empirical strategy 162
5.5 Results 163
5.5.1 Long-term effects on academic achievement 163
5.5.2 Effects on intermediate measures of academic achievement 171
5.5.3 Procrastinators and commitment 175
5.6 Discussion and conclusions 177
Appendix 180
A Additional tables and figures 180
B Experimental materials 187

6 General Discussion and Conclusions 194

Bibliography 198

List of Tables

2.1 Effect of invitation letter 20
2.2 Effect of invitation letter by timing of enrollment 22
2.3 Effect of invitation letter by endogenous strata 25
2.4 Effect of invitation letter on performance in math exam by timing of enrollment 28
2.5 Effect of invitation letter on overall performance by timing of enrollment 28
2.6 Effect of reminder letter 32
2.A1 Description of variables 36
2.A2 Regression of academic achievement on remedial math course participation – control group of invitation letter intervention (I0) . . 37
2.A3 Effect of invitation letter – heterogeneities 38
2.A4 Effect of invitation letter on performance in math exam – by endogenous strata 39
2.A5 Effect of invitation letter on overall performance – by endogenous strata 40
2.A6 Mean of outcomes of control group students by endogenous strata (invitation letter intervention) 41
2.A7 Effect of reminder letter – heterogeneities 42
2.B1 Number of observations by study program 46
2.B2 Number of observations by timing of invitation letter 46
2.B3 Descriptive statistics and balancing properties – invitation letter, summer term 47
2.B4 Descriptive statistics and balancing properties – invitation letter, winter term 48
2.B5 Descriptive statistics and balancing properties – reminder letter, summer term 49
2.B6 Descriptive statistics and balancing properties – reminder letter, winter term 50

2.C1 Minimum detectable effect sizes ($\alpha = 0.05$) . . . 52
2.C2 Expected effect sizes in the literature – based on DellaVigna and Linos (2020) . . . 52
2.D1 Effect of invitation letter . . . 53
2.D2 Effect of reminder letter . . . 54
2.D3 Effect of reminder letter by invitation letter treatment status . . . 54
2.D4 Effect of invitation letter – heterogeneities . . . 55
2.D5 Effect of reminder letter – heterogeneities . . . 55
2.D6 Effect of invitation letter . . . 56
2.D7 Effect of reminder letter . . . 57
2.D8 Effect of invitation letter by date of enrollment . . . 57
2.D9 Effect of invitation letter by first university . . . 58
2.D10 Effect of invitation letter by gender . . . 58
2.D11 Effect of reminder letter by first university . . . 59
2.D12 Effect of reminder letter by gender . . . 59

3.1 Descriptive Statistics and Balancing Properties . . . 66
3.2 Experimental Design and Number of Observations by Feedback Type 70
3.3 Effect of Feedback on Credits . . . 72
3.4 Effect of Feedback on Credits – By Feedback Type . . . 73
3.5 RD Estimates at Average – First Order Polynomial . . . 79
3.6 RD Estimates at 80th Percentile – First Order Polynomial . . . 81
3.7 Effect of Feedback on Credits – Repeated Treatment . . . 83
3.8 Effect of Feedback on Credits by Pre-Treatment Expectations – Above-Average Students, Experiment II: Replication . . . 87
3.9 Effect of Feedback on GPA and Dropout . . . 90
3.10 Effect of Feedback on Well-Being – Pooled Sample . . . 91
3.A1 Summary of Cost Incurred by the Relative Performance Feedback (in Euros) . . . 93
3.A2 Study Programs, Number of Students, and Treatment Rates . . . 93
3.A3 Description of Variables . . . 94
3.A4 Descriptive Statistics and Balancing Properties – Above-Average Students . . . 95
3.A5 Survey Questions on Students' Expectations – Experiment II . . . 96
3.A6 Effect of Feedback on GPA and Dropout – Third Semester . . . 96

3.A7 SURVEY QUESTIONS ON WELL-BEING – SECOND SEMESTER OF EXPERIMENT I AND II . . . 97
3.B1 RD ESTIMATES AT AVERAGE – DIFFERENT POLYNOMIALS AND DISCONTINUITY SAMPLES, POOLED SAMPLE . . . 100
3.B2 RD ESTIMATES AT 80TH PERCENTILE – DIFFERENT POLYNOMIALS AND DISCONTINUITY SAMPLES, POOLED SAMPLE . . . 104

4.1 DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES . . . 120
4.2 TREATMENT EFFECTS ON ACADEMIC ATTAINMENT . . . 124
4.3 TREATMENT EFFECT ON TIME TO DEGREE . . . 126
4.4 TREATMENT EFFECTS ON ACADEMIC PERFORMANCE . . . 127
4.A1 TREATMENT EFFECTS BY ABOVE-AVERAGE SUBGROUPS . . . 146
4.A2 TREATMENT EFFECTS BY NOT ABOVE-AVERAGE SUBGROUPS . . . 147
4.A3 DESCRIPTIVE STATISTICS BY PREDICTED PERFORMANCE STRATA . . . 148
4.B1 PERFORMANCE OF PREDICTION MODELS . . . 151

5.1 EXPERIMENTAL DESIGN . . . 159
5.2 DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES . . . 160
5.3 TREATMENT EFFECTS ON TWELFTH SEMESTER DROPOUT RATE AND FINAL GPA 166
5.4 TREATMENT EFFECTS ON TIME TO DEGREE – SENSITIVITY ANALYSES . . . 169
5.5 TREATMENT EFFECTS ON CREDITS PER SEMESTER DURING THE SCHEDULED STUDY DURATION . . . 174
5.6 TREATMENT EFFECTS BY APPLICATION DATE . . . 176
5.7 MAGNITUDE OF EFFECTS AND STATISTICAL POWER . . . 178
5.A1 SUMMARY OF COST (IN EUROS) . . . 180
5.A2 TREATMENT EFFECTS ON GRADUATION RATE . . . 181
5.A3 TREATMENT EFFECTS ON FINAL GPA – ROBUSTNESS . . . 182

List of Figures

2.1 Timeline and experimental design 13
2.2 Effect of invitation letter by endogenous strata 24
2.3 Histograms by timing of enrollment 30
2.A1 Invitation letter – social information highlighted in gray (translation) 43
2.A2 Reminder letter – social information highlighted in gray (translation) 44

3.1 Timeline of Experiments 67
3.2 Relative Feedback Graphs – Treatment Group (Examples) 68
3.3 Effect of Feedback on Credits across Feedback Types and Experiments 75
3.4 RD Plot at Average – First Order Polynomial, Pooled Sample 79
3.5 RD Plot at 80th Percentile – First Order Polynomial, Pooled Sample 80
3.6 Cumulative Distribution of Accumulated Credit Points at the End of the Second Semester – Pooled Sample 82
3.7 Shares of Students who Expected to Perform Above Average – Experiment II 85
3.A1 Feedback Letter I – Treatment Group (Example) 98
3.A2 Feedback Letter I – Control Group (Example) 99
3.B1 Distribution of the Running Variable at Average – Pooled Sample . 101
3.B2 RD Plot at Average for Covariates – First Order Polynomial, Pooled Sample 102
3.B3 RD Plot at Average for Covariates – First Order Polynomial, Pooled Sample (cont.) 103
3.B4 Distribution of the Running Variable at Average – Pooled Sample . 105

4.1 Timeline of intervention 116
4.2 Relative performance feedback – treatment group (examples from second semester) 118
4.3 Treatment effects on graduation rates 123

4.4 Treatment effects on dropout rates 125
4.5 Treatment effects on academic performance 128
4.6 Treatment effects on academic attainment, above versus not above average 130
4.7 Treatment effects on academic performance, above versus not above average 132
4.8 Heterogeneity and persistence of treatment effects 134
4.9 Predicted and observed second semester credits by predicted performance strata 137
4.10 Treatment effects on academic performance by predicted performance strata 139
4.11 Treatment effects on academic attainment by predicted performance strata 140
4.12 Post-treatment beliefs about relative performance by predicted performance strata 142
4.A1 Cumulative distributions of accumulated credit points by predicted performance strata – not-above-average students 149
4.A2 Cumulative distributions of accumulated credit points by predicted performance strata – above-average students 150

5.1 Timeline for the first semester 158
5.2 Commitment agreement (English) 161
5.3 Graduation rates, dropout rates, and treatment effects over time. 165
5.4 Average time to degree 167
5.5 Treatment effects on basic study stage completion 172
5.6 Treatment effects on credits per semester 173
5.7 Post-study probabilities 179
5.A1 Average time to basic study stage completion 183
5.A2 Distribution of application dates 184
5.A3 Time to degree distributions, 2019 – Germany 185
5.A4 Average time to degree – by application date 186
5.B1 Cover letter, introductory lecture all groups (English) 187
5.B2 Exam plan, introductory lecture all groups (English) 188
5.B3 Info material, introductory lecture Control and Reminder groups (English) 189

5.B4 Info material, introductory lecture Commitment group, text added to reminder text in grey (English) 190
5.B5 Sign-up letter 1st semester – Reminder, text added for Commitment in grey (English) 191
5.B6 Appendix to sign-up letter 7th semester – Reminder and Commitment (English) 192
5.B7 Study letter 1st semester – Reminder, text added for Commitment in grey (English) 193

List of Abbreviations

ACP	Accumulated Credit Points
ATE	Average Treatment Effect
BA	Business Administration
CGPA	Cumulative Grade Point Average
CI	Confidence Interval
CP	ECTS/Credit Points
CV	Cross-Validation
DID	Difference-In-Difference
EBIC	Extended Bayesian Information Criteria
ECTS	European Credit Transfer and Accumulation System
FE	Fixed Effects
GPA	Grade Point Average
HS	High School
IPW	Inverse-Probability Weighting
IPWRA	Inverse-Probability Weighted Regression Adjustment
ITT	Intention-to-Treat
LASSO	Least Absolute Shrinkage and Selection Operator
MDES	Minimal Detectable Effect Size
ME	Mechanical Engineering
NFE	Natural Field Experiment
OECD	Organisation for Economic Co-operation and Development
OLS	Ordinary Least Squares
pp	Percentage Points
PSP	Post-Study Probability
RCT	Randomized Controlled Trial
RD(D)	Regression Discontinuity (Design)
RMSE	Root Mean Square Error
SUTVA	Stable Unit Treatment Value Assumption

Chapter 1

General Introduction

1.1 Behavioral economics of higher education

Over the past decades, research has provided extensive evidence for the various pecuniary and non-pecuniary returns to educational investments (Psacharopoulos and Patrinos, 2018; Oreopoulos and Salvanes, 2011), and their crucial role in economic development (Hanushek and Woessmann, 2008; Hanushek and Woessmann, 2010).[1] These general findings also hold true for higher education: in addition to a variety of social returns (Hout, 2012), tertiary education is typically associated with a substantial wage premium, which has continued to increase in recent decades (Goldin and Katz, 2008; Oreopoulos and Petronijevic, 2013). This development is often attributed to the rise in demand for high-skilled labor due to the ongoing skill- and routine-biased technological change and the international division of labor (Dustmann et al., 2009; Goldin and Katz, 2008; Goos et al., 2014).

In light of the individual and societal returns, it is hardly surprising that enrollment in higher education and the share of individuals who obtain a tertiary degree has been growing in many developed countries: E.g., in Germany, the share of first-year students in the age-specific population has increased from 36.1% in 2005 to 52.3% in 2019 and the share of first-time graduates has risen from 19.9% to 32.4% in in the same period.[2] In OECD countries, the share of 25 to 34 year-olds with a tertiary degree has increased from 37% in 2010 to 45% in 2020 (OECD, 2021), while the percentage of 55 to 64 year-olds with tertiary education in OECD countries was only 29% in 2020.[3]

At the same time, however, the returns to higher education and the transition toward a more skilled workforce are jeopardized by the fact that a large proportion of students fails

[1] Bradley and Green (2020) provide a comprehensive introduction to the economics of education that includes several reviews on the different returns to education.

[2] See `https://www.datenportal.bmbf.de/portal/en/Tabelle-1.9.4.html` and `https://www.datenportal.bmbf.de/portal/en/Tabelle-1.9.5.html`, retrieved on November 8, 2021.

[3] See `https://data.oecd.org/eduatt/population-with-tertiary-education.htm`, retrieved on November 8, 2021.

to graduate or takes much longer than the scheduled number of semesters. Over the last 15 years, the share of students who leave the German tertiary education system without obtaining their bachelor's degree has consistently been around 28%[4], and only 40.3% (81.1%) of graduates finish their bachelor's degree within the nominal study duration (plus one year) (Statistisches Bundesamt, 2018). In other OECD countries, less than 40% of a cohort entering tertiary education complete their bachelor's degree within the nominal study time and three years later around 25% have left tertiary education without obtaining a degree (OECD, 2019).

When studying the underlying reasons for university students' academic success, or lack thereof, economists have traditionally focused on supply and demand side input factors such as the quality of teaching (Carrell and West, 2010; Feld et al., 2020), class size (Bandiera et al., 2010; Kara et al., 2021), the financial resources of institutions (Bound et al., 2010; Hoekstra, 2009), grant aid (Nguyen et al., 2019), and the academic preparedness of students (Bound et al., 2010). Although the study of those factors has contributed much to our understanding of the determinants of academic success, high dropout rates and delayed graduation remain a major concern, and there are still many aspects of the higher education production function that are not well understood (see Bound and Turner (2011) for a summary and discussion).

Against this background, the economics of (higher) education has recently been complemented by the perspective that lack of educational success can result from behavioral biases (see Koch et al. (2015), Lavecchia et al. (2016), and Leaver (2016) for general overviews): e.g., rather than making purely rational and fully informed decisions, many students struggle with self-control problems and limited attention, they often lack important information that is supposed to be readily available, and their decision-making can be affected by the behavior of others. In view of these insights, the literature has started to focus on developing and evaluating interventions aimed at counteracting lack of academic success by addressing information deficiencies and behavioral biases (see Damgaard and Nielsen (2018) and Escueta et al. (2020) for reviews). Because many of these behavioral measures are easy to scale and have little upfront costs, they can have very advantageous cost-benefit ratios and are thus generally considered to be attractive policy tools (Benartzi et al., 2017).

This thesis contributes to this young but growing body of research, by investigating the effectiveness of several low-cost and easy-to-implement measures that aim to increase students' success at university. By doing so, it provides a broad range of evidence demonstrating the versatility and potential of behaviorally informed interventions. For example, the thesis

[4]See `https://www.datenportal.bmbf.de/portal/de/Tabelle-2.5.90.html`, retrieved on November 8, 2021.

shows how social information can be employed to improve academic performance directly (Chapters 3 and 4) or indirectly (Chapter 2). Behavioral measures can also either be used to address specific biases such as procrastination and limited attention (Chapter 5), or to provide decision-relevant information that is not so easily observed by individuals, such as information about the behavior of others (Chapters 2 to 4). Given that the ultimate goal of interventions in the education sector is to improve students' final academic success, Chapters 4 and 5 also provide some much-needed evidence that behavioral interventions can have substantial positive effects in the long run as well. Yet, the chapters also paint a more nuanced picture and show that effects of interventions can often be heterogeneous. To prevent unintended effects, a targeted implementation may therefore be necessary in many situations. At the same time, the thesis also provides examples of how econometric methods can be used to identify the beneficiaries of an intervention more precisely (Chapters 2 and 4).

1.2 Randomized field experiments

As is the case for all policies and interventions, even well-intended behavioral measures can fail to achieve their promised goals or have adverse effects. Ideally, every policy and intervention should thus be subject to a rigorous and causal evaluation before it is introduced on a larger scale. Otherwise, policymakers run the risk of exposing individuals to negative consequences and wasting scarce resources on ineffective or even harmful policies.

To evaluate the behavioral interventions, all chapters of this thesis use randomized field experiments – also known as randomized controlled trials (RCTs), which are widely accepted as the gold standard for obtaining internally valid estimates (see, e.g., Angrist and Pischke, 2009, Athey and Imbens, 2017a, Duflo et al., 2007, Freedman, 2006, and Imbens, 2010).[5] In comparison to other methods for causal analysis, such as regression discontinuity and difference-in-difference designs, instrumental variable approaches, or simple regression adjustments, they rely on fewer and less strong assumptions. Additionally, recent research indicates that RCTs are also less subject to p-hacking and publication bias than other methods (Brodeur et al., 2020). RCTs therefore continue to contribute greatly to the so-called "credibility revolution" in empirical economics (Angrist and Pischke, 2010).[6]

[5] Following Shadish et al. (2002, p.38), internal validity can be defined as "the validity of inferences about whether observed covariation between A (the presumed treatment) and B (the presumed outcome) reflects a causal relationship".

[6] The growing emphasis on RCTs has also prompted criticism. In particular, there are concerns that research is becoming biased towards questions that can be evaluated via RCTs, leaving other important research questions unanswered (Deaton and Cartwright, 2018; Ravallion, 2020).

Following the taxonomy of Harrison and List (2004), the RCTs in this thesis can be categorized as natural field experiments (NFEs), because they directly study the population of interest (students) in its natural setting (university), and because subjects are not made aware of the fact that they are part of an experiment (see also Czibor et al., 2019). According to Czibor et al. (2019), NFEs are better suited to deal with threats to the external validity of the results compared to other types of experiments[7]: Due to being covert, they reduce concerns regarding experimenter demand effects – i.e., subjects behaving in a way that they think is expected of them – as well as Hawthorne and John Henry effects, which occur when the treatment or control group change their behavior simply because they are part of an experiment. By design, NFEs also prevent noncompliance and nonrandom selection into the experiment, and they offer greater generalizability, since they directly study the population of interest. This allows to apply the results to at least one natural setting, while also making it more likely that results generalize to contexts that are similar to the original one.

1.3 Outline of this thesis

This section provides a brief summary of the topics and research questions addressed in this thesis, as well as the main findings of each chapter.

Investments in education are often characterized by uncertainty, e.g., about the associated returns, which can lead to non-optimal decision-making. Providing individuals with the information necessary to resolve uncertainty may however not always be possible. Given the social nature of humans, information about the behavior of others could be a promising source of guidance in such situations. In the context of a voluntary remedial math course for university students, **Chapter 2** therefore studies whether social information about the past behavior of others can be used to counteract low attendance rates. Incoming students receive postal invitation and reminder letters that inform them about past sign-up rates and past evaluations on the usefulness of the course, respectively. It turns out that, on average, neither of the two interventions increases participation in the course. However, additional heterogeneity analyses provide evidence that a targeted provision of the invitation letters can be beneficial: First, by increasing the salience of the course, they increase attendance and subsequently also academic achievement in the first year of studies among late enrolling students, whose participation rates and performance are otherwise significantly lower compared to early enrollees. Second, the invitation letters with social information increase sign-up and participation in the remedial math course among students whose ex-ante sign-up

[7]Shadish et al. (2002, p.38) define external validity as "the validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables."

probability falls just short of the past sign-up rates, while the opposite is true for students whose ex-ante sign-up probability exceeds the past share of sign-ups.

But it is not only the past behavior of others that can influence individual decision-making. Due to competitive preferences and uncertainty about own abilities, the current performance of others is also an important benchmark against which students can compare their achievements and assess their abilities. Yet, others' performance may not always be easy to observe, in particular when individuals enter a new environment with unfamiliar peers. Providing students with the information necessary for social comparison may thus be a promising tool for universities to approach all students. Chapters 3 and 4 make use of two RCTs with consecutive cohorts of students at a university of applied sciences to evaluate an intervention in which treated students receive relative feedback on their accumulated course credits from the second semester until the end of their enrollment. **Chapter 3** studies the short-term impact and shows that the type of feedback matters for the behavioral response: relative feedback increases performance in the second semester only if it informs students that they placed above average in the first semester performance distribution. A regression discontinuity design is used to provide evidence that the effects are not driven by above-average students reacting particularly well to feedback due to underlying characteristics such as higher ability and motivation, or a better learning technology. Instead, the results suggest that the information about being above average makes feedback effective.

Chapter 4 studies the effects of the intervention and their persistence over a period of five years. The results provide evidence that relative feedback can be an effective tool for increasing academic success in the long run: It raises the eighth and ninth semester graduation rates by about four percentage points without affecting the dropout decision, and improves grades by 0.064 standard deviations. After five years, about one in ten treated students has graduated one semester earlier compared to controls. Consistent with the short-term results, the overall effects are driven by students who ranked above the average in the first semester credit distribution. Additional analyses based on the predicted second semester performance provide a more detailed picture of the heterogeneous effects. The findings indicate that students in the middle of the predicted performance distribution are most responsive to relative feedback. When feedback informs them that they did not rank above-average pre-treatment, they adjust their beliefs about relative performance downward and decrease their performance, while the opposite is true for students who are informed about an above-average performance. Students with the highest and lowest predicted performance, on the other hand, show no response. Descriptive evidence suggests that the former group is held back by institutional factors, while there is little room for the latter group to perform worse than their control group counterparts.

Chapter 5 leaves the topic of social information behind and studies the long-term effects of an intervention that aims to increase success at university by directly addressing behavioral biases. The main focus of the intervention is procrastination, a widespread phenomenon among students that can result in insufficient effort and lower academic success. Using a cohort of students at a university of applied sciences, Chapter 5 evaluates the effectiveness of a particularly simple and inexpensive measure to mitigate procrastination tendencies. University students are offered a non-binding agreement that allows them to commit to the recommended study structure. To also address inaction caused by limited attention, students additionally receive reminders during the scheduled study duration of seven semesters. The results show that this soft commitment device is highly effective in raising academic achievements: driven by less dropout and accelerated credit accumulation, it increases the five-year graduation rate by 15 percentage points and lessens time to degree by about 0.41 semesters. Importantly, treated students maintain the same GPA as controls. In line with theory, the findings additionally suggest that the commitment device is particularly effective for students with procrastinatory tendencies.

Chapter 6 concludes with a discussion of the limitations of this thesis and some recommendations for policy and further research.

Chapter 2

Social Information and Educational Investment – Nudging Remedial Math Course Participation

2.1 Introduction

Decisions about educational investment are often characterized by uncertainty about the associated returns, which can lead to non-optimal decision-making (Altonji, 1993; Altonji et al., 2012).[1] One way to address this is to provide individuals with the relevant information, which has, for example, been shown to change the amount of time individuals stay in school (Jensen 2010), which college major they choose (Wiswall and Zafar 2015b; Wiswall and Zafar 2015a), or their educational aspirations (Bleemer and Zafar 2018; Lergetporer et al. 2021). However, in many situations, the necessary information may not be available to policy makers, for example, because the educational investments have not yet been evaluated or because the returns also encompass non-pecuniary benefits that are difficult to measure.[2]

In such cases, providing information about the behavior of others could be a promising alternative: First, it may provide a signal about returns, if individuals believe that the decision of others is linked to the expected utility of the investment (see Coffman et al. (2015) and Coffman et al. (2017) for respective models). Second, and more general, the behavior of others may also be perceived as a descriptive norm, i.e., individuals might want to invest in education because they expect others to do the same (see, e.g., Bicchieri and Dimant 2019).

[1] In many contexts, beliefs about one's own academic ability and the probability to succeed also play an important role in deciding whether (further) investments in education should be made (see, e.g., Kunz and Staub 2020; Stinebrickner and Stinebrickner 2012; Stinebrickner and Stinebrickner 2014b; Stinebrickner and Stinebrickner 2014a).

[2] See e.g., Oreopoulos and Salvanes (2011) and Hout (2012) for reviews on the non-pecuniary benefits of education. Delavande and Zafar (2019) provide evidence that non-pecuniary benefits can play an important role in educational investment decisions.

This chapter studies if social information can indeed be used to influence an educational investment decision, and whether individuals benefit from their investment. The context is a voluntary remedial math course for economics and business students at a large German university that takes place before the beginning of the first semester; a setting characterized by the features described above: A considerable number of students does not participate in the course, even among those who initially signed up for it. Descriptively, these students perform worse in their first year of studies, suggesting that their decision is not optimal. This could be rooted in the fact that the course is an investment with uncertain returns. Students are usually not aware of the exact content of their study program in advance and whether it is really necessary to attend the course, given their prior knowledge.[3] Since the course has not been causally evaluated yet, students could not be informed of its returns directly.

To evaluate the effectiveness of two different social information interventions at increasing sign-up for and participation in the remedial math course, we conduct field experiments with an incoming summer cohort and the subsequent winter cohort: i) In both cohorts, after students enroll in their study program, we send a randomly selected group of students a postal invitation letter that includes the information that 85% of students signed up for the remedial math course in a previous cohort.[4] In both cohorts, we compare the sign-up and participation behavior of this treatment group with that of students who receive no invitation letter at all. To test whether the social information or salience drives potential effects, in the winter cohort, we add a treatment group that receives the same invitation letter, but without the social information. ii) To increase course attendance among students who initially sign up for it, we randomize a postal reminder letter that informs students that the course made it easier to get started with university mathematics for 95% of the students in the past. We compare the participation behavior of this group with one that receives no reminder letter at all, and, in the winter cohort, also a group that receives a reminder letter without the social information.

Our key findings are as follows: First, we find that, on average, neither the invitation nor the reminder letter intervention affect students' decision to sign up for or participate in the remedial math course, respectively. Second, further analyses provide evidence that the effects of the invitation letter intervention are heterogeneous along two dimensions: a) Both the letter with information on past sign-up rates and the letter without this information are more effective for students who enroll late in their degree program. They offset more than

[3]Information on the cost, i.e., the time it takes to participate, and a summary of the content of the course is available to students in advance. Furthermore, students do not have to pass a test as part of the course, which could otherwise create additional uncertainty.

[4]Throughout the chapter, we use the term "enrollment" when referring to study programs or the university and we use the term "sign-up" when referring to the remedial math course.

half of the roughly 9 and 16 percentage point (pp) lower sign-up and participation rates that we observe for these students in the absence of treatment relative to early enrollees. The fact that letters with and without social information are similarly effective suggests that low sign-up and participation among late enrollees is at least partly driven by a lack of (relative) salience of the course. b) The effect of the invitation letter with social information also depends on the ex-ante sign-up probability of students, which we predict using the control group, i.e., we use endogenous stratification (Abadie et al., 2018). Those with the highest ex-ante probability decrease sign-up and participation by about 10 pp, while the opposite is true for students whose probability falls just short of the social information on past sign-up rates; students with the lowest predicted sign-up probabilities show no behavioral response. We argue that this pattern is broadly consistent with the idea that treated students update their beliefs about the behavior of others, which in turn affects beliefs about the descriptive norm and the expected utility of the course.

Third, we investigate whether these heterogeneous effects on remedial math course participation carry over to academic achievement in the first year of studies. Our findings suggest that this only holds true for the timing of enrollment. For late enrolling students, whose overall performance is about 0.26 standard deviations worse compared to early enrollees, the increase in remedial math course participation is able to almost completely close the gap in academic performance. The remarkably large effects on first-year performance appear to be driven by large increases in average course participation among a small group of individuals. Our findings are in line with the notion that early engagement with their studies in form of the remedial math course prevents these students from dropping out early, thus explaining the large gains in academic achievement.

The context and results of our study contribute to the following strands of the literature. First, the chapter relates to the literature that studies whether and how the provision of social information influences individual decision-making. In contrast to this study, much of the existing evidence on positive effects comes from environments where the stakes for individuals are low, such as increasing charitable giving (e.g., Frey and Meier 2004; Croson and Shang 2008; Martin and Randal 2008; Shang and Croson 2009), public good contributions in the lab (e.g., Keser and Van Winden 2000; Fischbacher et al. 2001), or environmentally friendly behavior (e.g., Goldstein et al. 2008; Allcott and Rogers 2014; Byrne et al. 2018; Brent et al. 2020). Attempts to influence decisions with higher stakes have produced mixed results: Fellner et al. (2013) find no effect of social information on TV licensing fee compliance and Beshears et al. (2015) even find negative effects when trying to increase retirement savings. Hallsworth et al. (2017), Gee (2019), and Coffman et al., 2017 on the other hand show that social information can positively affect tax compliance, job application rates, and job take-up, respectively. Our study provides additional evidence for the effects of social information in environments with higher stakes, as the decision not to participate in the remedial math course can come at the cost of obtaining less course credits or dropping out of the study program. Moreover, to our knowledge, this is the first study that provides evidence on this type of social information from an educational context.[5][6]

Second, our study is related to the literature that uses nudges to improve decision-making in education (see, e.g., Damgaard and Nielsen 2018 for a review). More specifically, we contribute to research that aims to improve outcomes of students in higher education by providing information via low touch channels, such as text-messages, e-mails, and postal letters. Initial studies showed promising results, especially with respect to enrolling in college and applying for financial help (see French and Oreopoulos 2017 and Bird et al. 2019 for reviews). However, results from recent large-scale field experiments suggest that these interventions do not necessarily scale up, creating the need for alternative approaches (Bird et al., 2019; Bergman et al., 2019; Gurantz et al., 2021). Furthermore, attempts to improve persistence in college with the help of low-touch information interventions have so far not provided the desired results (Oreopoulos and Petronijevic, 2018; Oreopoulos and Petronijevic, 2019; Huntington-Klein and Gill, 2019). Our study extends this literature by adding the following: First, to the best of our knowledge, our study is the first to explicitly test whether information about the past behavior of others affects individual decision-making in higher education. Given the promising results in other areas, surprisingly little attention has been paid to this approach so far. Second, we are the first to study if (social) information can be used to influence smaller educational investment decisions. Previous studies have often focused on educational investments that are likely more difficult to influence, such as whether to enroll in college at all. Third, our results also illustrate that in some contexts a targeted provision of (social) information nudges may be necessary to achieve the desired results.

Last, the chapter contributes to the ongoing discussion on college remediation (e.g., Holzer and Baum 2017; Oreopoulos 2021). By studying whether the changes in remedial math course participation due to our intervention carry over to academic performance, we

[5] Darolia and Harper (2018) study the effects of student debt letters, which also include information on past borrowing behavior of others. However, they do not explicitly test the effects of the social information.

[6] The literature on relative performance feedback also makes use of social information (see Villeval 2020 for a review of the literature). However, there, information is usually given about the current performance of similar others, which can, for example, create positive effects through competitive preferences and learning about own ability. Such mechanisms are unlikely to play a role for the type of social information that we study in this chapter. In addition, relative performance feedback usually aims at affecting effort at the intensive and not on the extensive margin.

are the first who study the effects of remediation based on experimentally induced variation in attendance. Many existing studies on the effectiveness of remedial courses in higher education instead make use of natural experiments that occur when participation in remedial courses is based on performance in a placement test (see, e.g., Martorell and McFarlin Jr 2011; Boatman and Long 2018). Our study adds to this literature by providing evidence on the effectiveness of remediation from complier populations that have not been studied so far.

The chapter continues as follows. Section 2.2 describes the institutional background, data, and design of the two interventions as well as the empirical approach. In Sections 2.3 and 2.4, we present the results of the invitation and the reminder letter intervention, respectively. Section 2.5 concludes.

2.2 Institutional background and research design

2.2.1 Institutional background

We conducted our field experiments at one of the largest universities in Germany with a summer and a winter cohort of incoming first-year students who enroll in one of five bachelor's degree programs offered at the Faculty of Business and Economics.[7] Adequate knowledge of mathematics is an important prerequisite: students in all but one of the programs must pass a math exam by the end of the second semester in order to continue with their program.[8] In addition, all programs include other mathematically demanding subjects, such as microeconomics, that students have to pass at some point during their studies. To prepare students for these subjects, the Faculty of Business and Economics offers a voluntary remedial math course that usually takes place two to three weeks before the start of the first term. The course aims at refreshing the math knowledge that students should have acquired by the end of secondary education and filling potential gaps.[9] The course usually lasts eight days, with each day consisting of a lecture in the morning and two tutorials later in the day, which are held in groups of ten to twenty students. In order to facilitate the organization of

[7]Students can enroll both in the summer and the winter term, but most students enter the university in the winter after graduating from high school during the summer.

[8]Due to the COVID-19 pandemic, this requirement was relaxed and students in our field experiments were allowed to pass the math exam in a later semester. However, when students enrolled in their program, they were not aware of this.

[9]The content of the course covers subjects such as numbers, arithmetic, summation, binomial formulas, (in)equations and systems of linear equations, exponentiation, root extraction, logarithms, functions, and differential calculus.

the course, especially the prior formation of the tutorial groups, students are asked to sign up for the course in advance via a web portal.

In the absence of our interventions, students can access or receive information on the remedial math course via the following channels: First, information about the course is publicly available on a website. Second, incoming first-year students receive a letter from the student council that provides information about the (social) activities that are planned at the beginning of their studies, including information on the remedial math course. Third, the organizers of the course themselves email all newly enrolled students, inviting them to participate in the course. In principle, the remedial math course should thus be very salient to students. Still, a considerable number of students does not sign up for the course and the organizers informed us that attrition between sign-up and participation presents an additional problem. For example, in the control group of our first intervention, only 76% of the incoming students sign up for the course, and the participation rate in the first tutorial and the average participation across all tutorials are even lower at around 70 and 60%, respectively (see Table 2.1).

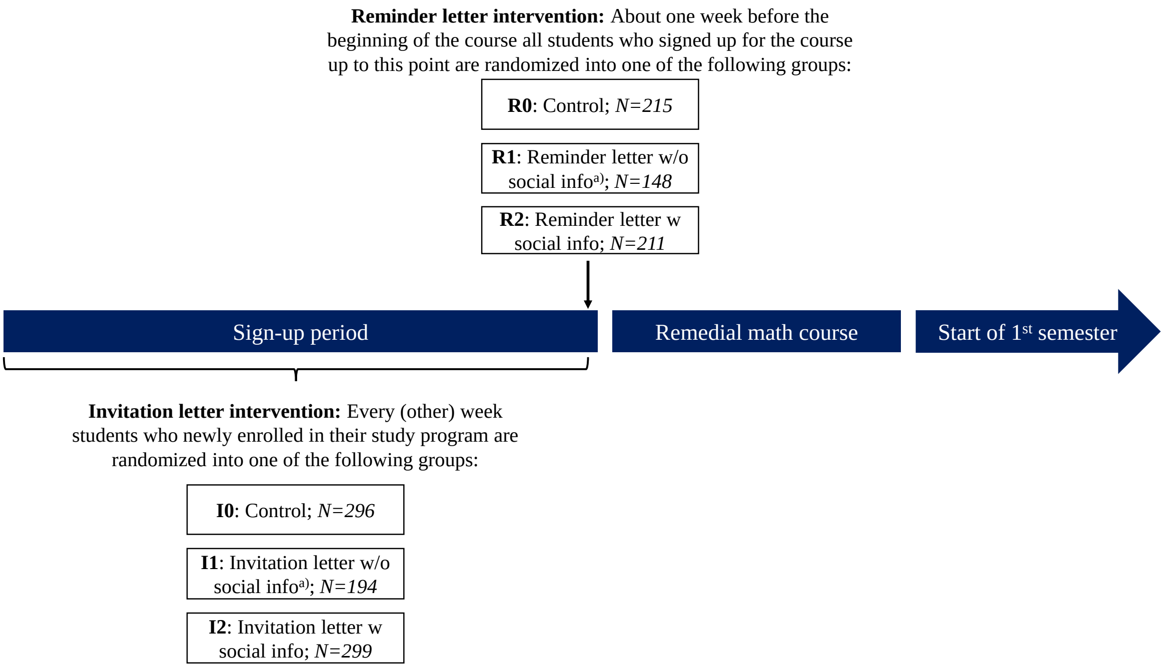
However, for students there may be good reasons not to participate in the course. Most importantly, they may think that they already possess the knowledge that is taught in the course. If that is indeed the case, then it might not even be desirable to persuade these students to participate. Since neither this nor – as far as we know – any other voluntary remedial math course has been evaluated to date, we do not know for certain whether attendance is actually beneficial and whether students are making a non-optimal decision. To provide some correlational evidence on this, in Table 2.A2, we use the control group of our first intervention and regress different measures of academic achievement on average participation, controlling for a large set of observables; effects of the high school GPA are shown as comparison. We find that going from 0 to 100% participation significantly increases the likelihood to attempt and pass the math exam within the first semester (year) by about 50 pp, the obtained credits by roughly 10 (19), and reduces the probability to drop out of the study program by around 14 (17) pp.

This raises the question as to why students fail to participate even though the course may have been beneficial for them. First and foremost, students may misjudge their own knowledge or how useful the course is, making it difficult to assess whether they actually need to participate in order to pass the math exam and other mathematically demanding subjects in their studies. Second, before the start of their studies, students are often still busy with organizational matters, in particular the search for an accommodation, which may reduce the salience or relative importance of the course. While there may certainly be other causes for a lack of attendance, e.g., students may simply not have time to participate, our interventions aim to address the salience of the course and uncertainty about its usefulness.

2.2.2 Experimental design

Against this background, we designed two social information interventions with the goal of increasing the share of students that sign up for and participate in the remedial math course: i) An invitation letter that includes information on past sign-up rates, and ii) A reminder letter for students who signed up for the course, containing information about how useful students have perceived the course in the past. To test their effectiveness, we conducted field experiments with a cohort of incoming first-year students who enrolled in the summer term and the subsequent cohort in the winter term. The general design and timing of the interventions was the same in both cohorts and is summarized in Figure 2.1.[10]

FIGURE 2.1: TIMELINE AND EXPERIMENTAL DESIGN



Note: [a)] As described in Section 2.2 the experimental groups who receive an invitation or reminder letter without the social information (I1 and R1) are only included in the winter cohort. The exact timing of the invitation letters and the respective number of observations are shown in Table 2.B2.

[10]The research design for both cohorts was approved by the IRB and the data protection officer of the university. The first experiment (summer term) is pre-registered under `https://osf.io/tm7k3` and the second experiment (winter term) under `https://osf.io/vqa84`. The interventions in both cohorts and the respective math courses took place before the Corona pandemic. For the first cohort, the pandemic started shortly before the exam period of the second semester, and for the second cohort it started shortly before the exam period of the first semester. Exams during the pandemic were in part held online and most lectures during the second semester of the second cohort took place online.

Invitation letter and information on past sign-up rates. In both cohorts, starting five to seven weeks before the beginning of the remedial math course until one week before, we used administrative data on the incoming students provided to us by the university to randomize students into a control and one (summer term) or two (winter term) treatment groups. Randomization was carried out using stratification and re-randomization (Morgan and Rubin, 2012); Appendix B provides details on the randomization and shows that the samples are well balanced.

In both cohorts, students who were randomized into the control group (**I0**) received no information about the remedial math course from us, but they could still receive information on the math course from the sources mentioned above.

Both in the summer and the winter term, students who were randomized into the treatment group **I2** received a letter on behalf of the organizers of the course (see Figure 2.A1). The letter stated that "[...] in order to help you get off to a good start in your studies, we would like to invite you to the remedial math course for students of business and economics. The course provides mathematical knowledge that is required in the mathematics lecture and in numerous other courses." and quoted the website where students could sign up for the course and get additional information. The letter continued by stating that "85% of the first-year students who, like you, were enrolled in a business or economics degree program in the last semester have signed up for the remedial math course. Only a small minority of students does not sign up for the remedial course". The figure of 85% was based on the sign-up rate for the remedial math course in the winter term that preceded our first experiment and was calculated among all students who enrolled in a study program at the Faculty of Business and Economics for the first time.[11] The aim of this information was to signal to students that the vast majority of students signs up for the math course and that they should thus do the same.

From a theoretical perspective, at least two arguments can be made why information on past sign-up rates should lead to an increase in sign-ups and participation. First, we can follow the model in Coffman et al. (2017) and assume that incoming students believe that higher sign-up rates signal higher returns to the remedial math course, and that individuals will only sign-up for and participate in the course if their beliefs about the expected sign-up rate are above some personal threshold. For individuals who are not going to sign up and participate, as their initial beliefs are below their threshold, treatment could lead to an

[11] Including students who did not enroll for a study program at the faculty for the first time results in a lower sign-up rate, as it includes for example students who simply switched programs and thus already had the possibility to participate in the math course at a previous point in time.

increase in sign-up and participation, if information on past sign-up rates shifts the beliefs above their personal threshold.[12]

Second, and more generally, the sign-up rate can be understood as a descriptive norm if students prefer to sign up and participate in the course when they expect other students to do the same (see, e.g., Bicchieri and Dimant 2019).[13] The social information in our letters could then increase the sign-up and participation rate if it leads to an upwards shift in the beliefs about the descriptive norm. The descriptive norm interpretation is also the reason why we included the sentence "Only a small minority of students does not sign up for the remedial course". Recent results in the literature on social norms have shown that presenting behavior as a minority activity can increase the effectiveness of descriptive norms (e.g., Hallsworth et al. 2017).

Due to all the available information the remedial math course should have been salient to students, even in the absence of the social information treatment. However, it could still be the case that the invitation letter and its personal nature increase the salience of the course or lead to an increase in the sign-up rate and ultimately participation through some channel that is unrelated to the information on past sign-up rates that we include in the letter. To study if this is the case, in the winter term, we included a second treatment group (**I1**), which also received an invitation letter, but without the descriptive social norm; i.e., it excluded the parts highlighted in gray in Figure 2.A1.

Reminder letter and information on the usefulness of the course. About one week before the start of the course (see Figure 2.1), we randomized all students who signed up for the course up to that point into a control and one (summer term) or two (winter term) treatment groups (see Appendix B for details and the balancing properties).

While students in the control group (**R0**) received no letter, students in treatment **R2** received a reminder letter on behalf of the organizers of the course (see Figure 2.A2). The letter stated that "[...] you have signed up for the remedial math course. We have therefore already reserved a seat for you and look forward to your participation. The course starts on $\langle$*date*$\rangle$ at $\langle$*location*$\rangle$" and mentioned the website where students could find the information on the tutorial group they were allocated to. Instead of social information similar to the one used in the invitation letter, the reminder included information about how useful students evaluated the course in the past, as it stated that "95% of students who, like you, are enrolled

[12] In principle, it is also possible that the social information shifts beliefs downwards, resulting in a decrease in sign-up and participation rates. A more general model on the effects of social information taking this into account is provided by Coffman et al. (2015).

[13] For example, students may mainly participate in the course for social reason such as meeting their future peers early.

in a business or economics degree program say that the remedial course in mathematics has made it easier for them to get started with university mathematics". This figure was based on one of the questions that was asked in a survey that was carried out a few years earlier by the course organizers among students who attended the mathematics lecture. Students were asked "[...] whether the remedial math course made it easier to get started with university mathematics?". On a scale from "1=no, not at all" to "7=yes, very much", 95% of 290 survey-participants had chosen answer category 5 or higher.

We expect that the information on how students have evaluated the course in the past leads to a decrease in attrition between sign-up for and participation in the course. A similar argument as before can be made. The treatment should provide a direct signal about the (subjective) returns to the remedial math course and should thus influence the participation decision of students who are unsure about the utility of the course, and for whom this signal leads to a sufficiently large upward shift in the expected utility.

Following the reasoning for the invitation letter, we again wanted to be able to test if potential effects of the reminder letter are driven by the social information or the letter itself. Therefore, in the winter term, we included a second treatment group (**R1**), which received a reminder letter without the social information (parts highlighted in gray in Figure 2.A2).

2.2.3 Data and estimation

Data. To analyze the effects of the two interventions, we use data from three sources: First, we use administrative information on background characteristics for our covariates. Second, we receive information from the organizers of the remedial math course about sign-ups for the course and participation in each of the tutorials, which we use as outcomes in our analyses. Third, we use administrative information from the student office about students' academic performance in the first year of studies to investigate whether potential effects on participation translate into higher academic achievement (Table 2.A1 describes all variables that we use).[14] For the analyses in the main part of the chapter, we pool the data from the summer and the winter cohort.[15]

Our main outcome variables are sign-up for the remedial math course, participation in the first tutorial, and average participation, i.e., the share of tutorials a student participated

[14]The analysis of the effects on academic achievement were pre-registered after the analyses of the effects on the sign-up and participation rates but before data on academic achievement was available to us. The pre-registration can be found under `https://osf.io/tv9yf`.

[15]Appendix D presents results separated by cohort – following the respective pre-registrations.

in.[16] In follow up analyses, we also study effects on academic achievement after the first semester and the first year of studies. Since the remedial course aims at improving math knowledge, we expect that attendance primarily affects whether students attempt and pass the math exam, and what grade they receive. In addition, we are interested in students' overall academic performance. For this, we consider the number of passed course credits[17], whether they dropped out of their study program[18], and their GPA. These variables were pre-registered as secondary outcomes, and we have no clear hypothesis as to which of these dimensions should be most influenced by remedial math course attendance. Therefore, and to reduce potential concerns regarding multiple hypothesis testing, we follow the approach suggested by Anderson (2008) and additionally construct an inverse-covariance weighted index of the three variables using the Stata program *swindex* by Schwab et al. (2020), which we use as an outcome when investigating the effects on academic achievement.[19]

Analysis of main effects. Regarding the main effects of our interventions, we provide ITT effects from OLS estimations that compare the average outcomes of the control group with the outcomes of the treatment groups. In the baseline specification, we control for the random assignment within blocks:

$$Y_i^k = \alpha_0 + \alpha_1 Letter_i + \alpha_2 LetterSocialInfo_i + \mathbf{x_i}\boldsymbol{\alpha_3} + \varepsilon_i, \tag{2.1}$$

where Y_i^k denotes the level of outcome measure k for individual i. $Letter_i$ is an indicator for being randomized into the treatment group that receives the invitation (reminder) letter without social information. $LetterSocialInfo_i$ is an indicator for being randomized into the treatment group that receives the invitation (reminder) letter including the respective social information. The vector $\mathbf{x_i}$ controls for the method of randomization by including study program fixed effects, a winter term dummy, the interaction between the study program

[16]We use both measures of participation, as the participation in later tutorials might be affected by the content of the course and its interaction with the treatments. For example, students may learn that the content of the course is not as useful as the information of the letters suggested. On the other hand, average participation is arguably the more relevant outcome with respect to the later performance in the study program.

[17] Europe-wide, universities use a standardized point system (European Credit Transfer and Accumulation System, ECTS), under which a full-time academic year consists of 60 credits, with the typical workload for one credit equaling 25-30 study hours. See also `https://ec.europa.eu/education/resources-and-tools`, retrieved on November 8, 2021.

[18]Dropout captures both students who left the university system completely and students who merely switched the study program and/or university. However, our data does not allow us to differentiate between those cases.

[19]The use of this index as an additional outcome was not included in the pre-registration for the effects on academic achievement. For the reasons stated above, we nevertheless believe that it is advisable to also consider effects on this aggregate performance measure.

fixed effects and the winter term dummy. Additionally, it includes invitation letter date fixed effects or invitation letter treatment status indicators (such that the two randomizations are orthogonal) when analyzing the effects of the invitation or reminder letter, respectively (see Appendix B for details on the randomizations).

In additional specifications, we add a vector $\mathbf{z_i}$, which includes further covariates:

$$Y_i^k = \alpha_0 + \alpha_1 Letter_i + \alpha_2 LetterSocialInfo_i + \mathbf{x_i}\boldsymbol{\alpha_3} + \mathbf{z_i}\boldsymbol{\alpha_4} + \varepsilon_i. \tag{2.2}$$

We follow two different approaches for selecting the variables that we include in this vector. First, we simply include all variables that were pre-registered as controls for the second experiment. This includes the first university and female dummies[20], the age at the beginning of the first semester, the high school GPA, an indicator if the high school degree was obtained within the last year before the beginning of the first semester, a dummy for the type of high school degree, indicators for the place where the high school degree was obtained, and the distance over which the letter was sent in kilometers (see Table 2.A1 for a detailed description of the variables).[21]

Second, we would like to employ an approach in which covariates that were not pre-registered are included in a non-arbitrary way and that furthermore leads to parsimonious specifications which only include covariates that either increase the precision of our treatment effect estimates or account for imbalances that are observed despite the randomization. For this purpose, we use the double-post LASSO approach suggested by Belloni et al. (2014) to select the covariates to be included in the estimation. In the covariate selection process, we consider the pre-registered covariates and for estimations of the effect of the invitation letter we additionally consider an indicator if the student signed up for the remedial math course before we sent, or would have sent in case of the control group, the invitation letter.

As we show in Appendix C and Section 2.5, our study is generally well-powered enough to detect effects sizes that have so far been published in the literature on social information nudges. There, we also discuss more recent evidence which suggests that previous results may suffer from publication bias and overstate the true effect.

[20]Although we were sometimes able to stratify on those variables as planned in the pre-registration, this was not possible in the majority of the randomizations. Thus, we include those variables in the control vector and not in the baseline specification.

[21]The distance over which the letter was sent was only pre-registered as a control variable in the second experiment.

Analysis of heterogeneities. In addition to the main analysis, we pre-registered to study heterogeneity along the following dimensions: First, based on the institutional background we expected that students who were previously enrolled at this or another university would be less likely to participate in the remedial math course, as they may have participated in a similar course previously. Treatment effects might therefore be concentrated among students for whom this is the first semester at any university. Second, we expected that students with an enrollment date closer to the remedial math course would be less likely to sign up for and participate in the course. These students may still be busy with organizational matters such as looking for accommodation, resulting in a lower (relative) salience or importance of the course. By increasing the salience or expected utility of the course, treatment may thus be particularly effective in this group.[22] Lastly, initial results from the field experiments in the summer term indicated that men may be more responsive to treatment, and we thus pre-registered gender as a potential source of heterogeneity for the field experiments in the winter term.

To estimate heterogeneous effects of our treatments for the pre-registered covariates, we employ the following regression specification:

$$\begin{aligned} Y_i^k = \alpha_0 + \alpha_1 Letter_i + Letter_i \mathbf{c_i} \boldsymbol{\alpha_2} + \alpha_3 LetterSocialInfo_i \\ + LetterSocialInfo_i \mathbf{c_i} \boldsymbol{\alpha_4} + \mathbf{c_i} \boldsymbol{\alpha_5} + \mathbf{x_i} \boldsymbol{\alpha_6} + \mathbf{z_i} \boldsymbol{\alpha_7} + \varepsilon_i. \end{aligned} \tag{2.3}$$

Where Y_i^k, $Letter_i$, and $LetterSocialInfo_i$ are defined as before. The vector $\mathbf{c_i}$ includes one or all of the covariates for which we want to study the treatment effect heterogeneity. In case of the invitation letter intervention this includes the first university dummy, the female dummy, and a dummy that indicates if a student received, or could have received in case of the control group, the invitation letter within the last month before the beginning of the course.[23] In case of the reminder letter, it includes the female and the first university dummy. The vector $\mathbf{x_i}$ again controls for the method of randomization. When we study the heterogeneity of the invitation letter intervention with respect to the timing of enrollment this vector now excludes the invitation letter date indicators. The vector $\mathbf{z_i}$ includes the pre-registered covariates, with the exception of those for which we investigate the heterogeneous effects of treatment.

[22] However, if external constraints are too strong, such as students not having accommodations at the beginning of the course, treatment effects could also be smaller in this group.

[23] We define the timing of the letter/enrollment in this way, because the respective cells get increasingly small in the last weeks before the beginning of the course, making a more fine-grained heterogeneity analysis difficult (see Table 2.B2).

2.3 Effects of the invitation letter intervention

In this section, we first report the main effects of the invitation letter intervention. We then study if the average treatment effects mask heterogeneous responses and whether effects on participation in the course carry over to students' academic performance in the first year of studies.

2.3.1 Main effects

Table 2.1 shows estimates for the effects of the invitation letters with and without social information about past sign-up rates. Columns (1) to (3) present the effects on sign-up. In the control group, 76% of the students sign up for the remedial math course. Among students who receive the invitation letter with social information, the sign-up rate is 1.5 pp ($p = 0.646$) higher. Adding controls in Columns (2) and (3) increases the coefficient to 2.4 pp ($p = 0.447$) and 2.1 pp ($p = 0.494$), respectively. Sending students the invitation letter without social information decreases the sign-up rate by about 3 pp across all three specifications ($p = 0.386$ in Column 3).

TABLE 2.1: EFFECT OF INVITATION LETTER

	Sign-up			Participation 1st tutorial			Average participation		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
I1: Letter	-0.030	-0.028	-0.031	-0.044	-0.041	-0.045	-0.034	-0.029	-0.036
	(0.040)	(0.038)	(0.036)	(0.044)	(0.041)	(0.039)	(0.041)	(0.038)	(0.037)
I2: Letter & social info	0.015	0.024	0.021	0.003	0.011	0.009	0.000	0.008	0.006
	(0.034)	(0.032)	(0.031)	(0.037)	(0.035)	(0.034)	(0.034)	(0.032)	(0.031)
Strata	yes	yes	yes	yes	yes	yes	yes	yes	yes
Controls	no	yes	d-p lasso	no	yes	d-p lasso	no	yes	d-p lasso
N	789	789	789	789	789	789	789	789	789
Control mean	0.76	0.76	0.76	0.70	0.70	0.70	0.60	0.60	0.60
(SD)	(0.43)	(0.43)	(0.43)	(0.46)	(0.46)	(0.46)	(0.43)	(0.43)	(0.43)

Note: Outcome variables: sign-up for remedial math course, participation in first tutorial of remedial math course, and average participation is the share of tutorials a student participated in; *strata:* study program FE, winter term dummy, interaction between study program FE and winter term dummy, and invitation letter date FE; *controls:* first university and female dummies, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent; the double-post LASSO specification considers all controls as well as a dummy if sign-up took place before the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Effects on participation follow a similar pattern. In the absence of treatment, 70% of the students participate in the first tutorial of the remedial math course. Students who receive the invitation letter with social information increase participation by 0.3 to 1.1 pp ($p = 0.943 – 0.745$; Columns 4 to 6). The pure invitation letter decreases participation in the first tutorial by 4.1 to 4.5 pp ($p = 0.251$ in Column 6). Regarding average participation, we observe that students in the control group participate in 60% of the tutorials (Columns 7 to 9). Even though this leaves more room for improvement, treatment effects are similar to those of the first two outcomes.

Overall, this suggests that, on average, the social information is not able to increase sign up for and participation in the remedial math course. Treatment effects are neither statistically significant at any conventional level, nor are they particularly large from a practitioners' point of view. Pure invitation letters may even decrease sign-up and participation in the course. One reason for the latter may be that sending students an invitation letter, in addition to all the other information on the course they already receive, could signal that participation in the course was too low in the past; thus acting similarly to social information about low sign-up rates.

2.3.2 Heterogeneity

Next, we study whether the main effects mask heterogeneity. Along the pre-registered dimensions, we find the most robust evidence for the timing of enrollment, which we discuss in detail below. With respect to students' gender and the first university dummy, we find the following (see Table 2.A3 for details): Students in the control group who were not enrolled at any university before are about 30 pp more likely to sign up for and participate in the remedial math course – which is in line with what we expected, but we find no evidence of heterogeneous treatment effects. Students' gender, on the other hand, is not predictive of sign up and participation in the absence of treatment, but we find that treatment effects on average participation are between 4.6 to 13.6 pp higher for females compared to males; these effects, however, are not estimated precisely.

Timing of enrollment. Panel a) in Table 2.2 presents effects by timing of enrollment. About 31% of the sample enrolled late and were sent the invitation letter within the last month before the beginning of the remedial course (see Table 2.B2 for the exact timing and number of observations). The first important observation is that students who enroll late are less likely to sign up for and participate in the course: their sign-up rate is 9.1 to 9.3 pp ($p = 0.102 - 0.085$) lower compared to early enrollees, and average participation is decreased by 15.7 to 16.8 pp ($p = 0.003 - 0.002$). Looking at the treatment effects and their interaction with the last month dummy, we find that both types of invitation letters are more effective for late enrollees: compared to students who receive the letter before the last month, the effects on sign up and, more importantly, average participation are 5.3 to 8.6 pp ($p = 0.494 - 0.321$) and 13.6 to 17.2 pp ($p = 0.098 - 0.013$) larger, respectively.

TABLE 2.2: EFFECT OF INVITATION LETTER BY TIMING OF ENROLLMENT

	Sign-up		Part. 1st tutorial		Avgerage part.	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel a)						
I1: Letter	-0.060	-0.051	-0.075	-0.064	-0.083*	-0.072
	(0.048)	(0.045)	(0.052)	(0.048)	(0.049)	(0.045)
I2: Letter & social info	-0.003	0.004	-0.026	-0.019	-0.050	-0.045
	(0.041)	(0.039)	(0.044)	(0.042)	(0.041)	(0.039)
Last month	-0.091	-0.093*	-0.140**	-0.132**	-0.168***	-0.157***
	(0.056)	(0.054)	(0.060)	(0.058)	(0.054)	(0.052)
I1*last month	0.086	0.072	0.090	0.072	0.148*	0.136*
	(0.087)	(0.085)	(0.094)	(0.089)	(0.087)	(0.082)
I2*last month	0.053	0.061	0.090	0.100	0.160**	0.172**
	(0.077)	(0.071)	(0.084)	(0.077)	(0.076)	(0.069)
Strata	yes	yes	yes	yes	yes	yes
Controls	no	yes	no	yes	no	yes
N	789	789	789	789	789	789
Panel b)						
Any letter (I1+I2)	-0.023	-0.016	-0.043	-0.035	-0.062*	-0.054
	(0.036)	(0.034)	(0.039)	(0.037)	(0.037)	(0.035)
Last month	-0.091	-0.093*	-0.140**	-0.132**	-0.168***	-0.157***
	(0.056)	(0.054)	(0.060)	(0.058)	(0.054)	(0.052)
(I1+I2)*last month	0.065	0.065	0.089	0.089	0.155**	0.158**
	(0.069)	(0.066)	(0.075)	(0.071)	(0.069)	(0.064)
(I1+I2)+(I1+I2)*last month	0.042	0.049	0.046	0.054	0.094	0.104*
	(0.059)	(0.056)	(0.064)	(0.061)	(0.058)	(0.054)
Strata	yes	yes	yes	yes	yes	yes
Controls	no	yes	no	yes	no	yes
N	789	789	789	789	789	789

Note: Last month indicates whether the invitation letter was sent within the last month before the beginning of the remedial math course. *Outcome variables:* sign-up for remedial math course, participation in first tutorial of remedial math course, and average participation is the share of tutorials a student participated in; *strata:* study program FE, winter term dummy, and interaction between study program FE and winter term dummy; *controls:* first university and female dummies, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The fact that the two invitation letters produce the same pattern and magnitude of results suggests that the social information provided in our main treatment plays little role. Rather, it is plausible that the remedial math course is not salient enough among students who enroll late, but that invitation letters with and without social information are able to mitigate this. In light of this finding, and to increase the statistical power of our analysis, in Panel b) of Table 2.2, we present results that pool observation from both treatment arms: sending students any of the two letters is 15.5 to 15.8 pp ($p = 0.024 - 0.014$) more effective at increasing average participation for students who enroll late compared to those that enroll early; treatment effects in the two groups are 9.4 to 10.4 pp ($p = 0.108 - 0.057$) and -5.4 to -6.2 pp ($p = 0.119 - 0.096$), respectively.

Endogenous stratification. So far, we have presented evidence that, on average, the invitation letter intervention does not affect sign up for and participation in the remedial math course, but that effects are heterogeneous with respect to the timing of enrollment. Going beyond the pre-registered dimensions, the goal of the following analysis is to explore heterogeneous effects in a more general way.[24] Specifically, it is conceivable that the effects of the invitation letters – in particular the one including information on past sign-up rates – depend on the ex-ante sign-up probability of students: First, students with a low sign-up probability simply have more room for improvement. Second, one might assume that some students refrain from participation because their beliefs about the sign-up rates of other students, and thus also about the utility of the course or the descriptive norm, are too low (see theoretical considerations in Section 2.2.2). In this case, information about past sign-up rates may provide a signal that shifts students' beliefs about the utility of the remedial math course or the descriptive norm upwards, thereby increasing sign-up and participation.

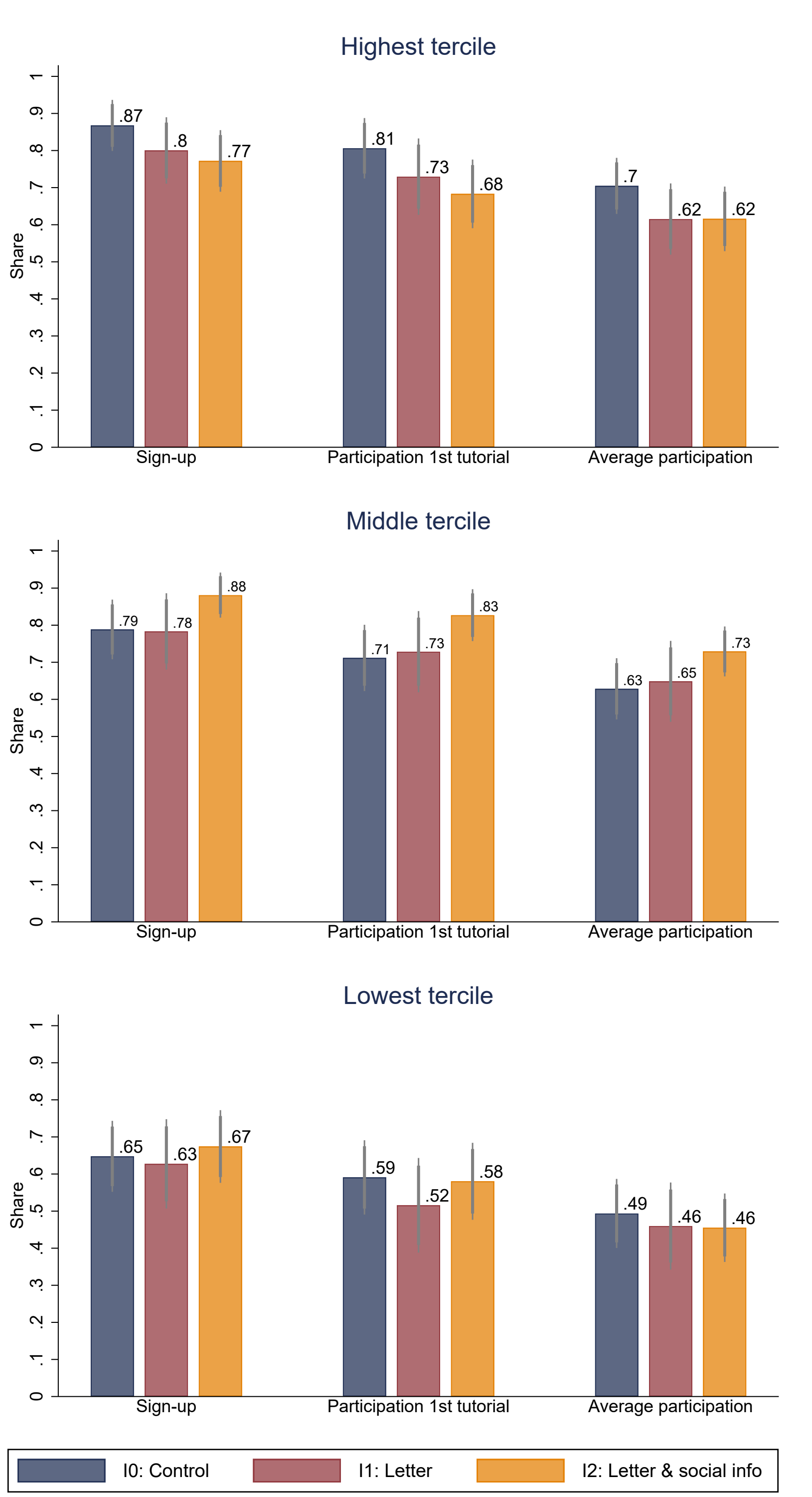
To study heterogeneity along this dimension, we construct endogenous strata employing an approach similar to Abadie et al. (2018): First, in the control group, we regress sign-up on all pre-registered controls, the strata variables, and their interactions with the winter cohort dummy. Next, we use the estimates to predict the sign-up probabilities in the control and treatment groups. For the control group, we use the leave-one-out predictions to avoid "overfitting bias" (see Abadie et al. 2018). Finally, within the two cohorts, we divide our sample into terciles to obtain three endogenous strata (low, middle, and high ex-ante sign-up probability). We then run the following regression specification to estimate treatment effects by strata:

$$\begin{aligned} Y_i^k = {} & \alpha_0 + \alpha_1 Letter_i + Letter_i \mathbf{e_i} \boldsymbol{\alpha_2} + \alpha_3 LetterSocialInfo_i \\ & + LetterSocialInfo_i \mathbf{e_i} \boldsymbol{\alpha_4} + \mathbf{e_i} \boldsymbol{\alpha_5} + \mathbf{x_i} \boldsymbol{\alpha_6} + \mathbf{z_i} \boldsymbol{\alpha_7} + \varepsilon_i, \end{aligned} \tag{2.4}$$

where Y_i^k, $Letter_i$, $LetterSocialInfo_i$, $\mathbf{x_i}$, and $\mathbf{z_i}$ are defined as in Equation 2.2. $\mathbf{e_i}$ includes the endogenous strata dummies. Based on this equation, we provide unadjusted estimates of the treatment effects in each endogenous strata by only including the vector $\mathbf{x_i}$, and adjusted estimates by also including the vector $\mathbf{z_i}$ with additional covariates.

[24] Studying heterogeneity through endogenous stratification was only included in the pre-registration for the effects on academic achievement.

FIGURE 2.2: EFFECT OF INVITATION LETTER BY ENDOGENOUS STRATA



Note: The figure depicts estimates for the three endogenous strata based on Equation 2.4. The endogenous strata group students into terciles of the predicted sign-up probability (see Section 2.3.2). *Outcome variables:* sign-up for remedial math course, participation in first tutorial of remedial math course, and average participation is the share of tutorials a student participated in; *strata:* study program FE, winter term dummy, interaction between study program FE and winter term dummy, and invitation letter date FE. 90% (thick) and 95% (thin) confidence intervals based on robust standard errrors are shown.

Besides allowing us to study heterogeneity with respect to students (counterfactual) sign-up probabilities, this approach has further advantages: First, compared to exploring heterogeneities along multiple covariates and sample splits, it reduces issues associated with multiple hypotheses testing and the selective presentation of significant results. Second, from a policy perspective, identifying heterogeneities on this dimension could provide an easy way to target the intervention in future cohorts. This might be necessary, for instance, if there are negative effects for some students, but it could also help to further reduce the already low cost of the intervention.

TABLE 2.3: EFFECT OF INVITATION LETTER BY ENDOGENOUS STRATA

	Tercile	Sign-up		Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I1: Letter	Lowest	-0.020	-0.012	-0.075	-0.068	-0.034	-0.030
		(0.077)	(0.075)	(0.081)	(0.077)	(0.075)	(0.070)
	Middle	-0.005	-0.026	0.017	-0.006	0.020	-0.002
		(0.066)	(0.062)	(0.072)	(0.068)	(0.070)	(0.067)
	Highest	-0.067	-0.028	-0.076	-0.029	-0.089	-0.039
		(0.055)	(0.055)	(0.065)	(0.064)	(0.060)	(0.059)
I2: Letter & social info	Lowest	0.027	0.059	-0.010	0.026	-0.038	-0.003
		(0.068)	(0.062)	(0.071)	(0.064)	(0.065)	(0.058)
	Middle	0.093*	0.087*	0.115**	0.109*	0.101*	0.089*
		(0.051)	(0.049)	(0.057)	(0.056)	(0.054)	(0.052)
	Highest	-0.096*	-0.066	-0.123**	-0.096*	-0.089	-0.061
		(0.053)	(0.051)	(0.061)	(0.058)	(0.057)	(0.055)
Strata		yes	yes	yes	yes	yes	yes
Controls		no	yes	no	yes	no	yes
N		789	789	789	789	789	789
I1: P-value F-test int. term		0.745	0.983	0.565	0.829	0.485	0.915
I2: P-value F-test int. term		0.038	0.078	0.018	0.042	0.046	0.137

Note: The table depicts treatment effect estimates for the three endogenous strata based on Equation 2.4. The endogenous strata group students into terciles of the predicted sign-up probability (see Section 2.3.2). F-tests in the bottom rows test the hypothesis that all interaction terms between the respective treatment indicator and the endogenous strata, i.e., $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_4}$ in Equation 2.4, are equal to zero. *Outcome variables:* sign-up for remedial math course, participation in first tutorial of remedial math course, and average participation is the share of tutorials a student participated in; *strata:* study program FE, winter term dummy, interaction between study program FE and winter term dummy, and invitation letter date FE; *controls:* first university and female dummies, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Results are reported in Figure 2.2 and Table 2.3. Among students in the highest tercile of predicted sign-up probabilities, 86.7% (88.4% without any controls) sign up for the remedial math course. In this group, both the letter without and the letter with social information decrease sign-up by 6.7 pp (p = 0.224) and 9.6 pp (p = 0.071), respectively (see Column 2 in Table 2.3). Importantly, these effects persist: students in the two treatment groups are 7.7 pp (p = 0.240) and 12.3 pp (p = 0.043) less likely to participate in the first tutorial of the course, and show a 8.9 pp (p = 0.140 and p = 0.120) lower average participation rate across all tutorials (see Columns 4 and 6 in Table 2.3). Adjusting for covariates in Columns (3), (5), and (7) leads to attenuated estimates, in particular for the pure invitation letter.

In the middle tercile, 78.9% (79.0% without any controls) of control group students sign up for the course. While the invitation letter without social information has no effect on either sign-up or participation in this group, our estimates indicate that the letter with social information is able to increase sign-up for the course by 9.3 pp ($p = 0.070$; see Column 2 in Table 2.3). This effect translates into a higher participation rate in the first tutorial and across all tutorials by 11.5 pp ($p = 0.045$) and 10.1 pp ($p = 0.062$), respectively (see Columns 4 and 6 in Table 2.3). Here, Columns (3), (5), and (7) indicate that the estimated effects are robust to the inclusion of covariates.

For students in the lowest tercile – among whom 64.7% (61.4% without any controls) sign up for the remedial math course – our estimates indicate that neither type of invitation letter is able to affect the decision to sign up for or participate in the math course. Both for the letter with and the letter without social information, we test if the interaction with the endogenous strata, i.e., $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_4}$ in Equation 2.4, are equal to zero. The p-values of the corresponding F-tests are depicted in the bottom rows of Table 2.3. For the invitation letter with social information, the null hypothesis can be rejected at the 5 to 10%-level in all but one of the specifications.

Turning back to the theoretical considerations, the pattern of results found for the invitation letter with information on past sign-up rates can be explained in the following way: For students in the lowest tercile, the social information may not be able to increase sign-up, because the signal does not increase beliefs sufficiently, or because these students simply do not expect to gain any value from the course, e.g., because they have previously participated in a similar course. The latter notion is supported by the observation that the share of students for whom this is the first semester at any university is particularly low in this strata (35.5%, compared to 77.8 and 95.6% in the middle and highest strata). The middle tercile, on the other hand, may consist of marginal students – i.e., students for whom the social information nudge leads to an increase in beliefs that is large enough to induce sign-up and participation. For the highest tercile, the opposite could be driving the results. Some students might have expected a higher sign-up rate than our letter suggests, and the signal may thus have led to a downward adjustment in beliefs, leading to lower sign-up and participation rates.

2.3.3 Effects on academic achievement

The goal of the remedial math course is to prepare students for the mathematically more demanding subjects of their studies. To provide some evidence on whether the course is successful in doing so, in this section, we study if the heterogeneous effects on average participation presented above translate into increased performance in the math exam and higher overall achievement in the first year of studies.[25]

Timing of enrollment. Tables 2.4 and 2.5 present heterogeneous treatment effects on academic achievement with respect to students' timing of enrollment. To increase statistical power, and because the effects on average participation were similar, we again pool observations from both treatment arms. Both tables show that control students who enroll late suffer from lower academic achievement compared to early enrollees: After the first semester, they are 13.9 pp ($p = 0.016$) and 12.4 pp ($p = 0.032$) less likely to have attempted and passed the math exam, respectively (Columns 1 and 2 of Table 2.4). Further, their overall performance index – i.e., the standardized inverse-covariance weighted average of obtained credits, dropout, and GPA – is 0.256 standard deviations lower ($p = 0.070$), they obtain 4.23 credits ($p = 0.008$) less, and are 6.2 pp ($p = 0.115$) more likely to have dropped out of their study program (Columns 1 to 3 in Table 2.5).[26] After the first year of studies, they are even less likely to have attempted or passed the math exam (−17.0 pp, $p = 0.003$ and −15.2 pp, $p = 0.009$; Columns 4 and 5 of Table 2.4), and their overall performance is still 0.257 standard deviations ($p = 0.046$) lower compared to students who enroll early (Column 5 of Table 2.5).

For all dimensions of academic achievement presented in Tables 2.4 and 2.5, we find that the invitation letters are able to offset all or almost all of the disadvantage in academic achievement that we observe for students who enroll late. Given that our intervention increased average participation of these students by about 10 pp, it may seem difficult at first to rationalize these large effects on performance: for example, if we assume that 10% of the students in the treatment group go from 0 to 100% participation in the course, the estimated effect on first year credits (Column 6 in Table 2.5) would imply that participation in the course increases obtained first year credits by about 56, which is close to the course load of a full academic year (see Footnote 17).

[25] As mentioned in Footnote 14, the analyses of the effects on academic achievement were pre-registered after the analysis of the effects on sign up for and participation in the course, but before data on academic achievement became available to us.

[26] Effects on the grade in the math exam and the GPA go in the same direction (in the German system 1.0 is the best, and 4.0 the worst passing grade). They should, however, be interpreted with caution, since the outcomes are only observed for students who have attempted the math exam at least once or passed at least one graded exam, respectively.

TABLE 2.4: EFFECT OF INVITATION LETTER ON PERFORMANCE IN MATH EXAM BY TIMING OF ENROLLMENT

	First semester			First year		
	Attempted	Passed	Grade	Attempted	Passed	Grade
	(1)	(2)	(3)	(4)	(5)	(6)
Any letter (I1+I2)	-0.048	-0.018	-0.005	-0.069*	-0.062	0.043
	(0.040)	(0.043)	(0.095)	(0.038)	(0.042)	(0.086)
Last month	-0.139**	-0.124**	0.201	-0.170***	-0.152***	0.119
	(0.058)	(0.058)	(0.144)	(0.057)	(0.058)	(0.130)
(I1+I2)*last month	0.160**	0.122*	-0.106	0.184***	0.128*	0.028
	(0.070)	(0.073)	(0.179)	(0.069)	(0.073)	(0.164)
(I1+I2)+(I1+I2)*last month	0.112*	0.105*	-0.110	0.115**	0.065	0.071
	(0.059)	(0.060)	(0.153)	(0.058)	(0.061)	(0.142)
Strata	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes
N	789	789	519	789	789	549

Note: Last month indicates whether the invitation letter was sent within the last month before the beginning of the remedial math course. *Outcome variables:* math exam attempted, math exam passed, grade in the math exam includes failing grades and is only observed for students who attempted the math exam at least once (highest passing grade is 1.0; lowest passing grade is 4.0; failing grade is 5.0); *strata:* study program FE, winter term dummy, and interaction between study program FE and winter term dummy; *controls:* first university and female dummies, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.5: EFFECT OF INVITATION LETTER ON OVERALL PERFORMANCE BY TIMING OF ENROLLMENT

	First semester				First year			
	Index	Credits	Dropout	GPA	Index	Credits	Dropout	GPA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Any letter (I1+I2)	-0.059	0.187	0.014	0.067	-0.118	-0.550	0.046	0.057
	(0.083)	(1.120)	(0.023)	(0.059)	(0.085)	(1.977)	(0.031)	(0.054)
Last month	-0.256*	-4.226***	0.062	0.096	-0.257**	-7.174***	0.071	0.092
	(0.141)	(1.582)	(0.039)	(0.098)	(0.129)	(2.731)	(0.048)	(0.080)
(I1+I2)*last month	0.361**	3.821*	-0.091**	-0.169	0.349**	6.170*	-0.112*	-0.145
	(0.159)	(1.970)	(0.044)	(0.119)	(0.158)	(3.461)	(0.058)	(0.102)
(I1+I2)+(I1+I2)*last month	0.303**	4.008**	-0.077**	-0.102	0.231*	5.620*	-0.067	-0.088
	(0.136)	(1.647)	(0.038)	(0.105)	(0.133)	(2.867)	(0.049)	(0.088)
Strata	yes	yes	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes	yes	yes
N	789	789	789	550	789	789	789	596

Note: Last month indicates whether the invitation letter was sent within the last month before the beginning of the remedial math course. *Outcome variables:* the index is the standardized inverse-covariance weighted average of the three overall performance measures (following Anderson (2008) and using the Stata program by Schwab et al. (2020)), obtained credits, dropout indicates if a student dropped out of their study program, grade point average includes passing grades only and is unobserved for students who have not obtained a passing grade yet (highest passing grade is 1.0, lowest passing grade is 4.0); *strata:* study program FE, winter term dummy, and interaction between study program FE and winter term dummy; *controls:* first university and female dummies, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

This raises the question as to how the large increase in achievement among late enrollees comes about. One possible explanation is that these students suffer from low motivation and engagement with their studies in the absence of treatment. The remedial math course may then be particularly beneficial, because it leads to higher engagement with the university, other students, and their studies, thereby leading to higher motivation and preventing students from dropping out of their program early on. This may be of particular relevance in the German context, where tuition fees are generally very low, and students thus face very low direct costs of studying.[27] The higher dropout rate among late enrollees in the control group and the decrease in dropout due to treatment presented in Table 2.5 already provides some evidence that is consistent with that notion.

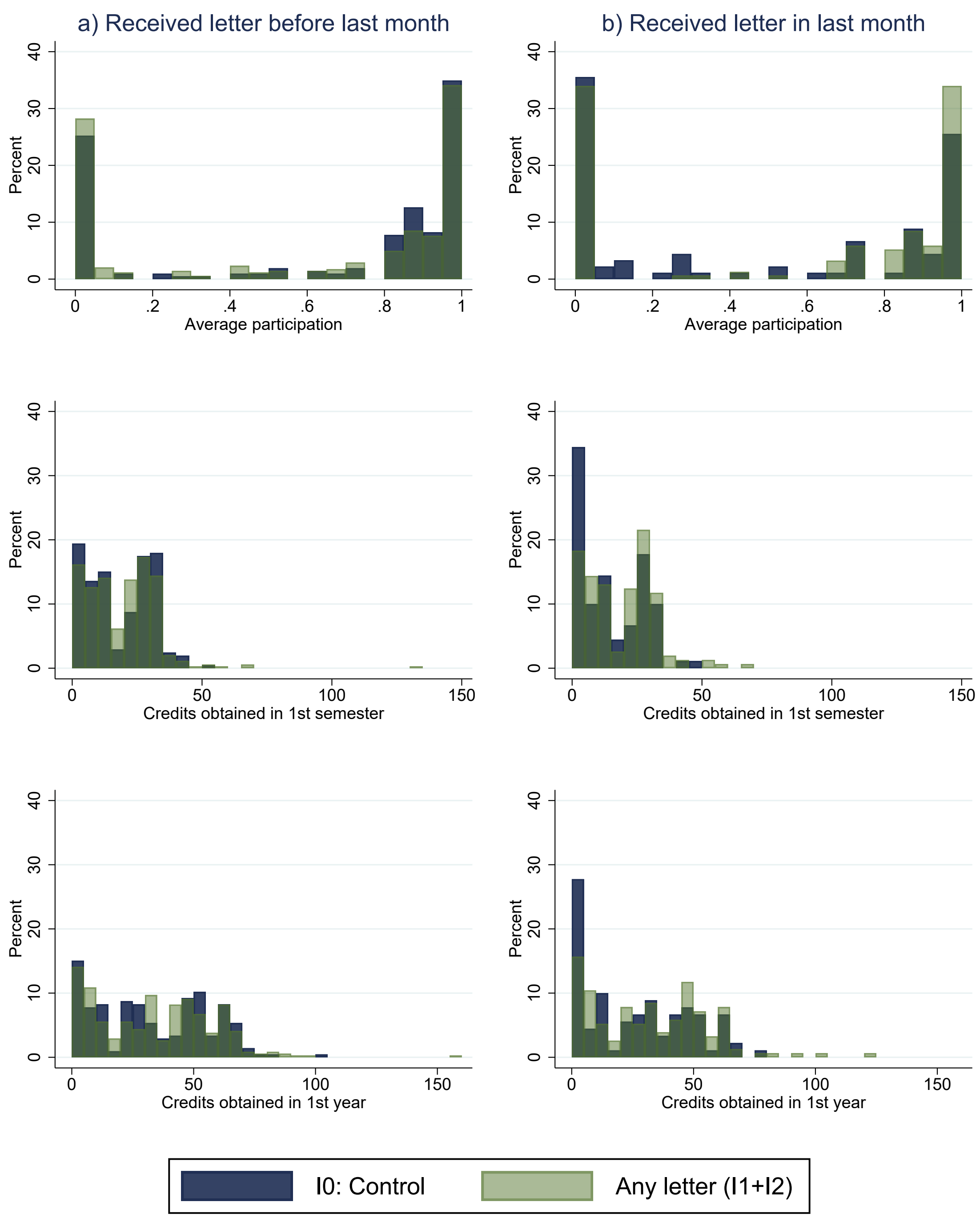
To study this idea further, Figure 2.3 depicts histograms of average participation and obtained credits by whether students received the invitation letter within the last month before the beginning of the course, separately by treatment status (the distribution of the control group is shown in dark blue and is overlaid by the distribution of the combined treatment group in transparent green). The top plot in Panel b) first provides evidence that – among late enrollees – the invitation letters do indeed lead to a large increase in average participation among few individuals rather than a small increase among many individuals. Second, the middle and the bottom plot in Panel b) provide evidence that in this group of students, about 35% (28%) of the controls obtain fewer than 5 credits in the first semester (year) – providing evidence for the low engagement with their studies. Among treated students, this share is decreased by roughly 15 pp and the distribution of credits instead is very similar to the ones that we observe for students that enroll early (shown in Panel a).

Overall, these results provide evidence that increasing remedial math course participation among late enrollees results in higher academic performance in the first year of studies. Given that late enrollees make up around 31% of the sample, our findings suggest that the remedial math course can be beneficial for a considerable number of students. In fact, since our intervention only includes students who enroll in their study program up to one week before the math course begins, our results may represent a lower bound.[28]

[27]At the time of the field experiments, the tuition fees at the university were around 350€ per semester. The tuition fees include free use of public transportation in the city where the university is located, and students can also ride regional trains in the federal state for free.

[28]Across both cohorts, about 100 additional students enrolled at an even later date, making it impossible for some of them to attend the remedial math course.

FIGURE 2.3: HISTOGRAMS BY TIMING OF ENROLLMENT



Note: Last month indicates whether the invitation letter was sent within the last month before the beginning of the remedial math course. The histograms of the control group (dark blue) are overlaid by the distribution of the combined treatment group (transparent green). Histograms start at 0 and have a binwidth of 0.05 (average participation) and 5 (obtained credits).

Endogenous strata. Studying treatment effect heterogeneity across the endogenous strata suggests that increased remedial math course participation does not translate into higher academic achievement among all subgroups. Based on Equation 2.4, Tables 2.A4 and 2.A5 report effects of the two types of invitation letters on the academic achievement dimensions across the three endogenous strata. Overall, we find little to no robust evidence that the effects on participation presented in Table 2.3 carry over to academic performance. For one, there is very little evidence for any significant effects in the first place. Given the large number of estimates, some of the few significant ones may simply arise from multiple hypothesis testing. In addition, given that the different dimensions of academic achievement are usually highly correlated with each other, we would expect changes in average participation to translate into consistent changes in academic achievement across the different outcomes – similar to what we found with respect to the timing of enrollment. However, this is not the case, further suggesting that the few significant estimates do not represent a robust pattern.

This raises the question as to why the effects on participation that we found for the invitation letter with social information do not carry over to academic achievement. One plausible explanation is that the academic performance of students who are identified by the endogenous strata is simply not so easily changed by remedial math course participation. This may for example be the case because the students in the middle tercile – who increase participation in response to treatment – already possess the knowledge that is necessary to pass the exams. Table 2.A6 depicts the means of our outcomes by endogenous strata among control group students and provides some evidence that supports this idea: with respect to almost all outcomes, students in the middle tercile have a higher performance compared to students in the highest tercile; they are, e.g., more likely to have passed the math exam after the first year of studies, they obtain more credits, and they are similarly likely to drop out of their program. More generally – and in contrast to the timing of enrollment – we observe that the heterogeneous pattern of sign-up and participation among control group students across the endogenous strata does not carry over to academic achievement.

2.4 Effects of the reminder letter intervention

In this section, we present the results of the reminder letter intervention that we conducted among students who signed up for the remedial math course.

Main effects. Table 2.6 reports the main effects of the reminder letter intervention. The bottom row shows that 90% of control group students who signed up for the course go on to participate in the first tutorial, implying an attrition of 10 pp. Sending students the reminder

TABLE 2.6: EFFECT OF REMINDER LETTER

	Participation 1st tutorial			Average participation		
	(1)	(2)	(3)	(4)	(5)	(6)
Letter	0.026	0.023	0.026	-0.018	-0.024	-0.019
	(0.033)	(0.032)	(0.032)	(0.036)	(0.034)	(0.034)
Letter & social info	0.028	0.028	0.028	0.013	0.016	0.014
	(0.027)	(0.027)	(0.026)	(0.030)	(0.029)	(0.028)
Strata	yes	yes	yes	yes	yes	yes
Controls	no	yes	d-p lasso	no	yes	d-p lasso
N	574	574	574	574	574	574
Control mean	0.90	0.90	0.90	0.79	0.79	0.79
(SD)	(0.30)	(0.30)	(0.30)	(0.32)	(0.32)	(0.32)

Note: Outcome variables: participation in first tutorial of remedial math course and average participation is the share of tutorials a student participated in; *strata:* study program FE, winter term dummy, interaction between study program FE and winter term dummy, and invitation letter treatment status FE; *controls:* first university and female dummies, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent; the double-post LASSO specification considers all controls. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

letter with social information about the usefulness of the course increases participation by 2.8 pp, independent of the exact specification ($p = 0.293$ in Column 3). The pure reminder letter without social information also increases participation by about 2.6 pp ($p = 0.407$ in Column 3), indicating that it is the reminder itself, and not the social information that leads to an increase in participation in the first tutorial. These estimates imply a substantial reduction in attrition of nearly 30%; however, they are not estimated precisely.

Next, we look at average participation in the math course across all tutorials (Columns 4 to 6). Among control group students, the average participation rate is 79%, leaving substantially more room for improvement. However, the effects on participation in the first tutorial do not translate into a higher participation across all tutorials. Students who receive a letter with social information increase their participation rate by 1.3 to 1.6 pp ($p = 0.658 - 0.582$), while students that receive a letter without social information decrease their participation rate by 1.8 to 2.4 pp ($p = 0.612 - 0.486$). A reason for this could be that students learn over time that the course is not as useful as the reminder letters suggested.

Overall, these results suggest that sending reminder letters, with and without social information on the usefulness of the course, to students who already signed up for the course, does, on average, not lead to a sustained change in participation rates.

Heterogeneity. In Table 2.A7, we investigate whether the average treatment effects presented above hide heterogeneous effects along the pre-registered dimensions (first university and gender). Across both dimensions, we find little to no evidence for heterogeneous treatment effects.[29] Similar to the invitation letter intervention, we also tried to employ endogenous stratification. However, since already 90% of control group students participate in the first tutorial, the predictive model based on the pre-registered covariates did not perform well and produced very little heterogeneity in estimated participation probabilities. To us, this indicates that among students who initially signed up for the course, reasons other than those that we can capture with our covariates drive the participation decision. For example, some students may not yet have moved to the city where the university is located or they may be subject to other external constraints.

2.5 Discussion and conclusions

In the context of a voluntary remedial math course at a large German university, we study whether social information nudges can be used to influence educational investment decisions. The results of our field experiments show that both an invitation letter with social information on past sign-up rates as well as a reminder letter with information on past evaluations of the course are, on average, not able to increase sign up for and participation in the course.

One concern with our main effects could be that our study is simply not well-powered enough to detect effects that are typically found in the literature. In their meta-analysis, DellaVigna and Linos (2020) report an average treatment effect of 13.81 pp for studies published in academic journals that make use of social cues. As we show in Appendix C, our study has enough power to detect these effects sizes, but also those that are more close to interventions that make use of reminders (5.02 pp). However, for studies conducted by nudge units, DellaVigna and Linos (2020) only report average treatment effects between 0.96 pp (social cues) and 2.56 (reminders), which is much closer to the main effects of this study. While the authors show that some of the differences in the effects can be explained by study characteristics, much of it appears to stem from publication bias. Given this knowledge, the results of our study may well be representative for the effect size that researchers should expect to find for the average treatment effect of social information nudges, although we do not have the power to estimate these kinds of effects precisely. Since this is the first study on this type of social information nudge in an educational context, our results may still serve as a basis for future studies in this area.

[29] For the pure reminder letter, the estimates suggest a very large interaction between treatment and the first university dummy (about 20 pp). However, this differential treatment effect is driven by a large negative treatment effect of about 19 pp among students who have already studied at this or any other university. Since this subgroup consists only of 24 students, these estimates should be interpreted with great caution.

Studying the heterogeneity of the main effects, we find evidence that both the invitation letter with and without social information increases participation among students who enroll late, suggesting that low salience may have been causing the low attendance that we observe for these students. In addition, we employ endogenous stratification and show that marginal students increase sign-up and attendance by 10 pp in response to the invitation letter with social information. Students who have a high estimated sign-up probability, on the other hand, decrease sign-up and attendance by about 10 pp after receiving the invitation letter with social information. Our study thus highlights that a targeted provision of social information may be necessary to prevent undesired results.

One explanation for those heterogeneities is that the behavioral response of individuals depends on their beliefs about the behavior of similar others. This has been observed for other contexts, such as the labor market (Coffman et al., 2017) and for antiauthoritarian protests (Cantoni et al., 2019). A caveat of our study is that we do not have direct information on students' beliefs. We try to address this with the help of endogenous stratification. However, this approach may be subject to concerns regarding the exact covariates and specifications that are used to construct the endogenous strata. We circumvent this by only including pre-registered covariates, even though this may result in us failing to find the specification that is most predictive of the outcome of interest. Future studies should elicit beliefs directly or study in advance which covariates are predictive of the outcome that is to be affected by the social information nudge.

Lastly, we also study whether the heterogeneous effects on course participation carry over to academic achievement in the first year of studies. By doing so, we provide first evidence on the effectiveness of (voluntary) remedial math courses based on experimentally induced variation in attendance. In situations where the intervention of interest cannot be evaluated directly, e.g., because it has to be offered to all students, such an encouragement design may be the only viable option. Because we only found increases in math course participation among subgroups, these analyses have limited statistical power, and should therefore be interpreted with caution.

We show that the increase in average participation among students who enroll late leads to robust changes in academic performance. Our results for late enrollees are consistent with the idea that an early engagement with their studies in the form of the remedial course may prevent these students from dropping out early, explaining their large gains in academic achievement. Considering that these students make up a substantial part of the sample and perform poorly in their first year of studies, they are an important population for further research. The heterogeneous effects of social information on participation for students who fall short off or exceed the descriptive norm, on the other hand, do not translate into changes in achievement. This highlights the importance of investigating whether influencing individual behavior, in this case participation in the remedial course, actually results in the desired outcomes later on.

Appendix

A Additional tables and figures

TABLE 2.A1: DESCRIPTION OF VARIABLES

Variable	Description
Treatment Variables	
Letter (I1 or R1)	Random assignment to the treatment group that receives an invitation (reminder) letter without the social information. Treatments only included in the winter cohort.
Letter & social info (I2 or R2)	Random assignment to the treatment group that receives an invitation (reminder) letter that includes the social information.
Stratification Variables	
Study program	BA=Business Administration, BIS=Business and Information Systems, BHRE=Business and Human Resource Education, E=Economics, 2SE=Two-Subject Economics.
Winter term	Dummy for the second cohort.
Invitation letter date	Date at which the invitation letter was sent to a student (see Table 2.B2). Only included in analyses of the effects of the invitation letter.
Invitation letter treatment status	Indicators for the treatment status in the invitation letter intervention, such that the two randomizations are orthogonal to each other. Only included in analyses of the effects of the reminder letter.
Control Variables	
First university$^{a)}$	Dummy that indicates if this is the first semester at any university.
Female$^{a)}$	Indicator for being female.
Age	Age in years at the beginning of the first semester.
HS GPA	Final high school grade point average (1.0=highest, 4.0=lowest).
Fresh HS degree	Dummy if the high school degree was obtained within the last year before the beginning of the first semester.
HS degree Abitur	Indicator for a general track degree (“Abitur”); reference category includes vocational track degree (“Fachhochschulreife”) and students who hold other degrees.
Place of HS degree	NI=Lower Saxony, NW=North Rhine-Westphalia, HE=Hesse, other=another federal state in Germany, and abroad.
Distance letter$^{b)}$	Distance over which the letter was sent (in kilometers).
Sign-up before letter$^{c)}$	Indicates if a student signed up for the remedial math course before the letter was sent to her, or could theoretically be sent to her in case of the control group.
Last month	Dummy that indicates whether the invitation letter was sent within the last month before the beginning of the remedial math course. Only used in the respective heterogeneity analysis.
Outcome Variables	
Sign-up	Indicates if a student signed up for the remedial math course.
Participation first tutorial	Indicates if a student participated in the first tutorial of the remedial math course.
Average participation	Share of tutorials that a student participated in.
Math attempted	Indicates whether a student attempted the math exam by the end of the first semester/year.
Math passed	Indicates whether a student passed the math exam by the end of the first semester/year.
Math grade	Grade in the math exam by the end of the first semester/year including failing grades (1.0 = highest and 4.0 = lowest passing grade, 5.0 = failing grade). Only observed if the exam was attempted at least once.
Credits	Number of credits obtained in the first semester/year.
Dropout	Indicates whether a student dropped out of her study program in the first semester/year.
GPA	Grade point average at end of first semester/year (passing grades only; 1.0=highest, 4.0=lowest). Only observed if a student passed at least one graded exam.
Performance index	Standardized inverse-covariance weighted average of credits, dropout, and GPA (following Anderson (2008) and using the Stata program by Schwab et al. (2020)).

Note: $^{a)}$ As explained in Section 2.2, and as it was intended in the pre-registrations, the first university and female dummies were in some cases used during stratification. However, due to the number of observations per cell, they could not be included in most of the randomizations, and we therefore include them with the other controls. $^{b)}$ The distance over which the letter was sent was not listed as a control variable in the pre-registration of the first experiment. $^{c)}$ Sign-up before letter was not listed as a control variable in the pre-registration of either experiment.

TABLE 2.A2: REGRESSION OF ACADEMIC ACHIEVEMENT ON REMEDIAL MATH COURSE PARTICIPATION – CONTROL GROUP OF INVITATION LETTER INTERVENTION (I0)

	Math exam			Overall performance		
	Attempted	Passed	Grade	Credits	Dropout	GPA
	(1)	(2)	(3)	(4)	(5)	(6)
a) First semester						
Average participation	0.530***	0.447***	-0.145	10.364***	-0.136***	0.012
	(0.069)	(0.068)	(0.203)	(1.766)	(0.046)	(0.137)
High school GPA	0.018	-0.108**	0.760***	-3.621**	-0.051	0.391***
	(0.052)	(0.054)	(0.132)	(1.553)	(0.035)	(0.098)
Strata	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes
N	296	296	195	296	296	197
a) First year						
Average participation	0.501***	0.467***	-0.255	19.027***	-0.167***	-0.109
	(0.065)	(0.070)	(0.192)	(3.156)	(0.058)	(0.108)
High school GPA	0.026	-0.085	0.761***	-6.263**	-0.025	0.353***
	(0.053)	(0.054)	(0.124)	(2.676)	(0.041)	(0.083)
Strata	yes	yes	yes	yes	yes	yes
Controls	yes	yes	yes	yes	yes	yes
N	296	296	209	296	296	224

Note: Average participation is the share of tutorials a student participated in. In the German system 1.0 is the best, and 4.0 the worst high school GPA. *Outcome variables:* math exam attempted, math exam passed, grade in the math exam includes failing grades and is only observed for students who attempted the math exam at least once (highest passing grade is 1.0; lowest passing grade is 4.0; failing grade is 5.0), obtained credits, dropout indicates if a student dropped out of their study program, grade point average includes passing grades only and is unobserved for students who have not obtained a passing grade yet (highest passing grade is 1.0, lowest passing grade is 4.0); *strata:* study program FE, winter term dummy, interaction between study program FE and winter term dummy, and invitation letter date FE; *controls:* age, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.A3: EFFECT OF INVITATION LETTER – HETEROGENEITIES

	Sign-up		Part. 1st tutorial		Avgerage part.	
	(1)	(2)	(3)	(4)	(5)	(6)
First university	0.279***	0.272***	0.313***	0.311***	0.300***	0.320***
	(0.059)	(0.061)	(0.060)	(0.063)	(0.055)	(0.057)
I1*first university	0.027	0.006	0.056	0.027	0.026	0.000
	(0.101)	(0.101)	(0.101)	(0.099)	(0.091)	(0.090)
I2*first university	0.005	-0.012	0.002	-0.014	0.064	0.043
	(0.080)	(0.079)	(0.082)	(0.082)	(0.073)	(0.072)
Female	0.032	0.015	-0.033	-0.061	-0.004	-0.030
	(0.050)	(0.051)	(0.055)	(0.056)	(0.051)	(0.052)
I1*female	-0.017	-0.005	0.106	0.126	0.116	0.136*
	(0.078)	(0.079)	(0.086)	(0.085)	(0.081)	(0.080)
I2*female	-0.015	0.021	0.031	0.061	0.046	0.067
	(0.069)	(0.068)	(0.076)	(0.075)	(0.070)	(0.069)
Last month	-0.082	-0.093*	-0.133**	-0.134**	-0.159***	-0.159***
	(0.054)	(0.054)	(0.058)	(0.058)	(0.053)	(0.053)
I1*last month	0.088	0.071	0.103	0.086	0.165*	0.150*
	(0.085)	(0.085)	(0.091)	(0.089)	(0.085)	(0.082)
I2*last month	0.056	0.061	0.096	0.102	0.168**	0.176**
	(0.072)	(0.071)	(0.078)	(0.077)	(0.071)	(0.069)
Strata	yes	yes	yes	yes	yes	yes
Controls	no	yes	no	yes	no	yes
N	789	789	789	789	789	789

Note: Outcome variables: sign-up for remedial math course, participation in first tutorial of remedial math course, and average participation is the share of tutorials a student participated in; *strata:* study program FE, winter term dummy, and interaction between study program FE and winter term dummy; *controls:* age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.A4: EFFECT OF INVITATION LETTER ON PERFORMANCE IN MATH EXAM – BY ENDOGENOUS STRATA

		First semester			First year		
	Tercile	Attempted	Passed	Grade	Attempted	Passed	Grade
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
I1: Letter	Lowest	-0.008	0.018	-0.175	-0.019	-0.031	-0.083
		(0.074)	(0.076)	(0.219)	(0.073)	(0.075)	(0.192)
	Middle	(0.073)	(0.079)	(0.182)	(0.071)	(0.080)	(0.176)
		0.056	0.126*	-0.179	0.064	0.082	-0.060
	Highest	(0.070)	(0.076)	(0.175)	(0.066)	(0.076)	(0.159)
		0.032	-0.053	0.400**	0.000	-0.069	0.413***
I2: Letter & social info	Lowest	0.027	-0.028	0.145	0.007	-0.035	0.109
		(0.064)	(0.065)	(0.164)	(0.063)	(0.065)	(0.147)
	Middle	-0.065	-0.048	-0.047	-0.058	-0.081	0.051
		(0.065)	(0.068)	(0.146)	(0.062)	(0.067)	(0.137)
	Highest	0.007	0.145**	-0.361**	-0.018	0.036	-0.193
		(0.058)	(0.066)	(0.151)	(0.055)	(0.064)	(0.137)
Strata		yes	yes	yes	yes	yes	yes
Controls		yes	yes	yes	yes	yes	yes
N		789	789	519	789	789	549
I1: P-value F-test int. term		0.812	0.337	0.438	0.664	0.458	0.719
I2: P-value F-test int. term		0.539	0.054	0.003	0.797	0.373	0.010

Note: The table depicts treatment effect estimates for the three endogenous strata based on Equation 2.4. The endogenous strata group students into terciles of the predicted sign-up probability (see Section 2.3.2). F-tests in the bottom rows test the hypothesis that all interaction terms between the respective treatment indicator and the endogenous strata, i.e., $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_4}$ in Equation 2.4, are equal to zero. *Outcome variables:* math exam attempted, math exam passed, grade in the math exam includes failing grades and is only observed for students who attempted the math exam at least once (highest passing grade is 1.0; lowest passing grade is 4.0; failing grade is 5.0); *strata:* study program FE, winter term dummy, interaction between study program FE and winter term dummy, and invitation letter date FE; *controls:* first university and female dummies, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.A5: EFFECT OF INVITATION LETTER ON OVERALL PERFORMANCE – BY ENDOGENOUS STRATA

		First semester				First year			
	Tercile	Index	Credits	Dropout	GPA	Index	Credits	Dropout	GPA
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
I1: Letter	Lowest	-0.093	-0.103	0.023	0.081	-0.028	-1.657	-0.002	0.116
		(0.138)	(2.203)	(0.039)	(0.112)	(0.135)	(3.848)	(0.047)	(0.105)
	Middle	0.101	1.366	-0.024	-0.119	0.027	2.160	-0.002	-0.092
		(0.170)	(2.004)	(0.048)	(0.105)	(0.184)	(3.683)	(0.066)	(0.093)
	Highest	0.078	1.945	-0.020	0.114	0.095	3.053	-0.027	0.004
		(0.156)	(1.751)	(0.044)	(0.105)	(0.159)	(3.330)	(0.060)	(0.093)
I2: Letter & social info	Lowest	-0.013	0.412	0.001	0.138	-0.104	-0.165	0.036	0.164
		(0.107)	(1.897)	(0.029)	(0.112)	(0.121)	(3.224)	(0.044)	(0.107)
	Middle	0.104	0.283	-0.029	-0.003	-0.048	-0.427	0.021	0.004
		(0.146)	(1.759)	(0.041)	(0.099)	(0.151)	(3.131)	(0.056)	(0.084)
	Highest	0.100	4.160**	-0.022	-0.088	0.046	5.154*	0.013	-0.116
		(0.147)	(1.789)	(0.041)	(0.101)	(0.149)	(3.066)	(0.056)	(0.088)
Strata		yes	yes	yes	yes	yes	yes	yes	yes
Controls		yes	yes	yes	yes	yes	yes	yes	yes
N		789	789	789	550	789	789	789	596
I1: P-value F-test int. term		0.575	0.767	0.654	0.228	0.837	0.634	0.937	0.318
I2: P-value F-test int. term		0.745	0.224	0.809	0.307	0.739	0.359	0.947	0.119

Note: The table depicts treatment effect estimates for the three endogenous strata based on Equation 2.4. The endogenous strata group students into terciles of the predicted sign-up probability (see Section 2.3.2). F-tests in the bottom rows test the hypothesis that all interaction terms between the respective treatment indicator and the endogenous strata, i.e., $\boldsymbol{\alpha_2}$ and $\boldsymbol{\alpha_4}$ in Equation 2.4, are equal to zero. *Outcome variables:* the index is the standardized inverse-covariance weighted average of the three overall performance measures (following Anderson (2008) and using the Stata program by Schwab et al. (2020)), obtained credits, dropout indicates if a student dropped out of their study program, grade point average includes passing grades only and is unobserved for students who have not obtained a passing grade yet (highest passing grade is 1.0, lowest passing grade is 4.0); *strata:* study program FE, winter term dummy, interaction between study program FE and winter term dummy, and invitation letter date FE; *controls:* first university and female dummies, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.A6: MEAN OF OUTCOMES OF CONTROL GROUP STUDENTS BY ENDOGENOUS STRATA (INVITATION LETTER INTERVENTION)

	(1)	(2)	(3)
Tercile	Lowest	Middle	Highest
Sign-up	0.614	0.790	0.884
Participation first tutorial	0.554	0.720	0.821
Average participation	0.461	0.631	0.721
Math attempted first semester	0.554	0.710	0.716
Math attempted first year	0.614	0.740	0.768
Math passed first semester	0.416	0.540	0.453
Math passed first year	0.495	0.630	0.589
Math grade first semester	3.436	3.662	3.850
Math grade first year	3.377	3.507	3.653
Performance index first semester	0.137	-0.060	-0.083
Performance index first year	0.138	-0.021	-0.125
Credits first semester	16.530	17.375	15.137
Credits first year	30.564	32.395	28.463
Dropout first semester	0.050	0.100	0.105
Dropout first year	0.099	0.180	0.189
GPA first semester	2.863	3.061	3.057
GPA first year	2.832	2.946	3.056

TABLE 2.A7: EFFECT OF REMINDER LETTER – HETEROGENEITIES

	Part. 1st tutorial		Avgerage part.	
	(1)	(2)	(3)	(4)
First university	0.052	0.059	0.108*	0.137**
	(0.055)	(0.057)	(0.059)	(0.058)
R1*first university	0.172	0.164	0.199*	0.201**
	(0.112)	(0.108)	(0.105)	(0.096)
R2*first university	-0.011	-0.007	0.004	0.017
	(0.072)	(0.070)	(0.078)	(0.075)
Female	-0.026	-0.036	0.060	0.046
	(0.046)	(0.046)	(0.046)	(0.045)
R1*female	-0.016	-0.025	-0.049	-0.047
	(0.068)	(0.068)	(0.069)	(0.067)
R2*female	0.060	0.048	-0.027	-0.047
	(0.058)	(0.058)	(0.062)	(0.060)
Strata	yes	yes	yes	yes
Controls	yes	yes	yes	yes
N	574	574	574	574

Note: Outcome variables: participation in first tutorial of remedial math course and average participation is the share of tutorials a student participated in; *strata:* study program FE, winter term dummy, interaction between study program FE and winter term dummy, and invitation letter treatment status FE; *controls:* age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

FIGURE 2.A1: INVITATION LETTER – SOCIAL INFORMATION HIGHLIGHTED IN GRAY (TRANSLATION)

University • postbox <number> • <zip code> <place>

Ms. / Mr.
<first name> <last name>
<street> <number>
<zip code> <place of residence>

Organizer of math course

Phone no. <number>
<e-mail>

<place>, <date>

Remedial math course for students of business and economics

Dear Ms./Mr. <last name>,

in order to help you get off to a good start in your studies, we would like to invite you to the remedial math course for students of business and economics. The course provides mathematical knowledge that is required in the mathematics lecture and in numerous other courses.

85% of the first-year students who, like you, were enrolled in a business or economics degree program in the last semester have signed up for the remedial math course.[1] Only a small minority of students does not sign up for the remedial course.

The remedial math course starts on <date> - we look forward to your participation! Please sign up at <website>.

Kind regards,

the remedial math course team

[1] The calculation of the sign-up rate is based on all students who enrolled in <year> for the winter term <year>.

FIGURE 2.A2: REMINDER LETTER – SOCIAL INFORMATION HIGHLIGHTED IN GRAY (TRANSLATION)

University • postbox <number> • <zip code> <place>

Ms. / Mr.
<first name> <last name>
<street> <number>
<zip code> <place of residence>

Organizer of math course

Phone no. <number>
<e-mail>

<place>, <date>

Remedial math course for students of business and economics

Dear Ms./Mr. <last name>,

you have signed up for the remedial math course. We have therefore already reserved a seat for you and look forward to your participation. The course starts on <date> at <location>.

95% of students who, like you, are enrolled in a business or economics degree program say that the remedial course in mathematics has made it easier for them to get started with university mathematics.[1]

Before the course starts, please inform yourself about the tutorial group you have been assigned to and the room in which your tutorial will take place at <website>.

Kind regards, see you on <date>!

the remedial math course team

[1] The data is taken from a survey among students attending the mathematics lecture, which was conducted in the winter term <year>.

B Randomization and balancing properties

Invitation letter intervention. In both cohorts, starting five to seven weeks before the beginning of the remedial math course until one week before, we used administrative data on the incoming students provided to us by the university to randomize students into a control and one (summer term) or two (winter term) treatment groups. Randomization was carried out using stratification and re-randomization (Morgan and Rubin, 2012). Each week in the summer term and about every other week in the winter term[30], we randomized within study programs and, if possible, i.e., if there were enough observations in the respective cells, within a dummy variable that indicates if this is students' first semester at any university and a female dummy. If the number of observations in the strata allowed it, we additionally re-randomized up to 5,000 times, keeping the randomization with the best balancing properties with respect to the age, the high school grade point average (GPA), and, if they were not used for stratification, the first university and female dummies. In total, we randomized 789 (208 in the summer and 581 in the winter term) students into the control and treatment groups (Tables 2.B1 and 2.B2 show the number of observations by study program and date of the randomization, respectively). Tables 2.B3 and 2.B4 show that in both cohorts the samples are well balanced.

Reminder letter intervention. About one week before the start of the course (see Figure 2.1), we randomized all students who signed up for the course up to that point into a control and one (summer term) or two (winter term) treatment groups. Again, we performed the randomization using stratification and re-randomization. Strata were constructed based on study program, the information about the treatment status in the invitation letter randomization – such that the two randomizations were orthogonal to each other – and, whenever possible, based on first university and female dummies. Re-randomization was conducted as before. Overall, 574 (129 in the summer and 445 in the winter term) students were randomized into treatment and control groups (Tables 2.B5 and 2.B6 show the respective balancing properties and Table 2.B1 the number of observations by study program).

[30] For the second experiment we moved to a larger interval between randomizations in order to have access to a larger number of observations at each point in time.

TABLE 2.B1: NUMBER OF OBSERVATIONS BY STUDY PROGRAM

	Summer Term					Winter Term					
Study program	BA	BIS	BHRE	E	N	BA	BIS	BHRE	E	2SE	N
Invitation letter											
I0: Control	60	18	12	14	104	101	25	18	16	32	192
I1: Letter	-	-	-	-	-	101	26	19	15	33	194
I2: Letter & social info	56	20	13	15	104	103	26	17	17	32	195
N	116	38	25	29	**208**	305	77	54	48	97	**581**
Reminder letter											
R0: Control	40	11	8	6	65	80	19	17	12	22	150
R1: Letter	-	-	-	-	-	79	21	14	14	20	148
R2: Letter & social info	37	12	9	6	64	79	19	14	13	22	147
N	77	23	17	12	**129**	238	59	45	39	64	**445**

Note: BA=Business Administration, BIS=Business Information Systems, BHRE=Business and Human Resource Education, E=Economics, 2SE=Two-Subject Economics.

TABLE 2.B2: NUMBER OF OBSERVATIONS BY TIMING OF INVITATION LETTER

	Summer Term						Winter Term					
Days until course	37	29	23	16	9	N	49	40	28	14	7	N
I0: Control	74	17	7	4	2	104	79	53	44	13	3	192
I1: Letter	-	-	-	-	-	-	84	49	44	13	4	194
I2: Letter & social info	74	16	6	5	3	104	83	50	46	12	4	195
N	148	33	13	9	5	**208**	246	152	134	38	11	**581**

TABLE 2.B3: DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES – INVITATION LETTER, SUMMER TERM

	(1) I0: Control mean (std. dev.)	(2) I2: Let. & soc. info coefficient (robust SE)	(3) p-Value
First university	0.510	-0.004	0.955
	(0.502)	(0.070)	
Female	0.298	-0.023	0.712
	(0.460)	(0.061)	
Age	21.654	-0.123	0.736
	(2.625)	(0.364)	
HS GPA	2.520	0.037	0.598
	(0.485)	(0.070)	
Fresh HS degree	0.423	0.008	0.911
	(0.496)	(0.069)	
HS degree Abitur	0.817	0.031	0.552
	(0.388)	(0.052)	
HS degree NI	0.577	-0.072	0.298
	(0.496)	(0.069)	
HS degree NW	0.125	-0.021	0.637
	(0.332)	(0.045)	
HS degree HE	0.096	0.088*	0.068
	(0.296)	(0.048)	
HS degree other	0.154	-0.045	0.327
	(0.363)	(0.046)	
HS degree abroad	0.019	0.039	0.163
	(0.138)	(0.028)	
Distance letter	172.433	-98.836	0.230
	(798.424)	(82.121)	
Sign-up before letter	0.067	-0.010	0.729
	(0.252)	(0.030)	
N	104	104	

Note: Column (1) presents the unadjusted control group means and standard deviations of the covariates. Column (2) presents the estimated coefficients of regressing the covariates on the treatment indicator using Equation 2.1. Column (3) tests the null hypothesis of no treatment effect. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.B4: DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES – INVITATION LETTER, WINTER TERM

	(1) I0: Control mean (std. dev.)	(2) I1: Letter coefficient (robust SE)	(3) p-Value	(4) I2: Let. & soc. info coefficient (robust SE)	(5) p-Value	(6) I1-I2 = 0 (robust SE)	(7) p-Value
First university	0.776	-0.000	0.991	-0.022	0.601	0.022	0.610
	(0.418)	(0.042)		(0.042)		(0.042)	
Female	0.375	-0.012	0.794	0.001	0.990	-0.013	0.785
	(0.485)	(0.048)		(0.048)		(0.048)	
Age	20.809	0.218	0.344	0.160	0.509	0.058	0.827
	(2.022)	(0.230)		(0.242)		(0.266)	
HS GPA	2.345	-0.003	0.959	0.000	0.998	-0.003	0.957
	(0.533)	(0.056)		(0.054)		(0.056)	
Fresh HS degree	0.432	-0.021	0.673	-0.070	0.157	0.049	0.320
	(0.497)	(0.050)		(0.049)		(0.049)	
HS degree Abitur	0.807	0.041	0.281	0.065*	0.070	-0.024	0.489
	(0.395)	(0.038)		(0.036)		(0.035)	
HS degree NI	0.552	-0.007	0.895	0.005	0.921	-0.012	0.813
	(0.499)	(0.051)		(0.050)		(0.049)	
HS degree NW	0.073	0.041	0.168	-0.010	0.681	0.051*	0.074
	(0.261)	(0.030)		(0.025)		(0.029)	
HS degree HE	0.115	-0.007	0.836	-0.001	0.976	-0.006	0.860
	(0.319)	(0.032)		(0.032)		(0.032)	
HS degree other	0.224	-0.026	0.530	-0.007	0.861	-0.019	0.641
	(0.418)	(0.042)		(0.042)		(0.040)	
HS degree abroad	0.036	-0.001	0.942	0.014	0.500	-0.015	0.458
	(0.188)	(0.019)		(0.020)		(0.020)	
Distance letter	170.863	-59.722	0.307	10.341	0.898	-70.064	0.195
	(814.427)	(58.351)		(80.627)		(54.031)	
Sign-up before letter	0.099	0.021	0.409	0.004	0.869	0.017	0.510
	(0.299)	(0.026)		(0.025)		(0.026)	
N	192	194		195			

Note: Column (1) presents the unadjusted control group means and standard deviations of the covariates. Columns (2) and (4) present the estimated coefficients of regressing the covariates on the treatment indicators using Equation 2.1. Columns (3) and (5) test the null hypotheses of no treatment effects. Columns (6) and (7) test for the equality of the two treatment effects. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.B5: DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES – REMINDER LETTER, SUMMER TERM

	(1) R0: Control mean (std. dev.)	(2) R2: Let. & soc. info coefficient (robust SE)	(3) p-Value
First university	0.646	-0.034	0.693
	(0.482)	(0.087)	
Female	0.277	0.001	0.991
	(0.451)	(0.078)	
Age	21.055	0.073	0.842
	(1.951)	(0.364)	
HS GPA	2.525	0.012	0.884
	(0.489)	(0.083)	
Fresh HS degree	0.508	-0.052	0.561
	(0.504)	(0.089)	
HS degree Abitur	0.908	-0.058	0.314
	(0.292)	(0.057)	
HS degree NI	0.538	0.022	0.808
	(0.502)	(0.089)	
HS degree NW	0.108	0.029	0.628
	(0.312)	(0.060)	
HS degree HE	0.108	0.053	0.372
	(0.312)	(0.059)	
HS degree other	0.185	-0.105*	0.076
	(0.391)	(0.059)	
HS degree abroad	0.015	-0.002	0.924
	(0.124)	(0.021)	
Distance letter	97.072	-14.394	0.351
	(94.000)	(15.363)	
N	65	64	

Note: Column (1) presents the unadjusted control group means and standard deviations of the covariates. Column (2) presents the estimated coefficients of regressing the covariates on the treatment indicator using Equation 2.1. Column (3) tests the null hypothesis of no treatment effect. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.B6: DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES – REMINDER LETTER, WINTER TERM

	(1) R0: Control mean (std. dev.)	(2) R1: Letter coefficient (robust SE)	(3) p-Value	(4) R2: Let. & soc. info coefficient (robust SE)	(5) p-Value	(6) R1 - R2 = 0 (robust SE)	(7) p-Value
First university	0.827	0.011	0.794	0.010	0.818	-0.001	0.977
	(0.380)	(0.043)		(0.043)		(0.043)	
Female	0.400	-0.004	0.948	0.018	0.752	-0.021	0.707
	(0.492)	(0.055)		(0.055)		(0.056)	
Age	20.667	0.056	0.810	-0.019	0.944	0.075	0.760
	(2.292)	(0.235)		(0.261)		(0.245)	
HS GPA	2.347	-0.006	0.921	-0.006	0.923	0.000	0.995
	(0.520)	(0.059)		(0.065)		(0.065)	
Fresh HS degree	0.427	-0.023	0.683	0.042	0.467	-0.066	0.258
	(0.496)	(0.057)		(0.058)		(0.058)	
HS degree Abitur	0.880	-0.042	0.286	-0.017	0.644	-0.024	0.547
	(0.326)	(0.039)		(0.038)		(0.040)	
HS degree NI	0.560	-0.027	0.633	0.021	0.712	-0.048	0.403
	(0.498)	(0.057)		(0.057)		(0.058)	
HS degree NW	0.120	-0.058*	0.080	-0.040	0.256	-0.018	0.547
	(0.326)	(0.033)		(0.035)		(0.030)	
HS degree HE	0.067	0.036	0.271	0.083**	0.021	-0.047	0.224
	(0.250)	(0.033)		(0.036)		(0.039)	
HS degree other	0.233	0.016	0.749	-0.065	0.156	0.081*	0.089
	(0.424)	(0.050)		(0.046)		(0.048)	
HS degree abroad	0.020	0.033	0.128	-0.000	0.985	0.033	0.134
	(0.140)	(0.022)		(0.017)		(0.022)	
Distance letter	138.545	111.244	0.280	-39.699	0.132	150.944	0.137
	(285.019)	(102.897)		(26.280)		(101.325)	
N	150	148		147			

Note: Column (1) presents the unadjusted control group means and standard deviations of the covariates. Columns (2) and (4) present the estimated coefficients of regressing the covariates on thetreatment indicators using Equation 2.1. Columns (3) and (5) test the null hypotheses of no treatment effects. Columns (6) and (7) test for the equality of the two treatment effects. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

C Statistical power

One potential concern with our study could be that it has not enough statistical power to detect effect sizes that are usually found in studies that employ social information. Table 2.C1 shows minimal detectable effect sizes (MDES) for $1-\beta = 0.6$ and $1-\beta = 0.8$, assuming $\alpha = 0.05$, $R^2 = 0.21$ ($R^2 = 0.10$), and $N = 600$ ($N = 400$) for the invitation (reminder) letter intervention. We performed the power calculations after the field experiment in the summer term, as it was clear that for the summer cohort alone we would not have enough power to detect reasonable effect sizes. The assumed R^2 are based on OLS estimations using the control group of the summer cohort that regress sign-up and participation in the first tutorial on the covariates that we pre-registered for the stratification procedure in the winter cohort (see Section 2.2.2). The assumed number of observations refers to the number of observations that we expected to have for the comparison of control versus the treatment group that received a letter with social information after pooling observations from both cohorts. In addition, we use information on the control group take-up rates in the summer cohort to calculate expected effect sizes in terms of percentage points.

For the invitation letter intervention our study has 60% power to detect effects sizes of 0.16 standard deviations, which, taking the observed sign-up and participation rates in the control group into account, corresponds to effects of 7.4 to 7.7 pp. With a power of 80%, we would be able to detect effects of 0.20 standard deviations or 9.3 to 9.6 pp. For the reminder letter intervention our MDES corresponds to 0.21 (0.27) standard deviations or 5.6 (7.2) pp, with a power of 60% (80%).

To evaluate these MDESs, we can compare them to effect sizes that are typically found in studies that make use of social information and reminders. As this is the first study that makes use of this type of social information in the context of education, we can only compare our study to other nudges that are employed in this context. For information on typical effect sizes we draw on the meta-analysis by DellaVigna and Linos (2020).[31] For studies that rely on social cues that are published in academic journals they report effect sizes of 13.81 pp or about 0.30 standard deviations (see also Table 2.C2). Our study is well powered to detect effects of this size. For studies that make use of pure reminders or that are set in the context of education, they report effect sizes of 5.02 and 2.56 pp (0.12 and 0.05 standard deviations), respectively. While our study does not necessarily have enough power to detect these kind of effect sizes, we should be able to detect effect sizes between those and the ones reported for social cues.

DellaVigna and Linos (2020) also report effect sizes for interventions that were conducted by Nudge Units in the US. The average effects sizes of these interventions are substantially smaller than those published in academic journals (see DellaVigna and Linos 2020 and Table 2.C2), and our study is not well-powered enough to detect these kind of effects sizes (see the related discussion in Section 2.5).

[31] The study by DellaVigna and Linos (2020) includes papers that are also covered in the meta-analysis by Hummel and Maedche (2019), which we initially used to evaluate our MDESs.

TABLE 2.C1: MINIMUM DETECTABLE EFFECT SIZES ($\alpha = 0.05$)

Invitation letter ($R^2 = 0.21$)				*Reminder letter* ($R^2 = 0.10$)			
$1-\beta$	N	MDES	in pp	$1-\beta$	N	MDES	in pp
Effect on sign up							
Control take-up=68.3%							
0.6	600	0.16	7.44				
0.8	600	0.20	9.31				
Effect on participation 1st tutorial							
Control take-up=63.5%				Control take-up=92.3%			
0.6	600	0.16	7.70	0.6	400	0.21	5.60
0.8	600	0.20	9.63	0.8	400	0.27	7.20

Note: The power calculations were performed after the results for the summer cohort were available. The assumed R^2 are based on control group OLS regressions of the outcome variables sign-up and participation in first tutorial on the covariates that we pre-registered for stratification in the field experiment with the winter cohort. For the invitation (reminder) letter $N = 600$ ($N = 400$) refers to the number of observations that we expected to have for the comparison of control versus letter with social information after pooling observations from both cohorts. The depicted control group take-up rates are from the summer cohort. The calculation of the MDES in pp is based on the following formula: $MDES \times \sqrt{(\text{control take-up}) \times (100 - \text{control take-up})}$. Power calculations were performed with Optimal Design (Spybrook et al., 2011).

TABLE 2.C2: EXPECTED EFFECT SIZES IN THE LITERATURE – BASED ON DELLAVIGNA AND LINOS (2020)

	Academic Journals			Nudge Units		
	Control take-up (%)	ATE (pp)	Effect size (sd)	Control take-up (%)	ATE (pp)	Effect size (sd)
Education	66.16	2.56	0.054	14.39	0.49	0.014
Reminders	25.17	5.02	0.116	27.29	2.56	0.058
Social cues	31.11	13.81	0.298	18.05	0.96	0.025

Note: Numbers from tables 1a and 1b in DellaVigna and Linos (2020). Effect size in standard deviations (sd) is calculated as according to $\frac{ATE}{\sqrt{(Take-up)\times(100-Take-up)}}$

D Pre-registered analyses by cohort

Since the main part of the chapter reports the results of the two interventions for the pooled samples only, in this appendix, we present results separated by cohorts, following the respective pre-registrations.

Summer cohort

The field experiment with the summer cohort is pre-registered under `https://osf.io/tm7k3`.

Main analyses. To address the main research questions, Table 2.D1 reports effects of the invitation letter with social information on sign up for the remedial math course and Table 2.D2 reports effects of the reminder letter on participation in the course.

Secondary analyses. Since we did not receive the respective information, we do not report effects on the performance in placement tests that took place at the beginning and the end of the remedial math course. Table 2.D3 reports results regarding the interaction of the two interventions. Tables 2.D4 and 2.D5 report heterogeneous effects of the interventions with respect to the time at which the invitation letter was sent and whether this is the first semester at any university.

TABLE 2.D1: EFFECT OF INVITATION LETTER

	Sign-up	
	(1)	(2)
I2: Letter & social info	0.052	0.057
	(0.058)	(0.057)
Strata	yes	yes
Controls	no	yes
N	208	208
Control mean	0.68	0.68
(SD)	(0.47)	(0.47)

Note: Outcome variable: sign-up for remedial math course; *strata:* study program FE, invitation letter date FE (= matriculation date FE), and first university dummy; *controls:* female dummy, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, and place of HS degree dummies. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D2: EFFECT OF REMINDER LETTER

	Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)
R2: Letter & social info	0.031	0.025	-0.038	-0.039
	(0.042)	(0.046)	(0.050)	(0.049)
Strata	yes	yes	yes	yes
Controls	no	yes	no	yes
N	129	129	129	129
Control mean	0.97	0.97	0.82	0.82
(SD)	(0.18)	(0.18)	(0.28)	(0.28)

Note: Outcome variables: participation in first tutorial of remedial math course and average participation is the share of tutorials a student participated in; *strata:* study program FE, invitation letter treatment status (the pre-registration mistakenly states matriculation date FE), and first university dummy; *controls:* female dummy, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, and place of HS degree dummies. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D3: EFFECT OF REMINDER LETTER BY INVITATION LETTER TREATMENT STATUS

	Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)
R2: Reminder & social info	0.067	0.030	-0.016	-0.065
	(0.046)	(0.040)	(0.075)	(0.072)
I2: Invitation & social info	-0.030	-0.058	-0.037	-0.066
	(0.068)	(0.065)	(0.071)	(0.071)
R2*I2	-0.042	0.018	-0.018	0.075
	(0.082)	(0.082)	(0.102)	(0.107)
Strata	yes	yes	yes	yes
Controls	no	yes	no	yes
N	127	127	127	127

Note: Only includes students that were part of both interventions. *Outcome variables:* participation in first tutorial of remedial math course and average participation is the share of tutorials a student participated in; *strata:* study program FE and first university dummy; *controls:* female dummy, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, and place of HS degree dummies. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D4: EFFECT OF INVITATION LETTER – HETEROGENEITIES

	Sign-up			
	Last month		First university	
	(1)	(2)	(3)	(4)
I2: Letter & social info	0.070	0.087	0.032	0.039
	(0.070)	(0.069)	(0.095)	(0.097)
Last month	-0.169*	-0.160*		
	(0.093)	(0.095)		
I2*last month	-0.052	-0.098		
	(0.125)	(0.120)		
First university			0.321***	0.283***
			(0.086)	(0.092)
I2*first university			0.041	0.036
			(0.117)	(0.119)
Strata	yes	yes	yes	yes
Controls	no	yes	no	yes
N	208	208	208	208

Note: Outcome variable: sign-up for remedial math course; *strata:* study program FE, invitation letter date FE (= matriculation date FE, only Columns 3 and 4), and first university dummy (only Columns 1 and 2); *controls:* female dummy, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, and place of HS degree dummies. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D5: EFFECT OF REMINDER LETTER – HETEROGENEITIES

	Last month				First university			
	Part. 1st tutorial		Average part.		Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
R2: Letter & social info	0.082*	0.072	0.022	0.007	0.042	0.033	-0.003	-0.016
	(0.048)	(0.049)	(0.058)	(0.056)	(0.072)	(0.071)	(0.092)	(0.087)
Last month	0.127**	0.131**	0.118*	0.105				
	(0.054)	(0.054)	(0.063)	(0.071)				
R2*last month	-0.236**	-0.217**	-0.279**	-0.212				
	(0.100)	(0.105)	(0.114)	(0.133)				
First university					0.017	0.036	0.146*	0.181**
					(0.071)	(0.076)	(0.081)	(0.084)
R2*first university					-0.017	-0.012	-0.056	-0.036
					(0.090)	(0.089)	(0.111)	(0.108)
Strata	yes	yes	yes	yes	yes	yes	yes	yes
Controls	no	yes	no	yes	no	yes	no	yes
N	129	129	129	129	129	129	129	129

Note: Outcome variables: participation in first tutorial of remedial math course and average participation is the share of tutorials a student participated in; *strata:* study program FE, invitation letter treatment status (the pre-registration mistakenly states matriculation date FE), and first university dummy (only Columns 1 to 4); *controls:* female dummy, age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, and place of HS degree dummies. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Winter cohort

The field experiment with the winter cohort is pre-registered under `https://osf.io/vqa84`.

Main analyses. To address the main research questions, Table 2.D6 reports effects of the invitation letter intervention on sign up for and participation in the remedial math course and Table 2.D7 reports effects of the reminder letter intervention on participation in the course.

Secondary analyses. Since we did not receive the respective information, we do not report effects on the performance in placement tests that took place at the beginning and the end of the remedial math course. Tables 2.D8, 2.D9, and 2.D10 report on the heterogeneous effects of the invitation letter intervention and Tables 2.D9 and 2.D10 on the heterogeneous effects of the reminder letter intervention.

TABLE 2.D6: EFFECT OF INVITATION LETTER

	Sign-up		Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)	(5)	(6)
I1: Letter	-0.038	-0.038	-0.049	-0.043	-0.042	-0.038
	(0.039)	(0.039)	(0.043)	(0.042)	(0.040)	(0.039)
I2: Letter & social info	0.002	0.002	-0.001	0.002	-0.010	-0.007
	(0.038)	(0.038)	(0.042)	(0.042)	(0.040)	(0.038)
I2-I1	0.041	0.041	0.048	0.045	0.032	0.030
	(0.039)	(0.039)	(0.043)	(0.041)	(0.040)	(0.039)
Strata	yes	yes	yes	yes	yes	yes
Controls	no	d.-p. LASSO	no	d.-p. LASSO	no	d.-p. LASSO
N	581	581	581	581	581	581
Control mean	0.80	0.80	0.73	0.73	0.64	0.64
(SD)	(0.40)	(0.40)	(0.45)	(0.45)	(0.42)	(0.42)

Note: Outcome variables: sign-up for remedial math course, participation in first tutorial of remedial math course, and average participation is the share of tutorials a student participated in; *strata:* study program FE, invitation letter date FE as well as first university and female dummies; *controls:* the double-post LASSO specification considers age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D7: EFFECT OF REMINDER LETTER

	Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)
R1: Letter	0.024	0.024	-0.008	-0.009
	(0.034)	(0.033)	(0.037)	(0.036)
R2: Letter & social info	0.025	0.025	0.034	0.034
	(0.035)	(0.033)	(0.036)	(0.035)
R2-R1	0.001	0.001	0.043	0.043
	(0.033)	(0.031)	(0.034)	(0.033)
Strata	yes	yes	yes	yes
Controls	no	d.-p. LASSO	no	d.-p. LASSO
N	445	445	445	445
Control mean	0.90	0.90	0.79	0.79
(SD)	(0.30)	(0.30)	(0.31)	(0.31)

Note: Outcome variables: participation in first tutorial of remedial math course and average participation is the share of tutorials a student participated in; *strata:* study program FE, invitation letter treatment status FE as well as first university and female dummies; *controls:* the double-post LASSO specification considers age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D8: EFFECT OF INVITATION LETTER BY DATE OF ENROLLMENT

	Sign-up		Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)	(5)	(6)
I1: Letter	-0.059	-0.057	-0.071	-0.064	-0.082*	-0.078
	(0.047)	(0.046)	(0.051)	(0.050)	(0.048)	(0.048)
I2: Letter & social info	-0.037	-0.043	-0.042	-0.046	-0.070	-0.079*
	(0.047)	(0.047)	(0.051)	(0.050)	(0.048)	(0.047)
Last month	-0.040	-0.061	-0.092	-0.105	-0.114*	-0.125*
	(0.066)	(0.066)	(0.073)	(0.074)	(0.066)	(0.066)
I1*last month	0.054	0.052	0.061	0.055	0.117	0.113
	(0.093)	(0.092)	(0.101)	(0.100)	(0.093)	(0.091)
I2*last month	0.111	0.130	0.121	0.135	0.184**	0.197**
	(0.086)	(0.085)	(0.097)	(0.095)	(0.088)	(0.086)
Strata	yes	yes	yes	yes	yes	yes
Controls	no	yes	no	yes	no	yes
N	581	581	581	581	581	581

Note: Outcome variables: sign-up for remedial math course, participation in first tutorial of remedial math course, and average participation is the share of tutorials a student participated in; *strata:* study program FE as well as first university and female dummies; *controls:* age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D9: EFFECT OF INVITATION LETTER BY FIRST UNIVERSITY

	Sign-up		Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)	(5)	(6)
I1: Letter	-0.098	-0.089	-0.096	-0.081	-0.049	-0.039
	(0.104)	(0.103)	(0.105)	(0.103)	(0.094)	(0.091)
I2: Letter & social info	0.011	0.020	0.006	0.010	-0.030	-0.030
	(0.099)	(0.098)	(0.104)	(0.102)	(0.090)	(0.088)
First university	0.248***	0.234***	0.321***	0.307***	0.323***	0.327***
	(0.078)	(0.080)	(0.084)	(0.086)	(0.075)	(0.079)
I1*first university	0.076	0.067	0.061	0.046	0.009	-0.002
	(0.112)	(0.111)	(0.115)	(0.113)	(0.104)	(0.102)
I2*first university	0.012	-0.025	-0.010	-0.016	0.027	0.021
	(0.107)	(0.106)	(0.113)	(0.112)	(0.100)	(0.098)
Strata	yes	yes	yes	yes	yes	yes
Controls	no	yes	no	yes	no	yes
N	581	581	581	581	581	581

Note: Outcome variables: sign-up for remedial math course, participation in first tutorial of remedial math course, and average participation is the share of tutorials a student participated in; *strata:* study program FE, invitation letter date FE, and female dummy; *controls:* age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D10: EFFECT OF INVITATION LETTER BY GENDER

	Sign-up		Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)	(5)	(6)
I1: Letter	-0.011	-0.015	-0.068	-0.072	-0.072	-0.079
	(0.049)	(0.049)	(0.054)	(0.054)	(0.049)	(0.049)
I2: Letter & social info	-0.006	-0.012	-0.023	-0.027	-0.044	-0.051
	(0.050)	(0.050)	(0.053)	(0.054)	(0.048)	(0.049)
Female	0.074	0.056	0.008	-0.023	0.017	-0.016
	(0.056)	(0.058)	(0.064)	(0.065)	(0.060)	(0.060)
I1*female	-0.075	-0.063	0.050	0.072	0.082	0.105
	(0.081)	(0.083)	(0.091)	(0.090)	(0.085)	(0.084)
I2*female	0.023	0.037	0.059	0.067	0.091	0.098
	(0.077)	(0.076)	(0.088)	(0.088)	(0.083)	(0.081)
Strata	yes	yes	yes	yes	yes	yes
Controls	no	yes	no	yes	no	yes
N	581	581	581	581	581	581

Note: Outcome variables: sign-up for remedial math course, participation in first tutorial of remedial math course, and average participation is the share of tutorials a student participated in; *strata:* study program FE, invitation letter date FE, and first university dummy; *controls:* age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D11: EFFECT OF REMINDER LETTER BY FIRST UNIVERSITY

	Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)
R1: Letter	-0.100	-0.102	-0.182	-0.189*
	(0.122)	(0.117)	(0.114)	(0.106)
R2: Letter & social info	0.025	0.022	0.024	0.018
	(0.107)	(0.104)	(0.109)	(0.105)
First university	0.081	0.081	0.103	0.125
	(0.083)	(0.083)	(0.086)	(0.084)
R1*first university	0.149	0.148	0.208*	0.214*
	(0.127)	(0.121)	(0.121)	(0.112)
R2*first university	0.001	0.005	0.013	0.027
	(0.114)	(0.111)	(0.116)	(0.112)
Strata	yes	yes	yes	yes
Controls	no	yes	no	yes
N	445	445	445	445

Note: Outcome variables: participation in first tutorial of remedial math course and average participation is the share of tutorials a student participated in; *strata:* study program FE, invitation letter treatment status FE, and female dummy; *controls:* age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 2.D12: EFFECT OF REMINDER LETTER BY GENDER

	Part. 1st tutorial		Average part.	
	(1)	(2)	(3)	(4)
R1: Letter	0.043	0.044	0.029	0.026
	(0.041)	(0.041)	(0.046)	(0.046)
R2: Letter & social info	0.008	0.014	0.057	0.073
	(0.046)	(0.047)	(0.048)	(0.048)
Female	0.004	-0.009	0.099*	0.083
	(0.054)	(0.054)	(0.056)	(0.055)
R1*female	-0.049	-0.060	-0.094	-0.093
	(0.072)	(0.074)	(0.076)	(0.074)
R2*female	0.042	0.026	-0.056	-0.082
	(0.071)	(0.072)	(0.074)	(0.072)
Strata	yes	yes	yes	yes
Controls	no	yes	no	yes
N	445	445	445	445

Note: Outcome variables: participation in first tutorial of remedial math course and average participation is the share of tutorials a student participated in; *strata:* study program FE, invitation letter treatment status FE, and first university dummy; *controls:* age, HS GPA, fresh HS degree dummy, HS degree abitur dummy, place of HS degree dummies, and the distance over which the letter was sent. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Chapter 3

Relative Performance Feedback and the Effects of Being Above Average – Field Experiment and Replication

*Joint work with Oliver Himmler and Robert Jäckle**

3.1 Introduction

Information about the behavior of similar others can provide an important benchmark against which individuals can compare their performance and gauge their abilities (see e.g., Bandura 1991, Corcoran et al. 2011, Festinger 1954, and Taylor et al. 1996). In practice, however, many situations are characterized by a lack of the information necessary for comparison. Knowledge of the underlying performance distribution may be absent, leaving individuals with no appropriate frame of reference. Providing feedback relative to a suitable peer group may facilitate comparison, thereby enhancing decision-making, motivation, and ultimately performance.

An economically and socially important setting where feedback may improve outcomes is higher education. Students at the start of their university careers face new tasks which are complex and challenging, and they are surrounded by new peers, making it difficult to assess their relative abilities. Providing relative performance feedback early on could therefore present a low-cost and easily scalable tool for universities to help students improve their performance.

*A previous version of this study was circulated under the title "Normatively Framed Relative Performance Feedback – Field Experiment and Replication". We gratefully acknowledge financial support from the German Federal Ministry of Education and Research under grant 01PX16003A, 01PX16003B, the Staedtler Stiftung, and administrative support from the TH Nürnberg.

To test this, we implement a relative performance feedback intervention at one of the largest universities of applied sciences in Germany. After the first semester, control group students receive letters from the university, informing them about how many credit points they managed to obtain in the previous semester. Students in the treatment group receive the same information, but the letters also inform them about how well, in terms of credits, they performed relative to the average student and the student on the 80th percentile.

As we give feedback on obtained credit points, we expect the effects of feedback to materialize on the same dimension. Obtained credits are a good standardized measure of academic progress: they reflect the number of passed exams, where exams are weighted by the required workload for each class.[1] Credit points therefore affect time to graduation, and earning more credit points has considerable payoffs: first, for the individual, it reduces forgone income by circumventing a longer study duration – and according to the literature on academic momentum, especially the initial progress in college is positively related to degree completion (see Doyle 2011; Attewell et al. 2012; Attewell and Monaghan 2016)[2]; second, from a social perspective, (faster) graduation can be expected to translate into, e.g., longer duration of contributions, in terms of taxes and payments into the social security systems.

The initial implementation of our feedback intervention has a positive, but statistically insignificant treatment effect on subsequent performance in the full sample. Based on the design of the feedback, we explore heterogeneous responses. Students whom the feedback informs that they have performed above average in the first semester increase their performance by a statistically significant 2 credits points (.16 standard deviations) in the second semester, relative to controls. For students at or below the average we do not find effects on behavior.

The initial results thus give rise to the hypothesis that above-average feedback enhances performance. We test this hypothesis with a replication experiment one year later with a new cohort of students and reproduce the full pattern of results. Most importantly, we replicate the result that above-average feedback significantly increases performance (1.6 credit points or .12 standard deviations). Across both experiments the effects are roughly equivalent to one third of an additional passed exam.

While our interventions target credits, we also show that those who benefit from the feedback on credit points maintain the same grade point average (GPA) as the individuals in the

[1] Europe-wide, universities use a standardized system (European Credit Transfer and Accumulation System, ECTS), under which a full-time academic year consists of 60 credits, with the typical workload for one credit equaling 25-30 study hours. See also `https://ec.europa.eu/education/resources-and-tools`, retrieved on November 8, 2021.

[2] Findings of this literature have motivated interventions, especially in the United States, that encourage students to aim for taking the prescribed number of credits per semester, such as *15 to Finish* (see `https://completecollege.org/strategy/15-to-finish/`, retrieved on November 8, 2021).

control group. This shows that the increase in earned credits can be interpreted as a net performance gain, as students do not buy gains on the (treated) credit points dimension with losses on the grade dimension. In addition, we also survey students on potential negative side effects of the intervention on their well-being, as for example, Celik Katreniak (2018) shows that stronger incentives can come at the cost of increased stress and reduced happiness. We find no such evidence: treated students are no different from controls in any of the well-being domains we observe.

We investigate potential mechanisms behind this causal reaction to above-average feedback: Employing a regression discontinuity design (RDD) based on the sharp cutoff at the average allows us to assess the causal effect of being informed about an above-average performance versus being informed about an average or below-average performance. Receiving above-average feedback increases subsequent performance by about six credits in comparison to receiving another type of feedback. This indicates that above-average students do not react so strongly to feedback due to different underlying characteristics – i.e., the RDD shows that among students with similar ability, motivation, or learning technology, only the above-average feedback produces effects. This suggests that it is the information about being above average itself that leads to the increase in performance.

In the replication experiment, we further explore mechanisms by analyzing the role of pre- and post-treatment expectations about relative performance. We find that students have inaccurate expectations about their relative performance pre-treatment and that students who receive above-average feedback subsequently update their beliefs. In the absence of treatment, those whose beliefs underestimate their actual relative performance obtain fewer credits. Feedback is able to offset this disadvantage associated with inaccurate beliefs. There is no evidence of updating for students at or below the average. We argue that the pattern of effects and the evidence that students selectively process feedback is consistent with theories on the management of self-confidence (see, e.g., Bénabou and Tirole 2002).

Finally, we show that repeated treatment in the third semester does not elicit additional performance gains. One reason may be that in the third semester, beliefs about relative performance held by the controls are almost as accurate as those of the treatment group.

Overall, our results show that providing relative performance feedback is beneficial for a large share of individuals entering university. When beliefs are inaccurate and the feedback is encouraging, performance can increase even in complex and challenging tasks such as passing exams in higher education. At the same time, the intervention is also inexpensive at a total cost of less than €2.5 per student and semester (see Table 3.A1). The results also highlight that feedback schemes can have distributional implications. While the effects may be considered pareto-improving (no negative effects), it is mostly students in the upper middle of the performance distribution who profit. The intervention may thus decrease performance equality in the education system. Inequality-averse policy-makers should take this into account and be aware that for the most precarious students, making their shortcomings salient via feedback may not generate any effects.

Relation to the literature. Our study contributes to research on the effects of relative performance feedback in higher education.[3] This setting is highly relevant, given that a large share of students takes much longer than the prescribed time to obtain their degree (see, e.g., Bound et al. 2012).[4] Increasing obtained credits is the only means to achieve a more timely graduation (or graduation at all), and there is still a dearth of low-cost and scalable interventions that can elicit changes in academic performance (see, e.g., Oreopoulos et al. forthcoming and Oreopoulos and Petronijevic 2019).

It is thus surprising that research on relative feedback in higher education is still sparse. In addition to our study, only Azmat et al. (2019) and Cabrera and Cid (2017) study the effects of relative performance feedback based on overall performance and consider both relevant outcome dimensions, grades and number of courses passed – a setup which allows to detect both substitution effects across courses and between outcome dimensions. Both studies find null or negative effects on performance. Our intervention differs in several ways from the feedback schemes used in their studies: rather than on GPA, we give feedback on obtained credits, we provide coarse instead of (rather) precise relative performance information, an approach that has proven effective in secondary education (Azmat and Iriberri, 2010), and students receive feedback by personalized physical letters and not via an online service portal of the university.[5]

Other studies give relative feedback on intermediate performance in a single course, i.e., performance in mid-term exams or online assignments, and evaluate effects on a single outcome dimension (grades). This approach can improve performance in the final exam (Kajitani et al., 2020; Tran and Zeckhauser, 2012) and may generate positive grade spillovers to other courses (Dobrescu et al., 2021). In contrast to those studies, our feedback does not require systematic measurement and collection of data on intermediate performance. We

[3]There is also a literature on the effects of relative performance feedback in primary and secondary education that mainly finds positive effects (Azmat and Iriberri, 2010; Fischer and Wagner, 2019; Goulas and Megalokonomou, 2021; Hermes et al., 2021). However, it is not clear whether these results extend to an adult population and in an environment that is much less structured than primary and secondary education.

[4]E.g., in OECD countries, only 39% of full-time students graduate within the planned duration of the program (OECD, 2019).

[5]Letters may generate more attention, as students nowadays often receive many emails but few letters from the university. This notion is consistent with interventions unable to increase academic performance using digital formats (e.g., Oreopoulos et al. forthcoming). Moreover, DellaVigna and Linos (2020) also provide evidence that interventions using physical letters may be more effective than interventions that rely on emails.

focus on overall performance from all courses taken, which is a simple and cost-effective approach, which can be easily implemented and scaled because the necessary information is readily available in existing administrative data.

Our study also contributes to the general literature on relative performance feedback. First, we focus on the transition into a new environment, a time when individuals face higher uncertainty about relative ability. This contrasts with existing research from the field, which often introduces relative performance feedback when individuals are already familiar with their tasks and peers (Ashraf, 2019; Barankay, 2012; Blanes i Vidal and Nossol, 2011; Delfgaauw et al., 2013). Second, while university presents a complex working environment with challenging, high-stakes tasks, much of the previous literature relies on real effort tasks in the lab (Azmat and Iriberri, 2016; Charness et al., 2013; Eriksson et al., 2009; Gill et al., 2019; Kuhnen and Tymula, 2012) or on rather repetitive tasks in the field (Ashraf, 2019; Bandiera et al., 2013; Barankay, 2012; Blanes i Vidal and Nossol, 2011; Delfgaauw et al., 2013).

Finally, the results of our study bear significance for the literature that studies the link between confidence and performance. Our findings are consistent with the idea that individuals try to maintain a positive self-assessment of their abilities by processing positive feedback while discarding negative feedback, which in turn leads to higher confidence in ability and motivates individuals to work harder and take beneficial risks (see, e.g., Bénabou and Tirole 2002 and Compte and Postlewaite 2004).[6][7] Empirical evidence from the lab shows that individuals who receive bad news indeed have little willingness to update their self-concept, whereas people who receive positive information are willing to incorporate the good news in their beliefs (Eil and Rao, 2011; Möbius et al., 2014). Our study thus complements the existing literature on motivated beliefs, by showing first tentative evidence for this type of behavior in the context of a relative performance feedback intervention.

The remainder of the chapter is structured as follows. Section 3.2 describes the institutional background, data, and design of our intervention as well as the empirical approach. Section 3.3 reports on the main results of our two field experiments. In Sections 3.4 to 3.7, we explore the drivers and mechanisms behind the main results, investigate effects of repeated treatment, and present the effects on auxiliary outcomes. Section 3.8 concludes.

[6] A strand of the psychological literature also has argued that individuals increase their efforts only after receiving positive feedback and underweigh adverse information about themselves (Ilgen et al., 1979; Ilgen and Davis, 2000; Pearce and Porter, 1986).

[7] Our findings are also in line with the literature on the effects of ordinal ranks on academic achievement, which argues that learning about a high rank can increase confidence in own ability (e.g., Elsner and Isphording 2017 and Murphy and Weinhardt 2020).

3.2 Institutional background and research design

We conduct our field experiments at one of the largest universities of applied sciences in Germany. The subjects of our interventions are two cohorts of students who enrolled in five bachelor's degree programs at the faculties of Business Administration (BA) and Mechanical Engineering (ME). All interventions were implemented and outcomes realised before any Corona-related restrictions.

3.2.1 Institutional background

Our sample is representative of a substantial portion of students who enroll at universities of applied sciences in Germany. For example, in the winter semester 2019, 15.2% of freshman students chose BA and 4.9% opted for ME (Statistisches Bundesamt, 2020b). Both fields are also among the most popular at the university where we ran the experiments: roughly 30% of all students study at the departments of BA and ME. Subjects in our trials are also representative in terms of students' final high school grades: they cover the entire grade distribution and have, on average, similar grades to the population of German high school graduates.[8]

The bachelor's programs at the university are organized according to the European Credit Transfer System (ECTS, see Footnote 1). To complete their degree, students have to accumulate 210 credits. Given the nominal study duration of seven semesters, students are thus expected to take courses worth 30 credits per semester.

Students can at all times access information on their individual study progress via a web portal maintained by the university. The portal provides data on absolute performance – credits and GPA. In the absence of our treatment, the university does not provide any information on a student's relative performance.

3.2.2 Field Experiment I

The initial field experiment is conducted with a cohort of first-year bachelor students. Treatment commences as soon as information on first semester performance is available, i.e., at the start of the second semester. At this point, 812 individuals are studying in the five bachelor's programs offered by the BA and ME faculties (Table 3.A2 provides an overview of all degree programs and the number of students in our intervention).

[8]The average high school GPA in the control group of our sample is about 2.56 (see Table 3.1), while it was around 2.42 among all German high school graduates in the years 2015 to 2019. See `https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html`, retrieved on November 8, 2021. In the German system 1.0 is the best and 4.0 the worst final high school GPA.

Randomization. Randomization was carried out after the first semester, using stratification and balancing (Morgan and Rubin, 2012). We built strata along bachelor's programs and obtained first-semester credits. Within these blocks we balanced on age, sex, high school GPA, time since high school graduation, pre-treatment GPA, and type of high school degree.[9] Table 3.1 shows that in the full sample all covariates are balanced. As we will explain later, an important subgroup in this chapter are above-average students. Table 3.A4 shows that all variables are balanced in this group as well.

TABLE 3.1: DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES

	Experiment I			Experiment II: Replication		
	(1) Control Mean (Std. Dev.)	(2) Treatment Coefficient (Robust SE)	(3) p-Value	(4) Control Mean (Std. Dev.)	(5) Treatment Coefficient (Robust SE)	(6) p-Value
Age	22.514	-0.086	0.696	22.417	0.084	0.689
	(3.376)	(0.220)		(3.078)	(0.210)	
Female	0.395	0.001	0.976	0.344	-0.001	0.973
	(0.489)	(0.030)		(0.476)	(0.029)	
HS Degree Abitur	0.430	-0.010	0.766	0.399	0.017	0.604
	(0.496)	(0.033)		(0.490)	(0.034)	
Time since HS Degree	1.341	-0.094	0.561	1.171	-0.005	0.968
	(2.523)	(0.161)		(1.885)	(0.133)	
HS GPA	2.567	-0.011	0.758	2.555	-0.042	0.225
	(0.563)	(0.036)		(0.622)	(0.035)	
% HS GPA Imputed$^{a)}$	0.012	0.002	0.754	0.020	-0.002	0.824
	(0.111)	(0.008)		(0.141)	(0.009)	
GPA 1st Semester	2.504	-0.057	0.168	2.602	-0.043	0.290
	(0.627)	(0.041)		(0.640)	(0.041)	
% GPA 1st Semester Imputed$^{a)}$	0.067	-0.001	0.968	0.088	-0.009	0.585
	(0.250)	(0.015)		(0.284)	(0.017)	
Credits 1st Semester	20.236	0.348	0.263	18.660	0.207	0.557
	(10.187)	(0.311)		(11.170)	(0.353)	
GPA at Randomization	2.491	-0.039	0.479	2.584	-0.031	0.493
	(0.713)	(0.054)		(0.683)	(0.046)	
% GPA at Randomization NA$^{a)}$	0.264	-0.014	0.426	0.116	-0.007	0.707
	(0.441)	(0.017)		(0.320)	(0.018)	
N	405	407		398	399	

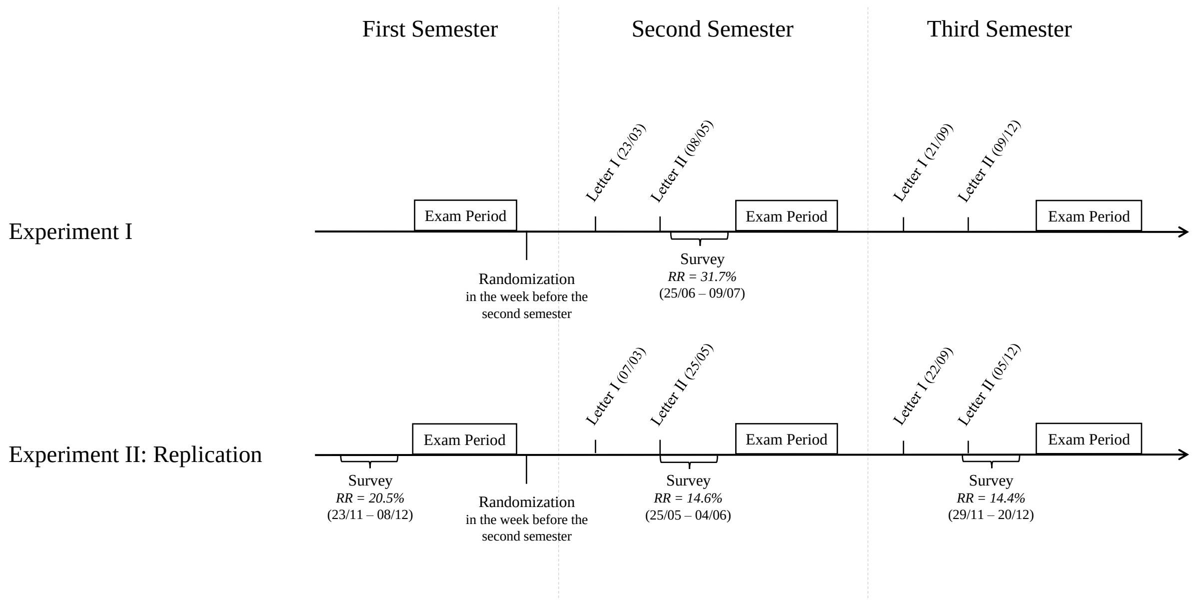
Note: Columns (1) and (4) present the unadjusted control group means and standard deviations of the covariates. For details on the variables see Table 3.A3 and Appendix C. Columns (2) and (5) present the estimated coefficients of regressing the covariates on the treatment indicator using Equation 3.1. Columns (3) and (6) test the null hypothesis of no treatment effect. $^{a)}$See Appendix C for details on the missing values and the imputation. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Feedback letter I. In the week before the second semester lectures start, students in the control and treatment group receive an unannounced letter in the mail, providing them with information on their accumulated credits and their cumulative GPA (see timeline in Figure 3.1). The letters thus include the same information that is available on the web portal

[9]See Appendix C for more details on the randomization procedure.

of the university. The treatment group additionally receives a graphical illustration that provides relative performance feedback on accumulated credit points. This feedback is shown in Figure 3.2, and we explain the design in detail below (the full letters are shown in Figures 3.A1 and 3.A2).

FIGURE 3.1: TIMELINE OF EXPERIMENTS



Feedback letter II. About four to five weeks before the exam period, students of both groups receive a second letter (see Figure 3.1). The letter design is identical to the first one, and for most students the contained information will also be identical to the first letter. In some cases the university updated the information on grades and credits (e.g., because first semester course results were not yet available when the first letter was composed), which can lead to different feedback compared to the first letter.[10] Apart from providing the most accurate information, the purpose of the second letter is to keep the feedback information salient as the exam period draws nearer. Consequently, we will use the content of the graph in the second letter when studying heterogeneity across the different types of feedback that students receive.

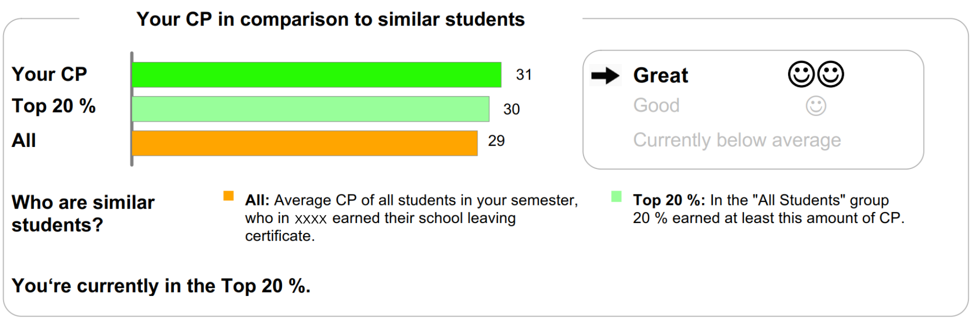
Relative performance feedback. The relative performance information in the treatment group is shown in Figure 3.2. It closely replicates social comparison approaches which have repeatedly been shown to reduce energy consumption, see e.g. Allcott (2011), Allcott and

[10]See Appendix C for details on the reasons and the number of observations that are affected.

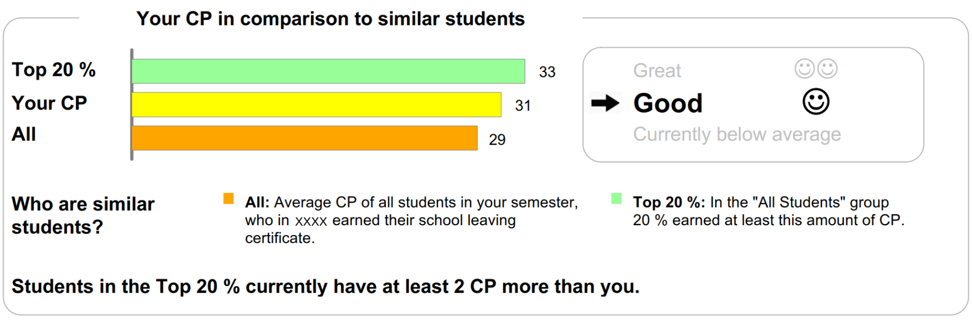
Rogers (2014), and Schultz et al. (2007). A bar chart compares the individual student's earned credits to the "Top 20%" and to the "Average" student enrolled in the same bachelor's program and from the same cohort as the student receiving the letter. "Average" performance is defined as the number of credits obtained by the median student(s), and "Top 20%" refers to the performance of the student(s) on the 80th percentile.

FIGURE 3.2: RELATIVE FEEDBACK GRAPHS – TREATMENT GROUP (EXAMPLES)

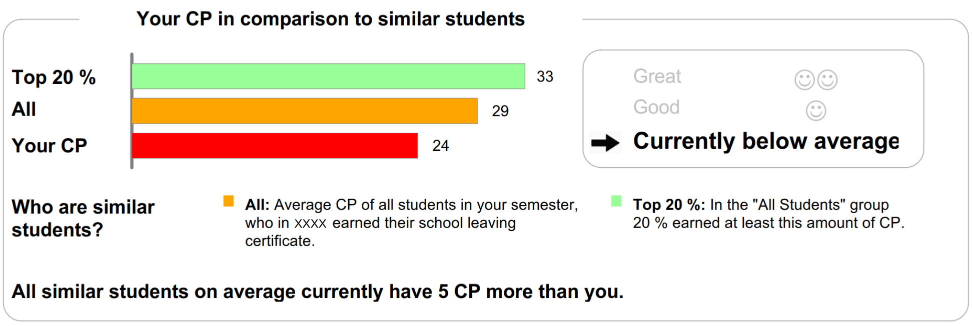
(A) Approving Frame *Great*



(B) Approving Frame *Good*



(C) No Approving Frame



To further personalize the performance feedback, we define several comparison groups for each program. This increases perceived similarity and minimizes the psychological distance to the reference group (Festinger 1954 and Trope and Liberman 2010). It also aims to reduce spillovers that could arise from sharing the feedback information.[11] In smaller programs the comparison group consists of students "who in/before ⟨*year*⟩ earned their school leaving certificate.", where *year* is the year in which the addressee of the letter received their school leaving certificate. In the two large bachelor programs, we use more fine-grained comparison groups by additionally distinguishing between the school-leaving certificates “vocational track degree (or below)” and “general track degree”.

According to the focus theory of normative conduct (Cialdini et al., 1990; Cialdini, 2011), this relative feedback can represent a descriptive norm, i.e., a reference point against which students can compare their performance. To prevent those above a norm from focusing downwards and reducing their effort (the so-called "boomerang effect"), the theory proposes to approve their performance using an injunctive norm. We thus follow the literature and add the following approving normative frames (Allcott, 2011; Cialdini, 2003; Schultz et al., 2007): performance that is at least average is categorized as *Good* (plus one "smiley" emoticon) and performance that is equal to or better than the 20% percentile is called *Great* (plus two "smiley" emoticons). Students who perform below the average do not receive an approving norm. Instead, the feedback includes the statement “currently below average” (and no emoticon).

Based on this design, in the heterogeneity analyses in Section 3.3, we consider the following feedback types: below-average, average, above-average (*Good*), and above-average (*Great*) feedback; see Table 3.2. Although the feedback for on-average students also includes a *Good* frame, we study this feedback type separately from the above-average type, because the literature suggests that normative frames may not have the intended effect if they are not aligned with the descriptive information (Cialdini et al., 2006). As we will discuss further in Sections 3.3 and 3.4, we also analyze the treatment effects for those who receive above-average feedback versus those who do not.

3.2.3 Field Experiment II: Replication

Using the same design as in the original experiment, we repeat the intervention one year later in the same programs and faculties (N = 797; Tables 3.1 and 3.A4 show the balancing properties[12]). The aim is to establish with a new cohort of students whether the results are replicable. In the taxonomy of Hunter (2001), Hamermesh (2007), and Czibor et al. (2019), ours is a statistical replication. Czibor et al. (2019) argue that early statistical replications are

[11] This rather “tailored” information in the letters should be shared less frequently because it may not appear to be of interest to other students, thus reducing downward-bias inducing spillovers.

[12] We made some small adjustments to the randomization procedure in Experiment II. See Appendix C for details.

TABLE 3.2: EXPERIMENTAL DESIGN AND NUMBER OF OBSERVATIONS BY FEEDBACK TYPE

Desc. Perf. Information	Not above average		Above average		
	Below average	On average	Below top 20%	Top 20%	
Normative Frame	No frame	*Good*		*Great*	*N* - Total
Experiment I	342	165	67	238	812
Control	174	76	34	121	405
Treatment	168	89	33	117	407
Experiment II: Replication	320	29	202	246	797
Control	163	12	97	126	398
Treatment	157	17	105	120	399

Note: Feedback type refers to the feedback students received in the first treatment semester.

"crucial" in experimental economics. Levitt and List (2009) call the difficulty of replication a "potential shortcoming" of field experiments, and so providing credible evidence that results can be reproduced is an important part of our study, especially given the inconclusive results in the literature on relative performance feedback and, more generally, the recent debate about replicability in economics and other fields (Camerer et al., 2016; Duvendack et al., 2017; Open Science Collaboration, 2015).

The only small change in the replication experiment is that we define average performance by using the mean instead of the median. This decision was made based on the result from the initial experiment that above-average feedback improves performance, as we will discuss in Sections 3.3 and 3.4. Table 3.2 shows that this tweak increases the share of students that receive above-average feedback from 37.5% in the initial experiment to 56.2% in the replication, and the share of students which exactly match the average is reduced from 20.3% to 3.6%. All design features of the original intervention remain unaltered.

3.2.4 Data and estimation

We use student-level data provided by the university's examination office and augment it with online surveys. Our main outcome of interest is the number of obtained credits, but as already mentioned, we will check for potential negative side effects on other domains, especially GPA and dropout behavior. We use demographic information and pre-treatment outcomes (first semester credits and first semester GPA) as covariates in our estimations. A detailed description of the data collection and processing as well as the use of the data for the randomization, feedback letters, and estimations can be found in Appendix C. Table 3.A3 provides a description of all variables.

Unless otherwise specified, we provide ITT effects from OLS estimations that compare the outcomes of the control and the treatment group. In the baseline specification, we follow the recommendations of Bruhn and McKenzie (2009) and control for the method of randomization:

$$Y_i^k = \alpha_0 + \alpha_1 Treatment_i + \mathbf{s_i}\boldsymbol{\alpha_2} + \varepsilon_i, \tag{3.1}$$

where Y_i^k denotes the level of outcome measure k for individual i. $Treatment_i$ is an indicator for being randomized into the treatment group. The vector $\mathbf{s_i}$ includes strata fixed effects which control for the random assignment of treatment and control units within blocks. In estimations with pooled data from both experiments, we also include a cohort dummy and its interaction with the strata variables.

In the second specification, we add a vector capturing baseline performance and further control variables that became available to us after randomization:

$$Y_i^k = \alpha_0 + \alpha_1 Treatment_i + \mathbf{s_i}\boldsymbol{\alpha_2} + \mathbf{x_i}\boldsymbol{\alpha_3} + \varepsilon_i. \tag{3.2}$$

$\mathbf{x_i}$ includes high school GPA, first semester credits, first semester GPA, age at randomization, an indicator for being female, time since high school graduation, and an indicator for the type of high school degree. The first semester GPA is missing for students who attempted no exams or failed all exams they attempted. In order to keep all observations in the sample, we impute values of the first semester GPA for students with a missing GPA.[13]

As discussed in Section 3.2.2 and shown in Table 3.2, building on the feedback design, we analyze heterogeneity across different feedback types:

$$Y_i^k = \alpha_0 + \alpha_1 Treatment_i + \mathbf{F_i}\boldsymbol{\alpha_2} + Treatment_i\mathbf{F_i}\boldsymbol{\alpha_{12}} + \mathbf{s_i}\boldsymbol{\alpha_3} + \varepsilon_i, \tag{3.3}$$

where $\mathbf{F_i}$ is a vector including either i) indicators for below-average (reference category), on-average, above-average (*Good*), and above-average (*Great*) feedback; or ii) an indicator for receiving above-average feedback (*Good* or *Great*). In a second specification, we again include the vector $\mathbf{x_i}$ with additional controls.

[13] See Appendix C for details.

3.3 Main results

3.3.1 Field Experiment I

Main effect. Table 3.3 shows the main effect of treatment on credits obtained in the second semester. Column (1) indicates that treated students obtain roughly .7 additional credits when controlling only for the method of randomization. Adding further control variables in Column (2) reduces the estimated effect to .3 credits. Although the parameters are positive, neither estimate is statistically significant and we therefore conclude that across the entire cohort, feedback does not increase performance.

TABLE 3.3: EFFECT OF FEEDBACK ON CREDITS

	Experiment I		Replication Experiment II		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.665	0.287	0.617	0.312	0.641	0.327
	(0.730)	(0.695)	(0.736)	(0.702)	(0.518)	(0.493)
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
N	812	812	797	797	1609	1609
Control Mean	21.07	21.07	19.75	19.75	20.42	20.42
(Std. Dev.)	(12.34)	(12.34)	(13.32)	(13.32)	(12.84)	(12.84)

Note: Outcome variable: credits second semester; *strata:* credit strata FE, study program FE, and in the pooled estimations a cohort dummy and its interaction with the study program FE; *controls:* HS GPA, credits first semester, GPA first semester, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Heterogeneity by feedback type. A characteristic of our experiment is that while all students in the treatment group receive feedback, the feedback slightly differs, depending on the position in the performance distribution (see Section 3.2.2). An important question to ask is thus whether students benefit from none of the feedback types or whether the treatment effects are, in fact, heterogeneous (Bryan et al. 2021 provide an overview of the treatment effect heterogeneity in earlier feedback experiments on energy conservation).

Panel (a) of Figure 3.3 visualizes the raw treatment effects with no control variables for the four different feedback types and Columns (1) and (2) in Panel (a) of Table 3.4 report estimates based on Equation 3.3 that control for the method of randomization and additional covariates, respectively.

Figure and table show positive treatment effects for the two above-average feedback types. The raw treatment effect for above-average students who receive a *Good* frame is 4.9 credits (4.7 when controlling for the method of randomization and 4.2 with further

TABLE 3.4: EFFECT OF FEEDBACK ON CREDITS – BY FEEDBACK TYPE

	Experiment I		Replication Experiment II		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a)						
Treatment (T)	-0.169	-0.459	-0.930	-1.412	-0.542	-0.893
	(1.196)	(1.196)	(1.147)	(1.146)	(0.829)	(0.826)
T*On-average	-1.088	-0.882	-3.470	-1.789	-1.145	-0.519
	(1.831)	(1.702)	(4.263)	(4.179)	(1.567)	(1.466)
T*Above-average *(Good)*	4.910*	4.609*	4.406**	4.553***	4.376***	4.245***
	(2.748)	(2.544)	(1.770)	(1.738)	(1.450)	(1.403)
T*Above-average *(Great)*	2.029	1.852	1.404	1.711	1.698	1.792
	(1.711)	(1.667)	(1.733)	(1.639)	(1.217)	(1.169)
T+T*On-average	-1.257	-1.341	-4.400	-3.200	-1.687	-1.411
	(1.390)	(1.238)	(4.110)	(4.030)	(1.333)	(1.219)
T+T*Above-average *(Good)*	4.741*	4.150*	3.475**	3.141**	3.834***	3.352***
	(2.471)	(2.237)	(1.346)	(1.299)	(1.188)	(1.130)
T+T*Above-average *(Great)*	1.859	1.393	0.474	0.299	1.156	0.900
	(1.222)	(1.155)	(1.300)	(1.179)	(0.891)	(0.828)
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
N	812	812	797	797	1609	1609
P-value of F-test	0.110	0.119	0.051	0.054	0.006	0.009
Panel (b)						
Treatment (T)	-0.334	-0.625	-0.968	-1.352	-0.587	-0.857
	(0.942)	(0.921)	(1.139)	(1.126)	(0.724)	(0.707)
T*Above-average	2.859**	2.613*	2.821*	2.984**	2.721***	2.644***
	(1.449)	(1.385)	(1.475)	(1.423)	(1.018)	(0.977)
T+T*Above-average	2.526**	1.988*	1.853**	1.632*	2.134***	1.787***
	(1.101)	(1.026)	(0.936)	(0.873)	(0.715)	(0.671)
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
N	812	812	797	797	1609	1609
P-value of F-test	0.049	0.059	0.056	0.036	0.008	0.007

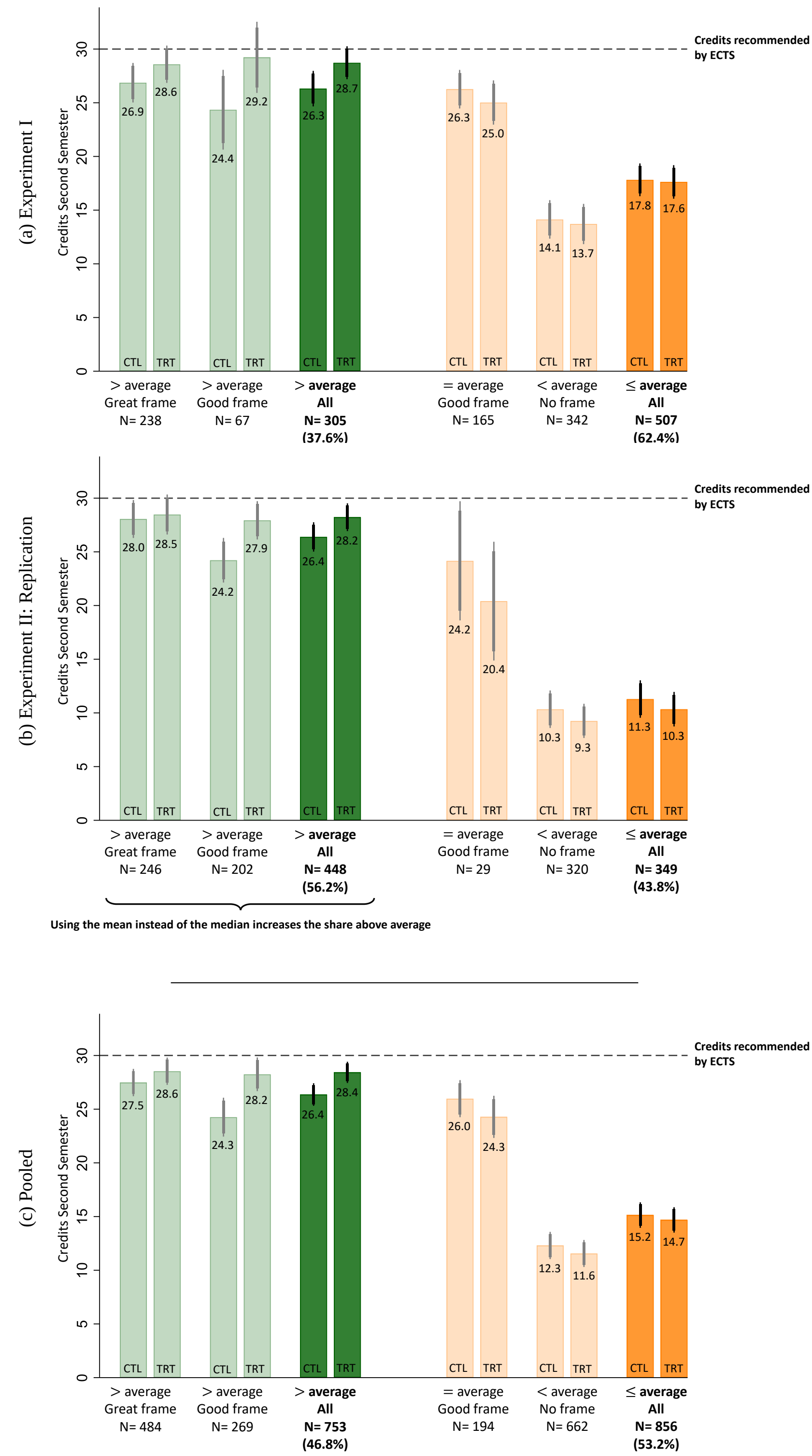
Note: Reference category in Panel (a) is below-average feedback. Reference category in Panel (b) are not-above-average students. P-values in the bottom row of the two panels are from F-tests that test the hypothesis that all interaction terms of treatment with the feedback types, i.e., $\boldsymbol{\alpha_{12}}$ in Equation 3.3, are equal to zero. *Outcome variable:* credits second semester; *strata:* credit strata FE, study program FE, and in the pooled estimations a cohort dummy and its interaction with the study program FE; *controls:* HS GPA, credits first semester, GPA first semester, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

controls) and significant at the 10%-level. Students in the top 20% received a *Great* frame, and we estimate a raw treatment effect of 1.7 credits (1.9 with strata controls and 1.4 with all controls) that is, however, not statistically significant (we discuss potential reasons for this smaller effect in Section 3.4.2). For both feedback types among students who are not above the average, on the other hand, we find no statistically significant effect on subsequent performance – though the point estimates are slightly negative. Still, there is an interesting distinction. For those below the average, receiving no approving normative frame is aligned with the information of having performed below the average. Yet, the students with an average performance receive an approving frame *(Good)*, so that frame and information can be considered misaligned in that the approving normative frame is not backed by factual information of being better than average. The lack of a (positive) treatment effect for this group is a tentative indication that simply attaching an approving normative frame to the information about a merely average performance is not able to raise subsequent performance. This is in line with research, which suggests that normative frames will not have the intended effect if they are not aligned with the descriptive information (Cialdini et al., 2006).

Overall, it appears that those who receive information of being above the average respond positively to feedback, while those who receive other types of feedback show no behavioral response. For all above-average-students, we find a significant treatment effect of 2.4 credits (2.5 with strata controls and 2.0 with all controls); see Figure 3.3 (dark green and dark orange bars) and Panel (b) of Table 3.4. This corresponds to an effect size of roughly .16 to .20 control group standard deviations. For those who do not receive above-average feedback we estimate effects between −0.2 and −0.6 credits that are far from being statistically significant at any conventional level. The difference in treatment effects between these two feedback types, above versus not above average, is statistically significant.

We use these insights from the original experiment when setting up the replication experiment. Given the evidence of performance gains for those receiving above-average feedback, our main hypothesis for the second experiment is that above-average feedback has performance enhancing effects (Section 3.4 shows more evidence from RDDs that this feedback type is driving the effects). It therefore seems reasonable to find ways of extending the above-average feedback to a larger share of students. This can be achieved by defining the average as the mean instead of the median, since the mean number of credits is lower than the median number of credits. As a consequence, in the replication experiment a much larger share of students receives above-average feedback (see Table 3.2 and Figure 3.3). While this small change leads to a much higher prevalence of above-average feedback, it preserves all design features of the original experiment.

FIGURE 3.3: EFFECT OF FEEDBACK ON CREDITS ACROSS FEEDBACK TYPES AND EXPERIMENTS



Note: The figures show the raw treatment effects without control variables across the four different treatment types in lighter shading. The bold print and darker shaded treatment effects for $>$*average* and $\leq$*average* combine the two *above-average* categories and *on-average* and *below-average*, respectively. Panel (a) shows treatment effects for the original experiment, Panel (b) for the replication experiment one year later, and Panel (c) for the pooled sample. In accordance with the ECTS (see Footnote 1) on average students are supposed to pass exams worth 30 credits per semester to finish in the regular study duration. 90% (thick) and 95% (thin) confidence intervals are shown.

3.3.2 Replication experiment and pooled results

Experiment II: Replication. In the replication experiment with a new cohort of students, we test the main hypothesis derived from the first experiment: receiving above-average feedback leads to improved performance. In addition, we investigate whether the entire pattern of results from the earlier experiment replicates. As can be seen in Columns (3) and (4) of Table 3.3 the full sample effects in the replication are very similar to the original experiment, at 0.3 to 0.6 credits, and not statistically significant. Looking at the effects for the different feedback types in Panel (b) of Figure 3.3 and Columns (3) and (4) of Table 3.4, we very closely replicate the results from the original experiment.

In line with our main hypothesis derived from Experiment I, we find that students increase their subsequent performance by 1.8 credits in response to feedback of being above the average (1.9 with strata controls and 1.6 with all controls). This corresponds to an effect size of roughly .12 to .14 control group standard deviations. As in the original experiment, the treatment effects for students who placed above-average are significantly different from those who did not (for whom we again find no statistically significant effects).

Pooled results. Since the two experiments share the same design, we also report results based on pooling the observations to increase the power of our statistical analysis. Panel (c) of Figure 3.3 shows the effects for the different feedback types, and Columns (5) and (6) of Tables 3.3 and 3.4 show the estimation results in the pooled sample. As expected, the estimated treatment effects fall between the original and the replication experiment and are more precise due to the larger sample size. In the pooled sample, above-average feedback increases subsequent performance by 1.8 to 2.1 credits (.14 to .17 control group standard deviations) and the estimate is statistically significant at the 1%-level. The effects can be interpreted as roughly one in three of these students passing an additional exam due to the treatment (on average, in our data an exam is worth 5.75 credits). Overall, the two experiments provide very robust evidence that above-average feedback increases performance. Students who are informed that they did not place above average show no behavioral response.

3.4 Mechanisms (i): regression discontinuity designs

In this section, we use regression discontinuity designs to show: (i) the behavioral response of above-average students is due to the information of being above average, and not driven by those students being more capable of responding to feedback than other students; (ii) the smaller treatment effect among above-average students who place in the top 20% is probably due to ceiling effects.

3.4.1 Characteristics of above-average students do not explain the response to above-average feedback

Given that the different feedback types are based on performance in the first semester, it is conceivable that the causal effect of above-average feedback is due to characteristics of above-average students and not due to the specific type of the feedback itself. For example, these students might have higher ability, a better learning technology, or they may be more motivated, enabling not only a higher first-semester performance, but also a better response to relative feedback.

Similar to the approach in Allcott (2011), we investigate this by employing a sharp regression discontinuity design (RDD) among treated students. We compare treated students just above the average to those on the average or just below, as they should not differ in their underlying characteristics. Instead, the only difference is the type of feedback that students receive: above-average feedback or not.

When implementing the RDD, we follow the suggestions of Lee and Lemieux (2010). If the usual RDD assumption holds, i.e., if there are no other discontinuities around the cutoff, it provides a causal local average treatment effect. To gather some intuition if this assumption is likely to hold, we study the control group, for which we should find no behavioral changes at the cutoff since they receive no relative performance feedback.[14] The running variable is the accumulated credits a student obtained in the first semester as depicted in the feedback letter, divided by the average credits of her respective comparison group (the corresponding distribution is shown in Figure 3.B1 in the Appendix).[15] Besides providing graphical depictions of the behavior of the outcome variable around the cutoff, we estimate the size of the jump by implementing a parametric RDD, using the following equation:

$$Y_i^k = \alpha_0 + \alpha_1 P_i + f(r_i) + f(r_i) P_i + \mathbf{s_i} \boldsymbol{\alpha_2} + \varepsilon_i, \tag{3.4}$$

[14] Another assumption is that individuals have no or only imprecise control over the running variable (Lee and Lemieux, 2010). This is very likely to hold in our case as when studying for their first semester exams, individuals do not know that they are going to receive feedback (let alone what form the feedback will have). Even if they did know, it would be virtually impossible to infer the exact value of the average performance in their comparison group or to precisely determine their position in the distribution of the assignment variable.

[15] This provides smoother distributions around the cutoff than using the raw distance to the cutoffs, because of differences in the credit point distributions within the different comparison groups.

where P_i indicates if a person placed on the right side of the cutoff, i.e., above the average. $f(r_i)$ is any smooth function of the running variable r_i that we allow to vary between the left and the right side of the cutoff and $\mathbf{s_i}$ is a vector including study program fixed effects and, in the pooled sample, a cohort fixed effect and its interaction with the study program fixed effects.[16]

The existence of a control group that does not receive relative performance feedback allows us to account for any potential jump in our outcome variable at the cutoff that is due to unobserved discontinuities that are the same in the treatment and the control group. We do this by estimating the following regression discontinuity difference-in-difference (RD-DID) specification:[17]

$$\begin{aligned} Y_i^k &= \alpha_0 + \alpha_1 Treatment_i + \alpha_2 P_i + \alpha_{12} Treatment_i P_i \\ &\quad + \alpha_3 r_i + \alpha_{13} Treatment_i r_i + \alpha_{23} P_i r_i + \alpha_{123} Treatment_i P_i r_i + \mathbf{s_i} \boldsymbol{\alpha_4} + \varepsilon_i, \end{aligned} \tag{3.5}$$

where we are interested in the parameter α_{12}.

Figure 3.4 visualizes the behavior of the outcome variable around the cutoff for the pooled sample. For the treatment group we observe a large jump of about seven credits when students receive above-average feedback, while we find no jump for the control group. Estimates of the respective coefficients based on two different discontinuity samples and a first order polynomial are shown in Table 3.5. Across the two experiments and the pooled sample, we estimate effects of 5.5 to 8.7 credits for students in the treatment group, all significant at the 1%-level. For students in the control group, on the other hand, we find no evidence for changes in second semester performance at the cutoff. Accordingly, coefficients based on the RD-DID specification shown in the bottom row confirm the effects estimated for the treatment group.[18] The absence of effects in the control group additionally provides some evidence against spillover effects, as we would expect that a high prevalence of information sharing should lead to jumps in the control group similar to those observed for the treatment group.

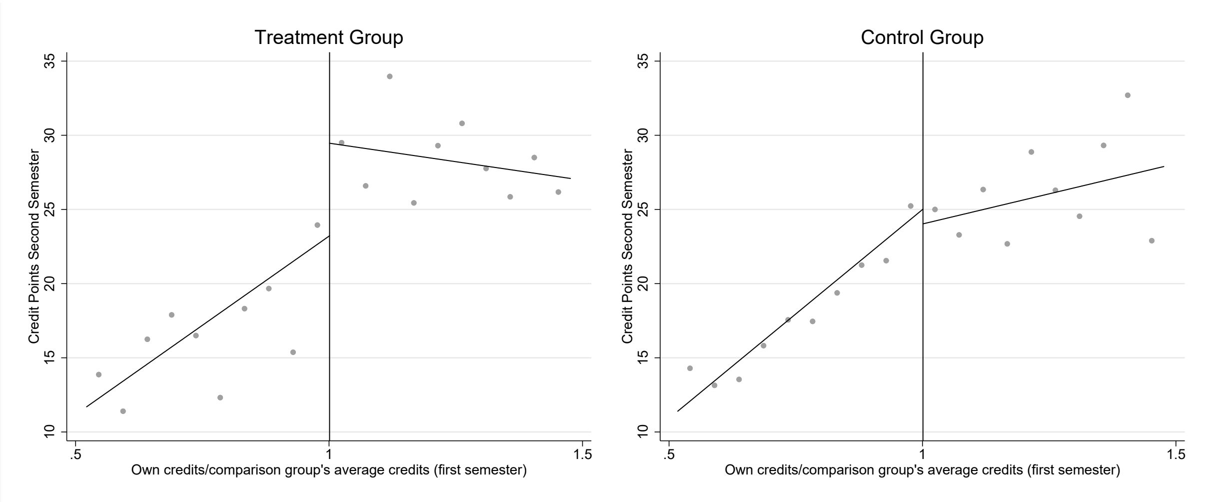
---

[16] We do not include the vector $\mathbf{x_i}$ as covariates, as this can make it difficult to differentiate between an inappropriate functional form and discontinuities in the covariates (Lee and Lemieux, 2010).

[17] See e.g., Danzer and Lavy (2018) or Dustmann and Schönberg (2012) for more papers that make use of similar RD-DID specifications.

[18] In Table 3.B1 in the Appendix, we show that the pooled sample estimates are robust to different polynomial specifications and further discontinuity samples. Estimated coefficients only become imprecise and unreliable when we increase the order of the polynomials while at the same time using small discontinuity samples. We are not concerned about this, as higher order polynomials can lead to overfitting, especially when the number of observations is low. As an additional robustness check, Figures 3.B2 to 3.B3 visualize the behavior of pre-treatment covariates around the cutoff. Most of them behave smoothly and in the few cases where we do observe small discontinuities, they behave similarly in the treatment and the control group. Any effects of those discontinuities should therefore be captured by our RD-DID specification.

FIGURE 3.4: RD PLOT AT AVERAGE – FIRST ORDER POLYNOMIAL, POOLED SAMPLE



Note: Binned scatterplots using first order polynomials. Running variable is the ratio of first semester credits as depicted in the feedback letter to the comparison group's average credits. Observations on the left side of the cutoff did not place above average. Observations on the right side placed above average.

TABLE 3.5: RD ESTIMATES AT AVERAGE – FIRST ORDER POLYNOMIAL

	Experiment I		Replication Experiment II		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity Sample	$0.25 < r < 1.75$	$0.5 < r < 1.5$	$0.25 < r < 1.75$	$0.5 < r < 1.5$	$0.25 < r < 1.75$	$0.5 < r < 1.5$
Treatment Group	7.874***	6.675***	6.346***	8.745***	5.492***	6.418***
	(1.776)	(1.991)	(2.028)	(2.531)	(1.243)	(1.450)
N	336	295	320	238	656	533
Control Group	0.515	-0.144	-0.689	-0.787	-0.887	-1.246
	(2.073)	(2.633)	(2.415)	(2.972)	(1.465)	(1.743)
N	344	302	313	238	657	540
Diff-in-Diff	7.163***	6.377**	6.488**	9.035**	5.742***	7.126***
	(2.704)	(3.225)	(3.153)	(3.878)	(1.889)	(2.230)
N	680	597	633	476	1313	1073
Study Program FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: Outcome variable: credits second semester; *study program FE:* study program FE and in the pooled estimations a cohort dummy and its interaction with the study program FE; *running variable (r):* ratio of first semester credits as depicted in the feedback letter to the comparison group's average credits. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

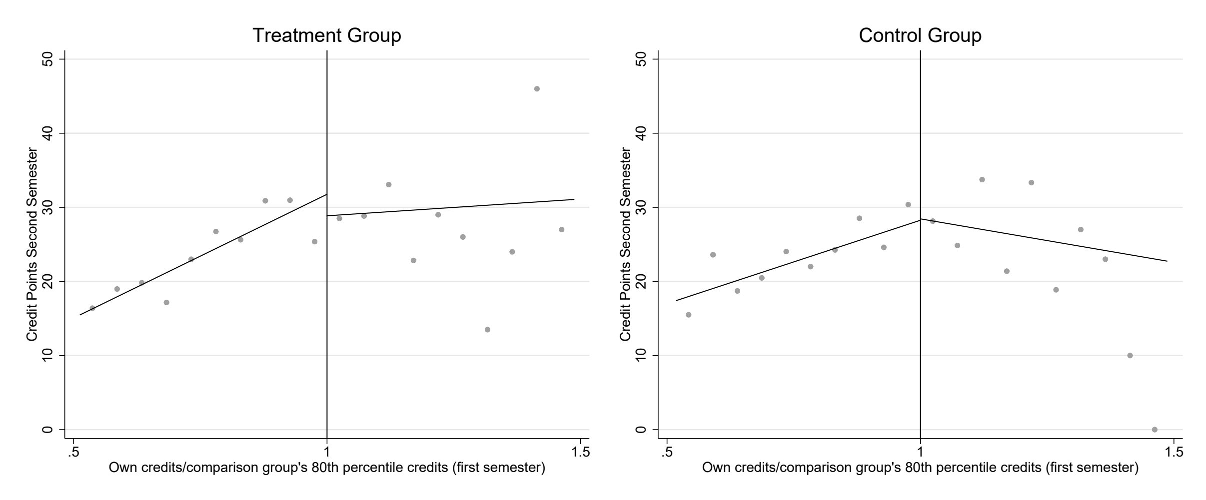
Overall, these RD estimates thus deliver robust evidence that relative performance feedback increases the subsequent performance of a student around the cutoff by roughly six credits if it informs her about an above-average instead of an at or below-average performance. This finding provides evidence that the large treatment effects for above average students are not due to their underlying characteristics. Rather, it suggests that the content of the relative performance feedback matters for the increase in subsequent performance.[19]

[19] However, the design of our experiment does not allow us to disentangle whether the information about being above average itself or the combination with the approving normative frame is driving the effect of

3.4.2 Why are the positive effects smaller for students in the top 20 percent?

While the treatment effects are positive for all students above the average, the top 20% of students react less to feedback. We want to understand what is behind this heterogeneity – specifically, we are interested in whether it is driven by the slight differences in feedback, or by the characteristics and specific circumstances of the top 20% students.

FIGURE 3.5: RD PLOT AT 80TH PERCENTILE – FIRST ORDER POLYNOMIAL, POOLED SAMPLE



Note: Binned scatterplots using first order polynomials. Runng variable is the ratio of first semester credits as depicted in the feedback letter to the 80th percentile of credits in the comparison group. Observations on the left side of the cutoff placed below the top 20%. Observations on the right side of placed in the top 20%.

This can be assessed with an RDD following the same approach as in the last Section. We now use the ratio of first semester credits to the 80th percentile of credits in the comparison group as our running variable and study the sharp cutoff at the 80th percentile.[20] If the attenuated treatment effects among students in the top 20% are indeed caused by the slightly different feedback type, we should find a negative jump in the outcome variable at the 80th percentile among treated students.

Figure 3.5 and the estimates in Table 3.6, show no significant jump at the cutoff.[21] We therefore conclude that there is no evidence of the differences in feedback type being behind the observed heterogeneity among above-average students.

above-average feedback. As noted in Section 3.3.1, the fact that we do not observe positive treatment effects for students with an average performance provides evidence that the inclusion of an approving normative frame alone is not sufficient to elicit performance gains.

[20]The distribution of the running variable is shown in Figure 3.B4 in the Appendix.

[21]Table 3.B2 in the Appendix shows that this result is robust to various polynomial specifications and discontinuity samples.

TABLE 3.6: RD ESTIMATES AT 80TH PERCENTILE – FIRST ORDER POLYNOMIAL

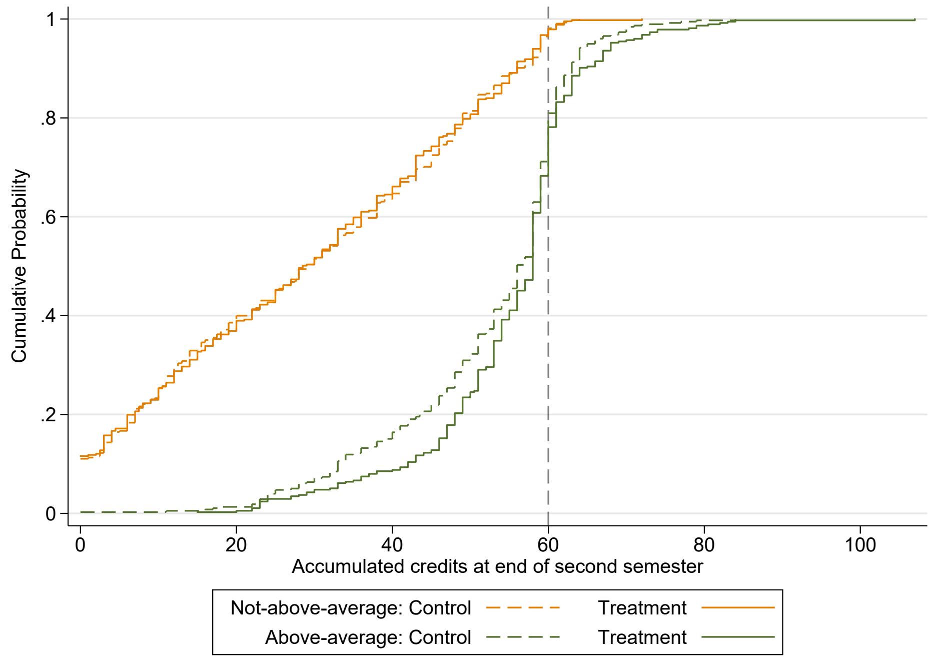
	Experiment I		Replication Experiment II		Pooled	
	(1)	(2)	(3)	(4)	(5)	(6)
Discontinuity Sample	$0.25 < r < 1.75$	$0.5 < r < 1.5$	$0.25 < r < 1.75$	$0.5 < r < 1.5$	$0.25 < r < 1.75$	$0.5 < r < 1.5$
Treatment Group	2.271	0.836	1.681	0.555	2.632	0.891
	(2.584)	(2.332)	(2.350)	(3.532)	(1.720)	(2.038)
N	348	285	319	267	667	552
Control Group	-0.108	0.926	1.135	5.094**	0.621	3.256
	(2.808)	(2.918)	(2.562)	(2.575)	(1.894)	(2.113)
N	355	291	314	267	669	558
Diff-in-Diff	2.485	0.012	0.734	-4.330	2.097	-2.233
	(3.778)	(3.639)	(3.368)	(4.200)	(2.501)	(2.848)
N	703	576	633	534	1336	1110
Study Program FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: Outcome variable: credits second semester; *study program FE:* study program FE and in the pooled estimations a cohort dummy and its interaction with the study program FE; *running variable (r):* ratio of first semester credits as depicted in the feedback letter to the 80th percentile of credits in the comparison group. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Instead, we argue that the heterogeneity is rooted in the characteristics and circumstances of those in the top 20%. Specifically, ceiling effects are a likely explanation. In Figure 3.3 it can be seen that treatment raises second semester performance for the above-average groups to roughly 29 credits, but no higher than that. Figure 3.6 provides additional evidence. It shows the cumulative distributions of total credit points at the end of the second semester for those above- and those not-above the average, separated by treatment status. Clearly, the treatment effects among above-average students are concentrated in the lower parts of the distribution, i.e., below 60 first-year credits.

Several mechanisms provide potential explanations for these observations. First, it could be the case that even top performing students do not have the capacity, cognitive or otherwise, to obtain more than roughly 30 credits per semester. Second, and perhaps more plausible, in the standardized European ECTS system, students are advised to collect 30 credits per semester, and curricula are set accordingly. This number may set an “artificial” ceiling or reference point that most students do not exceed. Finally, in Section 3.6.1, we will show that the effect of relative feedback is larger for students who did not expect to perform above average. It is therefore also possible that the best performing students are less surprised and, accordingly, less affected by the above-average feedback.

FIGURE 3.6: CUMULATIVE DISTRIBUTION OF ACCUMULATED CREDIT POINTS AT THE END OF THE SECOND SEMESTER – POOLED SAMPLE



Note: The figure plots the cumulative distribution of the accumulated credit points at the end of the second semester by treatment status for students who did not receive above-average feedback and students who received above-average feedback. The vertical dashed line indicates the number of accumulated credit points that students should have obtained at the end of the second semester in accordance with the ECTS (see Footnote 1).

3.5 Repeated treatment

In this section, we report results for the effects of repeated treatment, i.e., effects on performance in the third semester (= second treatment semester). We have to take into account that the treatment effects on second semester performance can affect the type of feedback that students receive in the third semester. To avoid this possible endogeneity problem, we therefore analyze the effects of receiving a certain type of feedback in the second semester on performance in the third semester.

Results based on estimations with all control variables are provided in Table 3.7. Panel (a) shows estimates for the treatment effects on credits obtained in the third semester. As in the second semester, we do not find evidence for effects on performance among all students (Columns 1 to 3) or for those who did not place above average (Columns 4 to 6) in either cohort. We also do not find additional positive effects for students who received above-average feedback in the second semester (Columns 4 to 6). In Panel (b), we report effects on the accumulated credits at the end of the third semester. In line with the previous results, the

TABLE 3.7: EFFECT OF FEEDBACK ON CREDITS – REPEATED TREATMENT

	(1) Exp. I	(2) Exp. II	(3) Pooled	(4) Exp. I	(5) Exp. II	(6) Pooled
(a) Credits in 3rd sem.						
Treatment (T)	0.032	-0.239	-0.061	-0.321	0.318	0.020
	(0.686)	(0.717)	(0.496)	(0.918)	(1.167)	(0.720)
T*Above-average				0.948	-0.974	-0.150
				(1.355)	(1.467)	(0.983)
T+T*Above-average				0.626	-0.656	-0.129
				(0.998)	(0.891)	(0.671)
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	812	797	1609	812	797	1609
Control Mean	20.35	17.93	19.15			
(Std. Dev.)	(12.54)	(13.23)	(12.93)			
(b) Total credits 3rd sem.						
Treatment (T)	0.319	0.074	0.266	-0.946	-1.034	-0.836
	(1.179)	(1.238)	(0.855)	(1.580)	(2.045)	(1.245)
T*Above-average				3.561	2.010	2.494
				(2.340)	(2.536)	(1.694)
T+T*Above-average				2.614	0.976	1.658
				(1.712)	(1.503)	(1.145)
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	812	797	1609	812	797	1609
Control Mean	61.66	56.35	59.03			
(Std. Dev.)	(29.70)	(32.29)	(31.10)			

Note: Above-average and not-above-average (the reference category) refers to the type of feedback students received in the second semester. *Outcome variables:* credits obtained in third semester (Panel a), sum of credits obtained in the first three semesters (Panel b); *strata:* credit strata FE, study program FE, and in the pooled estimations a cohort dummy and its interaction with the study program FE; *controls:* HS GPA, credits first semester, GPA first semester, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

full sample effect is close to zero. Regarding the effects on above-average students, we find that the estimated coefficients are close to the corresponding effects on second semester credits shown in Panel (b) of Table 3.4. For example, in the pooled sample, treated above-average students have accumulated 1.7 credits more by the end of the third semester, which is close to the second-semester treatment effect of 1.8 credits reported in Column (6) of Table 3.4. However, due to the higher variance of the accumulated credits at the end of the third semester – see the bottom row of Panel (b) in Table 3.7 – the coefficient is no longer statistically significant ($p = 0.148$).

These findings suggest that there are no additional effects of repeated treatment on top of those that we found for the second semester. In the next section, we provide evidence that the significant second semester effects, as well as the lack of additional effects of repeated treatment may plausibly be driven by beliefs about relative performance.[22]

3.6 Mechanisms (ii): beliefs about relative performance and theoretical considerations

3.6.1 Expectations and treatment effects

The goal of the second experiment was not only replication, but also to further investigate the mechanisms through which feedback changes behavior. One important condition for the effectiveness of feedback is that it provides new information and that individuals actually process it. To shed light on this, in the replication experiment, we conducted pre- and post-treatment surveys asking students about their expected relative performance in terms of credits; see Figure 3.1 for the timing of the surveys and Table 3.A5 for the questions.[23][24] We can then address the following questions. First, how do students' expectations line up with actual relative performance? Second, are the beliefs about relative performance influenced by the feedback? Third, does the treatment effect depend on the accuracy of the initial beliefs?

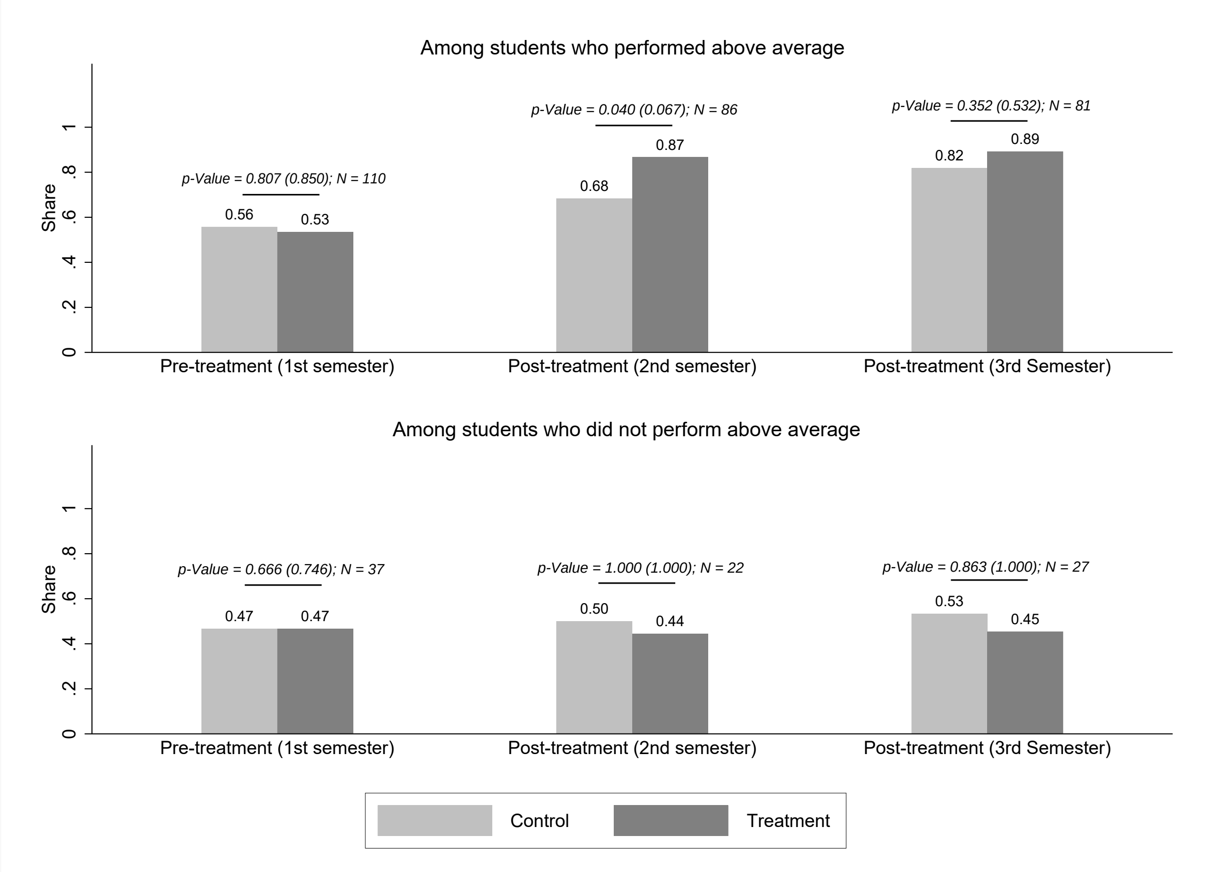
To address the first two questions, we use the survey questions to create a variable that indicates whether a student expects to place above average in the performance distribution at the end of the semester. Figure 3.7 visualizes the pattern of belief updating across the three semesters separately for students who performed above-average in the first semester (top panel) and students who did not (bottom panel). The two left panels provide evidence that pre-treatment students have very little intuition about their actual relative performance. Irrespective of actually placing above average or not, around 50% of treatment and control students expect to perform above average. In the second semester (middle panels), we find

[22] The institutional setup at the university may also play a role for the dynamics of the effects, and prevent treatment effects in the third semester. For more than half of the students in our sample, the mandatory internship semester is scheduled for the fourth semester. Therefore, students who wish to take additional exams in the third semester may have to choose from courses that are scheduled for the period after the internship semester. These courses are mostly electives, and because the internship period is supposed to help students figure out which electives to choose, they may decide against taking these exams early – thus precluding further effects.

[23] The wording in the pre- and post-treatment survey is not exactly the same. While this may have level effects on control and treatment groups simultaneously, it should not affect the argument about differential updating between treatment and control we make in this section.

[24] One caveat applies when considering the data from the surveys: in line with the general development in survey nonresponse rates (see e.g., Leeper 2019), the response rates in our surveys are between 15 and 30% (see Figure 3.1). Accordingly, our sample is rather small, and we find evidence that the respondents are a positively selected subpopulation.

FIGURE 3.7: SHARES OF STUDENTS WHO EXPECTED TO PERFORM ABOVE AVERAGE – EXPERIMENT II



Note: Above- and not above-average performance refer to the performance in the first semester and the respective type of feedback students received in the second semester. See Table 3.A5 for the survey questions on students' expectations and Figure 3.1 for the exact timing of the surveys. p-Values based on Pearson's chi-squared tests. p-Values in parenthesis based on Fisher's exact test.

that those who actually placed above average updated their beliefs: 87% in the treatment group and 68% in the control group now expect to be better than the average; the difference of 19 percentage points is significant at the 5%-level. For students who did not receive above-average feedback in the second semester, on the other hand, neither the control nor the treatment group appears to update their expected relative performance. The right panels show that in the third semester a large majority of above-average students, both in the treatment and the control group, expects to perform better than the average. For those who did not receive above-average feedback, we still find no evidence for updates in beliefs.

These results suggest the following: First and in line with the positive effects on performance, in the second semester above-average feedback leads to more accurate beliefs about relative performance compared to controls. Second, over time control students who performed above average learn about their relative performance, even in the absence of relative feedback. As a result, in the third semester, beliefs of students in the control group are almost as accurate as the beliefs of students in the treatment group. This disappearing informational gap between the two groups can plausibly explain the lack of additional treatment effects in the third semester. Last, we cautiously take the fact that those who did not place above average do not update beliefs as evidence that students discard or discount the feedback if they do not perform above average. This could provide an explanation why in the second semester we only observe a behavioral response for above-average students.

We can also study whether above-average students who received new or unexpected information respond more strongly – in contrast to students who already expected that they would perform above average.[25] To test this, we create a dummy U_i that is 1 if a student underestimated their performance in the first semester, i.e., they did not expect an above average performance although they then actually performed better than the average. We estimate the following equation among students that placed above average:

$$Y_i^k = \alpha_0 + \alpha_1 Treatment_i + \alpha_2 U_i + \alpha_{12} Treatment_i U_i + \mathbf{s_i} \boldsymbol{\alpha_3} + \varepsilon_i, \tag{3.6}$$

where α_1 gives the treatment effect for those who correctly expected to be above average, and α_{12} gives the difference in the treatment effect for those who underestimated their relative performance. Table 3.8 presents the results. Column (3) shows that the treatment effect for those who correctly estimated their position is roughly 1.4 credits, which is 1.3 points smaller than the 2.7 credit treatment effect that we find in the entire survey sample (Column 1). Control group students who underestimated their relative performance obtain on average 2.8 credits less in the second semester than students who did not underestimate their actual relative performance. The interaction suggests that informing those students that they actually are above the average can increase their performance to the level of those who correctly anticipated to be above average.

In Columns (2) and (4) we add control variables; especially the pre-treatment performance should be accounted for, as it is correlated with both the post-treatment performance and with the first semester expectations. We find that the negative effect of underestimating performance now becomes stronger. Given the covariates, this indicates that at equal ability, lower confidence can be detrimental to performance. Again, this negative effect is completely offset by receiving feedback. Overall, the treatment generates a significant effect of 5.7 credits for students who underestimated their relative performance (fourth row in Column 4), indicating that relative performance feedback will be especially helpful for this

[25] The number of survey respondents who perform (below-)average is too low to study their behavior in such detail.

TABLE 3.8: EFFECT OF FEEDBACK ON CREDITS BY PRE-TREATMENT EXPECTATIONS – ABOVE-AVERAGE STUDENTS, EXPERIMENT II: REPLICATION

	(1)	(2)	(3)	(4)
Treatment	2.678*	2.952**	1.426	0.701
	(1.528)	(1.274)	(1.634)	(1.147)
Underestimated Performance			-2.841	-3.457
			(2.704)	(2.133)
Treatment*Underestimated			2.797	5.012*
			(3.271)	(2.694)
Treatment+(Treatment*Underestimated)			4.223	5.713**
			(2.794)	(2.410)
Study Program FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
N	110	110	110	110

Note: Underestimated performance indicates students who expected a not-above-average performance in the first semester but then received above-average feedback. *Outcome variable:* credits second semester; *controls:* HS GPA, credits first semester, GPA first semester, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

group.

3.6.2 Theoretical considerations

How does our pattern of results and the evidence on students' beliefs about their relative performance line up with some of the most common theories on the effects of feedback and social comparison?

First, social comparison theory and competitive preferences do not fit our results particularly well: The focus theory of normative conduct suggests that individuals try to comply with descriptive norms, e.g., the average performance level, predicting positive treatment effects for those below the descriptive norm (Cialdini et al., 1990; Cialdini, 2011). For those above the norm, the focus is on preventing negative effects by adding an approving message, i.e., an injunctive norm (Allcott, 2011; Cialdini, 2003; Schultz et al., 2007). If individuals have competitive preferences, relative feedback is usually predicted to increase performance across the entire distribution (Azmat and Iriberri, 2010; Dobrescu et al., 2021). Given that we find no evidence for an increase in performance after below-average feedback, these theories do not provide a convincing explanation for our pattern of results.

Second, "self-perception theory" proposes that feedback can influence behavior by changing beliefs over ability (Azmat and Iriberri, 2010; Dobrescu et al., 2021; Ertac, 2005): performance is a function of both effort and ability, and ability and effort are assumed to be complements. Feedback will then affect beliefs about ability and thus the optimal choice of effort, leading to increased (decreased) effort and performance if the feedback signals a higher (lower) ability than individuals previously believed. In our context, above-average feedback provides a favorable signal about ability, explaining the positive treatment effects for this group of students – in particular among those who did not expect to perform above average. However, our results for students who do not place above average and who arguably receive an unfavorable signal about their ability, are in contrast with the theoretical predictions as we find no evidence for negative effects.

Instead, our pattern of results and belief updating can best be reconciled with theories on the management of self-confidence and the selective processing of information that often accompanies it (Villeval 2020 provides an overview on how relative performance feedback and confidence are connected). Similar to the mechanism in "self-perception theory", relative feedback provides a signal about ability. Because the signal can affect the confidence of individuals in their ability, it creates incentives to process feedback in a confidence-preserving way. For example, in the model by Bénabou and Tirole (2002) favorable signals serve individuals to form beliefs about their ability and maintain a positive self-image; positive effects on performance are expected, because effort and ability are assumed to be complements. To maintain confidence in one's own ability, adverse signals, on the other hand, do not adequately enter into beliefs: individuals selectively process good information. In Compte and Postlewaite (2004) the mechanism is similar: here, positive outcomes are attributed to own abilities or efforts whereas negative outcomes are attributed to, e.g., unfortunate circumstances and therefore do not appropriately depress self-confidence. In both models, the induced optimism and confidence in own abilities can then lead to better performance.[26]

The notion that individuals will update beliefs in ego-relevant domains such as relative performance asymmetrically is supported by evidence from laboratory experiments: individuals who receive good news about their rank in an IQ test are willing to incorporate this information in their beliefs, while individuals who receive negative news have little willingness to update their self-concept (Eil and Rao, 2011; Möbius et al., 2014).

Our result pattern is in accordance with the theoretical ideas in Bénabou and Tirole (2002), if we assume that students' beliefs about relative performance are linked to their beliefs about ability. In the second semester, above-average students who receive feedback have more favorable beliefs about their relative performance than controls and increase

[26]Beyond this mechanism, confidence may also have a direct effect on utility, i.e., individuals may simply enjoy feeling good about themselves (Bénabou and Tirole, 2002; Compte and Postlewaite, 2004; Köszegi, 2006). This may also be a motivational factor for effort allocation.

their subsequent performance. By the third semester, our results on students' beliefs suggest that beliefs about ability in the control and the treatment group have converged, which explains why there are no additional effects on performance. Students who did not receive above-average feedback show no sign of correcting their beliefs downward. Consistent with Bénabou and Tirole (2002), these students may hold optimistic beliefs about their abilities to overcome lack of willpower and stay motivated.[27] Feedback would inform these students that their ability is lower than initially believed, which would lead to a decline in motivation and effort and potentially worse outcomes. Disregarding feedback may then prevent this by preserving confidence in ability.

3.7 Negative spillovers on other domains?

We have found robust effects of above-average feedback on achieved second semester credits. An important question is whether students generate these gains in performance at the cost of losses in other domains. First, a concern could be that encouraging students to obtain more credits may come at the expense of worse grades because students may shift attention away from them. In Panel (a) of Table 3.9, we report treatment effects on students' GPA at the end of the second semester. We do not find any significant effects, neither in the full sample (Columns 1 to 3) nor among above-average students (Columns 4 to 6). Table 3.A6 in the Appendix shows that this also holds true in the third semester. It thus appears that feedback can raise performance in terms of obtained credits, without negatively affecting the other major performance dimension.

Second, it could be the case that relative performance feedback affects the dropout decision of students. Panel (b) of Table 3.9 reports effects on having dropped out of the study program by the end of the second semester. As with GPA, we find no evidence for statically significant effects on students' dropout behavior; this is also the case in the third semester (see Table 3.A6 in the Appendix).

One might also worry that the feedback affects other dimensions that are indirectly related to performance. For both our experiments we conducted a post-treatment survey in which we asked students how satisfied they are with their life, the degree to which they are satisfied with their study program, the degree to which they are satisfied with their performance, and how stressful they find their studies (see Figure 3.1 for the timing of the surveys and Table 3.A7 in the Appendix for the questions and the variables used in the estimations).

[27] Estimates suggest that up to 95% of college students may be subject to self-control problems (Ellis and Knaus, 1977; O'Brien, 2002); König et al. (2019) provide evidence that university students manage beliefs about return to effort in ways consistent with Bénabou and Tirole (2002).

TABLE 3.9: EFFECT OF FEEDBACK ON GPA AND DROPOUT

	(1) Exp. I	(2) Exp. II	(3) Pooled	(4) Exp. I	(5) Exp. II	(6) Pooled
(a) GPA						
Treatment (T)	-0.012	0.016	0.002	-0.021	0.043	0.002
	(0.021)	(0.020)	(0.014)	(0.029)	(0.038)	(0.023)
T*Above-average				0.025	-0.044	0.000
				(0.041)	(0.044)	(0.029)
T+T*Above-average				0.004	-0.001	0.002
				(0.029)	(0.021)	(0.017)
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	767	744	1511	767	744	1511
Control Mean	2.57	2.63	2.60			
(Std. Dev.)	(0.59)	(0.60)	(0.60)			
(b) Dropout						
Treatment (T)	-0.006	0.016	0.004	0.010	0.041	0.022
	(0.020)	(0.023)	(0.015)	(0.030)	(0.047)	(0.026)
T*Above-average				-0.041	-0.046	-0.039
				(0.037)	(0.051)	(0.030)
T+T*Above-average				-0.031	-0.005	-0.017
				(0.022)	(0.020)	(0.015)
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	812	797	1609	812	797	1609
Control Mean	0.13	0.16	0.15			

Note: Outcome variables: GPA by the end of the second semester (Panel a; passing grades only; highest passing grade is 1.0, lowest passing grade is 4.0), dropout during or before the second semester (Panel b); *strata:* credit strata FE, study program FE, and in the pooled estimations a cohort dummy and its interaction with the study program FE; *controls:* HS GPA, credits first semester, GPA first semester, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 3.10 shows the corresponding treatment effects for the pooled sample.[28] We find no statistically significant effects on any of the well-being dimensions we observe.

Taken together, these results show that the increase in the number of obtained credits does not come at the cost of negative effects on other outcomes, and can therefore be interpreted as a net positive effect of above-average feedback.

3.8 Conclusion

In a field experiment and a direct replication, we investigate the effects of relative feedback on academic performance. Our results show that students increase their subsequent performance when the feedback informs them about an above-average performance. With the

[28]The number of observations can vary between the outcomes, as students were allowed to give no answers to the questions in the survey.

TABLE 3.10: EFFECT OF FEEDBACK ON WELL-BEING – POOLED SAMPLE

	Satisfaction with				Satisfaction with			
	Life	Studies	Perform.	Stress	Life	Studies	Perform.	Stress
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment (T)	0.046	0.086	0.026	0.002	-0.087	0.082	-0.006	0.029
	(0.107)	(0.086)	(0.067)	(0.085)	(0.183)	(0.148)	(0.128)	(0.125)
T*Above-average					0.213	0.008	0.046	-0.039
					(0.221)	(0.185)	(0.150)	(0.172)
T+T*Above-average					0.126	0.090	0.040	-0.009
					0.128)	(0.108)	(0.075)	(0.115)
Study Program FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	353	353	356	354	353	353	356	354

Note: Outcome variables: Outcomes are standardized to have mean zero and standard deviation one; see Table 3.A7 for the survey questions that are used for the construction of the outcomes; *study program FE:* study program FE, a cohort dummy, and the interaction of the cohort dummy with the study program FE; *controls:* HS GPA, credits first semester, GPA first semester, age, female dummy, time since HS degree, and HS degree Abitur dummy. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

help of a regression discontinuity design we show that this is true irrespective of the underlying characteristics of the students.

In order to investigate the mechanism behind the behavioral reaction to above-average feedback, we survey individuals about their pre- and post-treatment expectations concerning relative performance and find suggestive evidence that the information about a below-average or average performance is not processed in the same way as the information about an above-average performance – which can explain the difference in the behavioral responses. In addition, our findings suggest that relative performance feedback is especially effective for those individuals with an above-average performance who initially underestimate their relative performance and this underestimation is linked to worse performance in the absence of relative feedback. The pattern of results is consistent with theoretical ideas suggesting that a higher confidence in ability motivates individuals, and that individuals try to maintain a positive self-assessment of their abilities by selectively processing information (see e.g., Bénabou and Tirole 2002).

Our results have important implications from a policy perspective. The intervention presents a low-cost and easy to implement tool which can increase the performance of a large share of students at a time that is crucial for habit-formation and getting on track to graduation. In addition, our findings suggest that the feedback is especially helpful to individuals who are held back by underestimation of their relative abilities. We also present tentative evidence that selective information processing may prevent undesirable effects of relative feedback when it threatens confidence in ability. While we find no negative effects of feedback on those in the lower parts of the performance distribution, policy-makers should be aware that this still implies that relative performance feedback can have distributional implications to the effect of widening achievement gaps. Future feedback schemes aimed at preventing this should take into account that feedback for weaker students may need to be designed in a way that does not jeopardize confidence but is able to enhance it. Another interesting avenue for future research is to investigate if and under which conditions feedback can generate beneficial effects in the long-run.

Appendix

A Additional tables and figures

TABLE 3.A1: SUMMARY OF COST INCURRED BY THE RELATIVE PERFORMANCE FEEDBACK (IN EUROS)

Cost calculation for relative performance feedback (cohort of 800)		
Student assistant	(60 hours per semester * €11.70)	€702
Postage	(2 letters * €0.48 * 800 students)	€768
Printing of letters	(2 letters * 2 pages * €0.12 * 800 students)	€384
Printing of letters 2nd language	(2 letters * 2 pages * €0.12 * 140 students)	€67.20
Envelopes	(2 letters * €0.02 * 800 students)	€32
Total cost per semester		€1,953.20
Cost per student per semester		€2.44

TABLE 3.A2: STUDY PROGRAMS, NUMBER OF STUDENTS, AND TREATMENT RATES

		Observations		Fraction in Treatment	
Study program	Faculty	Experiment I	Experiment II Replication	Experiment I	Experiment II Replication
Business Administration (BA)	BA	402	333	50.25%	50.15%
International Business (IB)	BA	63	59	49.21%	50.85%
Business Engineering[a] (BE)	BA	61	63	50.82%	50.79%
Mechanical Engineering (ME)	ME	235	298	50.21%	49.66%
Energy and Building Services Engineering (EBSE)	ME	51	44	49.02%	50.00%
N – Overall		812	797	50.12%	50.06%

Note: [a] BE is a joint degree program of the business and the tech faculty. During the first semesters most courses are related to business administration and economics. We therefore assign BE to the business faculty.

TABLE 3.A3: DESCRIPTION OF VARIABLES

Variable	Description
Treatment Variables	
Treatment	Random assignment to the treatment group.
Stratification Variables	
Study program	Indicators for study programs; for more information see Table 3.A2.
Credit strata	Indicating strata based on first-semester credit points.[a)]
Control Variables	
Age	Age in years at randomization.
Female	Indicator for being female.
HS degree Abitur	Indicator for a general track degree ("Abitur"); reference category includes vocational track degree ("Fachhochschulreife") and students who hold other degrees.
Time since HS degree	Time in years since high school graduation.
HS GPA	Final high school grade point average (1=best, 4=worst); missing values imputed.[a)]
GPA first semester	First semester grade point average (exam-level[b)]); (1=best, 4=worst); failed exams are not included in calculation. Missing values imputed.[a)]
Credits first semester	Number of credit points (exam-level[b)]) obtained in the first semester net of credits granted for an internship.[a)]
GPA at randomization	First semester grade point average provided to us by the university at the time of randomization (module-level[b)]); (1=best, 4=worst); only used in the randomization procedure.
Outcome Variables[c)]	
Credits	Credit points obtained in the respective semester net of credits granted for an internship.
Accumulated Credits	Total credit points accumulated until the end of the respective semester net of credits granted for an internship.
GPA	Grade point average at the end of the respective semester (1=best, 4=worst); failed exams are not included in calculation.
Dropout	Indicator for having dropped out of the study program before or in the respective semester.
Well-being	See Table 3.A7.

Note: [a)]For details see Appendix C. [b)]Exam-level: includes partly completed multiple-exam-modules (= passed sub-modules). Module-level: considers only fully completed modules. For more details see Appendix C. [c)]All outcome variables are measured on the exam-level.

TABLE 3.A4: DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES – ABOVE-AVERAGE STUDENTS

	Experiment I			Experiment II: Replication		
	(1) Control Mean (Std. Dev.)	(2) Treatment Coefficient (Robust SE)	(3) p-Value	(4) Control Mean (Std. Dev.)	(5) Treatment Coefficient (Robust SE)	(6) p-Value
Age	22.581	-0.412	0.175	22.085	-0.065	0.790
	(2.905)	(0.303)		(2.828)	(0.243)	
Female	0.394	-0.023	0.636	0.359	0.009	0.823
	(0.490)	(0.049)		(0.481)	(0.038)	
HS Degree Abitur	0.445	0.063	0.231	0.413	0.017	0.702
	(0.499)	(0.052)		(0.493)	(0.045)	
Time since HS Degree	1.426	-0.223	0.348	1.081	-0.035	0.819
	(2.433)	(0.237)		(1.699)	(0.152)	
HS GPA	2.445	-0.026	0.658	2.377	-0.037	0.435
	(0.515)	(0.060)		(0.581)	(0.048)	
% HS GPA Imputed[a)]	0.019	-0.008	0.489	0.004	-0.004	0.321
	(0.138)	(0.012)		(0.067)	(0.004)	
GPA 1st Semester	2.252	-0.088	0.190	2.383	-0.019	0.723
	(0.597)	(0.067)		(0.601)	(0.053)	
% GPA 1st Semester Imputed[a)]	0.006	-0.007	0.321	0.000	0.005	0.321
	(0.080)	(0.007)		(0.000)	(0.005)	
Credits 1st Semester	26.252	0.841	0.166	25.697	0.057	0.892
	(8.016)	(0.606)		(7.632)	(0.423)	
GPA at Randomization	2.286	-0.030	0.711	2.392	-0.017	0.753
	(0.711)	(0.081)		(0.624)	(0.054)	
% GPA at Randomization NA[a)]	0.071	-0.001	0.966	0.000	0.008	0.160
	(0.258)	(0.017)		(0.000)	(0.006)	
N	155	150		223	225	

Note: Columns (1) and (4) present the unadjusted control group means and standard deviations of the covariates. For details on the variables see Table 3.A3 and Appendix C. Columns (2) and (5) present the estimated coefficients of regressing the covariates on the treatment indicator using Equation 3.1. Columns (3) and (6) test the null hypothesis of no treatment effect. [a)]See Appendix C for details on the missing values and the imputation. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 3.A5: SURVEY QUESTIONS ON STUDENTS' EXPECTATIONS – EXPERIMENT II

First semester	*Assume that there are 100 students who have started studying at the same time and are enrolled in the same degree. If you were to rank all 100 students by their credit points (ECTS), such that rank 1 is the student with the highest number of credit points and 100 is the student with the lowest ECTS. In which position do you think you would be?*
Second semester	*What do you think? How many per cent of your fellow students will have achieved more credit points (ECTS) than you at the end of the current semester?*

Note: Questions provide the option to give no answer.

TABLE 3.A6: EFFECT OF FEEDBACK ON GPA AND DROPOUT – THIRD SEMESTER

	(1) Exp. I	(2) Exp. II	(3) Pooled	(4) Exp. I	(5) Exp. II	(6) Pooled
(a) GPA						
Treatment (T)	-0.012	-0.023	-0.016	-0.030	0.003	-0.017
	(0.023)	(0.021)	(0.016)	(0.032)	(0.039)	(0.024)
T*Above-average				0.047	-0.043	0.001
				(0.045)	(0.046)	(0.031)
T+T*Above-average				0.017	-0.040	-0.016
				(0.032)	(0.023)	(0.019)
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	771	746	1517	771	746	1517
Control Mean	2.59	2.68	2.64			
(Std. Dev.)	(0.60)	(0.58)	(0.59)			
(b) Dropout						
Treatment (T)	0.003	-0.000	0.001	0.012	0.010	0.010
	(0.022)	(0.025)	(0.017)	(0.032)	(0.048)	(0.027)
T*Above-average				-0.022	-0.018	-0.019
				(0.042)	(0.054)	(0.033)
T+T*Above-average				-0.010	-0.009	-0.009
				(0.027)	(0.024)	(0.018)
Strata	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	812	797	1609	812	797	1609
Control Mean	0.18	0.26	0.22			

Note: Above-average and not-above-average (the reference category) refers to the type of feedback students received in the second semester. *Outcome variables:* GPA by the end of the third semester (Panel a; passing grades only; highest passing grade is 1.0, lowest passing grade is 4.0), dropout during or before the third semester (Panel b); *strata:* credit strata FE, study program FE, and in the pooled estimations a cohort dummy and its interaction with the study program FE; *controls:* HS GPA, credits first semester, GPA first semester, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 3.A7: SURVEY QUESTIONS ON WELL-BEING – SECOND SEMESTER OF EXPERIMENT I AND II

Question	
1	*Now we would like to ask you about your overall satisfaction with your life: How satisfied are you currently with your life, all things considered?* [0 - completely dissatisfied; 10 - completely satisfied]
2	*During the last weeks, how often did you feel stressed out by our studies?* [never; rarely; sometimes; often; very often; always]
3	*Please think about the current semester. To what extent do you agree with the following statements about your studies: When thinking about my studies, I think of...*
3.1	*- not having enough time*
3.2	*- interesting lectures and curriculum*
3.3	*- pressure to perform well*
3.4	*- freedom in organizing my studies*
3.5	*- competition among students*
3.6	*- personal development and growth*
	[1 - completely disagree; 7 - completely agree]
4	*Now we would like to ask you about your overall satisfaction with your studies: How satisfied are you currently with your studies, all things considered?* [0 - completely dissatisfied; 10 - completely satisfied]
5	*More specifically: How satisfied are you so far with your performance in your studies?*
5.1	*- With my grades, I am...*
5.2	*- With my attained credit points (ECTS), I am...*
	[0 - completely dissatisfied; 10 - completely satisfied]
Estimation Outcomes	
	For the outcomes in Table 3.10 we ran exploratory factor analyses to see if there are variables that load on a common factor. Afterwards we standardized all survey questions within cohort and study program. In the cases where multiple questions captured the same latent construct, we constructed our outcomes by averaging across the corresponding questions:
Life Satisfaction	Question 1
Study Satisfaction	Questions 3.2, 3.6, and 4
Performance Satisfaction	Questions 5.1 and 5.2
Study Stress	Questions 2, 3.1, and 3.3

Note: All questions provide the option to give no answer. For this reason the number of observations in Table 3.10 can vary depending on the outcome of interest.

FIGURE 3.A1: FEEDBACK LETTER I – TREATMENT GROUP (EXAMPLE)

XXX
Postfach ▪ XXX XXX

Faculty of Business Administration

XXX XXX
XXX XXX
Access map at: XXX

Your reference:
Your message from:

Our reference:

Contact:
XXX XXX
xxx.xxx@xxx.de

Room: XXX

07/03/XXXX

Ms/Mr
XXX XXX
XXX XXX
XXX XXX

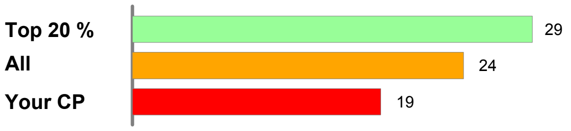
Feedback on your performance in the Bachelor's program International Business

Dear Ms/Mr XXX XXX,

the Department of Business Administration would like to assist you in the further organization and planning of your studies. To this end we provide you with feedback information about your current academic performance. So far you have earned **19 ECTS-Points (CP)** (as of 02/03/XXXX).

In order to allow you a better evaluation of your performance, the following figure compares you to students who are similar to you. Like you, they have been enrolled in International Business (Bachelor) at the XXX since the WS XXXX/XX.

Your CP in comparison to similar students



Great ☺☺
Good ☺
➜ **Currently below average**

Who are similar students?

All: Average CP of all students in your semester, who in XXXX earned their school leaving certificate.

Top 20 %: In the "All Students" group 20 % earned at least this amount of CP.

All similar students on average currently have 5 CP more than you.

Please also keep track of your grades when organizing and planning your studies. Your current grade point average is 2.55 (as of: 02/03/XXXX).

We wish you all the best for your studies and hope that you enjoy the time in XXX.

Yours sincerely

Prof. Dr. XXX XXX, Dean
Faculty of Business Administration

FIGURE 3.A2: FEEDBACK LETTER I – CONTROL GROUP (EXAMPLE)

XXX
Postfach ▪ XXX XXX

Ms/Mr
XXX XXX
XXX XXX
XXX XXX

Faculty of Business Administration

XXX XXX
XXX XXX
Access map at: XXX

Your reference:
Your message from:

Our reference:

Contact:
XXX XXX
xxx.xxx@xxx.de

Room: XXX

07/03/XXXX

Feedback on your performance in the Bachelor's program International Business

Dear Ms/Mr XXX XXX,

the Department of Business Administration would like to assist you in the further organization and planning of your studies. To this end we provide you with feedback information about your current academic performance. So far you have earned 23 ECTS-Points (CP), and your current grade point average is 3.43 (as of: 02/03/XXXX).

We wish you all the best for your studies and hope that you enjoy the time in XXX.

Yours sincerely

Prof. Dr. XXX XXX, Dean
Faculty of Business Administration

B Robustness of regression discontinuity designs

RDD at average

TABLE 3.B1: RD ESTIMATES AT AVERAGE – DIFFERENT POLYNOMIALS AND DISCONTINUITY SAMPLES, POOLED SAMPLE

(a) Treatment Group

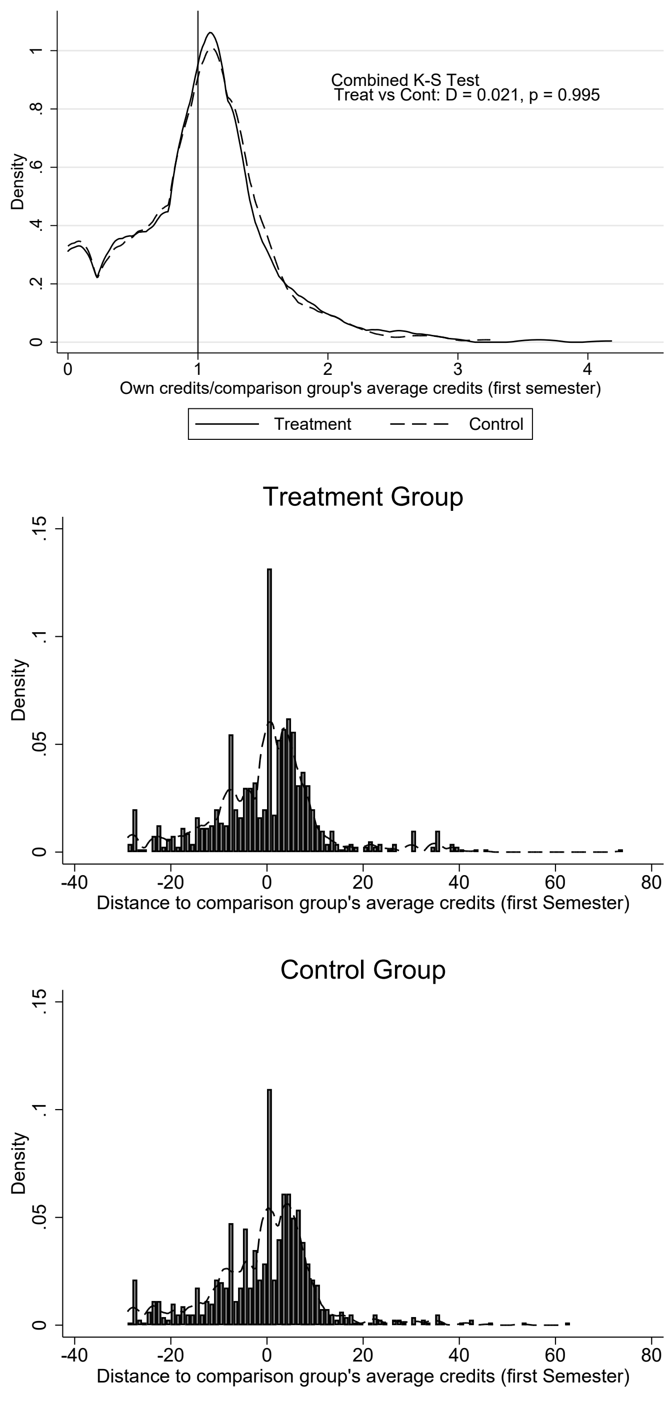
	(1)	(2)	(3)	(4)
	$0 < r < 2$	$0.25 < r < 1.75$	$0.5 < r < 1.5$	$0.75 < r < 1.25$
1st Order Polynomial	6.096***	5.492***	6.418***	8.570***
	(1.214)	(1.243)	(1.450)	(2.063)
2nd Order Polynomial	5.974***	7.379***	7.375***	10.517***
	(1.772)	(1.828)	(2.282)	(3.653)
3rd Order Polynomial	7.263***	6.138**	11.247***	-8.681
	(2.465)	(2.728)	(3.304)	(8.176)
4th Order Polynomial	7.504**	10.563***	12.308**	39.245**
	(3.223)	(3.555)	(5.550)	(18.682)
Study Program FE	Yes	Yes	Yes	Yes
N	700	656	533	352

(b) Control Group

	(1)	(2)	(3)	(4)
	$0 < r < 2$	$0.25 < r < 1.75$	$0.5 < r < 1.5$	$0.75 < r < 1.25$
1st Order Polynomial	0.189	-0.887	-1.246	-3.599
	(1.354)	(1.465)	(1.743)	(3.304)
2nd Order Polynomial	-2.499	-2.698	-5.206	0.214
	(2.023)	(2.502)	(3.431)	(5.650)
3rd Order Polynomial	-3.790	-4.535	1.476	-14.008
	(3.335)	(4.008)	(5.243)	(10.211)
4th Order Polynomial	-4.245	-2.311	-7.905	26.981
	(4.666)	(5.491)	(7.927)	(23.156)
Study Program FE	Yes	Yes	Yes	Yes
N	699	657	540	342

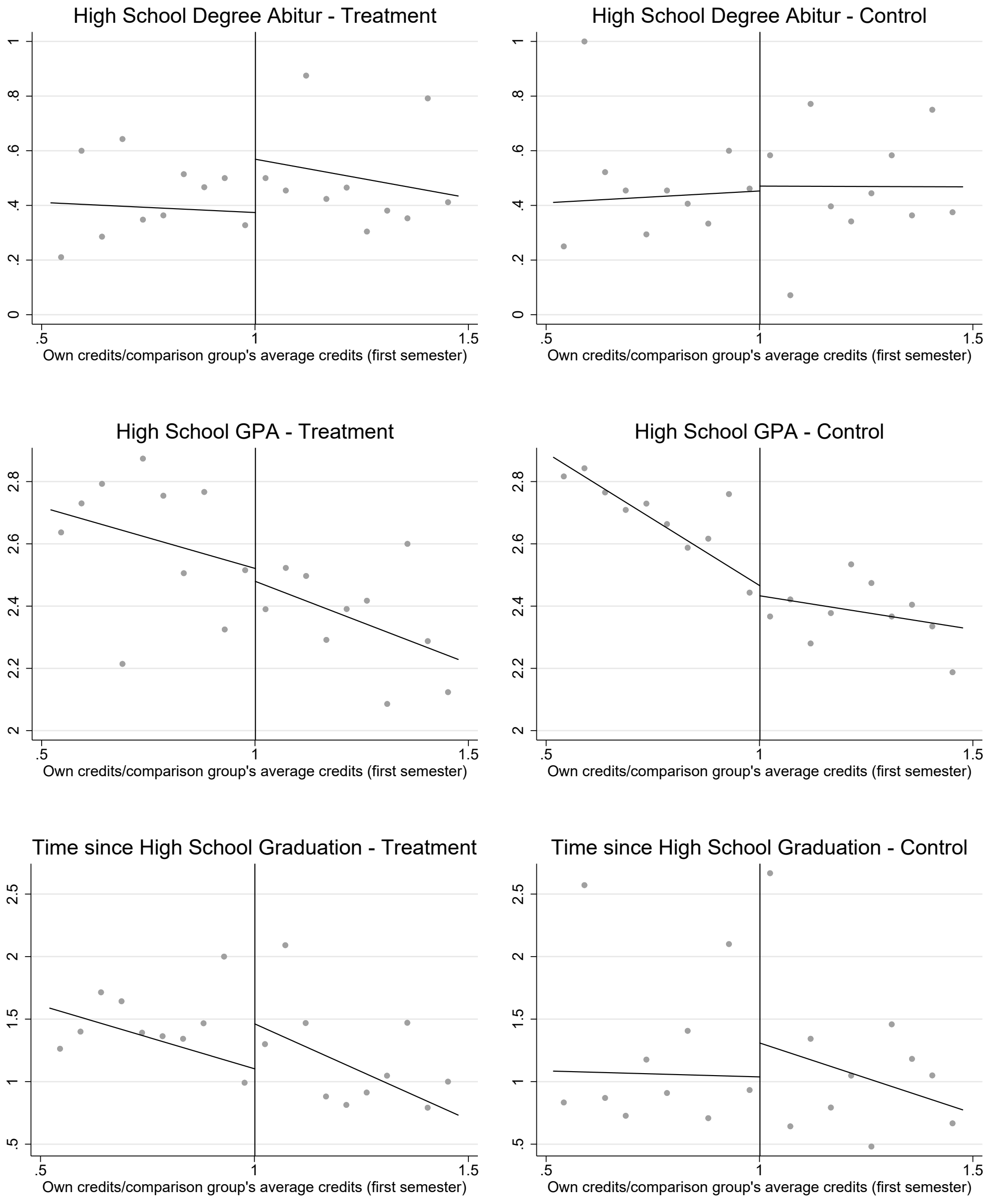
Note: Outcome variable: credits second semester; *study program FE:* study program FE, a cohort dummy, and the interaction of the cohort dummy with the study program FE; *running variable (r):* ratio of first semester credits as depicted in the feedback letter to the comparison group's average credits. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

FIGURE 3.B1: DISTRIBUTION OF THE RUNNING VARIABLE AT AVERAGE – POOLED SAMPLE



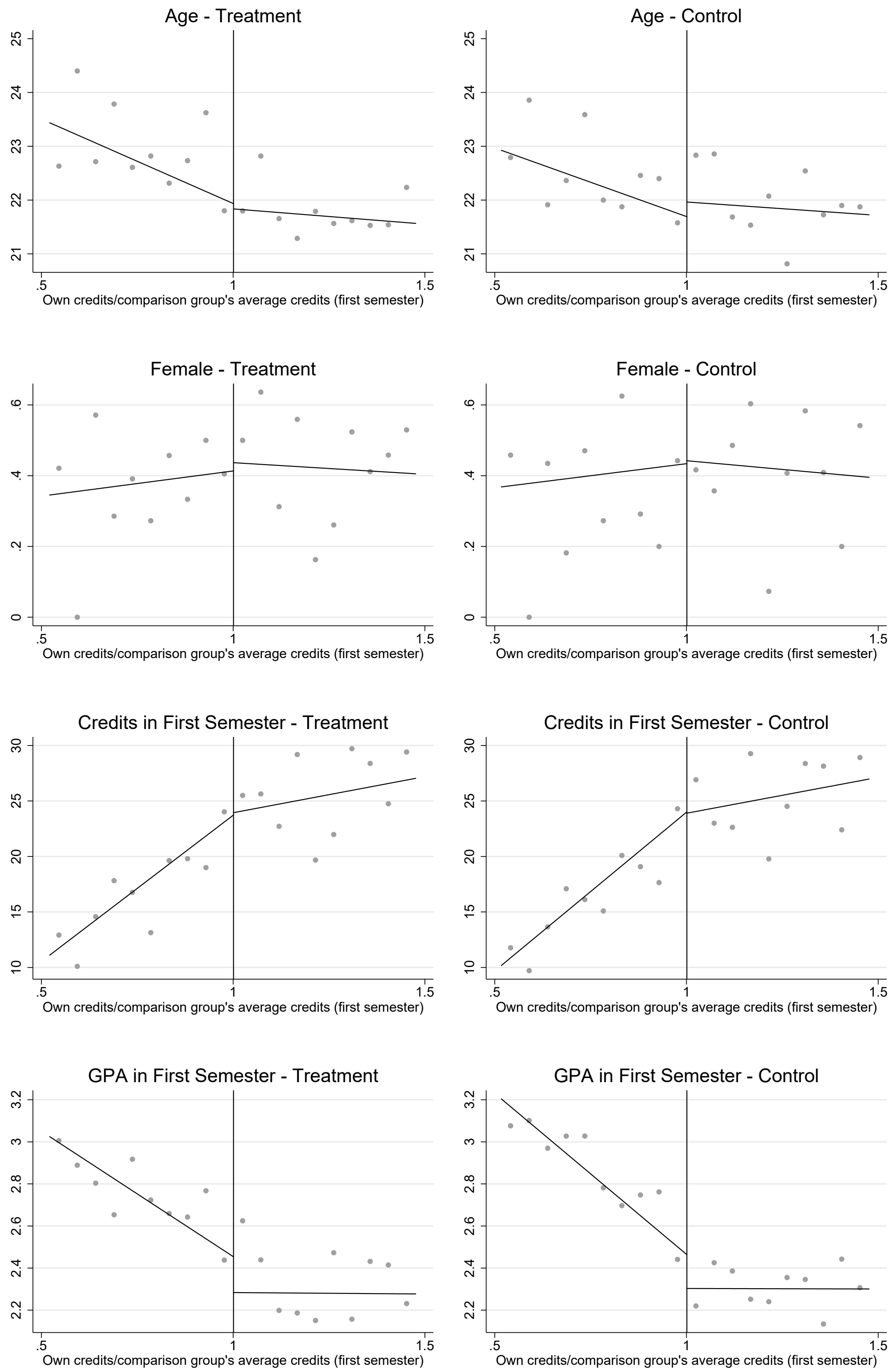
Note: The top panel shows the density of the running variable used for the RDD. Running variable is the ratio of first semester credits as depicted in the feedback letter to the comparison group's average credits. Observations with values lower or equal to 1 did not place above average and observations with values above 1 placed above average. The two bottom panels show the distribution of the distance to the comparison group's average in credit points. Observations with negative values or zero did not place above average and observations with positive values placed above average.

FIGURE 3.B2: RD PLOT AT AVERAGE FOR COVARIATES – FIRST ORDER POLYNOMIAL, POOLED SAMPLE



Note: Binned scatterplots using first order polynomials. Running variable is the ratio of first semester credits as depicted in the feedback letter to the comparison group's average credits. Observations on the left side of the cutoff did not place above average. Observations on the right side placed above average.

FIGURE 3.B3: RD PLOT AT AVERAGE FOR COVARIATES – FIRST ORDER POLYNOMIAL, POOLED SAMPLE (CONT.)



Note: Binned scatterplots using first order polynomials. Running variable is the ratio of first semester credits as depicted in the feedback letter to the comparison group's average credits. Observations on the left side of the cutoff did not place above average. Observations on the right side placed above average.

RDD at top 20%

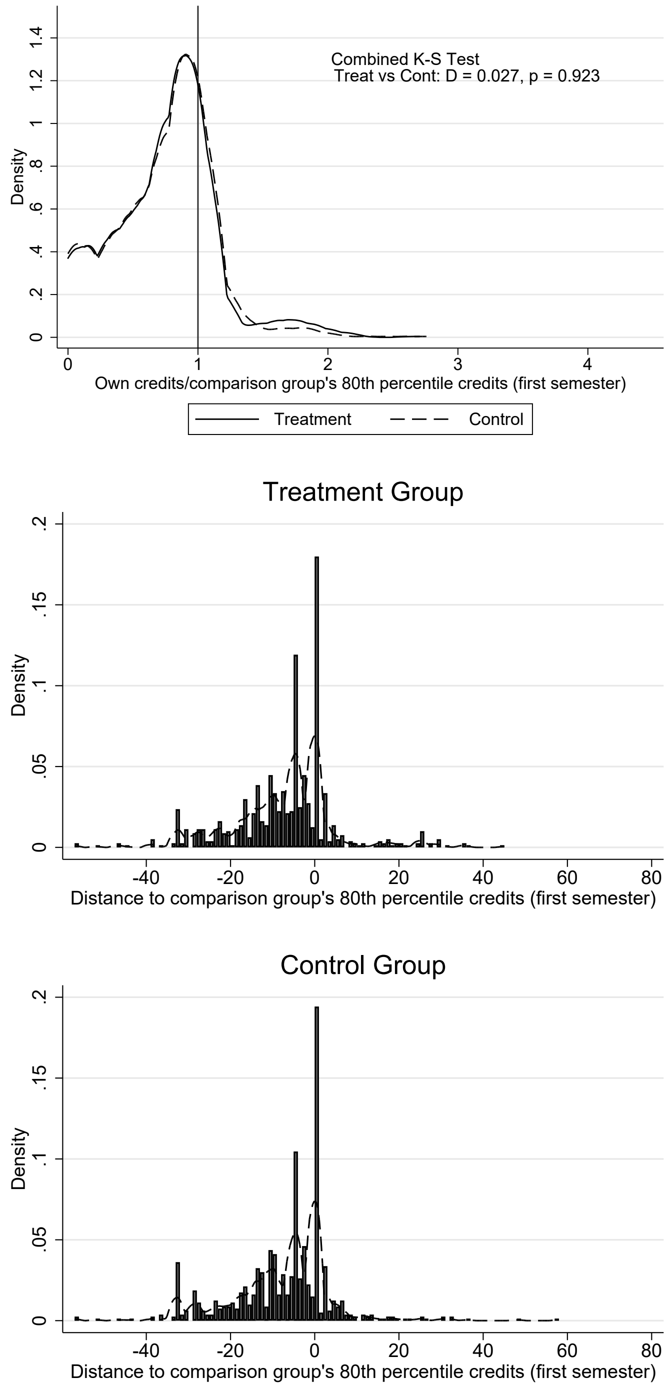
TABLE 3.B2: RD ESTIMATES AT 80TH PERCENTILE – DIFFERENT POLYNOMIALS AND DISCONTINUITY SAMPLES, POOLED SAMPLE

a) Treatment Group				
	(1)	(2)	(3)	(4)
	$0 < r < 2$	$0.25 < r < 1.75$	$0.5 < r < 1.5$	$0.75 < r < 1.25$
1st Order Polynomial	2.134	2.632	0.891	2.969
	(1.569)	(1.720)	(2.038)	(3.129)
2nd Order Polynomial	3.889	2.964	5.941	0.269
	(2.429)	(3.275)	(3.866)	(5.011)
3rd Order Polynomial	3.681	2.556	0.582	7.251
	(3.788)	(3.982)	(4.407)	(7.762)
4th Order Polynomial	0.148	1.990	-8.383	3.371
	(3.931)	(4.686)	(8.173)	(10.507)
Study Program FE	Yes	Yes	Yes	Yes
N	730	667	552	396

b) Control Group				
	(1)	(2)	(3)	(4)
	$0 < r < 2$	$0.25 < r < 1.75$	$0.5 < r < 1.5$	$0.75 < r < 1.25$
1st Order Polynomial	-0.365	0.621	3.256	0.943
	(1.604)	(1.894)	(2.113)	(2.327)
2nd Order Polynomial	1.786	1.657	-1.757	-2.875
	(2.547)	(2.776)	(3.011)	(3.808)
3rd Order Polynomial	1.701	-0.136	-1.709	-21.222***
	(3.457)	(4.250)	(4.159)	(7.934)
4th Order Polynomial	-4.984	-9.215*	-18.092***	-9.113
	(4.480)	(4.760)	(5.671)	(9.029)
Study Program FE	Yes	Yes	Yes	Yes
N	726	669	558	402

Note: Outcome variable: credits second semester; *study program FE:* study program FE, a cohort dummy, and the interaction of the cohort dummy with the study program FE; *running variable (r):* ratio of first semester credits as depicted in the feedback letter to the 80th percentile of credits in the comparison group. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

FIGURE 3.B4: DISTRIBUTION OF THE RUNNING VARIABLE AT AVERAGE – POOLED SAMPLE



Note: The top panel shows the density of the running variable used for the RDD. Running variable is theratio of first semester credits as depicted in the feedback letter to the 80th percentile of credits in the comparison group. Observations with values lower than 1 placed below the top 20% and observations with values above 1 placed in the top 20%. The two bottom panels show the distribution of the distance to the comparison group's 80th percentile in credit points. Observations with negative values placed below the top 20% and observations with zero or positive values placed in the top 20%.

C Data and methods appendix

The university provides student-level data on accumulated credit points (ACP), the cumulative grade point average (CGPA), individual exam performance, and demographic information. In this section, we describe how this data was used in the feedback letters and the randomization, how outcome variables and covariates are defined, and how we augment the administrative data with four online surveys.

Feedback data. As described in Section 3.2.1, we provide students with absolute and relative feedback on their ACP and their CGPA by sending postal letters twice per semester (see Figure 3.1). The information in the first and in the second letter was mostly identical but for some students changes occured (e.g., during the first treatment semester for 133 (21) students in Experiment I (II) the university updated the information on ACP).[29] These changes appear if (i) exam results were not yet available at the time the first letter was composed, (ii) grades were changed after students inspected their exams, or (iii) due to administrative problems at the university.[30] As a result, a small number of students received different types of relative feedback in the two letters: 15 (17) students in Experiment I (II) no longer had an above-average performance in the second letter although they did so in the first letter, and 10 (12) no longer had a below-average performance. We estimate heterogeneous treatment effects based on the relative feedback types of the second letter, as it provides the most accurate information and because it is more likely to be salient when students start to prepare for their exams.

The university awards credit points and grades on a module level. Modules can consist of a single exam or of several exams (sub-modules), all of which must be passed to complete the entire module. Module-level grades are based on the credit-weighted grades of the exams which make up a module. To compute the CGPA the university sums up the product of the grades and credit points of all modules and divides by the ACP.[31] Failing grades do not enter into the CGPA. It is important to note that the university only considers completed modules for the ACP and the CGPA.[32] We refer to the university's approach of accumulating performance measures as aggregation on the *module-level*.

Students can access their personal ACP and CGPA online on a website.[33] As mentioned before, we use the same variables to illustrate our feedback. Although information at the *exam-level* would have

[29]Updates during the semester also occur on a similar scale with respect to the CGPA. The changes in the ACP and CGPA do not necessarily coincide. The reasons for this are explained below.

[30] When the first letter of Experiment I was sent the university had not yet calculated the CGPA information for one of the smaller study programs (Business Engineering; N=61).

[31]For GPAs in the Business Administration program, the study regulations require the university to double weight each module scheduled after the first year.

[32] This procedure is in effect across all faculties for the CGPA, but not for the ACP. When calculating the ACP, the technical faculty also takes sub-modules into account, while the business faculty only counts completed modules.

[33]Importantly, in absence of the treatment, the university does not provide any information on the students' relative performance.

reflected their individual performance more accurately as it also includes partly completed multiple-exam-modules, we decided not to use it. The reason was that it would have led to conflicting numbers between the official information on the web and our letters, which could have caused questions and potentially also complaints from students.

The variable ACP is defined as zero and the variable CGPA contains a missing value if (i) students did not participate in an exam yet, (ii) students took exams but did not pass any of these, and (iii) students passed only sub-modules but did not yet complete a full module. For example, at the beginning of the second semester 64 (79) students had zero ACP in Experiment I (II) and 210 (89) students had missing values on the CGPA. In the feedback letters, the latter were replaced with an asterisk which refers to a footnote stating that "Due to technical reasons the grade point average is currently not available. Individual grades can be checked on [the online study platform of the university]". Regarding the ACP we printed the zero and no bar.

Randomization data. In both cohorts, randomization was carried out in the week before the second semester started using demographic information and the individual ACP and the CGPA.

We stratified on study program and ACP, and performed re-randomization (Morgan and Rubin 2012) based on CGPA, age, sex, high school grade, time since high school graduation, and (in Experiment II) type of high school degree. In Experiment I we defined five ACP strata for every study program ($ACP \leq 12$, $12 < ACP \leq 18$, $18 < ACP \leq 24$, $24 < ACP \leq 30$, $ACP > 30$). In Experiment II we defined ACP strata based on quantiles (Q); four ACP strata in the larger study programs BA and Mechanical Engineering ($ACP < Q_{0.25}$, $Q_{0.25} \leq ACP < Q_{0.5}$, $Q_{0.5} \leq ACP < Q_{0.75}$, $ACP \geq Q_{0.75}$) and two ACP strata in the other study programs ($ACP < Q_{0.5}$, $Q_{0.5} \leq ACP$). For the randomization in Experiment II, we filled missing values on the variables high school GPA (N=30) and CGPA (N=89) with a constant in order to avoid losing units in the randomization and to balance on the full sample.[34][35] Tables 3.1 and 3.A4 shows that missing data on both variables are balanced across the treatment and the control group.

Outcome variables and covariates. For the analysis in Section 3.3 we calculate credits and GPA based on semester-exam-level data. We use the following outcome variables: credit points per semester net of credits granted for internships, accumulated credits net of credits granted for internships, dropout, GPA (excluding failing grades), and survey variables on students' well-being.[36] In contrast to the ACP and the CGPA, the credit points and GPA are now measured on the exam-level,

[34] In Experiment I we balanced only for the study program Business Administration and only for observations without missing values.

[35] After the randomization, the university was able to provide us with information on the high school GPA of 15 of the 30 missing observations. To use high school GPA as a covariate, we thus only had to impute 15 missing values (see below).

[36] Internships are scheduled later in the study program (4th/5th semester). Some students are awarded these credits at the start of their studies because they completed an apprenticeship and have work experience. As we are interested in the effect of treatment on academic performance, we do not count these internship credits.

i.e., if students only partly completed a multiple-exam-module we still included the passed and failed sub-modules in our analyzes. Not only do these outcomes provide more accurate information on the students' performance in each semester, but using the ACP and the AGP as outcome variables could also result in an overstated treatment coefficient.[37]

In the regressions we include stratification fixed effects (study program dummies, ACP strata dummies, and a cohort dummy in pooled estimations and its interaction with the study programs), balancing variables (age, sex, high school grades[38], and time since high school graduation), and further control variables (type of HS degree, exam-level first semester credits) as covariates. To keep the number of observations constant across specifications we did not include the CGPA at randomization (210 (89) missing values in Experiment I (II)) in the vector of balancing variables. Instead, in the specifications using further control variables we complement the vector of ability controls by adding the individual GPA on the exam-level. The exam-level GPA still has missing values for students who attempted no exams or failed all exams they attempted (54 (66) in Experiment I (II) in the overall sample). We therefore predict the GPA of these students by running linear regressions of the first semester GPA on study program fixed effects, age dummies, gender, time since high school graduation, type of high school degree, and high school GPA to impute these missing values. The imputation allows us to keep the sample size constant across estimations.

Survey data. We also use data from four online surveys. They were conducted in the second half of the semesters, approximately at the time when we usually sent the second letter (see Figure 3.1). Three of the surveys were carried out after the treatment but in Experiment II we also conducted an additional survey prior to the treatment. The questionnaires included questions on outcome variables such as: how satisfied students are with their life, the degree to which they are satisfied with their study program, the degree to which they are satisfied with their performance, and how stressful they find studying. We only considered questions as potential outcomes of interest if they were asked the same way in the surveys of both experiments. Because some questions cover similar topics and to reduce the number of outcomes we ran exploratory factor analyses to see which questions load on a common factor. We then standardized all survey questions within cohorts and study programs and in the cases where multiple questions captured the same latent construct, we constructed our outcomes by averaging across the corresponding questions (see Table 3.A7 for the survey questions

[37]The upward bias occurs when a module consists of several exams which are taken in different semesters. To calculate the ACP and the CGPA the university records the credits and grades awarded for a module in the semester in which the last sub-module has been passed. Let's consider two sub-modules each worth five credits that constitute a composite module running over the first and second semester. Now compare two otherwise identical students – one in treatment and in the control group – both have already passed the first sub-module. If we assume that the feedback causes the treatment student to pass the second exam, the treatment effect in the cumulative data would be 10 credits. However, the actual performance difference between the two individuals in the treatment semester is only five credits.

[38]For some students, the university has no information on high school GPA. We therefore predict 11 (15) missing values on high school grades in Experiment I (II) from a linear regression of the HS GPA on study program fixed effects, age dummies, gender, time since high school graduation and type of high school degree.

and how they were aggregated to obtain the variables used in the estimations). Furthermore, in Experiment II we also gathered pre- and post-treatment information on students' beliefs about their relative performance (see Table 3.A5).

Chapter 4

Relative Performance Feedback and University Completion

*Joint work with Oliver Himmler and Robert Jäckle**

4.1 Introduction

During the past decades, in many developed countries the skill- and routine-biased technological change and the ongoing international division of labor have increased the proportion of jobs requiring a university degree (see, e.g., Dustmann et al. 2009, Goldin and Katz 2008, and Goos et al. 2014).[1] In the same period, the wage premium associated with college education has risen, leading to an increase in the returns to higher education (Oreopoulos and Petronijevic, 2013). As a result, enrollment rates have been growing: for example, in the U.S., college attendance among 18- to 24-year-olds rose by more than 50% since the 1970s, to a level of about 42%.[2] However, the transition towards a more skilled workforce is jeopardized by high rates of attrition, and those students who do graduate often exceed the nominal duration of their programs. In the U.S. and in other OECD countries, less than 40% of a cohort complete their bachelor's degree within the scheduled study time, and, even two or three years later, more than 30% still remain without a degree (OECD, 2019).[3]

*We gratefully acknowledge financial support from the German Federal Ministry of Education and Research under grant 01PX16003A, 01PX16003B, the Staedtler Stiftung, and administrative support from the TH Nürnberg.

[1] Gammarano (2018) provides a global overview.

[2] See https://nces.ed.gov/programs/digest/d20/tables/dt20_302.60.asp, retrieved on November 8, 2021. In other countries, such as Germany, the proportion of first-year students in a cohort has increased from 28.8% in 1997 to 57.0% in 2017; see https://www.datenportal.bmbf.de/portal/en/Table-2.5.73.html, retrieved on November 8, 2021.

[3] See Bound and Turner (2011) for a discussion of why collegiate attainment rates in the U.S. have stagnated for the last decades.

Faced with these challenges, the recent literature has increasingly focused on developing interventions to mitigate information deficiencies and behavioral biases that negatively affect academic achievement.[4] Compared to more traditional efforts, such as one-on-one coaching (Bettinger and Baker, 2014; Oreopoulos and Petronijevic, 2018), reducing class sizes (Bandiera et al., 2010; Kara et al., 2021), or comprehensive approaches that combine financial and academic assistance (see Dawson et al. 2020 for an overview), these interventions are often easier to scale and have an advantageous cost-benefit ratio. They are thus considered to be attractive policy tools (Benartzi et al., 2017).

Measures such as "low-touch" assistance and information provision (Castleman and Page, 2016), goal setting (Chase et al., 2013; Clark et al., 2020; Morisano et al., 2010), commitment devices (Ariely and Wertenbroch, 2002; Himmler et al., 2019; Patterson, 2018), exam sign-ups by default (Behlen et al., 2020), or social comparison (Dobrescu et al., 2021; Tran and Zeckhauser, 2012) have been shown to positively affect student performance in the short-run – i.e., in a single course or over the period of a semester or two.[5] However, studies that examine the impact on long-term outcomes like graduation rates, time to degree or final grades are still sparse – preventing more comprehensive comparisons with traditional measures. To our knowledge, only Azmat et al. (2019) and the study in Chapter 5 follow students until the completion of their program.[6]

A simple and promising tool for universities to approach all students is relative performance feedback. Providing such information is inexpensive and the necessary data should be readily available at most academic institutions. For students, the performance of others can be an important benchmark against which they can compare their achievements and assess their abilities.[7] The effects of relative feedback have been mainly studied in single courses, demonstrating its potential to enhance academic performance in the short run, but it is not clear whether it is also effective in the long run.[8] So far, the only long-term study

[4] For general overviews on the behavioral economics of education see, e.g., Damgaard and Nielsen (2018), Escueta et al. (2020), Koch et al. (2015), Lavecchia et al. (2016), Leaver (2016).

[5] There is also an important literature on interventions that target behavioral barriers during the college admission and enrollment process (see French and Oreopoulos 2017 for a review).

[6] The first paper studies a relative feedback intervention and finds null effects on long-term academic success. The latter shows that commitment devices can raise completion rates and decrease time to degree.

[7] Theory suggests that there can be positive effects of social comparisons on performance, for example, when individuals have competitive preferences (Azmat and Iriberri, 2010; Dobrescu et al., 2021), when the behavior of similar others conveys a social norm (Suls and Wheeler, 2000), or when a comparison with others changes the beliefs of individuals about their abilities (Azmat and Iriberri, 2010; Dobrescu et al., 2021; Ertac, 2005).

[8] Several short-run studies provide relative feedback on intermediate performance in a single course, i.e., on the performance in mid-term exams or online assignments, and evaluate effects on grades. This approach can improve performance in the final exam (Kajitani et al., 2020; Tran and Zeckhauser, 2012) and generate positive grade spillovers to other courses (Dobrescu et al., 2021). Chen et al. (2021) show that leaderboard

on relative feedback was conducted by Azmat et al. (2019). Over a period of four years, they provide repeated online-feedback about the students' decile rank in the grade point average (GPA) distribution and find no effects on final grades or university completion.

In this chapter, we study the long-term effects of providing relative feedback on accumulated credit points. At a German university of applied sciences, we conduct two field experiments with more than 1,600 students from consecutive cohorts. By sending them postal letters twice per semester, students are informed about their current academic progress in terms of accumulated credits.[9] Control group members receive no further information, while students in the treatment group additionally get feedback about how their performance compares to the average and the student ranked at the 80th percentile of their cohort. As is common in the social norms literature, the descriptive information is supplemented with approving normative cues for students with an at least average performance.

In the short-term, the intervention led to an update in beliefs and increased the obtained second semester credits among students whom the relative feedback informed about an above-average first semester performance (see Chapter 3). Importantly, a regression discontinuity design provided evidence that this "above-average effect" is not driven by, for example, higher ability and motivation, a better learning technology, or other individual characteristics of students above the average, but by the type of the feedback itself. The results reported here follow these students for five years, at which point more than 90% have either abandoned or completed their degrees.

The long-term analysis yields the following results: First, in the overall sample, treated students graduate significantly earlier than controls, they have better grades, and there is no difference in dropout rates. More specifically, the eighth- and ninth-semester graduation rates of treated students are significantly increased by about 4 pp compared to controls. After five years, about one in ten students in the treatment group has graduated one semester earlier than those in the control group. Students do not buy these performance gains with worse grades. On the contrary: after five years, the GPA of treated students has improved by about 0.064 standard deviations.

groups can improve outcomes in an online-assignment and the final exam grades of low performers. There is also a literature on the effects of relative performance feedback in primary and secondary education that mainly finds positive effects; for an overview and a review of the general literature on relative performance feedback see Villeval (2020).

[9] Europe-wide, universities use a standardized point system (European Credit Transfer and Accumulation System, ECTS), under which a full-time academic year consists of 60 credits. To complete a bachelor's degree at a university of applied sciences students usually need to earn 210 ECTS. See also `https://ec.europa.eu/education/resources-and-tools/european-credit-transfer-and-accumulation-system-ects_en`, retrieved on August 20, 2021.

Second, based on the short run results, we expected to find heterogeneous effects in the long run, too. Consequently, we use the pre-treatment credit point distribution and divide the sample into above- and not-above-average students. Our estimates show that the positive main effects on academic achievement are driven by students who placed above the average pre-treatment.

Third, we explore the heterogeneous effects of the relative feedback in more detail. Such analyses are desirable for education policies. They allow to precisely determine the beneficiaries of an intervention so that it can be provided only to them. We use pre-treatment performance and background characteristics of control group members to predict second semester performance across cohorts. For treated students, this can be thought of as the counterfactual performance if they had not been treated with the feedback. Among the subgroups of students with an above- and not-above-average first semester performance, we split students into a group of students with high predicted performance and a group with low predicted performance. The result are four "predicted performance strata" (the upper and lower half of above- and not-above-average students). We find that the overall effect is composed of positive effects for individuals in the lower 50% of those above the average (Stratum 3) and negative effects in the upper half of not-above-average students (Stratum 2). Students in Strata 1 and 4 show no response.

We discuss possible explanations for the heterogeneous results: (i) Regarding Strata 2 and 3, using data from three post-treatment surveys, we find tentative evidence that these students also update beliefs about their relative performance. Together, the effects on beliefs and performance in the two subgroups are broadly consistent with the predictions of the "self-perception theory" (Azmat and Iriberri, 2010; Dobrescu et al., 2021; Ertac, 2005). According to the theory, relative feedback can lead to positive (negative) updates of one's beliefs about ability, which – assuming that ability and effort are complements – results in higher (lower) effort and performance. (ii) The null effects in Stratum 1 may be explained by "floor effects". Applying the predictions of "self-perception theory" to Stratum 1 means that low-performing students either already know that their achievements are weak, or that it is simply not possible to fall further behind because of the already very low number of credits in the control group. (iii) For the high performers in our sample (Stratum 4) the European Credit Transfer and Accumulation System, which advises students to collect 30 credits points per semester, may create a reference point that hardly anyone wants to exceed ("ceiling effects").

The main contribution of this chapter is to provide robust evidence for an easily scalable, low-touch policy tool that can be applied to increase long-term academic success. In the field of higher education, our results thus allow for the first time to compare the long-term impact of an effective behavioral intervention with widely used traditional measures, such as grant aid. For example, in a meta-analysis including 43 studies, Nguyen et al. (2019) find that an additional $1,000 of grant aid increases on-time graduation, measured as completing the degree during the nominal study time or up to one quarter later (i.e., the 100% to 125% degree completion), by 1.8 pp.[10] Our relative feedback intervention increases the eight-semester graduation rate, i.e., the 114% degree completion, by about 3.5 to 4.3 pp, but at a much lower cost of €2.5 per student and semester.[11]

Beyond the context of higher education, this chapter provides two further contributions: First, there is generally little evidence on behavioral interventions that alter individual behavior in the long run: Beshears and Kosowsky (2020) analyze 174 studies that estimate the effects of behavioral interventions and find that only 21% collect follow up data at least once or attempt to measure the cumulative effect of a series of nudges. In their meta-analysis, DellaVigna and Linos (2020) find that the average intervention and data collection time among behavioral interventions published in academic journals is 6.7 months ($\sigma = 7.1$), and 4.5 months ($\sigma = 2.0$) in trials that were run by two large U.S. nudge units. They also report that the typical outcome variable is rather short-run: on average, the nudge units record outcome behavior 60.2 ($\sigma = 74.5$) days after conducting the intervention and the average time of measuring behavior after treatment in studies published in academic journals is after 68.7 days ($\sigma = 91.7$). The total time span of our intervention and collection of outcome data is 5 years – within this duration, 93% of all students who started their bachelors' program either dropped out or completed their degree. The time horizon of our intervention is thus beyond almost all of the studies considered in the above meta-analyses.

Finally, our study contributes to the literature showing that relative feedback and nudges in education often produce heterogeneous effects (see Villeval 2020 and Damgaard and Nielsen 2018 for overviews). By assessing heterogeneous effects along (counterfactual) out-of-sample predictions of outcomes, we apply an approach that is similar to risk stratification in medical research (Kent and Hayward 2007; Kent et al. 2018). The method provides an alternative to the often more "data-hungry" machine learning algorithms developed in recent years (see, e.g., Athey and Imbens 2017a; Athey and Imbens 2017b; Athey and Imbens 2019 for overviews).[12] Heterogeneity analyses like the one used in this

[10] Besides positively affecting on-time graduation, grant aid also increases college enrollment, persistence in college, and delayed graduation (Deming and Dynarski, 2010; Nguyen et al., 2019). Although our intervention primarily affects study duration rather than the overall rate of degree attainment, the cost-benefit ratio still compares favorably with that of more traditional measures.

[11] Cost calculations are included in Chapter 3.

[12] Our approach also shares similarities with "endogenous stratification", which makes use of in-sample outcome predictions (Abadie et al., 2018).

chapter are increasingly applied (e.g., Dynarski et al. 2021), and can be particularly useful when policymakers and researchers are concerned with identifying and targeting individuals who would have performed poorly in the absence of treatment or when interventions are presumed to affect behavior only among individuals who have sufficient scope for improvement.

The chapter proceeds as follows. In the next section, we introduce the institutional background, the design of the relative feedback intervention, and the data used for the analysis. Section 4.3 presents the empirical approach and the main results. In Section 4.4 we further investigate the heterogeneous effects and discuss potential mechanisms. Section 4.5 concludes.

4.2 Institutional background and research design

4.2.1 Institutional background

We conducted the relative feedback intervention at a large public university of applied sciences in Germany.[13] Our subjects were students of two consecutive cohorts who enrolled in one of five bachelor's degree programs at the departments of Business Administration and Mechanical Engineering. The context is representative for large parts of the German higher education system: for example, in the fall semester of 2019, around 40% of first-year students enrolled at a university of applied sciences and 12% started studying Business Administration or Mechanical Engineering (Statistisches Bundesamt, 2020b). The programs at our university are not selective: the average high school GPA in the control group of our sample is 2.56 (see Table 4.1), while it was 2.44 among all German high school graduates in 2014 and 2015, the years in which the students of our field experiments started studying.[14]

The study programs are structured in accordance with the European Credit Transfer and Accumulation System (ECTS).[15] The scheduled study duration is seven semesters and students are supposed to earn a total of 210 credit points in order to complete their degree. The programs follow a fairly clear structure: in the first and second year, a range of compulsory courses provides students with the fundamentals of their programs. Subsequently, all students are required to complete a mandatory internship, which lasts 20 weeks and is

[13]The university offers more than 20 bachelor's degree programs and various master's programs at 13 faculties. The university has a student population of more than 13,000 students and each year, more than 2,700 full-time, undergraduate students enroll.

[14]See `https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html`, retrieved on November 8, 2021. In the German system 1.0 is the best and 4.0 the worst final high school GPA.

[15]See `https://ec.europa.eu/education/resources-and-tools`, retrieved on November 8, 2021.

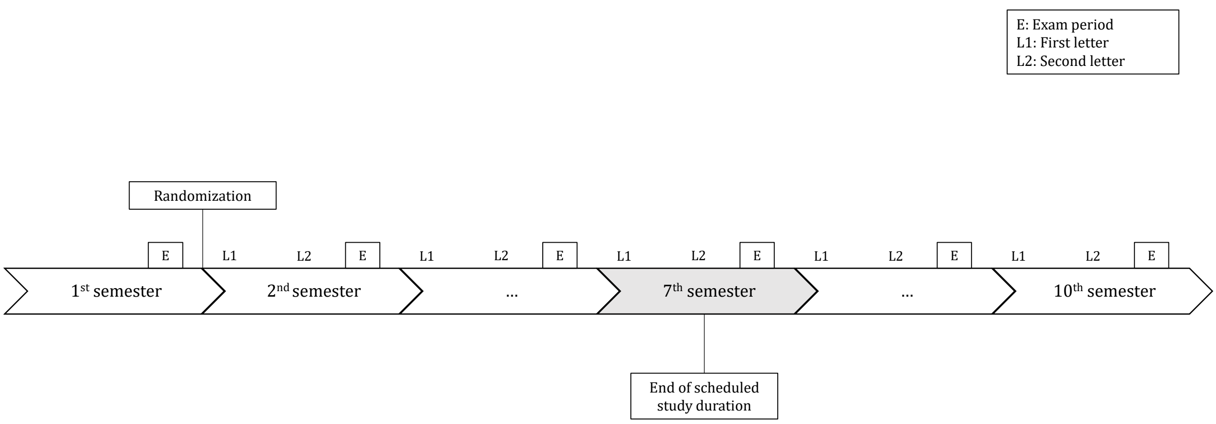
scheduled for the fourth or fifth semester. In the final year, students can typically choose from several electives which allows them to set individual majors.

To track their study progress, students can at all times access a web portal of the university that provides them with information on their current GPA and the number of credits they have earned so far. Importantly, the web portal only shows information on one's own performance without any comparison to the performance of other students.

4.2.2 Intervention

Against this background, we designed a relative performance feedback intervention aimed at raising students' academic success. To test its effectiveness, we ran two field experiments which share the same design. Students from the fall semester cohorts 2014 and 2015, who were still enrolled in their degree programs at the end of the first semester, were assigned to two groups: absolute (= control) and relative performance feedback (= treatment). Below, we summarize the implementation and the design of the intervention. Further details can be retrieved from Chapter 3.

FIGURE 4.1: TIMELINE OF INTERVENTION



Randomization. After the university provided us with information on the students' first semester performance (see Figure 4.1 for a timeline of the intervention), we randomized our subjects into two groups using stratification and re-randomization (Morgan and Rubin, 2012). Students were allocated to strata based on their study programs and their pre-treatment credit points. Within those strata, we carried out up to 500 random draws, keeping the allocation with the best balancing properties with respect to age, gender, high school

GPA, time since high school graduation, first semester GPA, and – in the second cohort – type of high school degree.

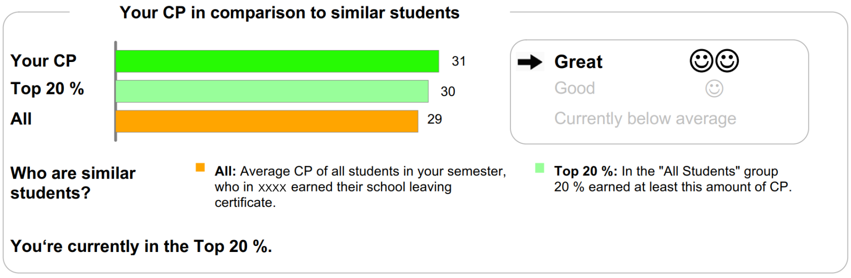
Feedback letters. Starting in the second semester and continuing throughout the observation period or until the end of their studies, both control and treatment group students received two unannounced postal letters per semester. The envelopes bore the official seal of the university and the letters were signed by the deans of the respective faculty. For both groups, the letters included the current GPA and the credits that students had earned so far – i.e., the same information that was also available via the web portal of the university. The treatment letters additionally contained a graphical illustration that provided students with relative feedback with respect to their credit points. Below, we explain the relative feedback in more detail.

In each semester, the first letter was sent in the week before the semester started to all students who re-enrolled for the upcoming term. About a month before the exam period, we sent the second letter. Its purpose was to keep the relative performance feedback salient at the time students usually start preparing for their exams. Except for the introductory paragraph, the design of the first and second letter was identical. In some cases, the information on students' academic progress had to be updated, e.g., because grades or credits from some courses were not yet available when the first letter was prepared. When determining students' placement in the first semester performance distribution based on the content of the feedback letters for our heterogeneity analysis, we rely on the information provided in the second letter. Since there are no midterms or other exams between the two letters, any differences between the first and second letter should not be affected by the treatment assignment.

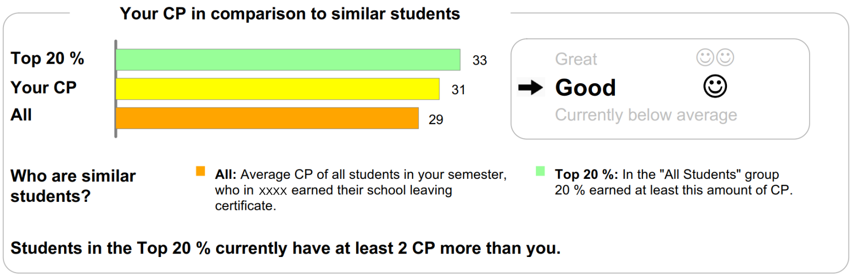
Relative feedback design. Figure 4.2 shows examples of the relative performance feedback included in the treatment letters. The design was inspired by the social comparison approaches of Schultz et al. (2007), Allcott (2011), and Allcott and Rogers (2014). A bar chart enabled students to compare their accumulated credit points to the "Top 20%" of their cohort – i.e., the performance of the student(s) on the 80th percentile – and the "Average". In the first cohort, the average was defined as the number of credits obtained by the median student(s), and in the second cohort, it was defined as the arithmetic mean. We came up with the little design tweak after analyzing the first experiment at the end of the second semester. Since the median of the performance distribution was located to the right of the arithmetic mean, the modification allowed more students to receive information about an above-average performance.[16] Importantly, however, modifying the average from median to arithmetic mean left the general design of the intervention unaltered.

FIGURE 4.2: RELATIVE PERFORMANCE FEEDBACK – TREATMENT GROUP (EXAMPLES FROM SECOND SEMESTER)

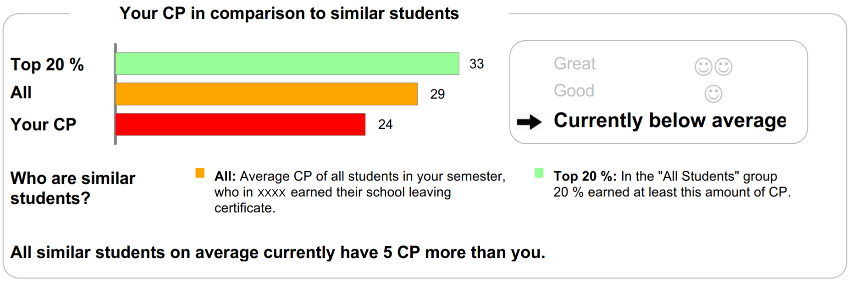
(A) Approving Frame *Great*



(B) Approving Frame *Good*



(C) No Approving Frame



The performance comparisons presented in the letters did not refer to all students in a degree program but were made within subgroups: In the three smaller programs, subjects were compared with others who earned their school leaving certificate in the same year. In the two larger programs, students were compared with others who obtained the same school leaving degree in the same year. These smaller comparison groups served two purposes: First, they aimed to increase perceived similarity among students. Second, they were

[16] E.g., in the first letter of the second semester, 56.2% of students in the second cohort received feedback of being better than average, while this was only the case for 37.5% of students in the first cohort.

intended to reduce spillovers that could potentially arise from sharing the feedback information.[17]

Because students might perceive the "Average" and the "Top 20%" bar as descriptive norms, and to prevent potential "boomerang" effects, i.e., negative effects for individuals who placed above the norms, we followed the literature on social comparison and augmented the descriptive information with normative frames (see, e.g., Schultz et al. 2007 and Allcott 2011). If the performance was at least average, the relative feedback included an approving normative frame. It categorized performance as *good* (plus one "smiley" emoticon) for students at or above the average, and *great* (plus two "smiley" emoticons) for students in the top 20%. For students below average, no approving frame was displayed. Instead, they received the statement "currently below average" (and no emoticon).

4.2.3 Data and descriptives

For most of the analyses, we employ student-level data provided by the university on students' background characteristics, i.e., age, gender, type of high school degree, time since the high school degree was obtained in years, and the high school GPA as well as current information on students' performance and progress towards their degree.[18]

The following outcome dimensions are considered: First, in order to assess whether the intervention raises academic attainment, we study effects on graduation rates, dropout, and time to degree (in semesters). Second, we investigate effects on students' accumulated credit points and credits obtained per semester. Because relative feedback was provided on credits, and since increased credit accumulation should translate into earlier graduation, we expect effects to materialize on these dimensions. However, students may buy performance gains on these outcomes with losses on another performance dimension, and we therefore also provide evidence regarding the effects on students' GPA.[19]

[17] Research assistants that were enrolled in the study programs at the time provided anecdotal evidence that sharing of this information was not common. E.g., students did not share pictures of the feedback graphs in social media. Moreover, our analysis of students' post-treatment beliefs about relative performance in Section 4.4.3 suggests that students' knowledge about their relative performance differs between treatment and control.

[18] Both trials began before the Corona pandemic. Throughout the first experiment (summer term 2015 until summer term 2019), all lectures and exams were held on campus and students could attend in person. Data collection for the second experiment ended in the summer semester 2020, the first semester of the pandemic. In this semester, all lectures took place online, and the exams were held partly online and partly face-to-face.

[19] To accurately reflect students' progress towards their degrees, we measure their performance (GPA and (accumulated) credits) on the module-level. Modules usually consist of one or sometimes also multiple exams that are spread across more than one semester. Students only receive credits if they pass the entire module. The university website and the feedback letters both provide information based on the module-level. Chapter 3 includes a detailed discussion of the differences between exam- and module-level performance information and why the exam-level data was used for the analyses of the short-term effects.

TABLE 4.1: DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES

	(1) Control Mean (Std. Dev.)	(2) Treatment Coefficient (Robust SE)	(3) p-Value
Age	22.466	−0.002	0.991
	(3.230)	(0.152)	
Female	0.370	−0.000	0.999
	(0.483)	(0.021)	
High school degree Abitur	0.415	0.004	0.875
	(0.493)	(0.024)	
Time since high school degree	1.257	−0.050	0.632
	(2.230)	(0.105)	
High school GPA[a)]	2.561	−0.027	0.280
	(0.593)	(0.025)	
First semester credits	19.219	0.073	0.800
	(12.039)	(0.289)	
First semester dropout rate	0.024	0.015*	0.071
	(0.152)	(0.008)	
First semester GPA	2.546	−0.038	0.233
	(0.659)	(0.032)	
First semester GPA N/A	0.096	−0.006	0.601
	(0.295)	(0.012)	
Imputed 1st semester GPA[b)]	2.593	−0.039	0.182
	(0.653)	(0.029)	
N	803	806	

Notes: Column (1) presents the unadjusted control group means and standard deviations of the covariates. Column (2) present the estimated coefficients of regressing the covariates on the treatment indicator controlling for credit strata FE, study program FE, and a cohort dummy and its interaction with the study program FE. Column (3) tests the null hypothesis of no treatment effect. *a*) For 13 students in the control and 13 students in the treatment group the university had no information on high school GPA. Within each cohort, we impute those values based on a linear regression of the high school GPA on age, a female dummy, time since high school graduation, a high school degree Abitur dummy as well as study program dummies and their interaction with the other variables. *b*) To keep all observations in the sample when using the first semester GPA as a control variable, we impute values for 77 (72) control (treatment) students, who obtained no grade in the first semester. Within each cohort, we impute the values based on a linear regression of the first semester GPA on first semester credits, high school GPA, age, a female dummy, time since high school graduation, a high school degree Abitur dummy as well as study program dummies and their interaction with the other variables. High school degree Abitur refers to the German general track degree. It is one of the two main secondary school degrees in the tracked school system in Germany that qualifies students to study at a university of applied sciences; the second being the vocational track degree (Fachhochschulreife). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

Table 4.1 presents descriptive statistics and balancing properties for students' background characteristics and measures of baseline performance. It provides evidence that our sample is well balanced. In the first semester, but also later, the GPA is missing if students had not yet passed a course. To keep all observations in the sample when using the first semester GPA as a covariate, we impute values for 149 students, who obtained no grade in the first semester.[20] We also report the imputed GPA in Table 4.1.[21] Similarly, the university did not have information on the high school GPA of 13 (13) treatment (control) students which we also imputed.[22]

4.3 Main results

We now discuss the main results of our feedback intervention. First, in Section 4.3.1 we describe the empirical approach. In Section 4.3.2, we present the main effects and show that, overall, treated students graduate significantly earlier, have better grades, and that there is no difference in dropout rates. Finally, in Section 4.3.3, we examine heterogeneous effects based on the feedback design and find that the results are driven by students who were performing above-average before treatment.

4.3.1 Empirical approach

Since the two experiments share the same design, we pool observations from both cohorts ($N = 1,609$) to increase the power of the statistical analyses.[23] We estimate ITT effects using the following OLS specification:[24]

$$Y_i^k = \alpha_0 + \alpha_1 Treatment_i + \mathbf{s_i} \boldsymbol{\alpha_2} + \varepsilon_i, \tag{4.1}$$

where Y_i^k denotes the outcome measure k for individual i. $Treatment_i$ is an indicator for being randomized into the treatment group. The vector $\mathbf{s_i}$ controls for stratified randomization by including strata fixed effects. Strata are defined by credits obtained in the first semester, study program dummies, a cohort dummy, and the interaction of the latter with the other strata variables. In additional specifications, we add the vector $\mathbf{x_i}$, which includes information on baseline performance, i.e. credits earned in the first semester, the

[20] We do not impute missing values when considering GPA as an outcome. Instead, we estimate effects among students for whom a GPA is observed and employ inverse probability weighting (see Section 4.3.2).

[21] Within each cohort, we impute missing values of the first semester GPA based on a linear regression on first semester credits, high school GPA, age, a female dummy, time since high school graduation, a high school degree dummy as well as study program fixed effects, and the interaction of those with the other variables.

[22] Within each cohort, we calculated those values based on a linear regression of the high school GPA on age, the female dummy, time since high school graduation, a high school degree dummy as well as study program dummies, and the interaction of the study program dummies with the other variables.

[23] In the short-run, the same pattern of results was found across both experiments (see Chapter 3).

[24] We estimate ITT effects, because we do not have information on whether students open and read the feedback letters.

first semester GPA (with missing values imputed, see Section 4.2.3), a dummy that indicates if the first semester GPA was imputed, and information on first semester dropout as well as further background characteristics (high school GPA, age at randomization, an indicator for being female, time since high school graduation, and an indicator for the type of high school degree).

We also examine heterogeneous effects for students who placed above average ($N = 753$) in the first semester performance distribution and students who did not ($N = 856$). First semester performance is used because later rank changes could endogenously depend on the intervention. We estimate the following specification:

$$Y_i^k = \alpha_0 + \alpha_1 Treatment_i + \alpha_2 A_i + \alpha_{12} Treatment_i A_i + \mathbf{s_i} \boldsymbol{\alpha_3} + \varepsilon_i, \tag{4.2}$$

where A_i indicates if a student placed above average and all other parameters are defined as before.

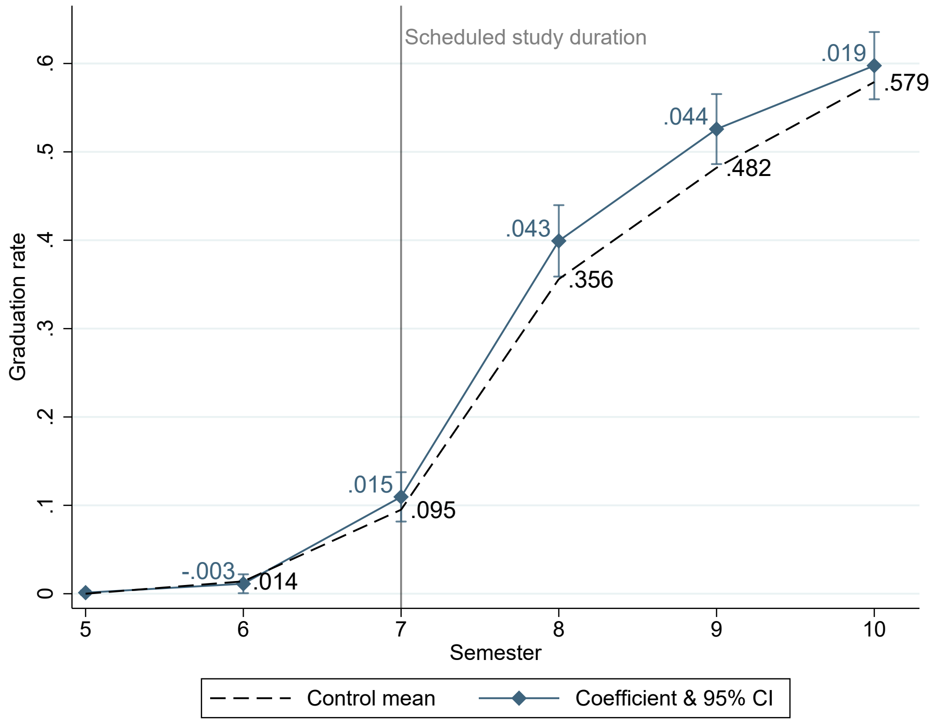
We focus on above- and not-above-average students due to the short-term results discussed in Chapter 3: In the short run, the relative feedback intervention increased performance among treated subjects who performed above the first semester average. A regression discontinuity design based on the sharp cutoff at the average provided evidence that the heterogeneous behavioral responses were not caused by differences in the underlying characteristics of those who placed above or not above average, e.g., ability or motivation. Instead, the results indicated that the information about being above average caused the effects. Further, the short-run findings did not indicate that the normative frames included in the feedback mattered for the effects on performance. Regarding the long-term results, we therefore hypothesize that the differential effects found in the short-term will persist in the long run.

4.3.2 Main effects

Graduation and dropout. Since graduation is the natural endpoint of a study program, we are primarily interested in studying whether the relative feedback intervention is successful in raising degree attainment. To this end, in Figure 4.3 and the top panel of Table 4.2 (Columns 1 to 8), we first report effects on graduation rates.

Indicating substantial room for improvement, our results show that only very few students manage to earn their degree within the scheduled study duration of seven semesters. In the control group, 9.5% graduated at this stage; in the treatment group the figure is slightly higher, although the effect is not estimated precisely (1.5 pp, $p = 0.307$). In the next two semesters (eighth and ninth), the majority of students in the control group complete their

FIGURE 4.3: TREATMENT EFFECTS ON GRADUATION RATES



Notes: The dashed line depicts the unadjusted mean. Coefficients are from regressions that are estimated separately for each semester and control for *strata*. Confidence intervals are based on robust standard errors. *Outcome variable:* indicates if a student graduated with a degree before or in the respective semester; *strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE. $N = 1609$.

studies.[25] In this period, our treatment significantly raises graduation rates. Columns (3) and (4) of Table 4.2 and Figure 4.3 show that the overall completion rate in the treatment group until the eighth semester is 4.3 pp ($p = 0.036$; 3.5 pp, $p = 0.073$ with covariates) higher than in the control group (35.6%). We find about the same difference in the ninth semester: at this stage 48.2% of the students in the control group have already graduated and the completion rate in the feedback group is up by 4.4 pp ($p = 0.030$; 3.7 pp, $p = 0.058$ with covariates). At the end of the observation period – i.e., the end of the tenth semester –, the control group catches up. At this stage its graduation rate is 57.9%, reducing the gap between the treatment and control group to 1.9 pp ($p = 0.339$).

Regarding dropout behavior, Columns (9) and (10) of Table 4.2 (top panel) and Figure 4.4 show that until the tenth semester, 34.8% of the control group subjects have dropped out without completing their degree; accordingly, at this stage, 7.3% are still enrolled in their

[25]Among students who do not drop out of their program during the observation period (65,2%), over 57% graduate in the eighth or ninth semester.

TABLE 4.2: TREATMENT EFFECTS ON ACADEMIC ATTAINMENT

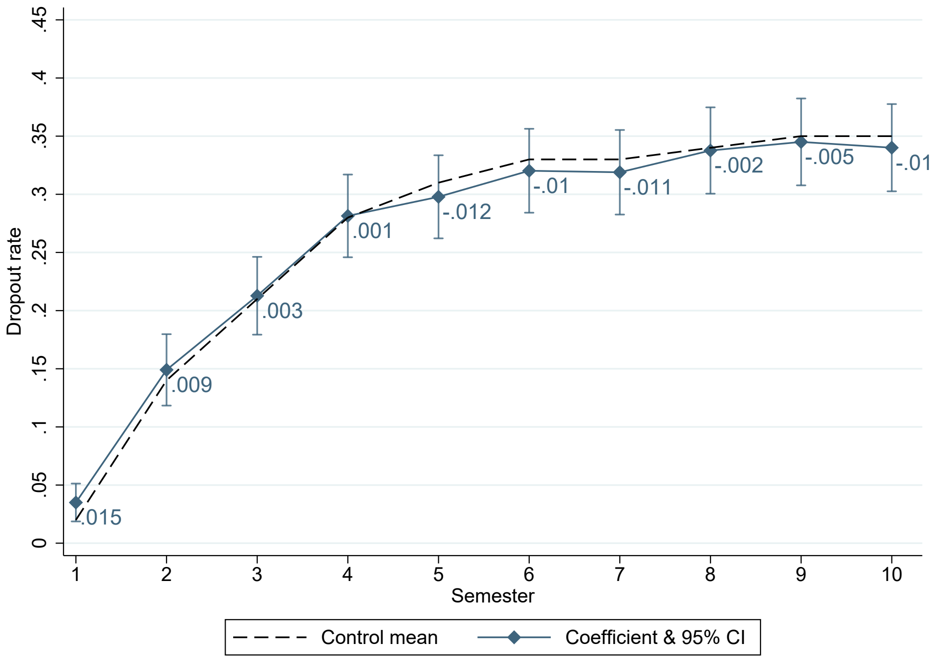
	Graduation rate								Dropout rate	
	7th semester		8th semester		9th semester		10th semester		10th semester	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Main effect										
Treatment	0.015	0.011	0.043**	0.035*	0.044**	0.037*	0.019	0.013	−0.010	−0.007
	(0.014)	(0.014)	(0.021)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
Control mean	0.095	0.095	0.356	0.356	0.482	0.482	0.579	0.579	0.348	0.348
Above average										
Treatment	0.025	0.022	0.088**	0.079**	0.087***	0.078***	0.062**	0.055**	−0.042*	−0.038
	(0.027)	(0.026)	(0.035)	(0.033)	(0.031)	(0.029)	(0.028)	(0.027)	(0.025)	(0.025)
Not above average										
Treatment	0.006	0.002	0.007	−0.001	0.008	0.003	−0.018	−0.022	0.017	0.018
	(0.012)	(0.012)	(0.023)	(0.023)	(0.026)	(0.026)	(0.027)	(0.026)	(0.028)	(0.028)
p-value int. term	0.520	0.475	0.054	0.048	0.053	0.055	0.038	0.039	0.124	0.129
Strata	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Covariates	no	yes	no	yes	no	yes	no	yes	no	yes
N	1609	1609	1609	1609	1609	1609	1609	1609	1609	1609

Notes: The coefficients for each semester are from separate regressions based on Equations 4.1 (full sample) and 4.2 (above average (N = 753) and not above average (N = 856)). *Graduation rate:* indicates if a student graduated with a degree before or in the respective semester; *dropout rate:* indicates if a student dropped out of the study program before or in the respective semester; *strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *covariates:* first semester credits, first semester GPA (missing values imputed), first semester GPA imputed dummy, first semester dropout, high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$. The reported p-values of the interaction term refer to α_{12} in Equation 4.2.

degree program. Our intervention does not cause a significant difference in dropouts between the control and treatment group (feedback decreases the dropout rate by about 1 pp; p = 0.604 and 0.700). Figure 4.4 further shows that there is no treatment effect on dropout rates throughout the entire study duration.

In sum, the difference in graduation rates between treatment and control as well as the absence of differential dropout suggest that the relative feedback intervention enables students to obtain their degree in less time. To quantify the effect on time to degree, in Table 4.3, we provide estimates for the treatment effect on the number of semesters it takes students to graduate. Among all students who have completed their degree by the end of the tenth semester, relative performance feedback reduces time to degree by 0.111 semesters (p = 0.058; 0.103 semesters, p = 0.062 with covariates). However, these estimates do not consider individuals who are still studying, because time to degree is not yet observed for them. In Columns (3) and (4) of Table 4.3, we therefore set time to degree to eleven semesters for all students who are still enrolled at the end of the tenth semester, thereby assuming no further dropouts. Furthermore, we estimate a right-censored Tobit model (at eleven semesters). The coefficients of 0.156 with strata (p = 0.034) and 0.136 with covariates (p = 0.045) suggest that not yet observing time to degree for all potential graduates leads to a downward bias in the initial estimates. Since time to graduation is measured in full semesters, the effect can, e.g.,

FIGURE 4.4: TREATMENT EFFECTS ON DROPOUT RATES



Notes: Dashed line depicts the unadjusted mean of the dropout rates in the control group. Coefficients are from regressions that are estimated separately for each semester and control for *strata*. Confidence intervals are based on robust standard errors. *Outcome variable:* indicates if a student dropped out of the study program before or in the respective semester; *strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE. $N = 1609$.

be interpreted as roughly one in ten students graduating one semester earlier. This suggests that the main effects mask heterogeneous responses to treatment.

Credit point accumulation. Since treated students graduate earlier, we expect this effect to already materialize in the number of credits they accumulate over the course of their studies. We investigate this in the top panel of Figure 4.5. By the end of the first treatment semester (=second semester), relative feedback increases the number of accumulated credits by 1.06 points ($p = 0.096$). In the following semesters, the gap between treatment and control rises further and peaks at the end of the scheduled study duration of seven semesters. At this stage, students in the treatment group have accumulated on average an additional 2.95 credits ($p = 0.278$). Including all covariates reduces the effect to 2.4 credits ($p = 0.349$, see Column 2 in Table 4.4).[26]

[26]After the peak in the seventh semester, the gap between treatment and control closes, because students who already completed their degree do not earn any further credits and graduation rates are higher among treated students.

TABLE 4.3: TREATMENT EFFECT ON TIME TO DEGREE

	OLS		Tobit	
	(1)	(2)	(3)	(4)
Treatment	-0.111^{*}	-0.103^{*}	-0.156^{**}	-0.136^{**}
	(0.058)	(0.055)	(0.073)	(0.067)
Strata	yes	yes	yes	yes
Covariates	no	yes	no	yes
N	948	948	1049	1049
Control mean	8.37	8.37	8.64	8.64
(Std. dev.)	(1.00)	(1.00)	(1.24)	(1.24)

Notes: Outcome variable: time to degree in semesters. OLS estimates in Columns (1) and (2) only include individuals who graduated with a degree by the end of the tenth semester. Tobit estimates in Columns (3) and (4) with right censoring at eleven semesters also include students who are still enrolled in their study program at the end of the tenth semester. Time to degree is set to eleven semesters for those students. *Strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *covariates:* first semester credits, first semester GPA (missing values imputed), first semester GPA imputed dummy, first semester dropout, high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

The parameter estimates of accumulated credits are less precise in later semesters, as the variance increases due to students dropping out of their program and therefore acquiring no further credit points. Since we found no evidence of differential dropout between treatment and controls, we also analyze feedback effects on the number of accumulated credits at the end of the seventh semester among all students who were still enrolled at that time. With coefficients ranging from 2.96 ($p = 0.066$) to 2.55 ($p = 0.079$), these conditional estimates are similar in size to the unconditional estimates. As expected, however, they are estimated more precisely, due to the lower variance of the conditional credits (see standard deviations in the bottom row of Table 4.4).

Taken together, these results suggest that treated students, on average, accumulated an additional three credits until the end of the scheduled period of study. Again, a possible interpretation of this is that one in ten treated students obtains 30 credits more. Since the regular course load of one semester is 30 credits, the effect on accumulated credits is consistent with the finding that one in ten students manages to graduate one semester faster. This indicates again that the effects of the relative feedback intervention are heterogeneous across subgroups.

Grade point average. We now study whether students buy gains in terms of accelerated credit accumulation and faster graduation with losses on another performance dimension. Specifically, it is conceivable that focusing on graduating earlier may result in students

TABLE 4.4: TREATMENT EFFECTS ON ACADEMIC PERFORMANCE

	Acc. credits 7th semester				GPA 10th semester			
	Uncoditional		Conditional		OLS		OLS with IPW	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	2.946	2.413	2.962*	2.554*	−0.058**	−0.037**	−0.037	−0.037**
	(2.715)	(2.574)	(1.612)	(1.452)	(0.026)	(0.015)	(0.026)	(0.017)
Strata	yes	yes	yes	yes	yes	yes	yes	yes
Covariates	no	yes	no	yes	no	yes	no	yes
N	1609	1609	1073	1073	1509	1509	1509	1509
Control mean	128.75	128.75	177.87	177.87	2.50	2.50		
(Std. dev.)	(77.41)	(77.41)	(30.96)	(30.96)	(0.58)	(0.58)		

Notes: Accumulated credits 7th semester: Number of accumulated credits at the end of the seventh semester. The conditional estimates in columns (3) and (4) only include students who were still enrolled in their study program at the beginning of the seventh semester. *GPA 10th semester:* grade point average at the end of the tenth semester including passing grades only; highest passing grade is 1.0, lowest passing grade is 4.0. GPA is unobserved for students who have never obtained a passing grade. Estimates in Colum (7) and (8) employ inverse probability weights based on a logit model that uses the *strata* and *covariates* to predict treatment assignment in the tenth semester among students for whom a GPA is observed. *Strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *covariates:* first semester credits, first semester GPA (missing values imputed), first semester GPA imputed dummy, first semester dropout, high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors clustered at the student level in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

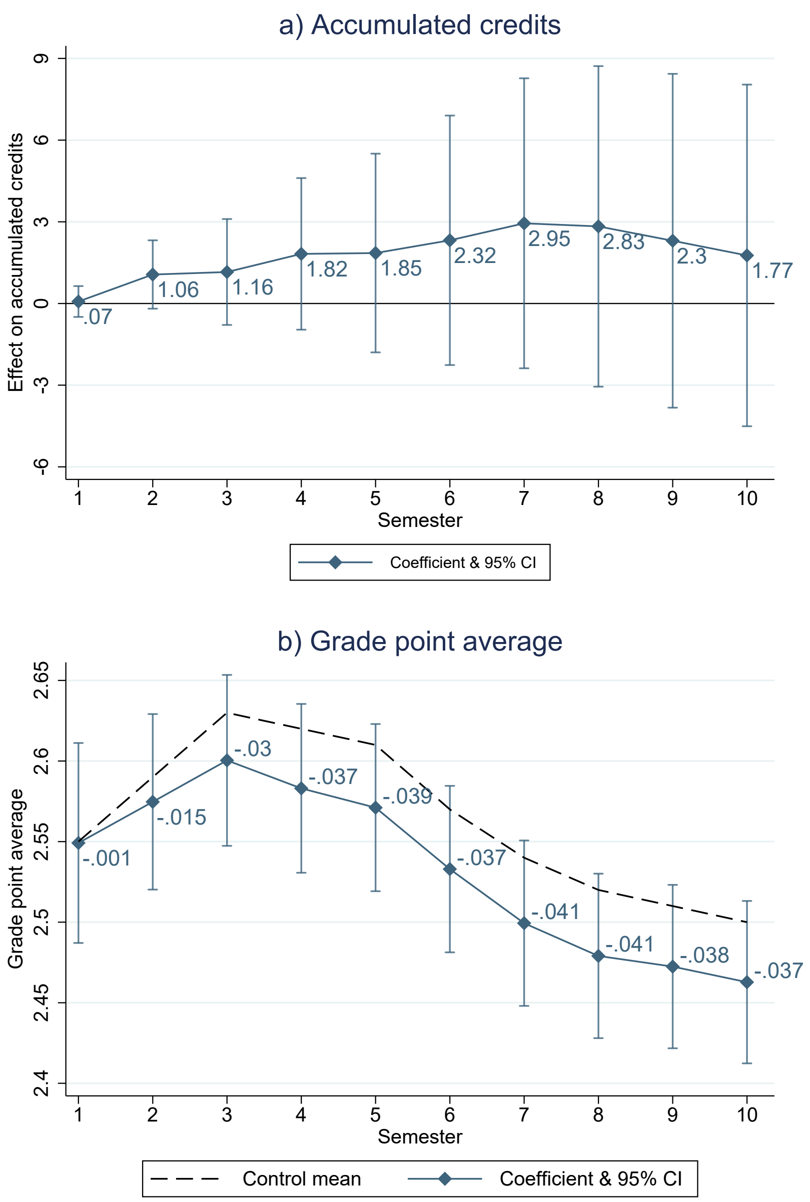
achieving poorer grades. This could lead to a deterioration in students' competence levels, and it may also reduce their labor market prospects (see, e.g., Piopiunik et al. 2020).

The bottom panel of Figure 4.5 plots effects on students' GPA over the whole study duration.[27] The graph provides no evidence that students reduce their performance in terms of GPA in order to graduate earlier (in Germany the highest passing grade is 1.0, while the lowest passing grade is 4.0). On the contrary, the estimates indicate that the GPA of treated students is 0.037 grade points ($p = 0.148$) better at the end of the tenth semester (Column 7 of Table 4.4) – corresponding to an effect size of 0.064 control group standard deviations. When we include the additional covariates in Column (8) of Table 4.4, the coefficient remains unchanged, but it is estimated more precisely ($p = 0.032$), mainly due to the inclusion of the baseline GPA.[28]

[27] GPA is unobserved for students who have not yet passed any exams. We take this into account by employing inverse probability weights based on a logit model that uses the strata and covariates to predict treatment assignment in each semester among students for whom a GPA is observed.

[28] Column (6) in Table 4.4 indicates that including all covariates without performing inverse probability weighting provides the same estimate.

FIGURE 4.5: TREATMENT EFFECTS ON ACADEMIC PERFORMANCE



Notes: Coefficients depicted in panel a) and b) are from from regressions that are estimated separately for each semester and control for *strata*. Confidence intervals are based on robust standard errors. Because GPA is unobserved for students who have not obtained a passing grade yet, regressions in panel b) also employ inverse probability weights based on a logit model that uses the *strata* and *covariates* to predict treatment assignment in each semester among students for whom a GPA is observed. The dashed line depicts the unadjusted mean. *Accumulated credits:* number of accumulated credits at the end of the respective semester; *grade point average:* grade point average at the end of the respective semester including passing grades only; highest passing grade is 1.0, lowest passing grade is 4.0; *strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *covariates:* first semester credits, first semester GPA (missing values imputed), first semester GPA imputed dummy, first semester dropout, high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. $N = 1609$.

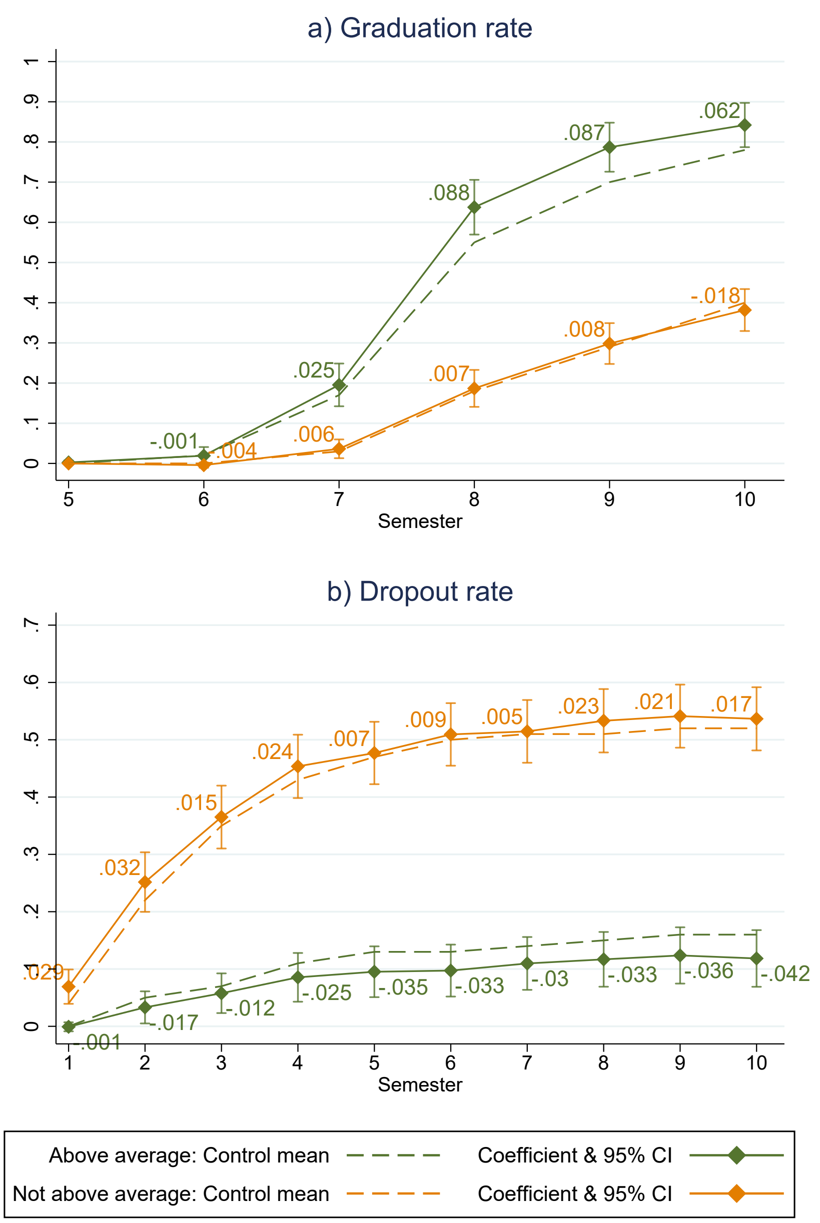
4.3.3 Heterogeneity: above- versus not-above-average feedback

The results presented so far provide evidence that relative performance feedback leads to faster graduation and even has a positive effect students' grades. In addition, we briefly discussed that our findings point to heterogeneous behavioral responses. As outlined in Section 4.3.1, we now use Equation 4.2 to investigate whether the treatment effects on long-term academic success depend on the type of feedback students initially received – i.e., whether students placed above average in the first semester or not.

Effects on graduation and dropout. The top panel of Figure 4.6 and the bottom panel of Table 4.2 (Columns 1 to 8) show treatment effects on graduation rates for the above- and not-above-average subgroups. For students who did not place above average, we find no evidence of changes in degree attainment. Instead, the main effect is completely driven by the subgroup of students who initially ranked above the average. For them, relative performance feedback increases graduation rates in the eighth and ninth semester by 8.8 pp ($p = 0.012$; 7.9 pp, $p = 0.016$ with covariates) and 8.7 pp ($p = 0.005$; 7.8 pp, $p = 0.008$ with covariates), respectively. By the end of the tenth semester, the difference between treatment and control is still 6.2 pp ($p = 0.027$; 5.5 pp, $p = 0.039$ with covariates). Moreover, the p-values of the interaction terms in Table 4.2 indicate that the treatment effects for above- and not-above-average students are significantly different from each other at the 5 or 10% level in all semesters except the seventh.

Next, in the bottom panels of Figure 4.6 and Table 4.2 (Columns 9 and 10), we study whether dropout behavior is also heterogeneous. Starting pre-treatment, the figure shows that the slight imbalance in the first semester dropout rate in Table 4.1 is fully driven by students who did not place above average (2.9 pp; $p = 0.056$). Until the end of the tenth semester, this gap decreases to 1.7 pp ($p = 0.556$; 1.8 pp $p = 0.509$ with covariates). For students with an above-average first semester performance, on the other hand, the baseline dropout rate is completely balanced. Post-treatment, the gap between treated students and controls gradually widens until it reaches -4.2 pp ($p = 0.100$, -3.8 pp, $p = 0.124$ with covariates) at the end of the observation period. This indicates that part of the positive effect on graduation rates for above-average students is driven by treated students dropping out at a lower rate, suggesting that they do not only take less time to graduate, but that the treatment also generally increases the likelihood to obtain a degree.

FIGURE 4.6: TREATMENT EFFECTS ON ACADEMIC ATTAINMENT, ABOVE VERSUS NOT ABOVE AVERAGE



Notes: Above average ($N = 753$) and not above average ($N = 856$) refer to the position in the performance distribution at the end of the first semester. Dashed lines depict unadjusted means. Coefficients are from regressions based on Equation 4.2 that control for *strata*, which are estimated separately for each semester. Confidence intervals are based on robust standard errors. *Graduation rate:* indicates if a student graduated with a degree before or in the respective semester; *dropout rate:* indicates if a student dropped out of the study program before or in the respective semester; *strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE.

Effects on credit point accumulation and grade point average. Figure 4.7 provides evidence that students' GPA and credit accumulation is subject to the same heterogeneous response. The top panel shows that above-average students earn more credits in each semester: starting with an additional 2.51 credits in the first treatment semester ($p = 0.006$) the difference accumulates to 8.09 credits ($p = 0.022$) until the end of the scheduled study duration, suggesting that these students are approximately 0.27 semesters ahead of the control group (remember that the scheduled course load of one semester is 30 credits). The bottom panel depicts effects on students' GPA. Similar to the other outcomes, the main effect is driven by students who initially performed above average. At the end of the tenth semester, their GPA is 0.06 points better ($p = 0.074$) compared to controls.

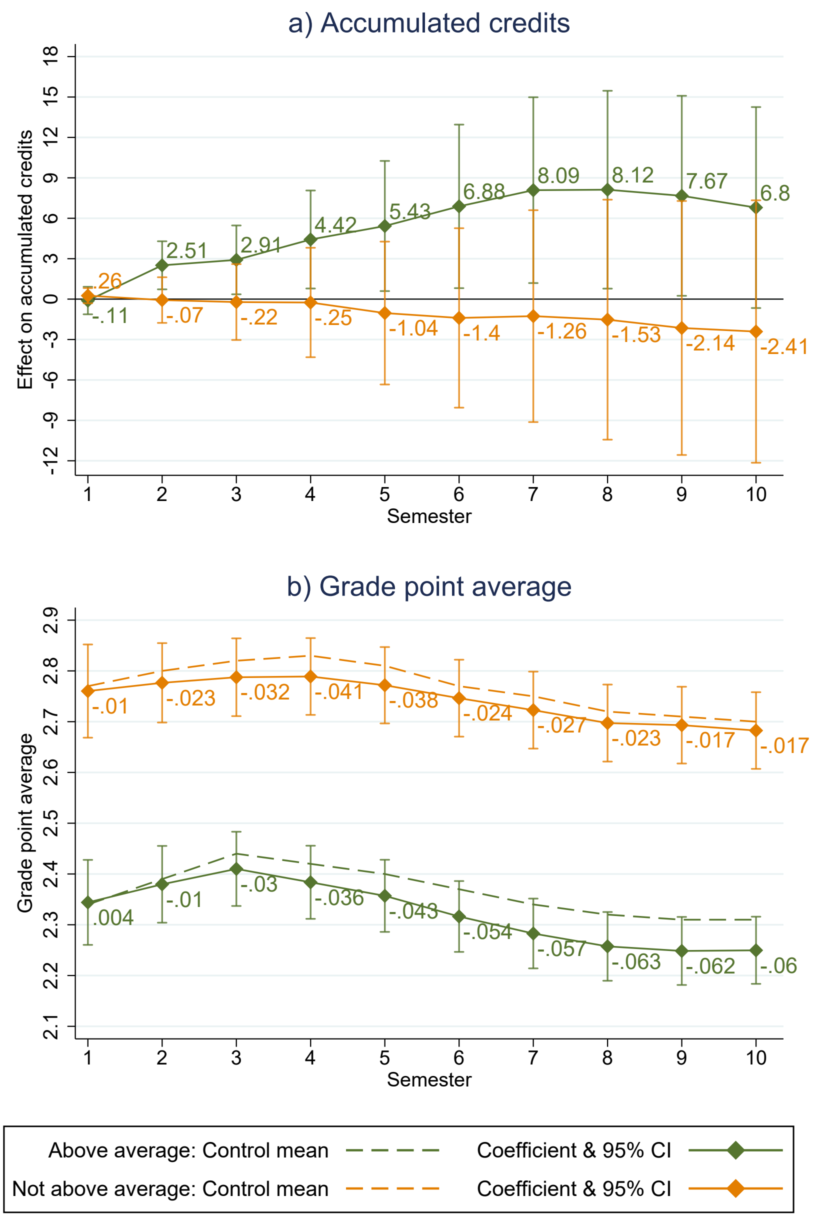
In sum, the results reported in this subsection provide evidence that the positive effects of the relative feedback intervention – with respect to all dimensions of academic success considered in this paper – are concentrated among students who were initially informed that they performed above-average.

4.4 Further heterogeneity and mechanisms

In this section, we further explore which subgroups drive the overall effect of relative feedback. Analyzing heterogeneities is important for two reasons: 1) From a political perspective, it is desirable to clearly identify the beneficiaries of an intervention. This can allow for a targeted provision, which not only saves resources, but also allows to support students for whom the intervention is not effective with programs that are better suited to their needs. 2) A closer look at heterogeneous effects helps to better understand possible mechanisms behind the differential responses and thus facilitates more comprehensive feedback designs in future research.

Our approach is as follows: First, in Section 4.4.1, based on the students' pre-treatment performance and background characteristics, we explore short- to long-term treatment effects across a broad range of subgroups. This aims to provide a better understanding of the extent and persistence of the heterogeneous effects. In Section 4.4.2, the differential effects are further examined by using the control group to predict the second semester performance in the absence of treatment. We employ these predictions to divide our sample into four "predicted performance strata" and investigate the treatment effects within these blocks. In Section 4.4.3 we discuss theoretical and institutional explanations for the different effects found across the strata.

FIGURE 4.7: TREATMENT EFFECTS ON ACADEMIC PERFORMANCE, ABOVE VERSUS NOT ABOVE AVERAGE



Notes: Above average (N = 753) and not above average (N = 856) refer to the position in the performance distribution at the end of the first semester. Coefficients depicted in panel a) and b) are from from regressions based on Equation 4.2 that are estimated separately for each semester and control for *strata*. Confidence intervals are based on robust standard errors. Because GPA is unobserved for students who have not obtained a passing grade yet, regressions in panel b) also employ inverse probability weights based on a logit model that uses the *strata* and *covariates* to predict treatment assignment in each semester among students for whom a GPA is observed. The dashed line depicts the unadjusted mean. *Accumulated credits:* number of accumulated credits at the end of the respective semester; *grade point average:* grade point average at the end of the respective semester including passing grades only; highest passing grade is 1.0, lowest passing grade is 4.0; *strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *covariates:* first semester credits, first semester GPA (missing values imputed), first semester GPA imputed dummy, first semester dropout, high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy.

4.4.1 Subgroup responses and persistence of treatment effects

We start by providing an overview of second semester treatment effects over a wide range of data splits. Based on the results in Section 4.3.3, we first divide the sample into above- and not-above-average students. Within these groups, we define further subgroups along student's background characteristics (age, gender, etc.), their first semester GPA[29], their faculty, and their study cohort.[30] Panel a) of Figure 4.8 plots the estimated parameters. The left (right) side depicts coefficients for the not-above-average (above-average) subgroups ordered according to the size of the effects.[31] As we discuss in more detail in the next section, individual estimates in this figure should be interpreted with caution. Rather, we are interested in the overall pattern of results which provides the following insights: First, the full sample effect is subject to substantial heterogeneity, with effects on second semester credits varying between minus two and four. Second, the estimates again confirm that there is a clear difference between above- and not-above-average feedback: while most treatment effects are close to zero for the latter, feedback raises performance by two or more credits in the above-average subgroups. Thus, it is evident that, although we find substantial heterogeneity, a single individual characteristic alone cannot explain the differential responses to feedback between above and not-above-average students.

Next, we study whether the heterogeneities persist. In Panel b) of Figure 4.8, we investigate the medium-term persistence, by plotting effects on third to seventh semester credits against the treatment effects reported in Panel a).[32] In addition, the graph includes the linear fit from a corresponding regression, weighted by subgroup sizes and excluding the three aggregate treatment effects (full-sample, above-average, not-above-average). The results suggest that treatment effects are serially correlated: a one credit treatment effect in the second semester is related to an average treatment effect of 0.515 credits in each of the following five semesters. To put this in perspective, in parentheses, we report the coefficient of a control-group OLS regression of third to seventh semester credits on the second semester credits. The coefficient of 0.45 provides a measure for the usual serial correlation of credits and is close to the correlation of the treatment effects.

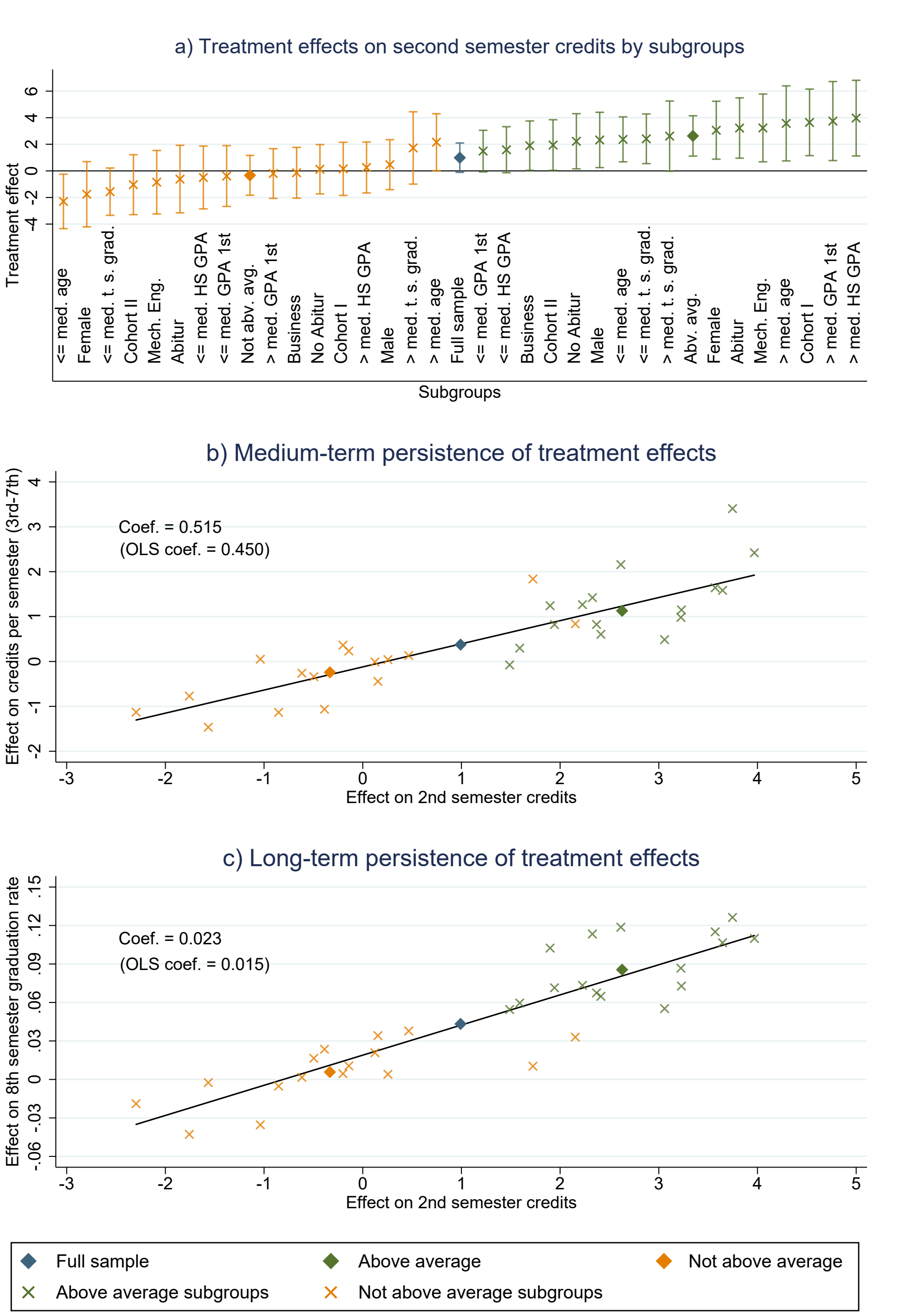
[29]We do not consider credits, as they are highly correlated with the above- versus not-above-average categorization

[30]To build subgroups based on continuous variables, we split them at the median within each cohort and program.

[31]The treatment effects and the respective sample sizes are reported in Columns (1) and (4) of Tables 4.A1 and 4.A2.

[32]Effects on third to seventh semester credits are estimated with pooled OLS using standard errors clustered at the individual level. The effects are reported in Column (2) of Tables 4.A1 and 4.A2.

FIGURE 4.8: HETEROGENEITY AND PERSISTENCE OF TREATMENT EFFECTS



Notes: Panel a) depicts coefficients for treatment effects on credits obtained in the second semester and correspoding 95% confidence intervals estimated in different subsamples. Coefficients are based on regressions in the respective subsamples that control for *strata* (credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE). Confidence intervals are based on robust standard errors. Panel b) and c) plot the treatment effects depicted in panel a) against the corresponding treatment effect estimates for the number of credits obtained per semester (third to seventh semester) and for the graduation rate in the eighth semester. The linear fits and coefficients are estimated by weighting the treatment effect estimates by the corresponding subgroup sample sizes while excluding the full sample, above-average, and not above-average average estimates. The OLS coefficients shown for comparison in parantheses are based on control group regressions that regress the number of credits obtained per semester (third to seventh semester) and the graduation rate in the eighth semester on the number of credits obtained in the second semester, controlling for *strata*. All treatment effects shown in this figure and the respective sample sizes are reported in Tables 4.A1 and 4.A2.

To provide evidence on long-term persistence, Panel c) of Figure 4.8 plots effects on the eighth semester graduation rate against the effects on the second semester credits. The results indicate that a one credit point increase in the second semester due to our feedback is related to a 2.3 pp treatment effect on the likelihood of having graduated by the end of the eighth semester.[33] Again, the control-group correlation of these outcomes is very similar: one additional credit in the second semester corresponds to a 1.5 pp increase in graduation probability.

Taken together these results indicate that the subgroup specific treatment effects are correlated over time, which provides further evidence for the persistent effects of relative feedback. There are two plausible explanations for this persistence: First, the ongoing delivery of relative feedback may lead to repeated treatment effects. Second, the initial feedback may prompt students to adopt performance-enhancing habits that also increase performance in later semesters. Unfortunately, we did not have the statistical power to include an additional group receiving the feedback just once, which would have allowed us to disentangle the two explanations.

4.4.2 Predicted performance stratification

Besides providing evidence on the persistence of treatment effects, the exploratory analysis in the last section indicates the existence of heterogeneities within the subgroups of above- and not-above-average students. In the absence of ex-ante hypotheses, however, these analyses have several limitations, making it difficult to draw robust and accurate conclusions about the beneficiaries of the intervention (see, e.g., Kent et al. 2020): First, seemingly arbitrary decisions about the inclusion and construction of subgroups may raise concerns regarding the selective presentation of results. Second, to prevent false positives, it is necessary to correct for multiple hypothesis testing. Depending on the method, however, this may lead to overly conservative estimates and produce false negatives. Last, the simple “one-variable-at-a-time” comparisons, such as low versus high ability or female versus male, may not adequately capture the full extent of heterogeneity across outcomes and treatment effects.

To address these limitations, we will now examine the effects of relative feedback along a single variable: second semester credit points. To deal with the problem that the counterfactual outcomes of treated students are unobserved, we use the control students’ first semester performance and background characteristics to train a prediction model. This collapses the various heterogeneity dimensions into a single performance index and allows us to estimate comparable outcomes for the control and treatment group. We then use the index – i.e., the

[33] The corresponding effects are reported in Column (3) of Tables 4.A1 and 4.A2.

predicted second semester credits – to construct subgroups, which we refer to as "predicted performance strata".

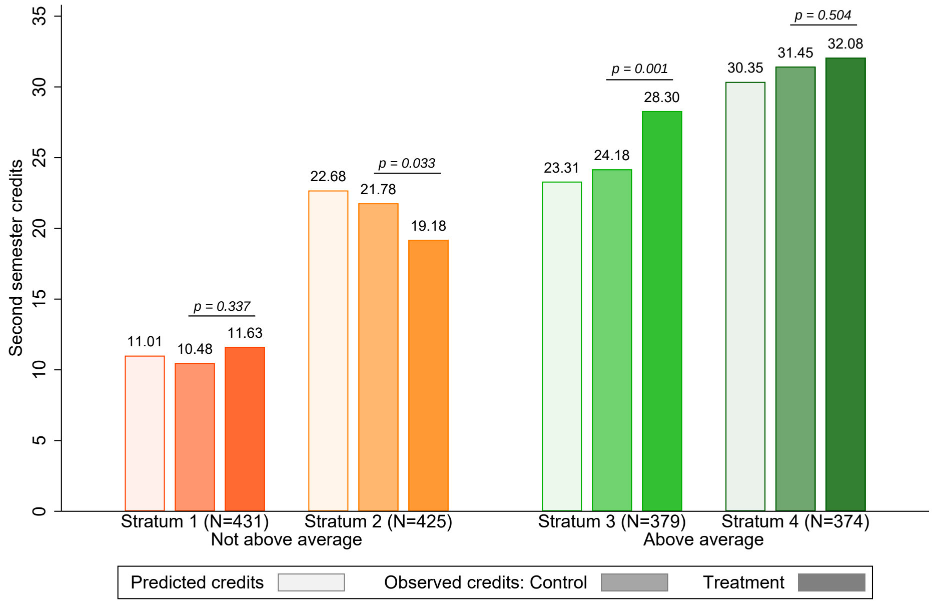
We use second semester credits to construct the index for several reasons: First, credits are the outcome dimension on which relative performance feedback is given and the second semester performance is the first realization of this outcome post-treatment. For long-term outcomes such as dropout or graduation rates, on the other hand, it would be less obvious which point in time should be selected. Second, given the evidence on treatment-effect persistence shown in Section 4.4.1, using the second semester credits should still provide a good proxy for heterogeneity in long-term outcomes. Third, in practical applications, using the second semester credits allows policymakers to use data from adjacent cohorts as prediction sample. Employing long-term outcomes would introduce a substantial lag between the cohorts that are used for prediction and the cohorts in which the intervention is supposed to be introduced. If study regulations or the composition of the student population change over time, larger lags could negatively affect the quality of the predictions.

Construction of predicted performance strata. Regarding the prediction sample, we take advantage of the two-cohort structure of our study and rely on out-of-sample predictions, i.e., we estimate the prediction model using the control group of the first cohort to predict performance among all students in the second cohort, and vice versa.[34] Our prefered prediction model includes baseline performance, background characteristics, and their interactions with the study program fixed effects as explanatory variables, and we employ a lasso approach with five-fold cross-validation (we use cubic functions of the continuous covariates). Appendix B describes the choice and the performance of the prediction model in more detail.

Next, within the two subgroups of above- and not-above-average students, we split the predicted values at the median of each cohort and study program, thereby creating four predicted performance strata: the bottom and top stratum of not-above-average students (Strata 1 and 2) and the bottom and top stratum of above-average students (Strata 3 and 4). Across the four blocks, Figure 4.9 displays predicted credits for the full student sample as well as the observed second semester credits separately for students of the control and treatment group. The comparison between the predictions and the observed credits of the control group shows that, on average, predictions and observed values are very similar. In addition, the figure also indicates that in terms of predicted and observed second semester

[34] Our approach is related to "risk stratification" in medical research (see, e.g., Kent and Hayward 2007; Kent et al. 2018). Another related method is "endogenous stratification", in which in-sample control group data is used for the predictions (Abadie et al., 2018).

FIGURE 4.9: PREDICTED AND OBSERVED SECOND SEMESTER CREDITS BY PREDICTED PERFORMANCE STRATA



Notes: The figure plots the predicted second semester credits for all students as well as the observed second semester credits for students in the control and in the treatment group separately for the four predicted performance strata. P-values are from t-tests on the equality of means.

performance, not-above-average control students in the top strata (Stratum 2) show hardly any differences compared to above-average control group subjects in the bottom strata (Stratum 3).[35] Thus, if the response to feedback would depend solely on the students' predicted performance (= an index of background characteristics and baseline performance), and not on the content of the feedback, i.e., receiving information about an above- or not-above-average performance, we would expect students in the two middle strata to behave similarly.

Treatment effects by predicted performance strata. The observed second semester credits of treated students in Figure 4.9 already provide strong evidence that feedback does not similarly affect students in Strata 2 and 3: in the second semester, treated students in Stratum 2 earn 2.6 credits less ($p = 0.033$) than controls, while treated individuals in Stratum 3 obtain 4.1 credits more ($p = 0.001$) than those in the control group. Students in Strata 1 and 4 show little or no response to feedback.

[35]This can also be seen from the descriptive statistics shown in Table 4.A3. With the exception of baseline performance (credits and GPA) the two groups are very similar. Interestingly, students in Stratum 3 have a worse baseline GPA than students in Stratum 2. This suggests that the general academic ability of students in the first group is not strictly worse compared to the second group.

Next, we study more formally whether the effects of relative performance feedback on long-term academic success differ across the strata. To determine the treatment effects of the four subgroups, we estimate the following equation:

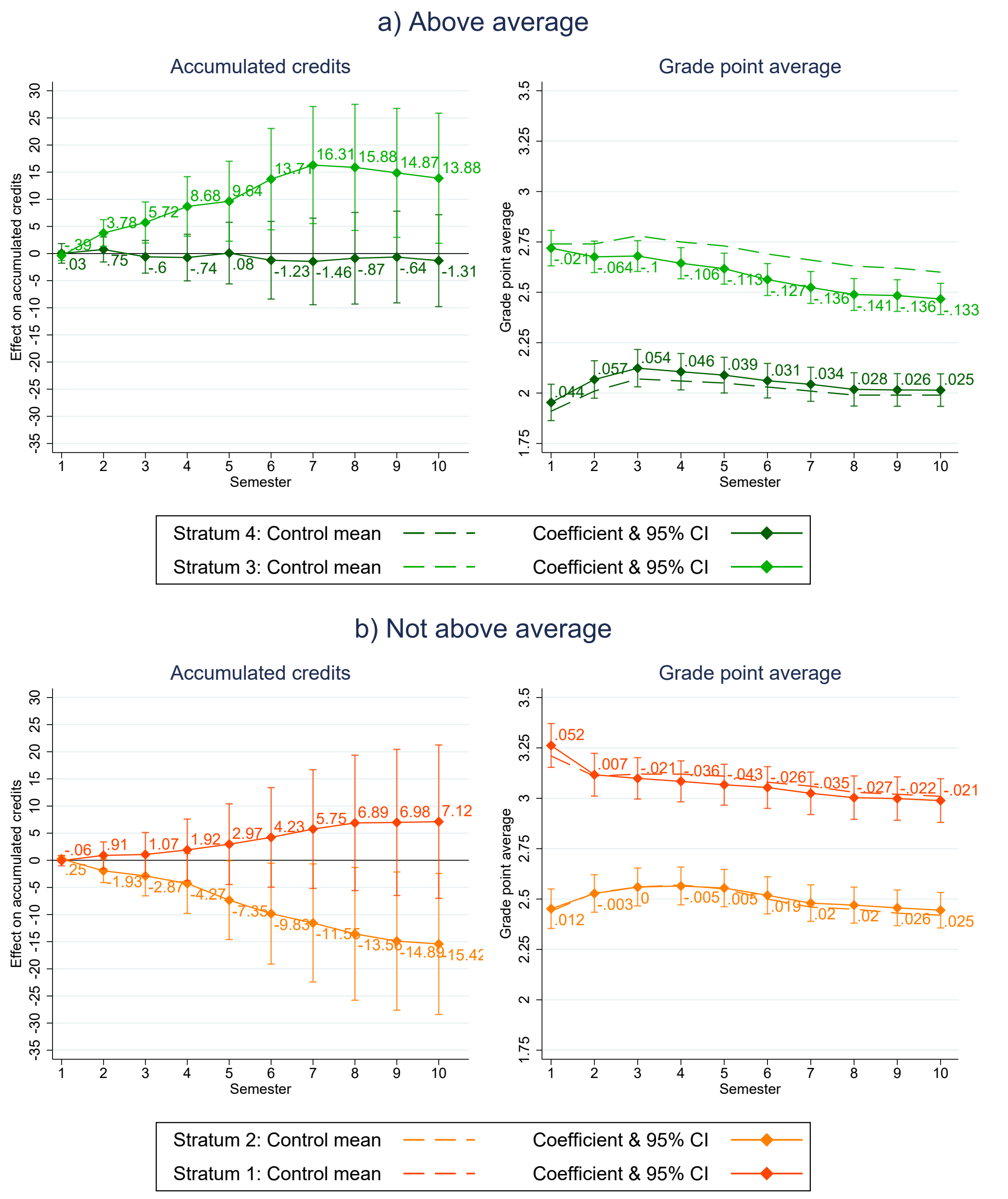
$$\begin{aligned} Y_i^k = {} & \alpha_0 + \alpha_1 Treatment_i + \alpha_2 A_i + \alpha_3 T_i + \alpha_{12} Treatment_i A_i \\ & + \alpha_{13} Treatment_i T_i + \alpha_{23} A_i T_i + \alpha_{123} Treatment_i A_i T_i + \mathbf{s_i} \boldsymbol{\alpha_3} + \varepsilon_i, \end{aligned} \tag{4.3}$$

where T_i is an indicator for placing in the top stratum among above- or not-above-average students, and all other parameters are defined as in Equations 4.1 and 4.2.

OLS estimates of this equation in Figure 4.10 show effects on academic performance (accumulated credits and GPA) over the whole study duration. The left plots in Panels a) and b) indicate that the pattern observed in Figure 4.9 persists. Effects on credits among above-average students are completely accounted for by Stratum 3. By the end of the scheduled study duration of seven semesters, students in this group have attained 16.3 credits more (p = 0.003), which corresponds to the course-load of over half a semester. Students in Stratum 2 continue to accumulate less credits, trailing the control group by 11.5 credits (p = 0.037) at the end of the regular study duration. The right plots of Panels a) and b) show that the positive main effect on GPA is also completely driven by the students in Stratum 3, whose GPA is .133 grade points better (p = 0.001) by the end of the tenth semester. For GPA, we find no evidence for negative effects among students in any other strata.

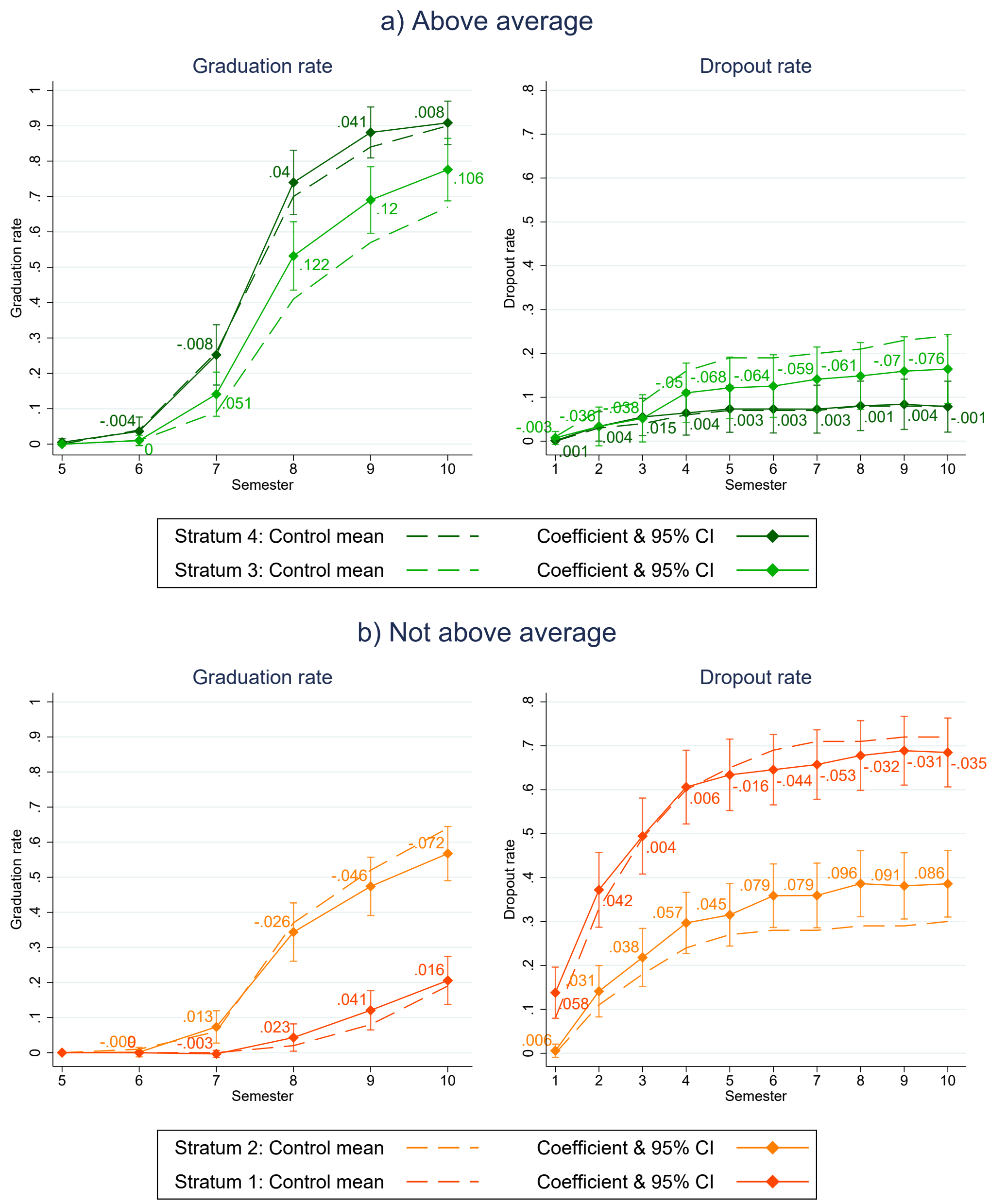
Figure 4.11 shows that the effects on students' academic attainment follow the same pattern as the effects on credit accumulation. According to the left plot in Panel a), graduation rates of students in Stratum 3 have risen by 12.2 pp (p = 0.014) in the eighth semester and are still up by 10.6 pp (p = 0.019) at the end of the tenth. Together with the 7.6 pp (p = 0.061) decrease in dropout rates evident from the right plot, this suggests that these students do not only graduate faster but are generally more likely to obtain their degree. In addition, the estimates indicate that the relative feedback intervention helps students with an intermediate predicted performance to catch up with the very top performers, reducing the difference in graduation and dropout rates observed in the control group by about 50%. Panel b) shows that relative feedback also leads to less unequal outcomes among not-above-average students, but only due to the unintended negative consequences for students in Stratum 2, who – by the end of the tenth semester – are 7.2 pp (p = 0.065) less likely to have graduated and 8.6 pp (p = 0.026) more likely to have dropped out of their study program.

FIGURE 4.10: TREATMENT EFFECTS ON ACADEMIC PERFORMANCE BY PREDICTED PERFORMANCE STRATA



Notes: Panel a) shows treatment effect estimates on academic performance for the predicted performance Strata 3 ($N = 379$) and 4 ($N = 374$) (see Section 4.4.2 for details) among students who placed above average in the performance distribution at the end of the first semester. Panel b) shows respective estimates for the predicted performance Strata 1 ($N = 431$) and 2 ($N = 425$) among students who did not place above average. Dashed lines depict unadjusted means. Coefficients in all panels are from full sample regressions based on Equation 4.3 that control for *strata*, which are estimated separately for each semester. Confidence intervals are based on robust standard errors. *Accumulated credits:* number of accumulated credits at the end of the respective semester; *grade point average:* grade point average at the end of the respective semester including passing grades only; highest passing grade is 1.0, lowest passing grade is 4.0; because GPA is unobserved for students who have not obtained a passing grade yet, the regressions also employ inverse probability weights based on a logit model that uses the *strata* and *covariates* (first semester credits, first semester GPA (missing values imputed), first semester GPA imputed dummy, first semester dropout, high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy) to predict treatment assignment in each semester among students for whom a GPA is observed.

FIGURE 4.11: TREATMENT EFFECTS ON ACADEMIC ATTAINMENT BY PREDICTED PERFORMANCE STRATA



Notes: Panel a) shows treatment effect estimates on academic attainment for the predicted performance Strata 3 (N = 379) and 4 (N = 374) (see Section 4.4.2 for details) among students who placed above average in the performance distribution at the end of the first semester. Panel b) shows respective estimates for the predicted performance Strata 1 (N = 431) and 2 (N = 425) among students who did not place above average. Dashed lines depict unadjusted means. Coefficients in all panels are from full sample regressions based on Equation 4.3 that control for *strata*, which are estimated separately for each semester. Confidence intervals are based on robust standard errors. *Graduation rate:* indicates if a student graduated with a degree before or in the respective semester; *dropout rate:* indicates if a student dropped out of the study program before or in the respective semester; *strata:* credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE.

4.4.3 Explaining the pattern of results

The pattern of results presented above raises two questions: i) What is behind the asymmetrical behavioral response of above- and not-above-average students, particularly those with an intermediate predicted performance? ii) How can we explain the absence of treatment effects among students with the lowest and the highest predicted performance?

Theoretical considerations. Addressing the first question, we start by considering how common theories on the effects of relative performance feedback can be aligned with our results. In the context of education, the two most prominent explanations are "competitive preferences" and "self-perception theory" (see Azmat and Iriberri 2010 and Dobrescu et al. 2021).

According to the theory of competitive preferences, individuals pursue to perform better than others, because their utility depends both on absolute as well as relative performance. While they have complete information about their own ability, they need to form expectations about the performance of others. Providing them with relative feedback improves the accuracy of these expectations, which in turn increases the weight that they put on the competitive part of their utility function. This leads to higher effort and performance across the entire distribution. Given that we observe positive treatment effects only for the subgroup of above-average students, it is unlikely that competitive preferences are the mechanism behind behavioral changes found in our study.

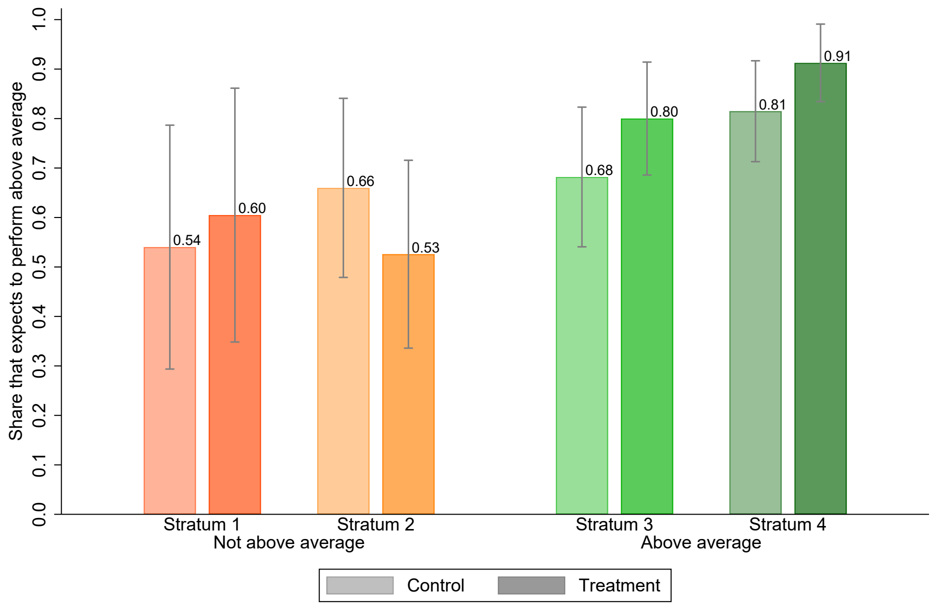
In contrast, self-perception theory suggests that students' utility depends only on their own performance. The idea is that individuals have incomplete information about their ability and therefore use task performance to form expectations. Since performance also depends on the difficulty of the task, its signal about ability is noisy. Therefore, the relative feedback information will additionally affect self-perceived ability. Intuitively, if the signal about the performance of others is high (low), it increases the likelihood that the task was easy (difficult), which in turn decreases (increases) the probability that one's own ability is high. Under the assumption that ability and effort are complements, a downward (upward) adjustment in beliefs about one's own ability leads to a reduction (increase) in effort and performance. Applied to our context, it seems conceivable that treated students adjust expectations about ability upward relative to controls when they receive information that they performed above average, and downward when feedback informs them about a not-above-average performance.[36] In turn, this differential updating of beliefs leads to the observed

[36]This simplified interpretation presumes that students process feedback in a rather discrete way, i.e., the exact distance to the average performance is only of minor importance compared to the information about being above or not above average. This notion is supported by the findings in Chapter 3, where an RDD is

negative or positive effects on performance.

Beliefs about relative performance. We explore the above idea further, by investigating effects on students' post-treatment beliefs regarding their relative performance. In three online surveys conducted in the fifth (Cohort I) as well as the second and third semester (Cohort II), students were asked "How many percent of your fellow students will have achieved more credit points (ECTS) than you at the end of the current semester?". Based on this question, we categorize students as expecting an above-average performance if they answered that less than 50% of the other students will achieve more credit points.

FIGURE 4.12: POST-TREATMENT BELIEFS ABOUT RELATIVE PERFORMANCE BY PREDICTED PERFORMANCE STRATA



Notes: The figure plots estimates for the share of students who expects to place above the average in the post-treatment credit point performance distribution separately for the two prediction strata among not above-average and above-average students, respectively. The data on expectations stems from three surveys that were conducted in the fifth semester (Cohort I) as well as the second and third semester (Cohort II). Estimates are from pooled OLS regression based on the three way intreaction depicted Equation 4.3 and control for study program FE and timing of survey FE. 95% confidence intervals are based on cluster robust standard errors clustered at the student level. $N = 338$ for 288 individual students.

Figure 4.12 depicts estimates for the share of students that expect to perform above average, separately for the treatment and control groups in the four performance strata.[37] The

used to provide evidence that receiving above- instead of not-above-average feedback leads to large increase in second semester performance for students around the average.

[37]The estimates are based on pooled OLS regressions using Equation 4.3, but controlling for study program fixed effects and survey fixed effects instead of the randomization strata. The survey response rate, measured as the share of students who answered the questions among all students in the respective cohorts, is about 15% for each of the surveys.

pattern of results which emerges for the two middle strata can be well aligned with the self-perception theory. Similar to the predicted and observed performance in the second semester shown in Figure 4.9, we find that the control group beliefs about relative performance are quite similar to each other, with about 67% of students in the two middle strata expecting to perform above average. The estimated shares for the treatment groups indicate that relative feedback leads students to adjust their beliefs in the expected direction: only 53% of treated not-above-average students in Stratum 2 expect to perform above average, while this share rises by 27 pp ($p = 0.015$) to 80% among treated above-average students in Stratum 3. Assuming that the expectations about relative performance also manifest in students' beliefs about their own ability, the survey results suggest that the mechanisms of self-perception theory can explain the effect of relative feedback on academic success.

For the worst and best performers, however, the beliefs have little explanatory power. Regarding the first group they provide little guidance as confidence intervals are large (see Figure 4.12).[38] For Stratum 4, we find tentative evidence of an adjustment of beliefs in the expected direction, which, following the self-perception theory, should have led to an increase in performance.

Floor and ceiling effects. Floor and ceiling effects offer an alternative explanation for the lack of a behavioral response at the lower and upper ends of the predicted performance distribution. Specifically, we argue that the performance of control group students in the very bottom (top) stratum is already so low (high) that treated students have little room to perform worse (better).

Regarding students in Stratum 1, the following descriptives illustrate the point: 1) In the second semester, they earn on average only half as many credits as students in Stratum 2 (see Figure 4.9). 2) About 50% of them have earned 15 credits or less in their first year (see the top left panel in Figure 4.A1), which is only a quarter of the credits recommended by the study curriculum; and 3) about 60% have dropped out of their program by the end of the fourth semester (see right plot in Panel b) of Figure 4.11). Given these numbers, however, it is also plausible that these students are well aware of their low performance and that our feedback therefore provides little new information about their (relative) abilities.

The cumulative distributions of the accumulated credit points shown in Figure 4.A2 provide descriptive evidence for the presumed ceiling effects among students in Stratum 4. In most semesters of the regular study duration a large share of these students has obtained almost exactly the number of accumulated credits that they are supposed to have earned

[38]The response rate among these students is only about 5%, which is in part driven by the fact that this subgroup also experiences the highest amount of dropout.

according to the study curriculum.[39] This suggests two things: 1) The study curriculum may set a reference point that students aim to adhere to and are unwilling to exceed. 2) Related to the first point, since the study programs, courses, and lectures are structured and scheduled in accordance with the official study duration of seven semesters, it can be difficult for students to take courses from future semesters. As a results, the top performers may have little opportunity to complete their study program faster. The comparison to the cumulative distributions of students in Stratum 3 additionally illustrates this: for them the treatment effects are almost exclusively generated in parts of the distribution that lie below the assumed ceiling. From an educational policy perspective, therefore, without removing these structural barriers, it might be difficult to support top students through relative feedback or other measures.

4.5 Conclusion

Based on two field experiments, this paper studies the long-term effects of providing university students with relative feedback on their accumulated credit points. On average, we find that the intervention helps students accumulate more credits, earn better grades, and graduate earlier without affecting dropout rates. Compared to more traditional approaches, such as financial aid, student coaching, or reducing class size, our intervention is easier to scale because of its low cost. In our view, however, classical and behavioral measures should be considered as complementary supply-side factors. For example, relative feedback cannot be effective if students drop out or do not even start studying because of a lack of financial resources. The interaction of the two approaches is thus an important field for future research.

Exploratory heterogeneity analyses provide evidence that students in the middle of the (predicted) performance distribution are most responsive to feedback: treatment effects are negative for not-above-average students with a high predicted performance, while above-average students with a low predicted performance respond positive to feedback, allowing them to catch up to the students with the highest predicted performance. The performance of students at the upper and lower end of the distribution, on the other hand, is left unchanged. Our findings suggest that for the best performing students the institutional setting may leave little scope for relative feedback to improve performance, as it creates barriers to

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

graduating before the nominal study duration. To help not-above-average students – in particular those with the lowest performance – succeed in university, other, more comprehensive measures may be necessary. Our approach of examining heterogeneity along students' predicted performance simultaneously provides a relatively simple way for the targeted delivery of interventions that are better tailored to the needs of those students.

Appendix

A Additional tables and figures

TABLE 4.A1: TREATMENT EFFECTS BY ABOVE-AVERAGE SUBGROUPS

	(1) Credits 2nd semester	(2) Credits per semester (3rd-7th)	(3) Graduation rate 8th semester	(4) N
Full sample	0.992	0.376	0.043	1609
	(0.563)	(0.467)	(0.021)	
Above average	2.628	1.130	0.086	753
	(0.772)	(0.599)	(0.035)	
Subgroups				
No Abitur	2.226	1.267	0.073	420
	(1.055)	(0.801)	(0.048)	
Abitur	3.224	0.983	0.087	333
	(1.152)	(0.902)	(0.052)	
Male	2.328	1.424	0.113	472
	(1.058)	(0.780)	(0.044)	
Female	3.058	0.488	0.055	281
	(1.106)	(0.866)	(0.056)	
Business Dept.	1.899	1.244	0.102	443
	(0.944)	(0.724)	(0.045)	
Mechanical Eng. Dept.	3.229	1.150	0.073	310
	(1.295)	(1.013)	(0.055)	
(Below) med. HS GPA	1.591	0.298	0.060	515
	(0.881)	(0.690)	(0.042)	
Above med. HS GPA	3.968	2.422	0.110	238
	(1.444)	(1.094)	(0.065)	
(Below) med. 1st sem. GPA	1.489	−0.077	0.054	505
	(0.796)	(0.589)	(0.040)	
Above med. 1st sem. GPA	3.747	3.405	0.126	248
	(1.509)	(1.259)	(0.063)	
Cohort I	3.646	1.585	0.107	305
	(1.272)	(0.891)	(0.053)	
Cohort II	1.943	0.823	0.071	448
	(0.968)	(0.803)	(0.046)	
(Below) med. Age	2.369	0.826	0.067	440
	(0.860)	(0.731)	(0.044)	
Above med. Age	3.571	1.643	0.115	313
	(1.435)	(0.999)	(0.057)	
(Below) med. t. s. grad.	2.414	0.605	0.065	488
	(0.950)	(0.723)	(0.043)	
Above med. t. s. grad.	2.615	2.159	0.119	265
	(1.340)	(1.047)	(0.060)	
Strata	yes	yes	yes	
Covariates	no	no	no	

Notes: Strata: credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *covariates:* first semester credits, first semester GPA (missing values imputed), first semester GPA imputed dummy, first semester dropout, high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses.

TABLE 4.A2: TREATMENT EFFECTS BY NOT ABOVE-AVERAGE SUBGROUPS

	(1) Credits 2nd semester	(2) Credits per semester (3rd-7th)	(3) Graduation rate 8th semester	(4) N
Full sample	0.992	0.376	0.043	1609
	(0.563)	(0.467)	(0.021)	
Not above average	−0.333	−0.240	0.006	856
	(0.762)	(0.677)	(0.023)	
Subgroups				
No Abitur	0.124	−0.009	0.021	519
	(0.944)	(0.834)	(0.028)	
Abitur	−0.617	−0.260	0.002	337
	(1.292)	(1.128)	(0.039)	
Male	0.466	0.133	0.038	541
	(0.956)	(0.826)	(0.027)	
Female	−1.757	−0.769	−0.043	315
	(1.247)	(1.148)	(0.043)	
Business Dept.	−0.141	0.233	0.011	538
	(0.972)	(0.848)	(0.032)	
Mechanical Eng. Dept.	−0.852	−1.133	−0.005	318
	(1.214)	(1.105)	(0.030)	
(Below) med. HS GPA	−0.496	−0.339	0.016	365
	(1.202)	(1.062)	(0.041)	
Above med. HS GPA	0.255	0.042	0.004	491
	(0.975)	(0.882)	(0.027)	
(Below) med. 1st sem. GPA	−0.387	−1.064	0.024	309
	(1.162)	(0.934)	(0.043)	
Above med. 1st sem. GPA	−0.200	0.364	0.005	547
	(0.951)	(0.909)	(0.027)	
Cohort I	0.154	−0.442	0.034	507
	(1.018)	(0.872)	(0.034)	
Cohort II	−1.038	0.052	−0.035	349
	(1.144)	(1.075)	(0.030)	
(Below) med. Age	−2.296	−1.128	−0.019	469
	(1.043)	(0.959)	(0.034)	
Above med. Age	2.154	0.838	0.033	387
	(1.090)	(0.938)	(0.031)	
(Below) med. t. s. grad.	−1.565	−1.463	−0.002	557
	(0.906)	(0.847)	(0.028)	
Above med. t. s. grad.	1.725	1.838	0.010	299
	(1.381)	(1.114)	(0.041)	
Strata	yes	yes	yes	
Covariates	no	no	no	

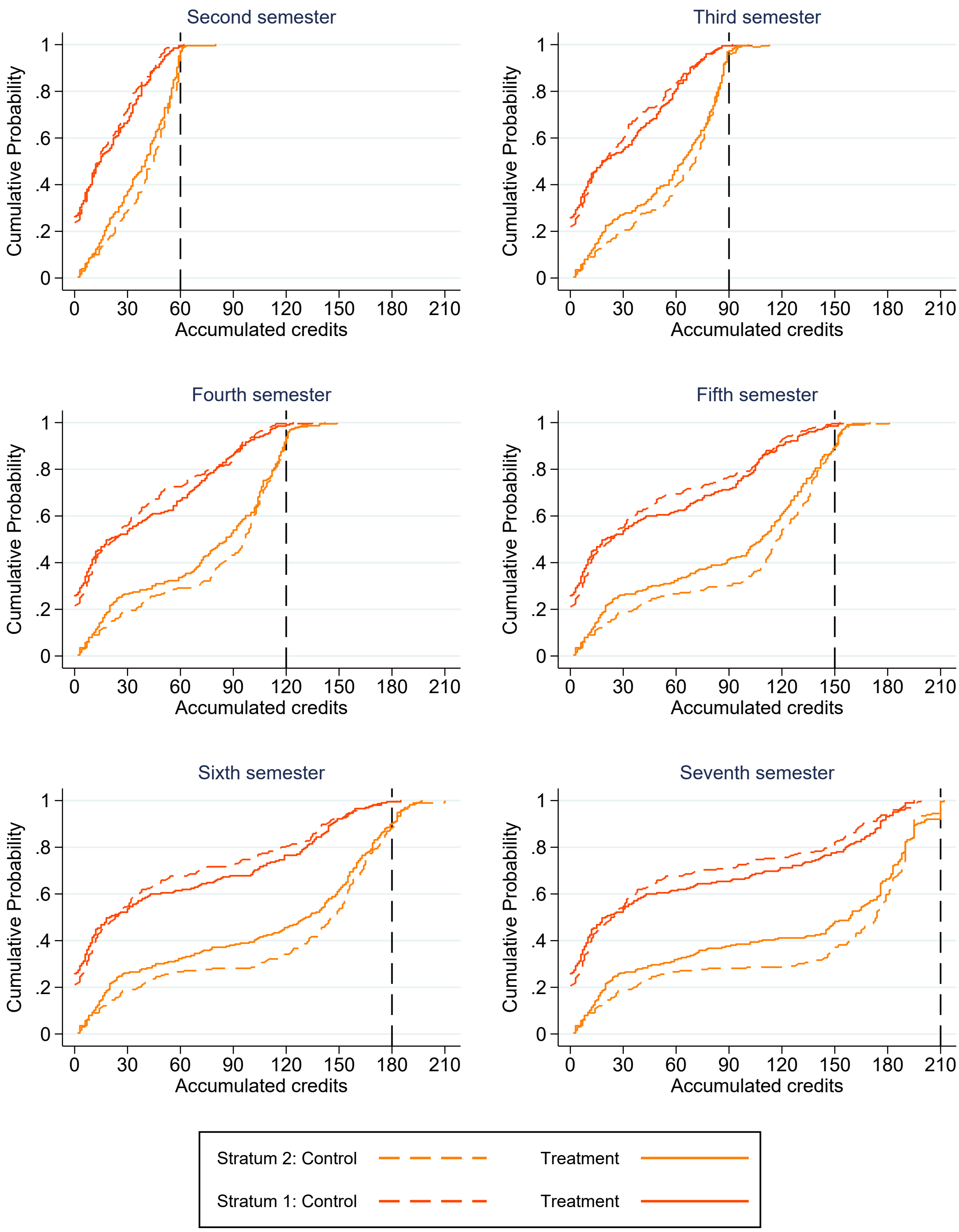
Notes: Strata: credit strata FE, study program FE, a cohort dummy, and its interaction with the study program FE; *covariates:* first semester credits, first semester GPA (missing values imputed), first semester GPA imputed dummy, first semester dropout, high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy. Robust standard errors in parentheses.

TABLE 4.A3: DESCRIPTIVE STATISTICS BY PREDICTED PERFORMANCE STRATA

	Not above average		Above average	
	(1)	(2)	(3)	(4)
	Stratum 1	Stratum 2	Stratum 3	Stratum 4
Age	23.200	22.207	22.309	22.075
	(3.849)	(2.952)	(2.745)	(2.667)
Female	0.350	0.386	0.325	0.422
High school degree Abitur	0.378	0.409	0.369	0.516
Time since high school degree	1.462	1.108	1.158	1.182
	(2.716)	(1.841)	(1.976)	(1.781)
High school GPA	2.819	2.546	2.619	2.157
	(0.631)	(0.530)	(0.504)	(0.488)
First semester credits	7.800	17.849	24.222	29.203
	(7.751)	(8.203)	(8.051)	(12.509)
First semester dropout rate	0.107	0.005	0.005	0.000
First semester GPA	3.215	2.420	2.720	1.929
	(0.509)	(0.558)	(0.467)	(0.461)
First semester GPA N/A	0.341	0.000	0.005	0.000
N	431	425	379	374

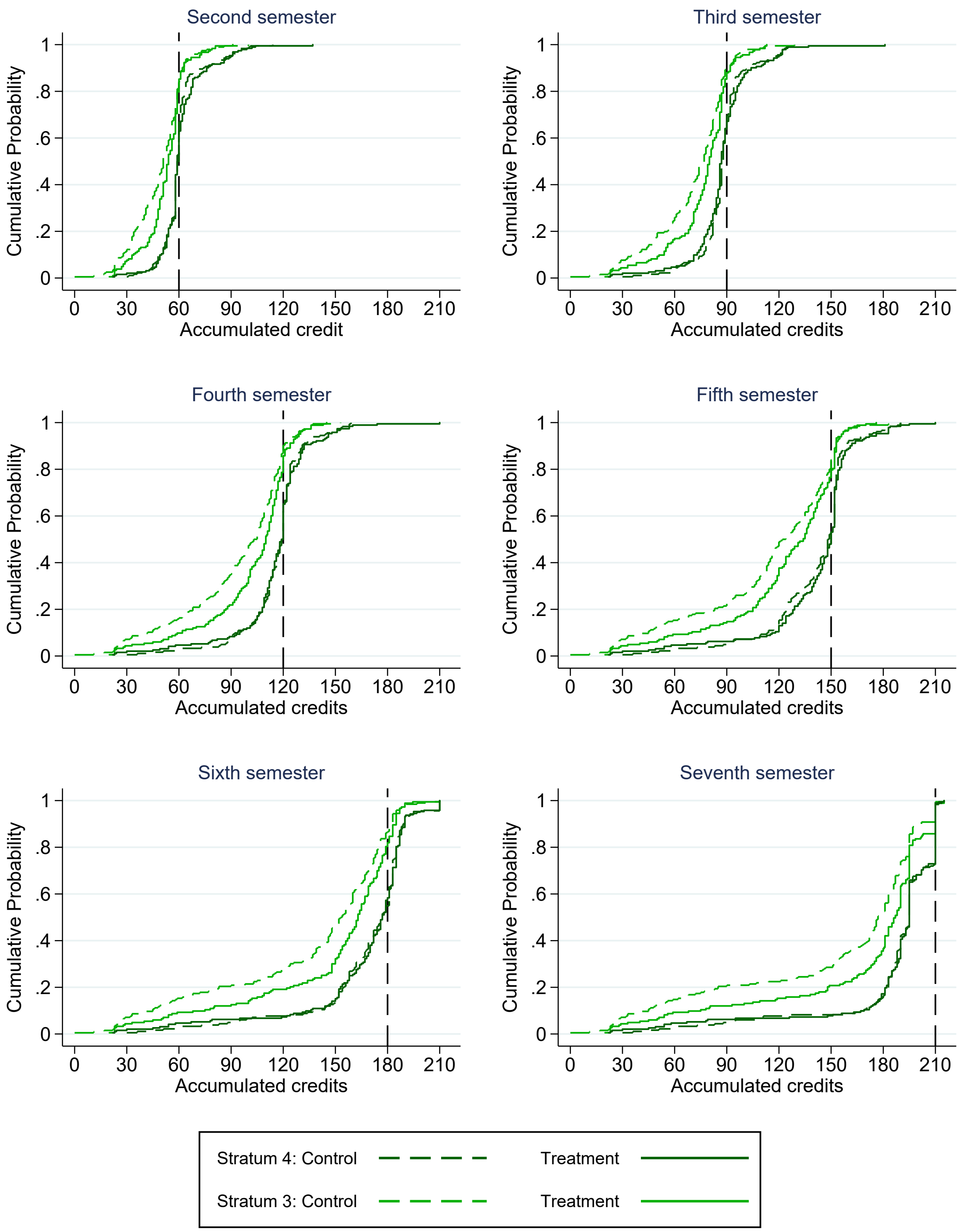
Notes: Above average and not above average refer to the position in the performance distribution at the end of the first semester. For continuous covariates the standard deviation is shown in parenthesis.

FIGURE 4.A1: CUMULATIVE DISTRIBUTIONS OF ACCUMULATED CREDIT POINTS BY PREDICTED PERFORMANCE STRATA – NOT-ABOVE-AVERAGE STUDENTS



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

FIGURE 4.A2: CUMULATIVE DISTRIBUTIONS OF ACCUMULATED CREDIT POINTS BY PREDICTED PERFORMANCE STRATA – ABOVE-AVERAGE STUDENTS



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

B Choice and performance of the prediction model

To choose a prediction model for our approach in Section 4.4.2, we rely on the following procedure. First, we estimate several models using the baseline performance, the background characteristics, and their interactions with the study program fixed effects as predictors. The prediction models differ in the way continuous covariates are included (linear or cubic functions) and in terms of the estimation method: we use OLS as well as three different lasso approaches, which vary in terms of tuning parameter selection (the extended Bayesian information criteria (EBIC), five-fold cross-validation (5-fold CV), and rigorous penalization). For the estimation of the lasso models we use Stata's *lassopack* command by Ahrens et al. (2020), who also provide an introduction to the different approaches. We then evaluate the performance of the models by comparing predicted with observed values in the control group of the cohort that is not used for the estimation of the prediction model. In terms of prediction accuracy, we find that the lasso with 5-fold CV and cubic terms for continuous covariates outperforms the other models in terms of RMSE, R^2, and Pearson's r (see Table 4.B1).

TABLE 4.B1: PERFORMANCE OF PREDICTION MODELS

			Lasso					
	OLS		EBIC		5-fold CV		Rigorous	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Linear	Cubic	Linear	Cubic	Linear	Cubic	Linear	Cubic
Cohort I								
RMSE	10.881	13.331	10.950	10.938	10.631	10.658	10.961	10.962
R^2	0.335	0.002	0.327	0.328	0.366	0.362	0.326	0.325
Pearson's r	0.579	0.048	0.572	0.573	0.605	0.602	0.571	0.570
Cohort II								
RMSE	10.729	13.520	11.076	11.190	10.451	10.365	11.134	11.130
R^2	0.422	0.082	0.384	0.371	0.451	0.460	0.377	0.378
Pearson's r	0.649	0.286	0.619	0.609	0.672	0.678	0.614	0.618

Notes: The table shows the prediction performance of various models used to predict the second semester credits in Cohort I (N = 812) and Cohort II (N = 797) based on a regression of all covariates (first semester credits, first semester GPA (missing values imputed), first semester GPA imputed dummy, first semester dropout, high school GPA, age, female dummy, time since HS degree, and HS degree Abitur dummy), study program FE, and their interaction with the other covariates using only control group data from the respective other cohort, e.g. Cohort I is used to predict the second semester credits in Cohort II. Prediction performance is evaluated based on control group regressions of observed second semester credits on predicted second semester credits. In Columns (1), (3), (5), and (7) linear functions of continuous covariates are included. In Columns (2), (4), (6), and (8) cubic functions are included instead. In Columns (1) and (2), OLS is used to estimate the prediction model. In Columns (3) to (8) Lasso models with different approaches for the selection of tuning parameters are used for estimation. In Columns (3) and (4) the extended Bayesian information criteria is used. In Columns (5) and (6) 5-fold cross-validation is used. In Columns (7) and (8) rigorous penalization is used. The lasso models are estimated with the *lassopack* commands by Ahrens et al. (2020).

Chapter 5

Helping Students to Succeed – The Long-Term Effects of Soft Commitments and Reminders

Joint work with Oliver Himmler, Robert Jäckle, and Philipp Weinschenk[*]

5.1 Introduction

How can we help students to succeed with their studies? This is an important question since (i) academic success is crucial for students themselves as well as for societies as a whole and (ii) many students fail to succeed.[1] The literature has identified two main types of failures. Many students either fail to complete their studies or take longer than scheduled.[2] We study whether we can help students to improve their success in both dimensions, i.e., reduce dropout rates and study times, by offering them a soft commitment device which is accompanied by regular reminders. The commitment device enables students to commit to the officially recommended study plan, which specifies when which course should be taken. The commitment device is soft, since failing to achieve the committed plan does not cause

[*]We gratefully acknowledge financial support from the German Federal Ministry of Education and Research under grant 01PX16003A, 01PX16003B, and administrative and financial support from the TH Nürnberg and the Max Planck Institute for Research on Collective Goods. Jäckle also gratefully acknowledges financial support from the Staedtler Stiftung.

[1]See Oreopoulos and Petronijevic (2013) for a review on the returns to higher education.

[2]For instance, in OECD countries, around 20% of students have dropped out of the higher education system completely by the end of the scheduled duration of the programs. In addition, less than 40% of a cohort complete their bachelor's degree within the scheduled study time, and, even two or three years later, more than 30% still remain without a degree (OECD, 2019). See also the evidence surveyed by Bound et al. (2010) and Bound and Turner (2011).

any material consequences.[3] This feature is quite important, since it ensures that many subjects are willing to commit. In our study, all students who got the opportunity signed the commitment.[4]

We chose our main treatment – a commitment device which is accompanied by reminders – for the following reasons. First, the literature has shown that many persons (including students) face self-control problems, which causes procrastination and thus insufficient efforts.[5] Students might be able to mitigate their self-control problems if they can use a commitment device. Second, the literature has shown that limited attention is often responsible for the inaction of subjects.[6] This problem could be tackled by reminders, who help activating subjects by bringing the task back to their minds. We conducted our study at a German university of applied sciences, where we followed all 392 incoming students of a Business Administration bachelor program over six years. Students were randomly assigned into three treatment groups: control, pure reminders, and commitment (accompanied by reminders).

We find that our soft commitment device is highly effective at raising students' success, as it increases the five-year graduation rate by 15 pp. This effect is driven both by lower dropout and earlier graduation: at the end of our observation period of six years, the dropout rate of students in the commitment group is 8 pp lower, compared to a control group mean of 35%, and they earn their degree about 0.41 semesters earlier than the control group. Pure reminders decrease the dropout rate by about 5 pp, but the estimates are not precise. For both the commitment and the pure reminders, we find no significant effects on final GPA, suggesting that students do not buy faster graduation with worse grades. If anything, our results suggest that students in the pure reminder condition have a worse final GPA than controls.

The long-run perspective of our study is important, since the ultimate goal of interventions in the educational sector is to improve students' final academic success, which one can only measure when accompanying students through their whole studies. Moreover, it is also theoretically by no means clear that a short-run success of an intervention automatically generates a long-run success. There are two reasons for this. First, the reactions towards an

[3] We follow the terminology of Bryan et al. (2010).

[4] Low take-up rates of commitment devices are often a major problem and could in some cases only be 10% (Giné et al., 2010; Royer et al., 2015).

[5] Limited self-control, for example in the form of present-bias (Laibson, 1997; O'Donoghue and Rabin, 1999), can lead to procrastination among students (Steel, 2007). Studies found that self-control problems are negatively correlated with academic success (Cadena and Keys, 2015; De Paola and Scoppa, 2015; Duckworth and Seligman, 2006; Himmler et al., 2019; Kim and Seo, 2015; Mischel et al., 1989; Wong, 2008).

[6] See Altmann and Traxler, 2014, Altmann et al., 2021, Dean et al., 2017, Ericson, 2017, and Taubinsky, 2014.

intervention might diminish quickly, such that a possible short-run success has only a limited effect on long-run outcomes. Second, after a certain time, subjects might take actions that counteract the effectiveness of an intervention. For instance, after students invested more efforts in their studies and improved their academic success due to the intervention, they might "rest on their laurels" and reduce efforts, which in turn lowers their academic success and mitigates or eliminates the long-run effectiveness of the intervention. Our study is the first that investigates the long-run effectiveness of commitment devices on academic outcomes and we show that these can work remarkably well.

The long-run perspective also allows us to study the dynamic effects of the intervention. We show that the effects start to emerge early, as students in the commitment group obtain on average about 2.5 credits more in each of the first seven semesters, i.e., during the scheduled study duration. They also complete the basic study stage, i.e., the exams scheduled for the first four semesters, at higher rates from the fourth semester onward.

Following Brown and Previtero (2020), De Paola and Scoppa (2015), Himmler et al. (2019), and Reuben et al. (2015), we consider the application date to university as a proxy for procrastinatory tendencies. We observe that students who apply in the second half of the application period have a 16 to 18 pp lower five-year graduation rate, drop out more often (by around 13 to 16 pp), and finish their degree 0.76 semesters later, which could be attributed to the higher share of procrastinators in this group. The commitment device increases graduation rates and reduces dropout among early and late appliers, but the effect is stronger for late appliers. Regarding the reduced time to graduation in the commitment group, we find that this is nearly completely driven by the effect on late appliers. In the presence of the commitment device, they graduate as quickly as early appliers. The commitment device thus completely offsets the negative effect of being a late applier. This indicates that the commitment device is a powerful tool to curb students' procrastinatory tendencies.

The commitment device is not only quite effective in improving students' academic success, but also particularly simple and cheap to implement. We implemented the device by giving students the opportunity to commit at the introductory lectures of their studies. This takes only a few minutes of time and is by no means complicated for any involved party. Moreover, the costs of the commitment device and the accompanied reminders are low. The total costs amount to less than €3.50 per student and semester. This is quite cheap, in particular in comparison to other interventions regularly studied in the educational sector, like the provision of higher grants or additional teaching staff.[7]

[7] For example, in their meta-analysis, Nguyen et al. (2019) find that an additional $1,000 of grant aid increases on-time degree completion by about 1.8 pp.

Related literature. This chapter contributes to the following strands of the literature. First, we contribute to the literature that studies the design and the effects of commitment devices. In educational contexts, deadlines are the most researched commitment device. Ariely and Wertenbroch (2002) provide evidence on the demand and effectiveness of deadlines when students can decide on the distribution of deadlines for different tasks. However, deadlines have also been found to have no (Bisin and Hyndman, 2020) or even negative (Burger et al., 2011) effects on students' performance. Anderberg et al. (2018) study deadlines that were not enforced and show that while there is a substantial demand for earlier deadlines, 74% of students fail to meet the self-imposed deadline. Baker et al. (2016) and Patterson (2018) study the effects of commitment devices in the context of massive open online courses. Baker et al. (2016) give students the possibility to commit to a certain point in time at which they would watch the course content, but find no effects on near-term engagement, and even negative effects on long-term course engagement, persistence, and performance. Patterson (2018) enables students to limit the time spent on distracting internet activities via a software tool and finds positive effects on the time spent working on the course and the final course performance. Robinson et al. (2018) give students the option to commit to improving their in-school conduct by putting money on the line and find a high demand for the commitment device that, however, fails to affect students' later behavior. Finally, Himmler et al. (2019) provide evidence on the short-run effects of the intervention that we study in this paper, and find that students who commit to the recommended study structure, sign up for, take part in, and pass more exams in the first semester of their studies. Contrary to other areas in which commitment devices were applied, there is no existing evidence in the educational context on the long-run effects of commitment devices.[8] This paper fills this gap by showing that soft commitment devices are a cheap and effective tool to improve students' long-run academic achievements.

Second, and more generally, we contribute some of the first long-term evidence to the literature studying the potential of behavioral interventions to increase academic achievements in higher education.[9] Measures such as goal setting (Chase et al., 2013; Clark et al., 2020; Morisano et al., 2010), commitment devices (Ariely and Wertenbroch, 2002; Himmler et al., 2019; Patterson, 2018), exam sign-ups by default (Behlen et al., 2020), or relative performance feedback (Brade et al., 2020; Dobrescu et al., 2021; Tran and Zeckhauser, 2012)

[8]Outside the educational sector, some papers have shown the long-run effectiveness of commitment devices (Giné et al., 2010; Kaur et al., 2015; Royer et al., 2015), while others (e.g., Ashraf et al. 2006a) have shown no long-run effects.

[9]See see, e.g., Damgaard and Nielsen (2018), Koch et al. (2015), Lavecchia et al. (2016), and Leaver (2016) for general reviews on the behavioral economics of education.

have been found to positively affect student performance in the short-run. However, contrary to traditional approaches such as grant aid (Nguyen et al., 2019) and comprehensive approaches that combine financial and academic assistance (Dawson et al., 2020; Weiss et al., 2019), studies that investigate the long-term effects of behavioral interventions on academic achievements are still sparse.[10]

The chapter proceeds as follows. In the next section we describe the key challenges for designing the commitment device. Section 5.3 explains the institutional background and our experimental design. In Section 5.4, we describe the data and the empirical approach. In Section 5.5, we present and discuss our results. Section 5.6 concludes.

5.2 Key design challenges

Improving academic outcomes is a difficult issue; see, for instance, the overview and discussion by Oreopoulos (2021). An appropriate design of the commitment device is therefore crucial. Against the backdrop that many students face self-control problems which severely interfere with their academic success (Cadena and Keys, 2015; De Paola and Scoppa, 2015; Duckworth and Seligman, 2006; Ellis and Knaus, 1977; Kim and Seo, 2015; Mischel et al., 1989; Semb et al., 1979; Solomon and Rothblum, 1984; Steel, 2007; Wong, 2008), we decided to implement a commitment device, which gives students the opportunity to commit to the recommended study structure. We faced several key challenges when designing the commitment device.

The first challenge is to ensure a high rate of participation. There should be a demand for commitment devices among individuals that are at least partially sophisticated about their self-control problems, and previous research provides evidence for this notion. For example, among students that show time-inconsistency in their behavior, Casari (2009) finds that 60% of them want to commit to a choice in the present, rather than facing the choice in the future. Augenblick et al. (2015) show that present-biasedness is positively correlated with a demand for commitment. In Houser et al. (2018), a substantial number of subjects is willing to remove a tempting choice from their choice set, even though it could be costly.

Yet, individuals might be (partially) naive about their present-biasedness (O'Donoghue and Rabin, 1999) and might therefore refuse to commit even if this would be optimal (Bryan

[10]To our knowledge, in addition to the study in Chapter 4, only Azmat et al. (2019), who study a relative feedback intervention and find null effects on long-term academic success, follow students until the completion of their program.

et al., 2010). Consistent with the idea of partial naiveté, Acland and Levy (2015) and DellaVigna and Malmendier (2006) find that many individuals overpredict their future gym attendance when signing a membership contract. Augenblick and Rabin (2019) provide evidence against substantial sophistication and estimate that no more than 24% of participants understand their present bias. In line with theoretical predictions, John (2020) finds that individuals who are not sophisticated about their self-control issues adopt commitments that are too weak. Given that many individuals are not aware of their self-control problems, it is little surprising that the willingness to use commitment devices is often rather low, which restricts the amount of individuals that can benefit from them. In Ashraf et al. (2006b), for example, only 28% of the individuals take up the offered commitment product, while in the studies by Giné et al. (2010) and Royer et al. (2015) this number is even lower, with 11% and 12%, respectively.[11] To make sure that many students are willing to use our commitment device, we decided to make it soft. In our study, all students who got the opportunity signed the commitment.[12]

The second challenge is to ensure the long-term success of the intervention. According to standard economic theory, an individual will consider all alternatives when making a decision. If the individual exhibits limited attention, it might however not consider the full set of choices. In particular, the individual might forget to take an action (Taubinsky, 2014), forget to complete a planned task (Ericson, 2017), or only focus attention on a particular set of choices (Dean et al., 2017). In education, an individual might not consider all alternatives when applying to college or forget to start the exam preparation as early as initially planned. Consistent with these theoretical ideas, students fail to (re-)apply for financial aid (King, 2004) and lack accurate information about the costs and benefits of education (Hoxby and Turner, 2015; McGuigan et al., 2016; Oreopoulos and Dunn, 2013), even though this information is often readily available. Additional evidence comes from health contexts, where forgetting is among the top reasons that individuals mention when they fail to adhere to prescribed medication (MacDonell et al., 2013; Vyankandondera et al., 2013).

Reminders may help in cases of limited attention, by refocusing attention on a decision problem or task, by making specific choices or actions more salient, or by providing previously unknown information (Ericson, 2017; Karlan et al., 2016; Taubinsky, 2014). Repeated

[11] Individuals might also decide against commitment because they value flexibility, for example due to uncertainty about future shocks. The potential trade-off between commitment and flexibility (Amador et al., 2006; Ambrus and Egorov, 2013; Bond and Sigurdsson, 2018), makes it non-trivial to design an optimal commitment device when individuals value both (Galperti, 2015).

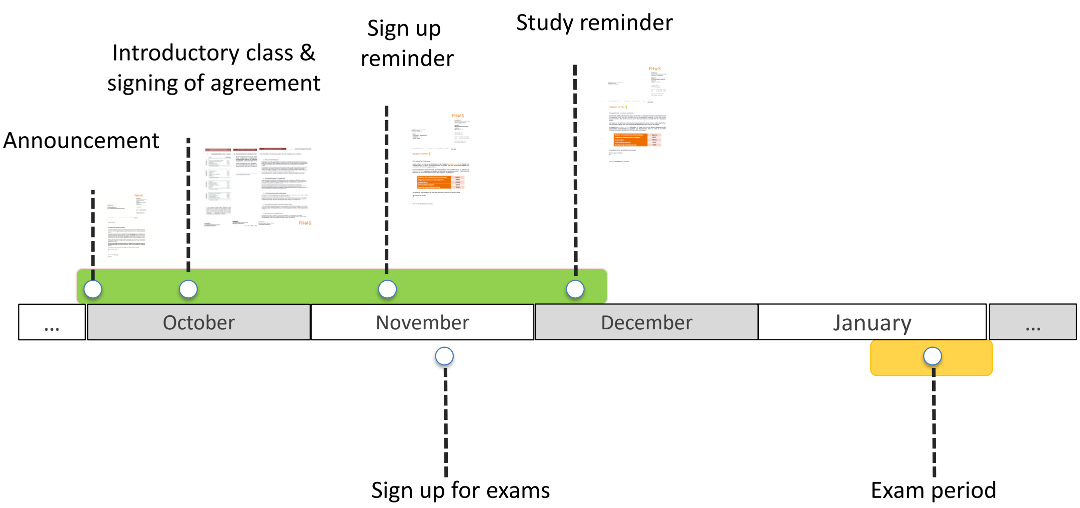
[12] The commitment device also did not raise any complaints of students when they could sign the commitment or anytime afterwards.

reminders might also help to form habits (Taubinsky, 2014). Since we gave students the opportunity to sign the commitment directly at the start of their studies, the effect of the commitment could quickly erode over time. To ensure the long-term success of the intervention, we therefore decided to remind students repeatedly on the upcoming exams in the reminder treatment and additionally on their commitment in the commitment treatment.

5.3 Institutional background and experimental design

Institutional background. The field experiment was conducted with an incoming fall semester cohort of undergraduate business administration students at a German university of applied sciences that currently serves around 9,000 students. The setting is representative for a substantial portion of the current German higher education system. For example, in the fall term of 2019, around 38% of freshman students enrolled at universities of applied sciences, and business administration ranked as the most popular degree program at German universities, accounting for 8.5% of freshman students (Statistisches Bundesamt, 2020b). The business administration program at the university where we conduct the intervention is also not particular selective. In our sample the average final high school GPA is around 2.7 (see Table 5.2), while it was 2.46 among all German high school graduates in 2013, the year in which our cohort started studying.[13]

FIGURE 5.1: TIMELINE FOR THE FIRST SEMESTER



[13]See https://www.kmk.org/dokumentation-statistik/statistik/schulstatistik/abiturnoten.html, retrieved on November 8, 2021. In the German system 1.0 is the best and 4.0 the worst final high school GPA.

TABLE 5.1: EXPERIMENTAL DESIGN

Group	Control	Reminder	Commitment
Introductory lecture	yes	yes	yes
Information folder	yes	yes	yes
Postal reminders	-	yes	yes
Commitment agreement	-	-	yes
N	131	132	129

The regular duration of the study program is seven semesters and, in accordance with the European Credit Transfer and Accumulation System (ECTS)[14], students have to obtain 210 credit points in total. According to the study structure that is recommended to finish in seven semesters (see Figure 5.B6), the first four semesters of the program form the *basic study stage*, which consists mainly of compulsory subjects and only a few electives. As is common at universities of applied sciences, the program includes a mandatory internship that is scheduled for the fifth semester. In the last two semesters students can specialize in different areas of business administration and they write their bachelor thesis.

In total, 392 students enrolled in the program and were randomly assigned to three experimental groups: control, reminder, and commitment. Table 5.1 gives an overview of the experimental design and Figure 5.1 shows the timing of the events in the first semester.

Randomization. Randomization was carried out before the first semester started using stratification and re-randomization (Morgan and Rubin, 2012). We used the final high school GPA to build four strata and within those strata we re-randomized 30 times, keeping the randomization with the best balancing properties with respect to final high school GPA, age, and gender.[15] Table 5.2 presents descriptive statistics and balancing properties for the variables that were used during the randomization and some additional background characteristics that we received from the student office only after the randomization was carried out. Overall, the covariates are well balanced across the three experimental groups.[16]

Introductory lecture. The field experiment commences with an introductory lecture in one of the first classes of the semester. Students were encouraged to attend this lecture and informed in advance that they would receive important information on how to organize their studies. Right before the introductory lecture took place, students received a personalized information folder which displayed the number of one of the three lecture halls students

[14]See `https://ec.europa.eu/education/resources-and-tools`, retrieved on November 8, 2021.

[15]The four strata based on the high school GPA are defined as follows: $1.0 \leq GPA < 2.0, 2.0 \leq GPA < 2.5, 2.5 \leq GPA < 3.0, 3.0 \leq GPA < 4.0$.

[16]Only 3 out of the 42 coefficients depicted in the table are significant at the 5%-level, which is close to the expected number of incorrect rejections of the null hypothesis when performing multiple comparisons.

TABLE 5.2: DESCRIPTIVE STATISTICS AND BALANCING PROPERTIES

	(1) Control Mean (Std. Dev.)	(2) Reminder Coefficient (Robust SE)	(3) p-Value	(4) Commitment Coefficient (Robust SE)	(5) p-Value	(6) Reminder− Commitment= 0 (Robust SE)	(7) p-Value
Male$^{a)}$	0.504	−0.049	0.426	0.000	0.999	−0.049	0.425
	(0.502)	(0.061)		(0.061)		(0.061)	
HS GPA$^{a)}$	2.659	0.002	0.927	0.005	0.829	−0.003	0.903
	(0.407)	(0.020)		(0.021)		(0.022)	
Age (first Semester)$^{a)}$	21.718	−0.172	0.699	−0.371	0.388	0.199	0.655
	(3.537)	(0.446)		(0.429)		(0.444)	
Foreigner	0.069	−0.031	0.269	−0.014	0.630	−0.016	0.529
	(0.254)	(0.028)		(0.030)		(0.026)	
Days left in application period	35.832	−4.102	0.193	−2.560	0.415	−1.542	0.623
	(25.829)	(3.149)		(3.137)		(3.133)	
Fresh HS degree	0.527	−0.050	0.423	−0.077	0.211	0.027	0.659
	(0.501)	(0.062)		(0.061)		(0.062)	
HS degree FOS	0.542	−0.042	0.491	0.038	0.529	−0.080	0.185
	(0.500)	(0.061)		(0.061)		(0.060)	
HS degree Abitur	0.412	0.042	0.470	−0.046	0.430	0.088	0.128
	(0.494)	(0.058)		(0.059)		(0.058)	
Other degree	0.046	−0.001	0.984	0.008	0.769	−0.008	0.753
	(0.210)	(0.025)		(0.027)		(0.027)	
HS degree in BY	0.626	0.003	0.962	0.001	0.980	0.001	0.982
	(0.486)	(0.060)		(0.060)		(0.060)	
HS degree in BW	0.229	−0.009	0.859	−0.112**	0.016	0.103**	0.025
	(0.422)	(0.052)		(0.046)		(0.046)	
HS degree in HE	0.061	−0.023	0.386	0.055	0.120	−0.078**	0.018
	(0.240)	(0.027)		(0.035)		(0.033)	
HS degree in other state	0.084	0.030	0.422	0.056	0.156	−0.026	0.527
	(0.278)	(0.037)		(0.039)		(0.041)	
Transferred credits	2.405	−0.099	0.943	0.279	0.863	−0.378	0.807
	(11.754)	(1.376)		(1.620)		(1.545)	
N	131	132		129			

Notes: $^{a)}$ Variables that were used during re-randomization. Column (1) presents the unadjusted control group means and standard deviations of the covariates. Columns (2) and (4) present the estimated coefficients of regressing the covariates on the reminder and the commitment indicators controlling for *strata FE*. Columns (3) and (5) test the null hypothesis of no reminder and commitment effect. Column (6) tests for the equality of the reminder and the commitment indicators, respectively. Column (7) presents the corresponding p-Value. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

were randomly assigned to. Helpers made sure that all students entered the hall they were assigned to. The content of the lecture – and of the corresponding informational folder – was standardized and gave students an overview of what was expected of them in the first four semesters, i.e., the basic study stage, and advised them to stick to the officially recommended exam schedule.

For the control and the reminder group this introductory lecture and the corresponding information folder were exactly the same (see Figures 5.B1 to 5.B3). The folder of the commitment group on the other hand contained two unsigned copies of an agreement (see Figure 5.2), which students could sign to voluntarily commit to the officially recommended exam schedule. Both in the introductory lecture and in the folder it was emphasized that the agreement is non-binding and failure to comply would carry no further consequences for the students than those that arise from the official examination regulations (see Figure

FIGURE 5.2: COMMITMENT AGREEMENT (ENGLISH)

Target agreement for <<first name>> <<last name>> **Bachelor Program Business Administration**

II. Target Agreement „Study with a Plan".

By signing this target agreement I declare that in order to successfully graduate I will adhere to the study plan. ***In particular, I will take the exams according to the „Exam plan for successful studies", as shown in section I.***

This agreement is between ***«first name» «last name»****, and the Economics Department at the University of Applied Sciences* [redacted]

Prof. Dr. [redacted], Vice Dean

Date, «first name» «last name»

5.B4). All 115 of the 129 students in the commitment group who participated in the introductory lecture signed the agreement.

Exam sign-up reminders. Each semester, students have a period of around two weeks in which they can sign up online for the exams they want to take. Students have to sign up to participate in an exam, but they can later withdraw from the participation. The official study structure recommends taking exams that sum up to 30 credits in most semesters, but students are free to sign up for fewer or more exams. In each term during the regular study duration of seven semesters, students in the commitment group received a letter shortly before the sign-up period that reminded them to sign up for the recommended exams.[17] For students who had signed the agreement at the beginning of their studies, these letters also included a sentence reminding them of the commitment they had made (see Figure 5.B5).[18]

It is conceivable that the potential treatment effects in the commitment group arise solely from reminding students. We therefore also established a treatment group which only received reminders.[19] Since students in the reminder group were not offered the commitment

[17]While the letters are initially unannounced, from the second semester onward students are likely to expect the letters and their content.

[18]This figure shows the sign-up letter students received in the first semester, which is representative of the letters students received in later semester.

[19]Students in the reminder and the commitment group who were not present at the introductory lecture did not receive reminder letters.

device, the reminder letters make no reference to commitments. This treatment group allows us to study whether pure reminder letters are able to elicit gains in academic achievement.

From the second semester onward, the letters in both groups contained a second page that showed a schedule of all exams students should have completed up to this point and the exams students were recommended to write in the current semester. It also gave them the option to check boxes next to exams that they already passed or signed up for, and to calculate the total amounts of credits they already obtained (Figure 5.B6 shows the second page for the seventh semester, the end of the regular study duration).

Study reminders. Around one month before the exam period started, students in the reminder and the commitment group received another letter, which advised them to start preparing for the upcoming exams (Figure 5.B7). For the commitment group it again mentioned the agreement, and in each of those letters we enclosed a copy of the signed agreement. These letters also contained a second page with the recommended exam schedule for the past and the current semesters. They were sent in all seven semesters of the regular study duration, too.

5.4 Empirical strategy

To analyze the long-run effects of the intervention, we use administrative data provided by the university spanning the six years that followed the start of the treatments. The primary outcomes of interest are the graduation rate (respectively the dropout rate) of students and the number of semesters it takes students to graduate.[20] We additionally look at the effects on the final GPA of graduated students to determine potential side effects of the intervention. Intermediate outcomes, such as the obtained credit points[21] and the semester by which students completed the basic study stage, are analyzed later. Unless explicitly stated differently, we provide the ITT effects from OLS estimations that compare the outcomes of the control group with the outcomes of the reminder and the commitment group.

[20]Once all students have either graduated or dropped out of the program, the graduation rate is equal to (1−dropout rate). However, this is not the case as long as some students are still enrolled.

[21]For the short-run analyses, Himmler et al. (2019) measured the number of passed exams instead of the credits. This was possible, as the exams scheduled in the first semester are each worth six credits. In later semesters, however, students also take exams that are worth more or less credits, such that the obtained credits are the more appropriate outcome measure for the long-run analysis.

The baseline specification is

$$Y_i^k = \alpha_0 + \alpha_1 Reminder_i + \alpha_2 Commitment_i + \mathbf{s_i}\boldsymbol{\alpha_3} + \varepsilon_i, \tag{5.1}$$

where Y_i^k denotes the level of outcome measure k for individual i. $Reminder_i$ and $Commitment_i$ are indicators for being randomized into the respective treatment group. The vector $\mathbf{s_i}$ includes strata fixed effects which control for the random assignment of treatment and control units within blocks. Strata are defined by the final high school GPA.

In a second specification, we add a vector $\mathbf{x_i}$ with additional covariates:

$$Y_i^k = \alpha_0 + \alpha_1 Reminder_i + \alpha_2 Commitment_i + \mathbf{s_i}\boldsymbol{\alpha_3} + \mathbf{x_i}\boldsymbol{\alpha_4} + \varepsilon_i. \tag{5.2}$$

The vector $\mathbf{x_i}$ includes the covariates that were used during re-randomization, i.e., the final high school GPA, age, and a dummy for being male. It also contains individual characteristics that were made available to us after the randomization. Namely, a dummy for being a foreigner, the number of days left in the application period when applying to the study program, a dummy for having freshly obtained the high school degree – i.e., in the year when the cohort starts their studies, dummies for high school degree *Abitur* or other degree (the reference group being a degree from a *Fachoberschule*), dummies for originating from the federal states of Baden-Württemberg, Hesse, or another state (the reference group being Bavaria), and transferred credits.[22]

5.5 Results

In this section, we present and discuss the results of the intervention. We first provide evidence on the effects on long-term academic achievement. Afterwards, we briefly study intermediate outcomes to provide further evidence on the dynamic effects of the intervention. Finally, and in accordance with the results reported in Himmler et al. (2019), we investigate whether the commitment device is particularly beneficial for students with procrastinatory tendencies.

5.5.1 Long-term effects on academic achievement

We now first investigate whether the treatments increase academic attainment, as measured by graduation and dropout rates. We then examine whether students graduate in a shorter

[22]Transfer credits are credits students obtained in previous study programs that could be transferred to the new program.

amount of time. Finally, we explore whether the intervention has an impact on final GPA.

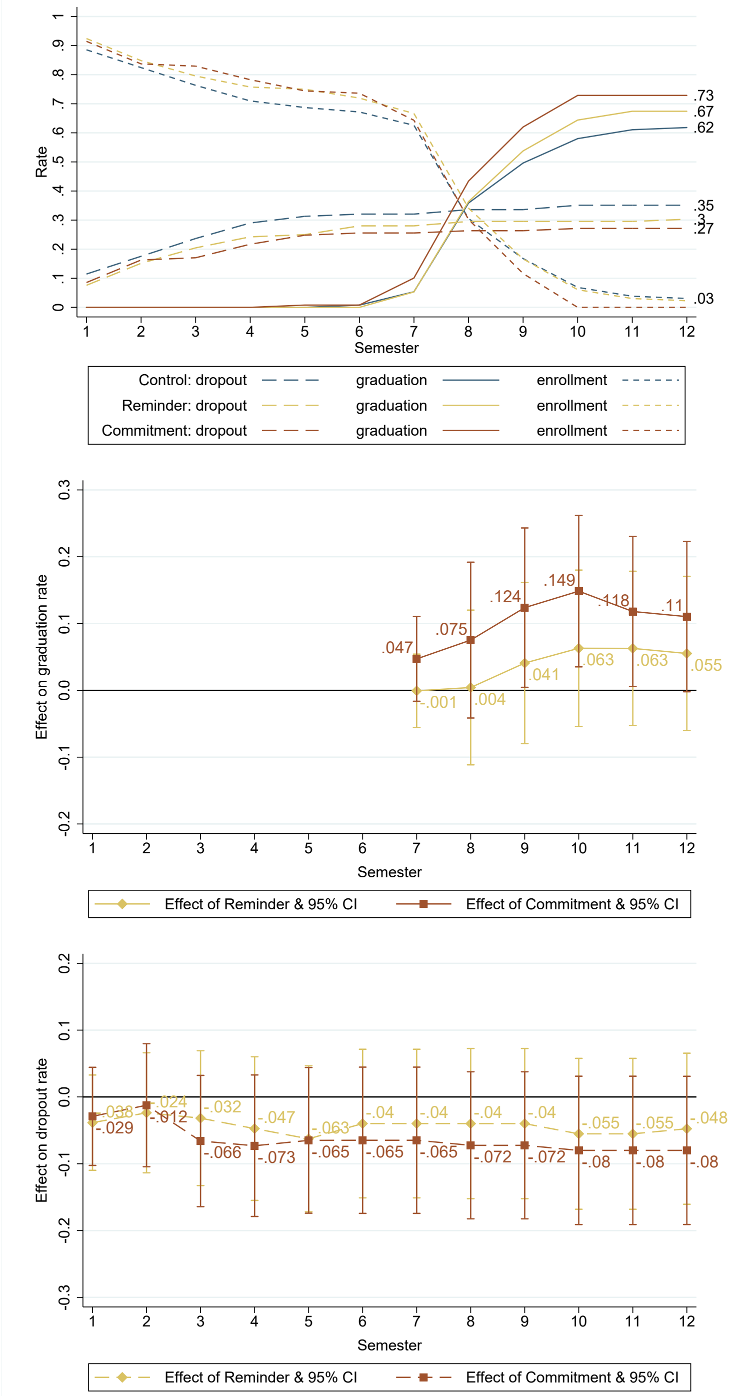
Degree attainment. The solid lines in the top panel of Figure 5.3 depict graduation rates for the three experimental groups over time. The middle panel plots treatment effects and 95% confidence intervals based on Equation 5.1 from the seventh semester onward. In the control group, only around 5% of the students manage to obtain their degree in the regular study duration of seven semesters, implying substantial room for improvement. By the end of the twelfth semester, 62% percent of control group students have graduated with a degree.

In the commitment group, we find that students are 4.7 pp (p = 0.129) more likely to have graduated by the seventh semester. The gap in graduation rates steadily increases until the tenth semester. At this point, the graduation rate in the commitment group is 14.9 pp ($p = 0.011$) higher compared to the control group (58%) and all students in the commitment group have either graduated or dropped out of the program. Consequentially, in the two following semesters, the gap between the two groups shrinks slightly, as some students in the control group still manage to graduate (the top panel shows that 3% of the students in the control group are still enrolled at the end of the twelfth semester). Table 5.A2 shows that the estimates are robust to the inclusion of the covariates.

The pure reminder treatment does not increase graduation rates to the same extent. Students in this group only start graduating at higher rates from the ninth semester onward and the biggest difference in comparison to controls is in the tenth semester with 6.3 pp ($p = 0.277$).

Taken together, these results provide first evidence that a commitment device is able to elicit increases in long-term academic attainment. They also provide tentative evidence that pure reminders, while not as effective as the commitment device, may still have an effect on academic achievement. The results also pose the question as to how the effects on the graduation rate come about. Do they arise from students dropping out at lower rates, from students simply graduating faster, or a combination of the two?

FIGURE 5.3: GRADUATION RATES, DROPOUT RATES, AND TREATMENT EFFECTS OVER TIME.



Notes: The top panel shows unadjusted dropout, graduation, and enrollment rates over time by treatment status. The middle panel shows treatment effects on the graduation rate, i.e., whether a student graduated before or in the respective semester. The bottom panel shows treatment effects on the dropout rate, i.e., whether a student dropped out of the study program before or in the respective semester. Coefficients in the middle and the bottom panel are based on OLS regressions with strata fixed effects estimated separately for each semester. The 95% confidence intervals are based on robust standard errors.

Dropout behavior. The top panel of Figure 5.3 also visualizes the dropout behavior, i.e., the share of students in each experimental group that dropped out of the study program before or in the respective semester; the corresponding treatment effects and 95% confidence intervals are shown in the bottom panel.

TABLE 5.3: TREATMENT EFFECTS ON TWELFTH SEMESTER DROPOUT RATE AND FINAL GPA

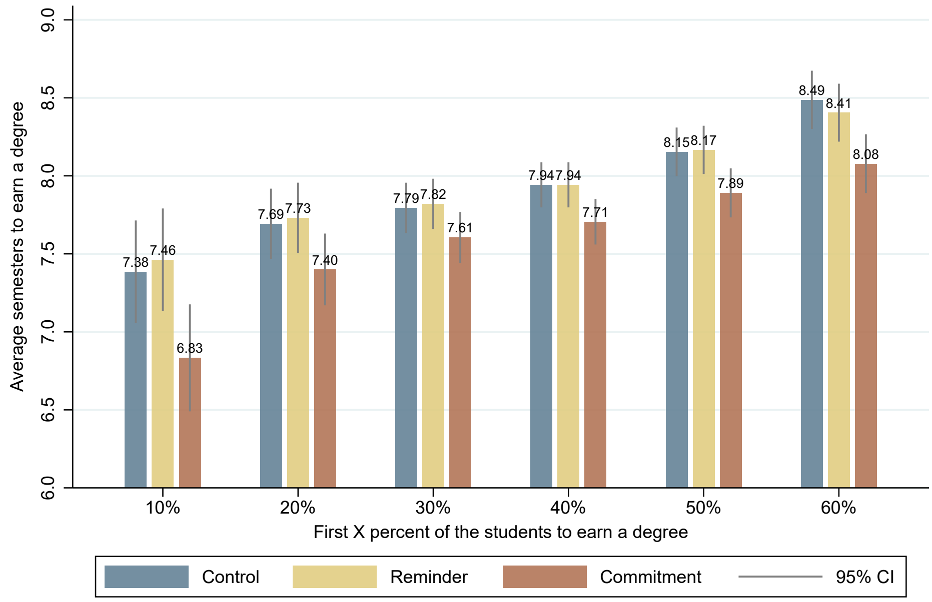
			Final GPA			
	Dropout		OLS		IPW	IPWRA
	(1)	(2)	(3)	(4)	(5)	(6)
Reminder	−0.048	−0.061	0.062	0.077	0.091*	0.074
	(0.058)	(0.057)	(0.054)	(0.049)	(0.048)	(0.047)
Commitment	−0.080	−0.088	−0.040	−0.035	−0.020	−0.036
	(0.056)	(0.056)	(0.059)	(0.057)	(0.054)	(0.054)
Strata	yes	yes	yes	yes	see notes	
Covariates	no	yes	no	yes	see notes	
N	392	392	264	264	264	264
Control mean	0.35	0.35	2.19	2.19	2.18	2.19
(Std. deviation)	-	-	(0.40)	(0.40)	-	-

Notes: Dropout: indicates if a student dropped out of the study program before or in the twelfth semester; *final GPA:* only includes students that have obtained a degree by the end of the twelfth semester; includes passing grades only; highest passing grade is 1.0, lowest passing grade is 4.0; *strata:* HS GPA strata FE; *covariates:* HS GPA, age, male dummy, foreigner dummy, application date (days left), fresh hs degree dummy, HS degree Abitur dummy, other degree dummy, HS degree in BW dummy, HS degree in HE dummy, HS degree in other state dummy, credits transferred from previous studies; *IPW:* Inverse probability weighting using the strata and covariates to predict treatment assignment among students that obtain a degree by the end of the twelfth semester. *IPWRA:* inverse-probability-weighted regression adjustment, using the strata and covariates for the regression adjustment. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

In all three groups, most of the dropouts are in the first four semesters. Descriptively, the dropout rates in the commitment and the reminder group start to be considerably lower from the third and fourth semester onward. By the end of the twelfth semester, 35% of the students in the control group have dropped out of the study program without a degree. For students in the commitment group, the dropout rate is decreased by 8 pp ($p = 0.157$), and for students who receive pure reminders it is 4.8 pp lower ($p = 0.409$; see Column 1 of Table 5.3). Including the covariates in Column (2) slightly changes the coefficients to 8.8 pp ($p = 0.118$) and 6.1 pp ($p = 0.280$), respectively.

While none of the reported coefficients in Figure 5.3 and Table 5.3 are significant at any conventional level, they are still sizable. For the reminder group they are, in absolute numbers, pretty close to the effect on the graduation rate and for the commitment group they correspond to about half of the effect. These results thus suggest that the effects on graduation rates are at least partly driven by lower dropout rates and not only by students graduating faster.

FIGURE 5.4: AVERAGE TIME TO DEGREE



Notes: Average semesters to earn a degree is calculated as the average number of semesters it took for the first X percent of degree earners to earn their degree. The 95% confidence intervals are based on robust standard errors from unadjusted OLS regressions.

Time to degree. Investigating the effects on time to degree is complicated by the fact that it is only observed for students who have already graduated, i.e., students who have not dropped out and who are no longer studying. A simple comparison of time to degree among all students who have graduated by the end of the twelfth semester would thus likely be biased.

To overcome this challenge, we employ an approach similar to that in Weiss et al. (2019) and focus on parts of the time to degree distribution that are observed for all groups. The approach makes use of the fact that the remaining students who will manage to graduate will do so in a greater amount of time than students who have already graduated. Thus, for all three experimental groups the “fastest” or first X% of degree earners are already observed, where X% is the share of students who graduated with a degree by the end of the twelfth semester. With 62%, the control group has the lowest share of graduates, and we will therefore only look at time to degree among the first 62% of graduates in the treatment groups.

Figure 5.4 depicts the average time to degree for the first 10 to 60% of degree earners, separately by treatment status. The figure provides the following insights: First, the commitment device reduces the time to degree across the complete distribution. The largest differences are observed for the first 10% of degree earners, for whom it reduces time to degree by .55 semesters ($p = 0.024$) compared to controls (7.38 semester), and for the first 60%

of degree earners, for whom it reduces time to degree by .41 semesters ($p = 0.003$) compared to controls (8.49 semesters). Second, the pure reminder treatment does not have any effects on average time to degree along the parts of the distribution that are observed for all experimental groups. This suggests that the effects of the reminder on graduation rates observed in Figure 5.3 are only driven by reduced dropout and not by faster graduation.

The analysis of the effects on time to graduation thus provides evidence that the soft commitment device is not only able to increase the number of students who graduate with a degree, but it also enables these students to do so faster, giving them earlier access to the returns to their education.

Sensitivity of effects on time to degree. One concern regarding the analyses and interpretation of the results presented so far may be that they are influenced in large parts by the definitions of dropout and graduation that we have to rely on. Namely, students are considered to be dropouts when they drop out of the business program under observation, since we no longer observe them. However, many of those students may actually not drop out of the higher education system completely. Instead, they may go on to pursue and obtain a degree at another university. Taking this perspective, it is conceivable that the results presented so far present an upper bound of the more general effects on dropout and graduation. In the most pessimistic case, all dropouts would go on to obtain a degree at another university and our intervention would then have no effect on the overall graduation rate. However, even in this case, our intervention could still affect the time it takes to obtain a degree, which can have substantial benefits on its own.

For simplicity, and as it presents the most pessimistic case for the effects of our intervention, we assume below that all students who drop out obtain a degree at another university. One might be concerned that our treatments somehow make dropouts less likely to continue their studies at another institution and that the assumption thus leads us to overstate the effect. The administrative data includes some information on this, as it indicates whether dropouts are planning to enroll at another university. The data indicates that 23.9% of control, 20.0% of reminder, and 31.4% of commitment dropouts plan to study at another university, suggesting that students in the commitment group are, if anything, more likely to do so.

To investigate the effects on time to degree including students who drop out, we employ the following procedure that uses information on time to graduation for all university degrees obtained in Germany in 2019 from the Federal Statistical Office (Statistisches Bundesamt, 2020a). For students who are still enrolled, we simply assume that they finish in the thirteenth semester. This will work against us finding effects, since in reality some of them will likely be slower and because time to degree is already observed for all graduates in the commitment group. To impute the time to degree for students who dropped out, we randomly draw values from time to degree distributions based on the data from the Federal Statistical Office and add this to the number of semesters until dropout. By doing so, we obtain a time to degree value for every student in our sample, and we therefore no longer need to be concerned about sample selection. Using the imputed time to degree as outcome, we then estimate treatment effects based on Equations 5.1 and 5.2. We repeat this process 10,000 times, recovering the average time to degree in the control group and the treatment effect coefficients from each iteration.

TABLE 5.4: TREATMENT EFFECTS ON TIME TO DEGREE – SENSITIVITY ANALYSES

	Assumption I		*Assumption II*	
	(1)	(2)	(3)	(4)
Control				
Median time to degree	9.359		9.069	
95% interval	[9.092, 9.626]		[8.898, 9.259]	
99% interval	[9.023, 9.725]		[8.844, 9.319]	
Reminder				
Median effect	-0.047	-0.016	0.012	0.040
95% interval	[-0.412, 0.320]	[-0.380, 0.354]	[-0.235, 0.264]	[-0.209, 0.293]
99% interval	[-0.533, 0.441]	[-0.499, 0.469]	[-0.303, 0.337]	[-0.280, 0.371]
Commitment				
Median effect	-0.460	-0.457	-0.420	-0.433
95% interval	[-0.815, -0.099]	[-0.823, -0.093]	[-0.668, -0.178]	[-0.686,-0.186]
99% interval	[-0.921, 0.009]	[-0.934, 0.013]	[-0.743, -0.104]	[-0.760,-0.114]
Strata	yes	yes	yes	yes
Covariates	no	yes	no	yes
N	392	392	392	392

Notes: The table depicts the median average time to degree estimate in the control group, the medians of the treatment effect estimates for the reminder and the commitment group, and intervals that span 95% (99%) of the respective estimates, which are recovered from 10,000 iterations of the following procedure. For students who are still enrolled after the twelfth semester, time to degree is set to thirteen semesters. For all students who drop out, it is assumed that they subsequently earn a degree at another institution, and their total time to graduation is hence the sum of the time until dropout and the time until they obtain the new degree. For *Assumption I*, we assume that dropouts go on to obtain any degree at another university. The top part of Figure 5.A3 shows the corresponding time to degree distribution from which we randomly draw in each of the 10,000 iterations. For *Assumption II*, we assume that dropouts go on to obtain a business or economics degree, which allows them to transfer credits. To account for this, we reduce the time to degree at the new university by their study progress before dropping out, but we assume that it takes them at least one semester at the new university to obtain their degree. To calculate the progress in semesters, we divide the number of obtained credits at dropout by 30, which – in accordance with the ECTS – is the supposed course load of one semester. The bottom part of Figure 5.A3 shows the corresponding time to degree distribution from which we randomly draw in each of the 10,000 iterations. *Strata:* HS GPA strata FE; *covariates:* HS GPA, age, male dummy, foreigner dummy, application date (days left), fresh hs degree dummy, HS degree Abitur dummy, other degree dummy, HS degree in BW dummy, HS degree in HE dummy, HS degree in other state dummy, credits transferred from previous studies.

We report results for two different assumptions regarding the time to degree of dropouts: For *Assumption I*, we suppose that all students who drop out will go on to pursue any degree at a German university (the corresponding time to degree distribution is shown in the upper part of Figure 5.A3). This means that they could, for example, also pursue master programs which usually take less time then undergraduate degrees. As this mostly includes study programs that are unrelated to business studies, we assume that students are not able to transfer any credits to their new program that they obtained before dropping out. For *Assumption II*, we suppose that students continue to pursue a business or economics undergraduate degree at another university, which will allow them to transfer credits, thereby shortening the study duration (the distribution is shown in the bottom part of Figure 5.A3). To account for this, we reduce the study duration drawn from the distribution by the study progress at the time of dropout, assuming that it takes them at least one semester at the new university to graduate. We calculate the study progress at dropout by dividing the number of obtained credits by 30, i.e., the supposed course load of one semester according to the ECTS.[23] By doing so, *Assumption II* should lead to smaller estimates for the effects on time to degree.

Table 5.4 reports the median estimates of all iterations as well as respective 95 and 99% intervals. The top row shows that the median average time to degree estimates among students in the control group are 9.36 and 9.06 for *Assumption I* and *II*, respectively. As expected, assuming that dropouts go on to study another business or economics degree but can transfer credits leads to a lower estimated time to degree. In line with the results presented above, we find no evidence that students in the reminder group obtain their degrees faster (see middle rows). The commitment device, on the other hand, reduces time to degree by around 0.46 semesters under *Assumption I* (bottom rows in Columns 1 and 2) and by about 0.42 semesters under *Assumption II* (Columns 3 and 4). Both estimates are thus very close to the effect on time to degree that we estimated among the first 60% of degree earners.

Taken together, these results provide evidence that even in the scenario in which all dropouts manage to complete a degree at another university, the commitment device still shortens time to graduation by 0.42 to 0.46 semesters, giving students considerable faster access to their educational returns.

Final GPA. It is conceivable that reduced time to graduation could come at the cost of worse grades. Therefore, in Columns (3) to (6) of Table 5.3, we show the treatment effects on students' final GPA. The estimates in Column (1) indicate that students in the reminder group have a final GPA that is 0.062 grade points worse ($p = 0.250$) compared to controls (the best grade is 1.0, while the worst grade is 4.0). For students in the commitment group the final GPA is 0.040 grade points better ($p = 0.504$). Adding covariates in Column (2) leaves those estimates largely unchanged.

[23]The two assumptions we make (students being able to study any degree, even graduate programs that take less time, and assuming that students can transfer all credits that they have previously obtained), should both work against us finding results. Both reduce the time of the next degree students attempt. As there are more dropouts in the control group, this group will be more affected by those reductions in time to degree.

One caveat with these estimates is that they could be affected by a selection bias due to the different graduation rates. To address this, in Column (3), we employ inverse probability weights to reweigh observations according to their treatment assignment probability. In Column (4), we extend on this by also performing regression adjustments, giving the estimates the double-robust property. Inverse probability weighting alone increases the negative effect of the reminder letter on GPA to 0.091 grade points ($p = 0.060$) and reduces the positive effect of the commitment device to 0.020 ($p = 0.711$). Additionally performing regression adjustment produces roughly the same estimates as the OLS specification with all controls in Column (2). Another caveat is that we do not observe the final GPA for students who are still studying, yet. In Table 5.A3, we include those students and replace the unobserved final GPA with their GPA from the end of the twelfth semester. Doing so keeps the estimates for the reminder letter mostly unchanged and for the commitment group, if anything, it increases the magnitude of the effects to about 0.060 grade points.

In sum, these estimates suggest that the commitment device may increase the final GPA by 0.020 to 0.060 grade points (0.050 to 0.150 control group standard deviations), albeit none of these estimates are significant on any conventional level. More importantly, they provide no evidence that the soft commitment device leads to a decrease in the final GPA and the positive effects on students' graduation rate, the reduction in dropout, as well as the reduced time to degree can therefore be interpreted as net gains. The pure reminder, on the other hand, may worsen the final GPA by 0.060 to 0.090 grade points (0.150 to 0.225 control group standard deviations), although most of these estimates are not precise.

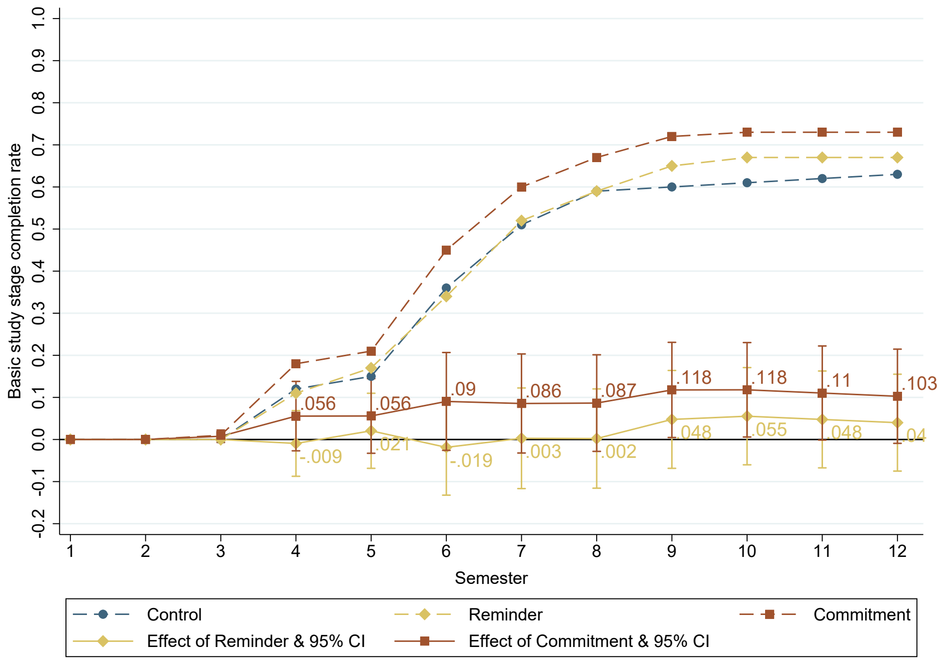
5.5.2 Effects on intermediate measures of academic achievement

How do the effects on long-term academic achievement come about? We first investigate how the treatments affect the basic study stage completion. Afterwards, we study students' credit accumulation over time.

Basic study stage completion. The basic study stage consists of 120 credits and encompasses all courses that are regularly scheduled for the first four semesters (see Figure 5.B6). Almost all of these courses are mandatory, leaving no room for treated students to select into easier electives. The dashed lines in Figure 5.5 indicate the shares of students who completed the basic study stage before or in the respective semester, separately by treatment. The figure also depicts coefficients and corresponding 95% confidence intervals from estimating Equation 5.1 separately for each semester. For the control group, it emerges that only 12.2% (15.3%) of students complete the basic study stage by the end of the fourth (fifth) semester.

After that, the share increases steadily. However, even by the end of the eighth semester, only 58.8% managed to complete it.

FIGURE 5.5: TREATMENT EFFECTS ON BASIC STUDY STAGE COMPLETION

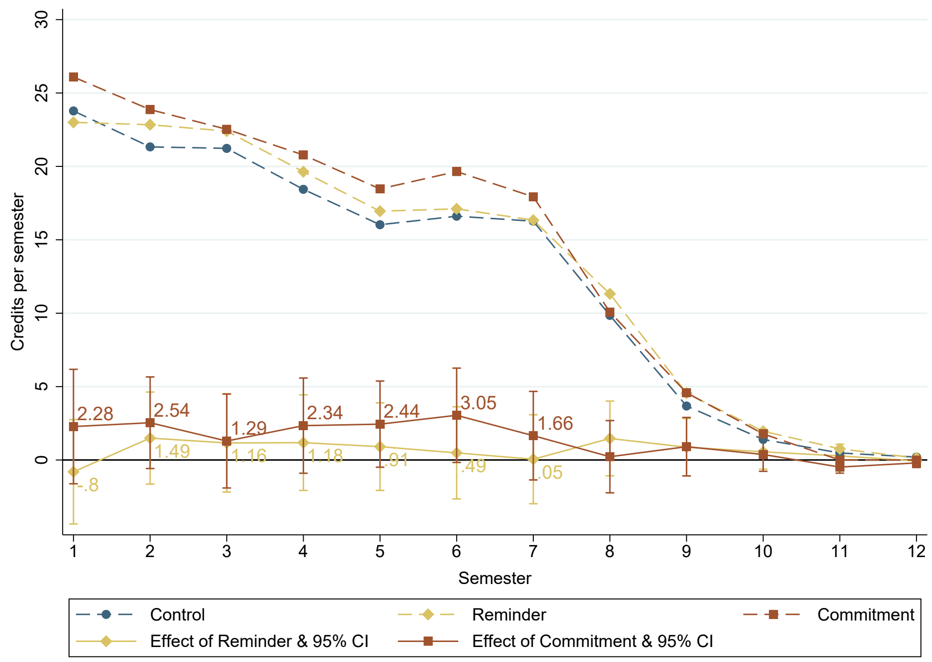


Notes: Dashed lines depict the unadjusted mean of the basic study stage completion rate. Coefficients based on OLS regressions with strata fixed effects estimated separately for each semester. *Outcome variable:* indicates if a student completed the basic study stage before or in the respective semester. The 95% confidence intervals are based on robust standard errors.

In line with earlier graduation, students in the commitment group complete this key part of their studies earlier. Already by the fourth semester, students in the commitment group are 5.6 pp (p = 0.186) more likely to have completed the basic study stage and the gap gets larger afterwards. Figure 5.A1 shows that the first 60% of students in the commitment group who complete the basic study stage do so .51 semesters (p = 0.015) earlier than than their counterparts in the control group (6.1 semesters). For students in the reminder group, we only find a small and not statistically significant increase in the completion rate from the ninth semester onward. Figure 5.A1 suggests that, as in the case of the graduation rate, this is only due to reduced dropout and not due to faster completion of the basic study stage.

Credit accumulation. The analysis of the credit accumulation allows a closer look at what time during the study the treatments affect students' academic achievement. Figure 5.6 depicts the average credits the students of the three experimental groups obtain in each semester and the treatment effects based on Equation 5.1 and their 95% confidence intervals. It shows that during the regular study duration of seven semesters, students in the commitment group consistently earn about 2.5 credits more per semester. As large shares of students start to graduate from the eighth semester onward, and because graduates no longer obtain credits, the gap shrinks to zero in the semesters after the regular study duration. For the group that receives pure reminders, the data provide little evidence of increased credit accumulation, which is in line with the observation that these students do not graduate faster.

FIGURE 5.6: TREATMENT EFFECTS ON CREDITS PER SEMESTER



Notes: Dashed lines depict the unadjusted mean of the credits per semester. Coefficients based on OLS regressions with strata fixed effects estimated separately for each semester. *Outcome variable:* credits per semester. The 95% confidence intervals are based on robust standard errors.

Column (1) in Table 5.5 shows estimates for the effects on credit accumulation during the regular study duration. Students in the commitment group earn on average 2.23 credits ($p = 0.096$; 0.162 control group standard deviations) more per semester compared to students in the control group (19.10). The inclusion of covariates in Column (2) increases the effect to 2.65 credits ($p = 0.046$; 0.191 standard deviations). For the reminder letter group, the estimates are not very precise, but they indicate that students in this group may earn between 0.64 to 0.97 credits more per semester (0.046 to 0.070 standard deviations).

One caveat of the previous analysis is that it includes students who are no longer enrolled in the program and who therefore earn zero credits in the respective semesters. While this has the advantage of keeping the initial sample unchanged and thus the estimates unbiased,

TABLE 5.5: TREATMENT EFFECTS ON CREDITS PER SEMESTER DURING THE SCHEDULED STUDY DURATION

	Unconditional		Conditional			
	OLS		OLS		IPW	IPWRA
	(1)	(2)	(3)	(4)	(5)	(6)
Reminder	0.640	0.970	−0.196	−0.236	−0.138	−0.033
	(1.349)	(1.307)	(0.766)	(0.767)	(0.716)	(0.749)
Commitment	2.228*	2.651**	1.461**	1.688**	1.772***	1.659**
	(1.334)	(1.323)	(0.720)	(0.703)	(0.668)	(0.703)
Strata	yes	yes	yes	yes	see notes	
Covariates	no	yes	no	yes	see notes	
N	2744	2744	2245	2245	2245	2245
Control mean	19.10	19.10	24.12	24.12	24.14	24.07
(Std. deviation)	(13.86)	(13.86)	(11.02)	(11.02)	-	-

Notes: The scheduled study duration spans the first seven semesters. Estimates from pooled OLS estimations. For each semester, the conditional estimates include only students who are still enrolled at the beginning of that semester. *Outcome variable:* credits per semester; *strata:* HS GPA strata FE; *covariates:* HS GPA, age, male dummy, foreigner dummy, application date (days left), fresh hs degree dummy, HS degree Abitur dummy, other degree dummy, HS degree in BW dummy, HS degree in HE dummy, HS degree in other state dummy, credits transferred from previous studies. *IPW:* Inverse probability weighting using the strata and covariates to predict treatment assignment in each semester among students that are still enrolled at the beginning of the semester. *IPWRA:* inverse-probability-weighted regression adjustment, using the strata and covariates for the regression adjustment. Robust standard errors clustered at the student level in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

the estimates are likely influenced by the differences in dropout behavior. It is thus interesting to also investigate the credits earned conditional on still being enrolled at the beginning of the respective semester. Columns (3) and (4) in Table 5.5 provide the corresponding estimates of the effects on average credits per semester during the regular study duration. Conditional on still being enrolled at the beginning of the semester, students in the commitment group earn 1.46 credits ($p = 0.043$; 0.133 control group standard deviations) more credits in each semester of the regular study duration compared to the controls (24.12). Including all covariates increases the coefficient to 1.69 credits ($p = 0.017$; 0.153 standard deviations). The attenuated estimates suggest that part of the effect on credit accumulation is indeed driven by differences in dropouts. In line with the finding that the pure reminder does not lower the time to graduation, the estimated effects are close to zero. Because the conditional estimates are based on a selected sample, in Columns (5) and (6), we also provide estimates using inverse probability weighting and inverse probability weighting in combination with regression adjustment. The corresponding estimates are very close to the OLS estimates with all covariates in Column (4).

In sum, these results suggest that students in the commitment group accumulate more credits per semester during the whole span of the regular study duration, which is in line with the substantial reduction in the time to degree and the large increase in the graduation rate by the end of the tenth semester.

5.5.3 Procrastinators and commitment

Identifying procrastinators. To identify procrastinators, we follow the reasoning of Brown and Previtero (2020), De Paola and Scoppa (2015), and Reuben et al. (2015) and consider the application date to university as a proxy for procrastinatory tendencies (see also Himmler et al. (2019) for a detailed discussion of this approach). While not all individuals who apply late are necessarily procrastinators, the approach assumes that procrastinators are more likely to apply towards the end of the application period compared to non-procrastinators, and the application date thus serves as a proxy for procrastination tendencies. The distribution of the application date is depicted in Figure 5.A2. We split the sample at the median of the application date distribution and group individuals into early and late appliers. Analyzing effects across the two subgroups provides the following insights:

Academic achievements by application date. First, we observe that late appliers' academic achievements are substantially lower in the absence of treatment. The top row of Column (1) in Table 5.6 shows that students who apply early are 17.9 pp ($p = 0.037$) more likely to have graduated by the end of the tenth semester compared to students who apply late (estimates for the other semesters are reported in Table 5.A2). Column (2) provides evidence that this gap between early and late appliers is robust to the inclusion of covariates (16.1 pp; $p = 0.061$). As Columns (3) and (4) show, early appliers are also 12.6 to 15.7 pp ($p = 0.132$ to 0.057) less likely to drop out of their study program before or in the twelfth semester. Additionally, they obtain on average 3.68 to 4.16 credits ($p = 0.062$ to 0.035) more than late appliers in each semester of the scheduled study duration (Columns 7 and 8). In Figure 5.A4, we visualize time to degree separately for the two subgroups: among late appliers, it takes the first 50% on average 8.70 semesters to earn their degree, while the first 50% among early appliers only take 7.94 semesters to graduate. The final GPA appears to be the only performance dimension on which early appliers do not outperform students who apply late (see Columns 5 and 6 of Table 5.6).

Commitment device. Second, consistent with the notion that commitment devices are particularly effective for procrastinators, we find that our soft commitment device is able to offset most of the disadvantages that late appliers face: the bottom part of the respective Columns in Table 5.6 provides evidence that commitment increases the tenth semester graduation rate by 19.8 to 20.2 pp ($p = 0.021$ to 0.015), reduces dropout by the end of the twelfth

TABLE 5.6: TREATMENT EFFECTS BY APPLICATION DATE

	Degree 10th sem.		Dropout 12th sem.		Final GPA		Credits 1st-7th sem.	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Early applier	0.179**	0.161*	−0.157*	−0.126	−0.014	−0.012	4.155**	3.682*
	(0.086)	(0.086)	(0.083)	(0.084)	(0.082)	(0.078)	(1.966)	(1.965)
Reminder	0.127	0.126	−0.145*	−0.142*	0.066	0.054	2.118	2.103
	(0.087)	(0.086)	(0.084)	(0.084)	(0.083)	(0.078)	(2.003)	(1.974)
Rem.*early app.	−0.107	−0.111	0.184	0.170	−0.013	0.039	−2.474	−2.376
	(0.119)	(0.117)	(0.115)	(0.114)	(0.108)	(0.101)	(2.678)	(2.619)
Rem.+Rem.*early app.	0.020	0.015	0.039	0.027	0.054	0.093	−0.356	−0.273
	(0.081)	(0.080)	(0.079)	(0.077)	(0.070)	(0.062)	(1.788)	(1.729)
Commitment	0.198**	0.202**	−0.117	−0.106	−0.080	−0.110	3.510*	3.422*
	(0.086)	(0.083)	(0.085)	(0.084)	(0.090)	(0.089)	(2.025)	(1.956)
Com.*early app.	−0.081	−0.064	0.056	0.030	0.075	0.140	−2.132	−1.432
	(0.115)	(0.114)	(0.113)	(0.113)	(0.122)	(0.114)	(2.655)	(2.595)
Com.+Com.*early app.	0.118	0.138*	−0.061	−0.076	−0.005	0.029	1.378	1.989
	(0.076)	(0.078)	(0.074)	(0.076)	(0.081)	(0.073)	(1.717)	(1.762)
Strata	yes	yes	yes	yes	yes	yes	yes	yes
Covariates	no	yes	no	yes	no	yes	no	yes
N	392	392	392	392	264	264	2744	2744

Notes: Degree 10th semester: indicates if a student graduated before or in the tenth semester; *dropout 12th semester:* indicates if a student dropped out of the study program before or in the twelfth semester; *final GPA:* only includes students that have obtained a degree by the end of the twelfth semester; includes passing grades only; highest passing grade is 1.0, lowest passing grade is 4.0; *credits 1st-7th semester:* credits per semester during the scheduled study duration of seven semesters; estimates from pooled OLS estimations; *strata:* HS GPA strata FE; *covariates:* HS GPA, age, male dummy, foreigner dummy, fresh hs degree dummy, HS degree Abitur dummy, other degree dummy, HS degree in BW dummy, HS degree in HE dummy, HS degree in other state dummy, number of credits transferred from previous studies. Robust standard errors in parentheses (clustered at the student level in Columns 7 and 8). * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

semester by 10.6 to 11.7 pp (p = 0.207 to 0.172), and increases the number of credits obtained per semester during the regular study duration by 3.4 credits (p = 0.081, Column 8) among students who apply late. In line with this, time to graduation is reduced by 0.73 semesters (p = 0.001) among the first 50% to earn a degree (see bottom panel of Figure 5.A4). The final GPA is, if anything, also better, and not worse.

Third, we find some evidence that the commitment is also effective for early appliers, but to a lesser extent. The graduation rate in the tenth semester is increased by 11.8 to 13.8 pp (p = 0.124 to 0.079), the twelfth semester dropout rate is decreased by 6.1 to 7.6 pp (p = 0.413 to 0.316), the number of obtained credits is increased by 1.38 to 1.99 credits (p = 0.413 to 0.260), and among the first 70% to earn a degree, time to graduation is reduced by 0.25 semesters (p = 0.185, see Figure 5.A4). Given that the application date is only a proxy, this may simply arise from there also being some procrastinators among early appliers, who also benefit from the commitment. While most of these effects are not estimated precisely, they are still sizable from a practitioner's perspective.

Pure reminders. Last, we find suggestive evidence that the pure reminder is also able to offset some of the disadvantages that late appliers face (see middle part of Table 5.6): they increase the graduation rate in the tenth semester by 12.6 (p = 0.142), reduce the twelfth semester dropout rate by 14.5 pp (p = 0.086), and increase credits per semester during the first seven semesters by about 2.1 (p = 0.287, Column 8), lessening time to graduation among the first 50% of degree earners by about 0.33 semesters (p = 0.109, see Figure 5.A4). We find little evidence for effects on final GPA, and for students who apply in the first half of the application period, the pure reminder letters have no effect. A potential explanation for these findings could be that students who apply late also suffer from lower motivation or lower engagement with their studies and are thus more likely students who are on the margin of dropout. The constant communication through the university by way of personalized reminder letters – something the university does not otherwise engage in – could lead to some of the students staying enrolled and finally graduating.

5.6 Discussion and conclusions

This chapter studies whether a commitment device can help students to succeed. We conduct a field experiment, where we follow students over a span of six years. We show that the commitment device improves students' long-run academic achievements: the five-year graduation rate is increased by 15 pp, dropouts are reduced by 8 pp, and the time to graduation is shortened by about 0.41 semesters. The commitment device helps students throughout their studies by raising the credit points earned in each semester by roughly 2.5 or 0.18 standard deviations. Pure reminders decrease dropout by around 5 pp, though the estimates are not precise. For both treatments we do not find significant effects on the final GPA, which indicates the absence of negative side effects. If anything, students in the pure reminder condition have a worse GPA than controls.

These findings are policy relevant, since they show that a commitment device increases the number of students who can benefit from the returns to higher education and it lets them do so earlier. An important feature of our commitment device is that it is soft. This ensures a high rate of participation and therefore secures that many students can benefit. In addition, the commitment device is not only effective, but also simple and cheap to implement, with costs of less than €3.50 per student and semester.

To put these results in a more general perspective, we can compare them to effects that are commonly found in causal evaluations of education interventions. Based on 1,942 effect

TABLE 5.7: MAGNITUDE OF EFFECTS AND STATISTICAL POWER

d	Degree 10th sem.	Dropout 12th sem.	Credits per sem. 1st-7th sem.	$1-\beta$	Time to deg. first 60%	$1-\beta$
(1)	(2)	(3)	(4)	(5)	(6)	(7)
0.05	2.47	-2.38	0.693	0.069	0.048	0.061
0.20	9.87	-9.54	2.772	0.362	0.193	0.237
0.30	14.81	-14.31	4.158	0.674	0.289	0.461

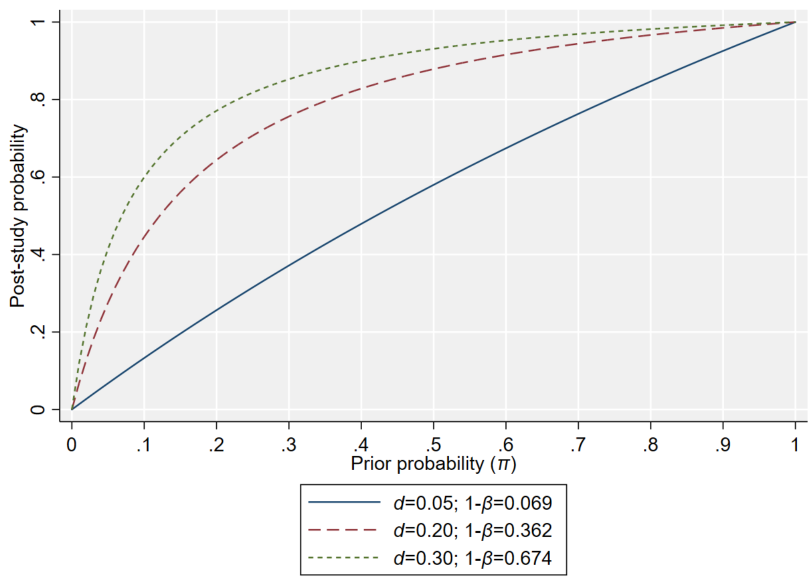
Notes: This table depicts power calculations for three effect sizes (d) and the corresponding changes in terms of the main outcome variables. Power calculations are performed with *Stata's power twomeans* command, assuming $\alpha = 0.05$ and $N = 260$ ($N = 0.6 \times 260 = 156$) in Column 5 (7). Effect sizes in terms of percentage points in Columns 2 and 3 are calculated based on the formula $d \times \sqrt{(\text{control take-up}) \times (1 - \text{control take-up})}$ and the control take up rates, i.e., the graduation and dropout rate in the control group, shown in Tables 5.A2 and 5.3, respectively. For credits per semester the effects are expressed in terms of the control group standard deviation shown in Column 1 of Table 5.5 (13.86 credits) and for time to degree we use the control group standard deviation among the first 60% of graduates, which is 0.964 semesters.

estimates from 747 randomized controlled trials, Kraft (2020) proposes the following benchmarks for effect sizes from educational interventions that target student achievement.[24] Effects of less than 0.05 standard deviations are small, effects from 0.05 to less than 0.20 standard deviations are medium, and effects of 0.20 or more standard deviations are large effects. Given this categorization, Table 5.7 depicts respective changes in our outcomes of interest that translate into effect sizes of 0.05, 0.20, and 0.30 control group standard deviations. Comparing these with our results, we can conclude that the soft commitment device has large positive effects on students' academic achievements.

Given the comparatively modest sample size of about 130 students per experimental group and the relation between statistical power and the post-study probability (PSP) that a statistically significant result is actually true (Maniadis et al., 2014), one might be concerned about the relevance of our findings. In Table 5.7, we therefore also provide power calculations for the different effects sizes. This shows that the study is not well powered to detect small or medium effects, but has around 67% power for the detection of large effect sizes, such as the effect on the graduation rate in the tenth semester. Based on the power calculations, in Figure 5.7, we follow the approach by Maniadis et al. (2014) and assess the PSP for the three effect sizes and varying priors about the probability that the soft commitment device has an effect on academic achievements. For a small effect, the power of the study is generally too low to elicit a meaningful change in the PSP across the whole range of prior probabilities. For supposed effects on academic achievement of 0.20 and 0.30 standard deviations, on the other hand, the PSPs are 64.4% and 77.1%, respectively, even for a rather low prior probability of 20% – i.e., an 80% prior probability that the soft commitment has no

[24] Kraft (2020) mainly refers to preK-12 education interventions. However, our reading of the literature is that effects of interventions in higher education tend to fall within a similar range.

FIGURE 5.7: POST-STUDY PROBABILITIES



Notes: This figure depicts post-study probabilities for the effect sizes and power calculations shown in Table 5.7 for varying priors π about the probability that the soft commitment device affects academic achievement and assuming $\alpha = 0.05$. Following Maniadis et al. (2014), the post-study probabilities are calculated as $((1-\beta)\times\pi)/((1-\beta)\times\pi+\alpha\times(1-\pi))$.

effect. This indicates that the significant results presented in this study lead to a meaningful revision of the PSP that a soft commitment device has large positive effects on students' academic achievements.

Our analysis further reveals that the commitment device is particularly helpful for students with procrastinatory tendencies. In fact, it enables procrastinators to graduate as quickly as non-procrastinators. A potential caveat of this analysis is that we rely on the date of application to university to proxy procrastination tendencies. As hinted at in the last section, it is possible that this variable also correlates with other characteristics that matter for the effectiveness of the treatments, e.g., students' motivation for their study program, and we are thus not able to provide conclusive evidence on the underlying mechanisms. Nevertheless, our results still provide evidence that the soft commitment device – and also the pure reminder – are particularly beneficial for a large share of students who otherwise suffers from substantially lower academic achievements.

Appendix

A Additional tables and figures

TABLE 5.A1: SUMMARY OF COST (IN EUROS)

Cost calculation for commitment device (cohort of 400)		
Student assistant	(7 semesters*40 hours per semester*€13.00)	€3,640.00
Introductory lecture		
Printing of information folders	(400 students*€2.30)	€920.00
Printing of information folder content	(5 pages*400 students*€0.10)	€200.00
Transparent envelopes	(400 students*€0.10)	€40.00
Reminders		
Printing of letters	(7 semesters*4 pages*400 students*€0.10)	€1,120.00
Envelopes	(7 semesters*2 letters*400 students*€0.02)	€112.00
Postage	(7 semesters*2 letters*400 students*€0.63)	€3,528.00
Total cost per semester		€1,365.71
Cost per student per semester		**€3.41**

Notes: This table summarizes the cost of the measure in Euros, in total and per student (for a cohort of 400). The postage in this table is the average postage paid during the intervention.

TABLE 5.A2: TREATMENT EFFECTS ON GRADUATION RATE

	7th semester		8th semester		9th semester		10th semester		11th semester		12th semester	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
a) Main effect												
Reminder	−0.001	0.007	0.004	0.003	0.041	0.039	0.063	0.072	0.063	0.072	0.055	0.064
	(0.028)	(0.029)	(0.059)	(0.058)	(0.061)	(0.060)	(0.060)	(0.058)	(0.059)	(0.057)	(0.059)	(0.058)
Commitment	0.047	0.051	0.075	0.078	0.124**	0.135**	0.149**	0.166***	0.118**	0.127**	0.110*	0.120**
	(0.032)	(0.034)	(0.059)	(0.059)	(0.061)	(0.059)	(0.058)	(0.056)	(0.057)	(0.056)	(0.057)	(0.56)
Strata	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Covariates	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
b) By application date												
Early applier	0.025	0.031	0.142*	0.106	0.178**	0.158*	0.179**	0.161*	0.177**	0.150*	0.161*	0.134
	(0.038)	(0.040)	(0.083)	(0.082)	(0.087)	(0.085)	(0.086)	(0.086)	(0.084)	(0.085)	(0.085)	(0.085)
Reminder	0.009	0.016	−0.003	−0.022	0.112	0.109	0.127	0.126	0.151*	0.146*	0.135	0.131
	(0.032)	(0.033)	(0.079)	(0.078)	(0.086)	(0.083)	(0.087)	(0.086)	(0.086)	(0.085)	(0.086)	(0.085)
Rem.*early app.	−0.017	−0.022	0.041	0.046	−0.124	−0.143	−0.107	−0.111	−0.159	−0.157	−0.143	−0.141
	(0.056)	(0.054)	(0.117)	(0.116)	(0.123)	(0.120)	(0.119)	(0.117)	(0.117)	(0.115)	(0.118)	(0.116)
Rem.+Rem.*early app.	−0.008	−0.006	0.038	0.024	−0.011	−0.034	0.020	0.015	−0.009	−0.010	−0.009	−0.010
	(0.046)	(0.045)	(0.086)	(0.087)	(0.087)	(0.087)	(0.081)	(0.080)	(0.080)	(0.078)	(0.080)	(0.078)
Commitment	0.029	0.025	0.152*	0.124	0.178**	0.173**	0.198**	0.202**	0.165*	0.156*	0.149*	0.141*
	(0.036)	(0.037)	(0.083)	(0.083)	(0.087)	(0.084)	(0.086)	(0.083)	(0.085)	(0.083)	(0.086)	(0.084)
Com.*early app.	0.040	0.054	−0.141	−0.098	−0.090	−0.070	−0.081	−0.064	−0.075	−0.052	−0.060	−0.037
	(0.063)	(0.063)	(0.119)	(0.118)	(0.121)	(0.120)	(0.115)	(0.114)	(0.114)	(0.114)	(0.114)	(0.114)
Com.+Com.*early app.	0.070	0.079	0.011	0.026	0.088	0.104	0.118	0.138*	0.090	0.104	0.089	0.104
	(0.053)	(0.054)	(0.084)	(0.085)	(0.084)	(0.085)	(0.076)	(0.078)	(0.075)	(0.077)	(0.075)	(0.077)
Strata	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Covariates	no	yes	no	yes	no	yes	no	yes	no	yes	no	yes
N	392	392	392	392	392	392	392	392	392	392	392	392

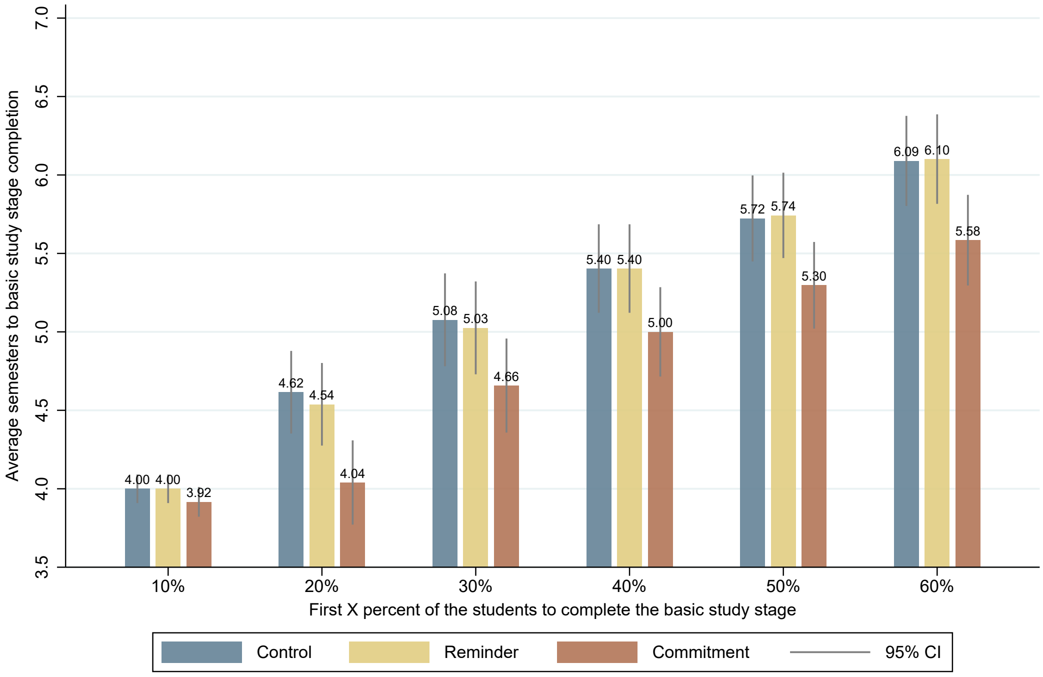
Notes: The coefficients for each semester are from separate regressions. *Outcome variable:* indicates if a student graduated before or in the respective semester; *strata:* HS GPA strata FE; *covariates:* HS GPA, age, male dummy, foreigner dummy, application date (days left; only in Panel a), fresh hs degree dummy, HS degree Abitur dummy, other degree dummy, HS degree in BW dummy, HS degree in HE dummy, HS degree in other state dummy, number of credits transferred from previous studies. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

TABLE 5.A3: TREATMENT EFFECTS ON FINAL GPA – ROBUSTNESS

	OLS		IPW	IPWRA
	(1)	(2)	(3)	(4)
Reminder	0.059	0.073	0.084*	0.064
	(0.056)	(0.049)	(0.048)	(0.047)
Commitment	−0.066	−0.056	−0.035	−0.058
	(0.061)	(0.056)	(0.054)	(0.054)
Strata	yes	yes	see notes	
Covariates	no	yes	see notes	
N	271	271	271	271
Control Mean	2.21	2.21	2.20	2.22
(SD)	(0.43)	(0.43)	-	-

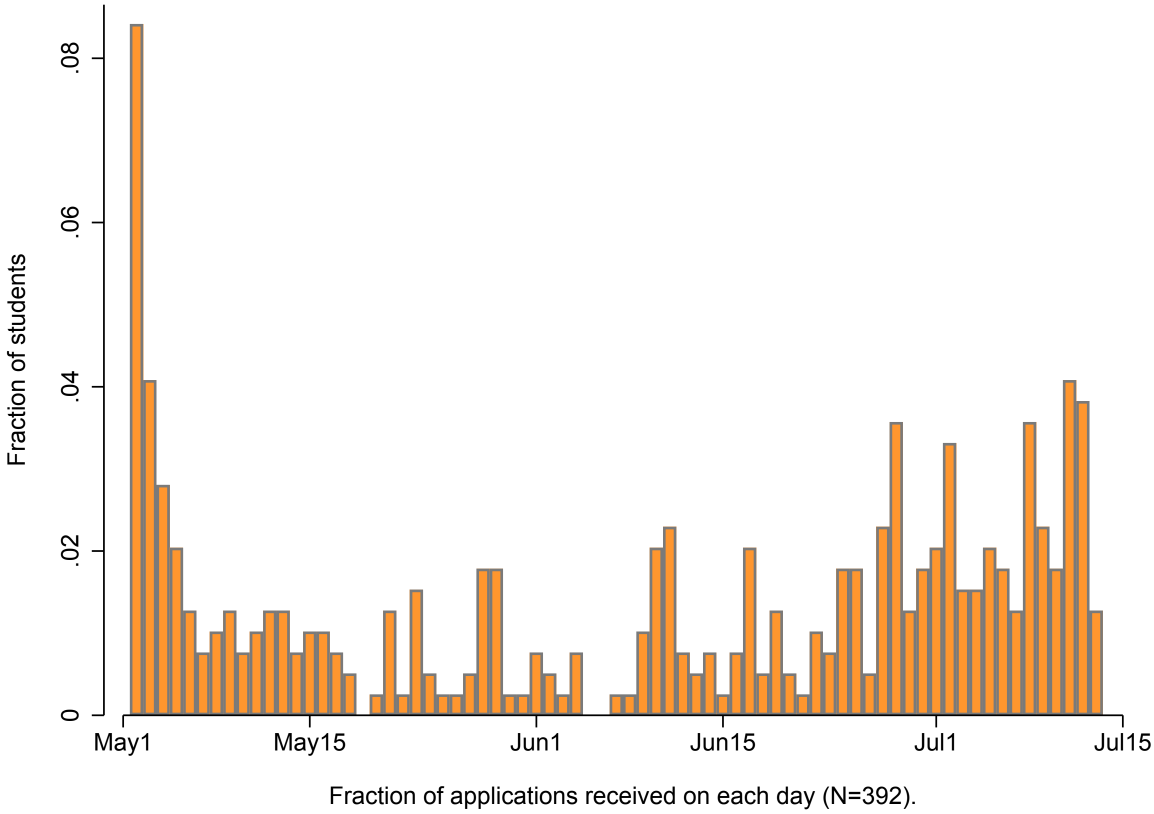
Notes: Only includes students that have obtained a degree by the end of the twelfth semester or who are still enrolled. *Outcome variable:* final GPA, including passing grades only; highest passing grade is 1.0, lowest passing grade is 4.0; for students who are still enrolled the final GPA is set to the GPA at the end of the twelfth semester; *strata:* HS GPA strata FE; *covariates:* HS GPA, age, male dummy, foreigner dummy, application date (days left), fresh hs degree dummy, HS degree Abitur dummy, other degree dummy, HS degree in BW dummy, HS degree in HE dummy, HS degree in other state dummy, credits transferred from previous studies; *IPW:* Inverse probability weighting using the strata and covariates to predict treatment assignment among students that obtain a degree by the end of the twelfth semester or who are still enrolled. *IPWRA:* inverse-probability-weighted regression adjustment, using the strata and covariates for the regression adjustment. Robust standard errors in parentheses. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

FIGURE 5.A1: AVERAGE TIME TO BASIC STUDY STAGE COMPLETION



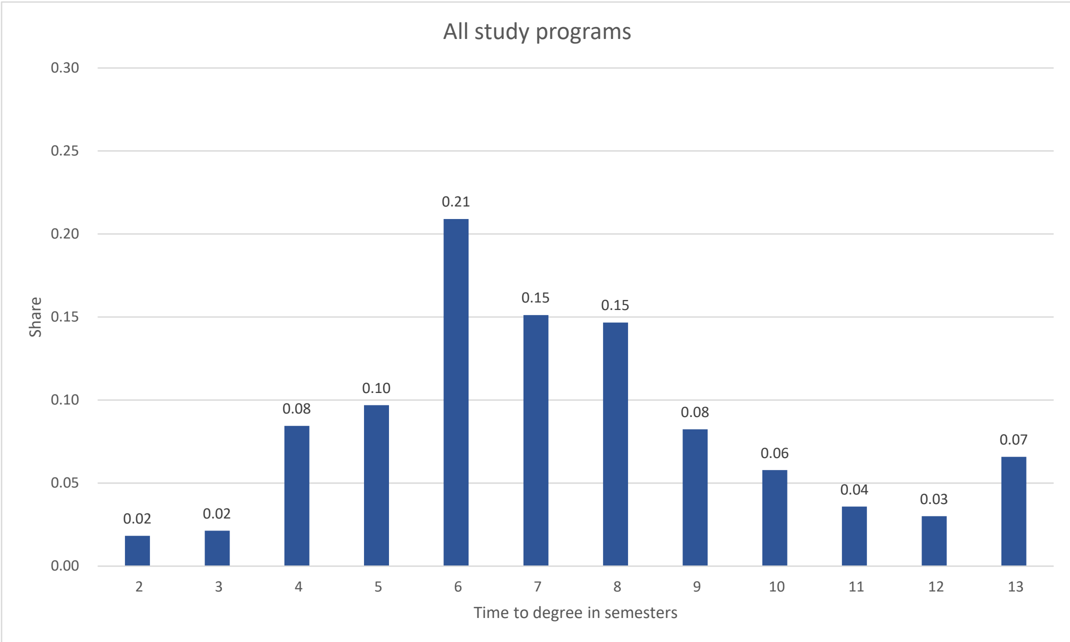
Notes: Average semesters to complete the basic study stage is calculated as the average number of semesters it took for the first X percent to complete the basic study stage. The 95% confidence intervals are based on robust standard errors from unadjusted OLS regressions.

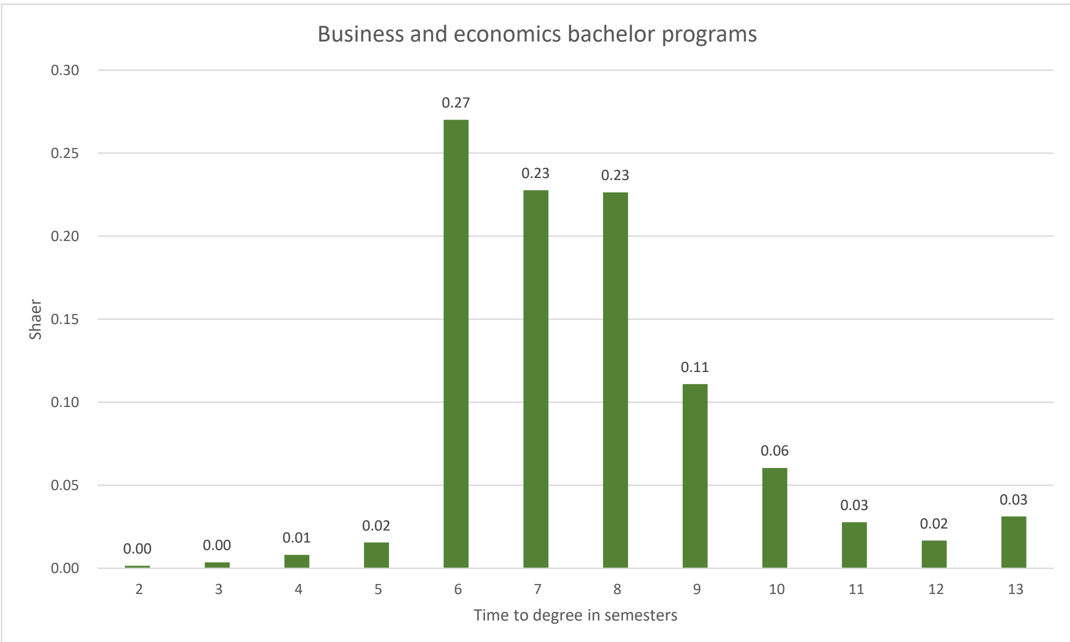
FIGURE 5.A2: DISTRIBUTION OF APPLICATION DATES



Notes: Application dates of the incoming Business Administration students. The application period was May 2 to July 12.

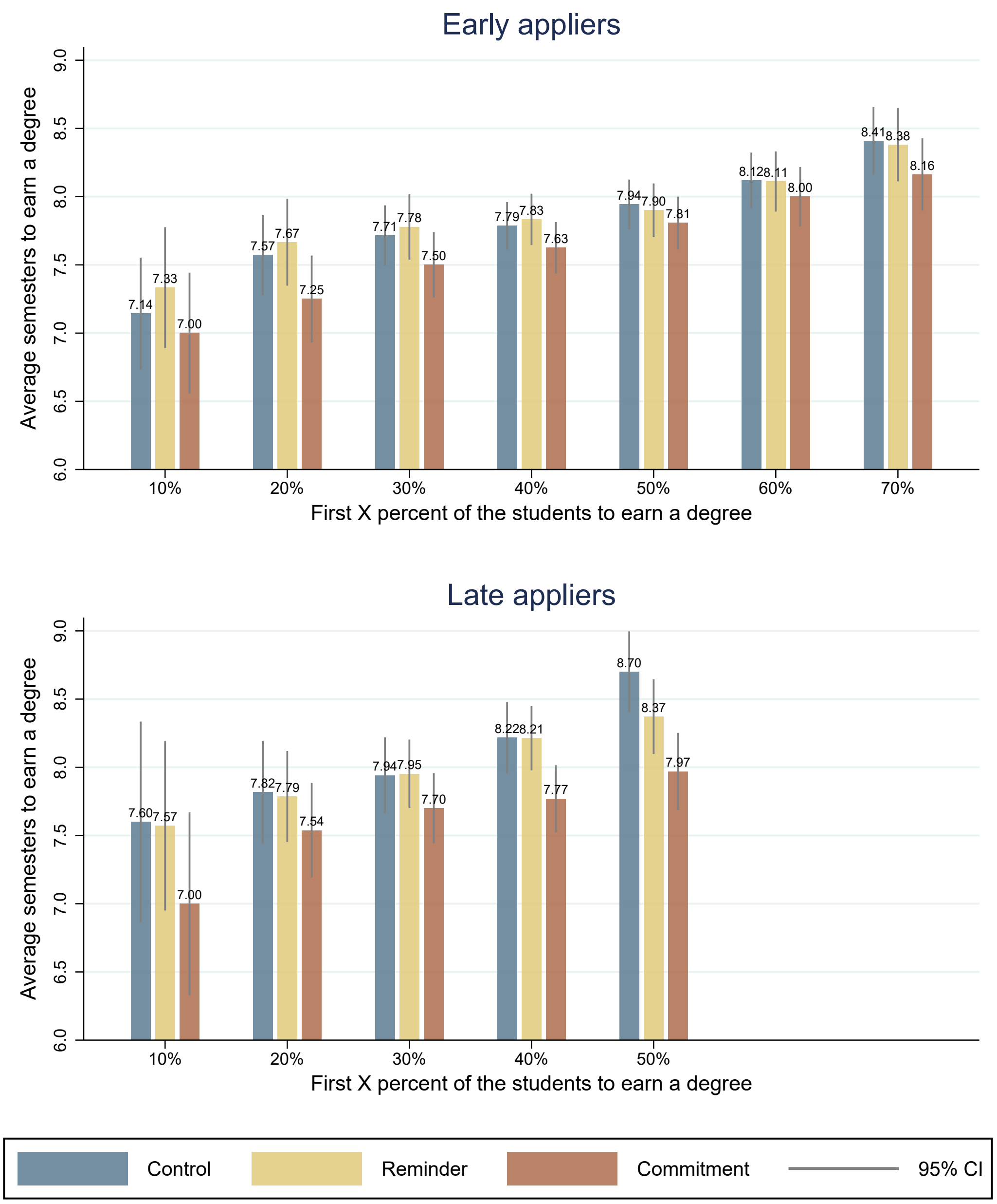
FIGURE 5.A3: TIME TO DEGREE DISTRIBUTIONS, 2019 – GERMANY





Notes: The figure depicts time to degree distrubtions for degrees obtained in Germany in the year 2019 (Statistisches Bundesamt, 2020a).

FIGURE 5.A4: AVERAGE TIME TO DEGREE – BY APPLICATION DATE



Notes: Early and late appliers are defined by splitting the sample at the median application date (see Figure 5.A2). Average semesters to earn a degree is calculated as the average number of semesters it took for the first X percent of degree earners to earn their degree. The 95% confidence intervals are based on robust standard errors from unadjusted OLS regressions.

B Experimental materials

FIGURE 5.B1: COVER LETTER, INTRODUCTORY LECTURE ALL GROUPS (ENGLISH)

[redacted]

Hochschule
für angewandte Wissenschaften
[redacted]

Fakultät
Wirtschaftswissenschaften
Dekanat
[redacted]

[redacted]

STUDY WITH A PLAN
at the Department of Economics and Business

[redacted]

Ihre Zeichen/Ihre Nachricht vom Unser Zeichen [redacted] Zimmer [redacted]

Study with a Plan

Dear Mr/Ms «first name» «last name»,

we are delighted that you have chosen to study Business Administration at our university, and we welcome you in [redacted]. Today we want to introduce to you the program ***Study with a Plan***. As the name gives away, it is aimed at helping you to optimally organize your studies.

To this end you will find on the next page an **Exam Plan**, which will facilitate studying successfully and without overlapping courses during the first four semesters. After the fourth semester, the intern semester marks an important milestone in the course of your studies. ***Study with a Plan*** means that you are **„exam free“** after the fourth semester. Therefore, you can enter **without „obligations“** into your internship in the fifth semester, and the following second study period.

Studying with a plan has a number of additional advantages, which we will point out to you in today's lecture. We hope that this information will assist you in your studies from the start.

In addition you can today give us feedback on your first two weeks at the university. We would appreciate it if you could fill out the attached questionnaire.

We wish you a good and successful start at our department!

With kind regards
Your

Prof. Dr. [redacted]

Vice Dean

FIGURE 5.B2: EXAM PLAN, INTRODUCTORY LECTURE ALL GROUPS (ENGLISH)

Target agreement for «vorname» «nachname» — **Bachelor Program Business Administration**

I. Study with a Plan – Exam plan for successful studies.

	Class	Acronym	Hours	Credit Points (CP)	Passed	
1. Semester	Statistik und mathematische Grundlagen	MATH	4	6 CP	☐	
	Allgemeine Betriebswirtschaftslehre	ABWL	4	6 CP	☐	
	Organisation	ORGA	4	6 CP	☐	
	Wirtschaftsprivatrecht	WIPR	4	6 CP	☐	
	Buchführung und Bilanzierung	BUBI	4	6 CP	☐	
	CP after 1st Semester			Target: 30 CP	Actual:	CP
2. Semester	Volkswirtschaftslehre	VOWL	4	6 CP	☐	
	Arbeitsrecht	ARBR	4	6 CP	☐	
	Personal	PERS	4	6 CP	☐	
	Kosten- und Leistungsrechnung	KOLR	4	6 CP	☐	
	Steuern	STEU	4	6 CP	☐	
	PC Praktikum	PCP		Has to be passed before end of 3rd Semester	☐	
	CP after 2nd Semester			Target: 60 CP	Actual:	CP
3. Semester	Marketing	MARK	6	6 CP	☐	
	Operation Management	OPMG	4	6 CP	☐	
	Wirtschaftsinformatik	WINF	4	6 CP	☐	
	Finanzierung und Investition	FINI	6	6 CP	☐	
	Controlling	CONT	4	6 CP	☐	
	Englisch	ENGL		Proof of English proficiency before end of 3rd Sem.	☐	
	CP after 3rd Semester			Target: 90 CP	Actual:	CP
4. Semester	Angewandte Volkswirtschaftslehre	AVWL	4	6 CP	☐	
	Seminar / Planspiel	SEMA/PLSP	4	6 CP	☐	
	Anwendung BWL-Methoden	ABWM	4	6 CP	☐	
	Wirtschaftsenglisch	WENG	4	6 CP	☐	
	Fachbezogene Wahlpflichtfächer	FWPF	4	6 CP	☐	
	CP after 4th Semester			Target: 120 CP	Actual:	CP

In the rightmost column of this summary you can document your progress in your studies. Check off the exams you have already passed and note the obtained Credit Points. Over the course of your studies this allows you to evaluate whether you are still "on track". If you are, this should motivate you to continue to „Study with a Plan". Otherwise, if needed you can correct your course in time by (re)taking the missing exams.

Hochschule
für angewandte Wissenschaften
[redacted]

[redacted]

FIGURE 5.B3: INFO MATERIAL, INTRODUCTORY LECTURE CONTROL AND REMINDER GROUPS (ENGLISH)

Target agreement for <<first name>> <<last name>> | **Bachelor Program Business Administration**

III. Information about the Exam Plan for successful studies.

a) Adhering to the study plan – You can do it!

According to the study plan you should expect a workload of 30 Credit Points or 900 hours in every semester (1 CP = 30 hours). This amounts to roughly 45 hours per week over a period of 20 weeks, and it includes the weeks with classes as well as the period of preparing for exams. The workload therefore approximates that of a "regular employee". Under typical circumstances, this is doable!

Accordingly, a 6 CP class comes with a workload of 180 hours, of which 45 hours (15 weeks*3hours class) are reserved for the classes. The remaining 135 hours are your responsibility and are spent independently studying. This corresponds to 6-7 hours per week (assuming 20 weeks including exam preparation). Of course this can only give some rough orientation, and the individual workload can deviate from these numbers.

b) The study plan takes into account the sequence of classes.

Some classes build on other classes, i.e. the lecturers will assume that you have attended introductory classes. This means that certain fundamentals will not be covered again in the more advanced classes.

Some advanced classes (especially in later semesters) can only be taken after certain credit point thresholds have been cleared, or after certain prerequisite exams have been passed. The reason is that only if you have understood the basics these classes can proceed to convey advanced knowledge in an efficient manner.

c) The study plan allows for studying without overlapping classes

Only if you adhere to the study plane it is guaranteed that there will be no overlapping classes. All classes in a specific semester are scheduled in a way that they do not overlap. As soon as you take classes from different semesters, e.g. because you have to retake a class, the times of your classes will almost certainly clash.

d) Plan ahead: core classes.

Core classes are only offered once a year. It is important that you consider this when planning your studies. The best course of action is to adhere to the study plan – the plan makes sure that there will be no lost time, which would ultimately delay your graduation.

Hochschule
für angewandte Wissenschaften
[redacted]

[redacted]

FIGURE 5.B4: INFO MATERIAL, INTRODUCTORY LECTURE COMMITMENT GROUP, TEXT ADDED TO REMINDER TEXT IN GREY (ENGLISH)

Target agreement for <<first name>> <<last name>> **Bachelor Program Business Administration**

III. Information about the Exam Plan for successful studies.

a) Why a target agreement?

Agreeing on targets is an instrument that is widely used in business contexts. Specifying and confirming goals in writing leads to a higher probability of reaching them. With the attached target agreement you set for yourself the target of "Studying with a Plan". Specifically, this means: with the target agreement you and the department aim at studying successfully, according to the Study Plan.

The agreement is therefore a measure by which you make clear to yourself that you really „take your studies seriously". You get to keep one copy of the agreement for your records, so that you can later bring to mind the targets that you set for yourself when you started your studies. The intention is to motivate yourself to check the progress you are making in your studies against the study plan, evaluate your progress, and take appropriate actions if things are not going as planned.

By signing the agreement you are subject to the same consequences that already arise from the official examination regulations. In this respect, you can only be better off by signing the agreement.

b) Adhering to the study plan – You can do it!

According to the study plan you should expect a workload of 30 Credit Points or 900 hours in every semester (1 CP = 30 hours). This amounts to roughly 45 hours per week over a period of 20 weeks, and it includes the weeks with classes as well as the period of preparing for exams. The workload therefore approximates that of a "regular employee". Under typical circumstances, this is doable!

Accordingly, a 6 CP class comes with a workload of 180 hours, of which 45 hours (15 weeks*3hours class) are reserved for the classes. The remaining 135 hours are your responsibility and are spent independently studying. This corresponds to 6-7 hours per week (assuming 20 weeks including exam preparation). Of course this can only give some rough orientation, and the individual workload can deviate from these numbers.

c) The study plan takes into account the sequence of classes.

Some classes build on other classes, i.e. the lecturers will assume that you have attended introductory classes. This means that certain fundamentals will not be covered again in the more advanced classes.

Some advanced classes (especially in later semesters) can only be taken after certain credit point thresholds have been cleared, or after certain prerequisite exams have been passed. The reason is that only if you have understood the basics these classes can proceed to convey advanced knowledge in an efficient manner.

d) The study plan allows for studying without overlapping classes

Only if you adhere to the study plane it is guaranteed that there will be no overlapping classes. All classes in a specific semester are scheduled in a way that they do not overlap. As soon as you take classes from different semesters, e.g. because you have to retake a class, the times of your classes will almost certainly clash.

e) Plan ahead: core classes.

Core classes are only offered once a year. It is important that you consider this when planning your studies. The best course of action is to adhere to the study plan – the plan makes sure that there will be no lost time, which would ultimately delay your graduation.

Hochschule
für angewandte Wissenschaften
[redacted]

[redacted]

FIGURE 5.B5: SIGN-UP LETTER 1ST SEMESTER – REMINDER, TEXT ADDED FOR COMMITMENT IN GREY (ENGLISH)

[redacted]

Hochschule
für angewandte Wissenschaften
[redacted]

Fakultät
Wirtschaftswissenschaften
Dekanat
[redacted]

[redacted]

[redacted]

[redacted]

Hochschule für angewandte Wissenschaften
[redacted]

Frau
«Vorname» «Nachname»
«Strasse»
«PLZ» «Ort»

Ihre Zeichen/Ihre Nachricht vom Unser Zeichen Telefon [redacted] Zimmer [redacted]

Study with a Plan

Dear Mr/Ms «Last Name»,

surely you remember the introductory lecture to ***Study with a Plan*** which took place in the context of the statistics class. In this lecture you were given important information on how to best organize your studies and you have signed a target agreement with us.

The exam sign-up period for the winter semester is coming up shortly. **In the context of *Study with a Plan* we recommend: In the period from 18.11. - 29.11. please use the university website to sign up for at least the following exams:**

Statistics and Mathematics	MATH
Business Administration	ABWL
Organization	ORGA
Civil law	WIPR
Accounting	BUBI

We hope you enjoy your time at our department and wish you all the best.

Kind regards
Your

Prof. Dr. [redacted], *Vice Dean*

FIGURE 5.B6: APPENDIX TO SIGN-UP LETTER 7TH SEMESTER – REMINDER AND COMMITMENT (ENGLISH)

I. Curriculum und requirements 1st to 6th semester

Curriculum 1st semester		Passed
Statistics and mathematics	MATH	☐
Business administration	ABWL	☐
Organization	ORGA	☐
Civil law	WIPR	☐
Accounting	BUBI	☐
Total CP (6CP each, target=30CP)	∑	

Curriculum 2nd semester		Passed
Economics	VOWL	☐
Labor law	ARBR	☐
Human resources	PERS	☐
Cost and activity accounting	KOLR	☐
Taxation	STEU	☐
Total CP (6CP each, target=30CP)	∑	

Curriculum 3rd semester		Passed
Marketing	MARK	☐
Operations management	OPMG	☐
Information systems	WINF	☐
Financing and investment	FINI	☐
Controlling	CONT	☐
Total CP (6CP each, target=30CP)	∑	

Requirements after 3rd semester	Attained
72 CP after the 3rd semester	☐
Computer internship completed successfully	☐
Command of English proven	☐

Curriculum 4th semester		Passed
Applied economics	AVWL	☐
Seminar/business game	SEMA/PLSP	☐
Applied business methods	ABWM	☐
Commercial English	WENG	☐
Subject-specific electives	FWPF	☐
Total CP (6CP each, target=30CP)	∑	

Curriculum 5th semester		Passed
Internship[2]	PRAK	☐
Preparation/reflection of internship	PRAV/PRAR	☐
General electives	AWFP	☐
Total CP (22+3+5 CP, target=30CP)	∑	

Curriculum 6th semester		Passed
Business management/ethics	UNTF/UETH	☐
Major 1a	SCHW	☐
Major 2a	SCHW	☐
Total CP (7+10+10 CP, target=27CP)	∑	

II. Curriculum 7th semester

Curriculum 7th semester		Signed Up
Major 1b	SCHW	☐
Major 2b	SCHW	☐
Bachelor seminar	BAGS	☐
Bachelor thesis	BACA	☐
Total CP (10+10+1+12 CP, target=33CP)	∑	

[1] 66 instead of 72 CP from the first three semesters are required if the 3rd semester was not completed at the FHWS.
[2] Requirements for internship: 95 CP, minimum duration of 20 weeks, timely registration.

FIGURE 5.B7: STUDY LETTER 1ST SEMESTER – REMINDER, TEXT ADDED FOR COMMITMENT IN GREY (ENGLISH)

[redacted]

Hochschule
für angewandte Wissenschaften
[redacted]

Fakultät
Wirtschaftswissenschaften
Dekanat
[redacted]

[redacted]

Hochschule für angewandte Wissenschaften
[redacted]

Ihre Zeichen/Ihre Nachricht vom | Unser Zeichen | Telefon [redacted] | Zimmer [redacted]

Study with a Plan

Dear Mr/Ms «First Name» «Last Name»,

Christmas and the New Year are just around the corner. We wish you a peaceful time and a happy new year. Enjoy the upcoming holidays with your family and friends. Please also remember that the exam period starts shortly after the turn of the year

On October 10th you have signed a target agreement by which you commit to taking exams according to the „Exam plan for successful studies“.

In the context of *Study with a Plan* we recommend that you start the necessary preparations before the Christmas break. The exam plan prescribes that in the first semester you successfully participate in at least the following exams:

Statistics and Mathematics	MATH
Business Administration	ABWL
Organization	ORGA
Civil law	WIPR
Accounting	BUBI

We wish you all the best for the exams!

Kind regards
Your

Prof. Dr. [redacted], *Vice Dean*

Chapter 6

General Discussion and Conclusions

This thesis investigates the effectiveness of several low-cost and easy-to-scale behavioral measures at increasing students' success at university. Although there remains much to be learned about the intricacies of such interventions and how they should best be employed to benefit students across the entire distribution, this thesis nonetheless provides compelling evidence for the potential of such interventions and the wide range of possible applications.

But before using the results of this thesis for further research and policy, it is essential to discuss potential limitations. First, it is necessary to consider threats to the **internal validity** of the presented evidence and the consequences for the interpretation of the results. In the potential outcomes framework (see, e.g., Imbens and Rubin (2015) for an introduction), three conditions have to be met to identify a causal Average Treatment Effect (ATE) based on an experiment (Czibor et al., 2019): treatment has to be assigned randomly – which holds true for all studies in this thesis, subjects have to not opt in or out of the treatment that was assigned to them, and the Stable Unit Treatment Value Assumption (SUTVA) has to hold, i.e., the potential outcomes of an individual are assumed to be unrelated to the treatment status of other individuals (see Angrist et al., 1996).

Due to the covert nature of the experiments, students can generally not opt in or out of their assigned treatment.[1] However, with the exception of the introductory lecture and the corresponding offer of the non-binding agreement in Chapter 5, all treatments were administered via letters, and it remains unobserved whether students actually open and read the letters. The estimates presented throughout the thesis thus have to be interpreted as ITT effects, which are, however, often considered to be the more policy-relevant estimates anyways (Czibor et al., 2019).

Among the most common violations of the SUTVA are spillover effects, and, in particular, cases in which untreated individuals are also affected by the treatment. If the true effect is positive and there are positive spillovers to the control group, the ITT effect will generally be

[1] In the relative performance feedback intervention studied in Chapters 3 and 4, students were offered the option to opt out of receiving the feedback letters, but very few of them actually made use of this.

underestimated (Duflo et al., 2007). It is conceivable that students in the studies in Chapters 3 to 5 share the information included in the letters with students of the control group, who may thus also benefit from treatment.[2] If this was indeed the case, then the presented ITT effects would be downward biased and the positive effects of the interventions would be understated. For Chapters 3 and 4, two arguments can be made against widespread information spillovers. First, treated students' beliefs about relative performance are more accurate compared to controls. Second, in the case of spillovers, the RDD in Chapter 3 should have provided evidence for positive effects at the cutoff for students in the control group as well.

A second important concern that is commonly raised is whether studies have sufficient **statistical power**. There are three important implications of low statistical power (Czibor et al., 2019): First, a high rate of false negatives, i.e., the null hypothesis of no effect is not rejected, despite being false. Second, low statistical power can lead to effect inflation, since only large treatment effect estimates will pass the threshold for statistical significance. Third, results from low-powered studies contribute less to the general scientific knowledge, because they only have a small effect on the PSP that a research finding which has achieved statistical significance is actually true.

To increase the power of the statistical analysis, all RCTs in this thesis use stratified randomization. In addition, Chapter 3 and 4 are based on an original study and a replication experiment, which improves the power and robustness of the findings substantially. Chapters 2 and 5 are most prone to the critique of low statistical power. In Chapter 2, power calculations were performed after the pilot study. Comparisons to the literature showed that the study was well-powered enough to detect effects of social cues that are generally found in the literature. In view of newer research, which suggests that previous results may have been overstated due to publication bias, the study was likely not well-powered enough, and the insignificant main effects could well be false negatives.

For Chapter 5, ex-post power calculations provided the following insights: while the study was not well-powered enough to reliably detect effects sizes smaller than those that were actually found for the graduation rate and time to degree, the subsequent evidence on the PSP showed that the study still leads to a meaningful revision of the PSP that a soft commitment device can indeed have a large positive effect on academic achievements. Considering the very promising results and comparatively small number of observations, the intervention described in Chapter 5 is nevertheless the one with the biggest need for replication.

[2]This is less likely to be of concern in the intervention studied in Chapter 2. Since it takes place before the beginning of the first semester, it is unlikely that students already know each other and share information. However, there could be information spillovers during the remedial math course, which may affect the treatment effects on average participation across all tutorials. In this case, the direction of the potential bias is expected to be the same as in the other RCTs.

Last but not least are considerations regarding the **generalizability** or **external validity** of the results. Shadish et al. (2002, p.38) define external validity as “the validity of inferences about whether the cause-effect relationship holds over variation in persons, settings, treatment variables, and measurement variables.” It is beyond the scope of this thesis to discuss generalizability in all its details. Nevertheless, a few general considerations can be made.

Given that the RCTs in all Chapters can be categorized as NFEs, it is very likely that the presented evidence is at least externally valid for the setting in which they were conducted (see Chapter 1) – i.e., results are likely to replicate with another cohort of the same study programs at the same institutions. In the case of Chapters 3 and 4, a replication experiment provides direct evidence for this notion.

Going beyond the original settings, the first question relates to the representativeness of the study populations. With respect to the high school GPA students in all interventions are representative for the general population of high school graduates, i.e., the study programs are not selective. However, all RCTs were conducted in BA and Economics or ME programs. Studies show that the characteristics of students in these subjects can be different from the general student population. In ME, females – who are argued to be less competitive (Niederle and Vesterlund, 2011) – are often underrepresented, and studies suggest that students of Economics behave more rational than the general population (Hellmich, 2019). It is therefore an open question to what extent the results of the interventions generalize to other study programs. At the same time, the aforementioned programs still make up a substantial part of the German higher education system. In the winter semester of 2019, 14.7% of freshman students enrolled in an Economics, BA, or ME bachelor’s program (Statistisches Bundesamt, 2020b). Consequently, there remains a sizable portion of the student population to which the results of this thesis can be more readily applied.

Yet, results may also vary depending on the (type of) institutions. E.g., while there is often a considerable amount of similarity in the structure of study programs, universities of applied sciences regularly feature an internship semester, tend to be more practice-oriented, and often have a scheduled study duration of seven instead of six semesters. In addition, the student populations can differ between institutions, e.g., due to different degrees of selectivity. Lastly, it is even more difficult to assess to what extent results generalize beyond the German higher education system. This is particularly true when considering systems with large tuition fees such as in the U.S. In these settings the direct costs of studying and the related incentives to finish the study program in the scheduled duration may leave less room to improve performance. Nevertheless, delayed graduation and high dropout rates are not limited to Germany, but are also of major concern in other countries (see Chapter 1), which indicates the need for further measures.

What are the implications of these considerations going forward? In combination with the evidence presented throughout this thesis, the following key recommendations can be derived: i) Given their large potential for helping students at university, researchers and policymakers alike should continue to develop, evaluate, and implement behavioral interventions. ii) A particular emphasis should be put on replicating promising interventions and studying whether they hold up at scale (see Al-Ubaydli et al. (2017) for a discussion of the threats to scalability). This step is essential to find out whether and for whom an intervention is actually beneficial. iii) The potential of heterogeneous effects needs to be taken into account in the design stage. Connected to the last point, taking heterogeneity more seriously may go a long way in helping understand whether and why interventions replicate and scale up successfully (Bryan et al., 2021). iv) Finally, and applying both to behavioral and traditional approaches, the interaction between interventions needs to be studied more closely to understand which inputs in the higher education production function complement each other and which are actually substitutes.

Bibliography

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

Acland, Dan and Matthew R. Levy (2015). “Naiveté, Projection Bias, and Habit Formation in Gym Attendance”. *Management Science* 61 (1), 146–160.

Ahrens, Achim, Christian B. Hansen, and Mark E. Schaffer (2020). “lassopack: Model Selection and Prediction With Regularized Regression in Stata”. *Stata Journal* 20 (1), 176–235.

Al-Ubaydli, Omar, John A. List, and Dana L. Suskind (2017). “What Can We Learn from Experiments? Understanding the Threats to the Scalability of Experimental Results”. *American Economic Review* 107 (5), 282–86.

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

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

Altmann, Steffen and Christian Traxler (2014). “Nudges at the Dentist”. *European Economic Review* 72, 19–38.

Altmann, Steffen, Christian Traxler, and Philip Weinschenk (2021). “Deadlines and Memory Limitations”. *mimeo*.

Altonji, Joseph G. (1993). “The Demand for and Return to Education When Education Outcomes are Uncertain”. *Journal of Labor Economics* 11 (1), 48–83.

Altonji, Joseph G., Erica Blom, and Costas Meghir (2012). “Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers”. *Annual Review of Economics* 4 (1), 185–223.

Amador, Manuel, Iván Werning, and George-Marios Angeletos (2006). “Commitment vs. Flexibility”. *Econometrica* 74 (2), 365–396.

Ambrus, Attila and Georgy Egorov (2013). “Comment on “Commitment vs. Flexibility””. *Econometrica* 81 (5), 2113–2124.

Anderberg, Dan, Claudia Cerrone, and Arnaud Chevalier (2018). “Soft Commitment: A Study on Demand and Compliance”. *Applied Economics Letters* 25 (16), 1140–1146.

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

Angrist, Joshua D, Guido W. Imbens, and Donald B. Rubin (1996). "Identification of Causal Effects Using Instrumental Variables". *Journal of the American Statistical Association* 91 (434), 444–455.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

— (2010). "The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics". *Journal of Economic Perspectives* 24 (2), 3–30.

Ariely, Dan and Klaus Wertenbroch (2002). "Procrastination, Deadlines, and Performance: Self-Control by Precommitment". *Psychological Science* 13 (3), 219–224.

Ashraf, Anik (2019). "Do Performance Ranks Increase Productivity? Evidence from a Field Experiment". *mimeo*.

Ashraf, Nava, Dean S. Karlan, and Wesley Yin (2006a). "Household Decision Making and Savings Impacts: Further Evidence from a Commitment Savings Product in the Philippines". *Yale University Economic Growth Center Discussion Paper No. 939*.

— (2006b). "Tying Odysseus to the Mast: Evidence From a Commitment Savings Product in the Philippines". *Quarterly Journal of Economics* 121 (2), 635–672.

Athey, Susan and Guido W. Imbens (2017a). "The Econometrics of Randomized Experiments". *Handbook of Economic Field Experiments*. Vol. 1. Elsevier, 73–140.

— (2017b). "The State of Applied Econometrics: Causality and Policy Evaluation". *Journal of Economic Perspectives* 31 (2), 3–32.

— (2019). "Machine Learning Methods That Economists Should Know About". *Annual Review of Economics* 11, 685–725.

Attewell, Paul, Scott Heil, and Liza Reisel (2012). "What is Academic Momentum? And Does it Matter?" *Educational Evaluation and Policy Analysis* 34 (1), 27–44.

Attewell, Paul and David Monaghan (2016). "How Many Credits Should an Undergraduate Take?" *Research in Higher Education* 57 (6), 682–713.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger (2015). "Working over Time: Dynamic Inconsistency in Real Effort Tasks". *Quarterly Journal of Economics* 130 (3), 1067–1115.

Augenblick, Ned and Matthew Rabin (2019). "An Experiment on Time Preference and Misprediction in Unpleasant Tasks". *Review of Economic Studies* 86 (3), 941–975.

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

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

— (2016). “The Provision of Relative Performance Feedback: An Analysis of Performance and Satisfaction”. *Journal of Economics & Management Strategy* 25 (1), 77–110.

Baker, Rachel, Brent Evans, and Thomas Dee (2016). “A Randomized Experiment Testing the Efficacy of a Scheduling Nudge in a Massive Open Online Course (MOOC)”. *AERA Open* 2 (4).

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2013). “Team Incentives: Evidence from a Firm Level Experiment”. *Journal of the European Economic Association* 11 (5), 1079–1114.

Bandiera, Oriana, Valentino Larcinese, and Imran Rasul (2010). “Heterogeneous Class Size Effects: New Evidence from a Panel of University Students”. *Economic Journal* 120 (549), 1365–1398.

Bandura, Albert (1991). “Social Cognitive Theory of Self-Regulation”. *Organizational Behavior and Human Decision Processes* 50 (2), 248 –287.

Barankay, Iwan (2012). “Rank Incentives: Evidence from a Randomized Workplace Experiment”. *mimeo*.

Behlen, Lars, Oliver Himmler, and Robert Jaeckle (2020). “Defaults in Education”. *mimeo*.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014). “Inference on Treatment Effects after Selection among High-Dimensional Controls”. *Review of Economic Studies* 81 (2), 608–650.

Bénabou, Roland and Jean Tirole (2002). “Self-Confidence and Personal Motivation”. *Quarterly Journal of Economics* 117 (3), 871–915.

Benartzi, Shlomo, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon, and Steven Galing (2017). “Should Governments Invest More in Nudging?” *Psychological Science* 28 (8), 1041–1055.

Bergman, Peter, Jeffrey T. Denning, and Dayanand Manoli (2019). “Is Information Enough? The Effect of Information about Education Tax Benefits on Student Outcomes”. *Journal of Policy Analysis and Management* 38 (3), 706–731.

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine L. Milkman (2015). “The Effect of Providing Peer Information on Retirement Savings Decisions”. *Journal of Finance* 70 (3), 1161–1201.

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

Bettinger, Eric P. and Rachel B. Baker (2014). "The Effects of Student Coaching: An Evaluation of a Randomized Experiment in Student Advising". *Educational Evaluation and Policy Analysis* 36 (1), 3–19.

Bicchieri, Cristina and Eugen Dimant (2019). "Nudging With Care: The Risks and Benefits of Social Information". *Public Choice*.

Bird, Kelli A., Benjamin L. Castleman, Jeffrey T. Denning, Joshua Goodman, Cait Lamberton, and Kelly Ochs Rosinger (2019). "Nudging at Scale: Experimental Evidence from FAFSA Completion Campaigns". *NBER Working Paper 26158*.

Bisin, Alberto and Kyle Hyndman (2020). "Present-Bias, Procrastination and Deadlines in a Field Experiment". *Games and Economic Behavior* 119, 339–357.

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

Bleemer, Zachary and Basit Zafar (2018). "Intended College Attendance: Evidence from an Experiment on College Returns and Costs". *Journal of Public Economics* 157, 184–211.

Boatman, Angela and Bridget Terry Long (2018). "Does Remediation Work for All Students? How the Effects of Postsecondary Remedial and Developmental Courses Vary by Level of Academic Preparation". *Educational Evaluation and Policy Analysis* 40 (1), 29–58.

Bond, Philip and Gustav Sigurdsson (2018). "Commitment Contracts". *Review of Economic Studies* 85 (1), 194–222.

Bound, John, Michael F. Lovenheim, and Sarah Turner (2010). "Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources". *American Economic Journal: Applied Economics* 2 (3), 129–57.

— (2012). "Increasing Time to Baccalaureate Degree in the United States". *Education Finance and Policy* 7 (4), 375–424.

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

Brade, Raphael, Oliver Himmler, and Robert Jäckle (2020). "Relative Performance Feedback and the Effects of Being Above Average – Field Experiment and Replication". *mimeo*.

Bradley, Steve and Colin Green (2020). *The Economics of Education: A Comprehensive Overview*. Academic Press.

Brent, Daniel A., Corey Lott, Michael Taylor, Joseph Cook, Kimberly Rollins, and Shawn Stoddard (2020). "What Causes Heterogeneous Responses to Social Comparison Messages for Water Conservation?" *Environmental and Resource Economics*, 1–35.

Brodeur, Abel, Nikolai Cook, and Anthony Heyes (2020). "Methods Matter: p-Hacking and Publication Bias in Causal Analysis in Economics". *American Economic Review* 110 (11), 3634–60.

Brown, Jeffrey R. and Alessandro Previtero (2020). "Saving for Retirement, Annuities and Procrastination". *mimeo*.

Bruhn, Miriam and David McKenzie (2009). "In Pursuit of Balance: Randomization in Practice in Development Field Experiments". *American Economic Journal: Applied Economics* 1 (4), 200–232.

Bryan, Christopher J., Elizabeth Tipton, and David S. Yeager (2021). "Behavioural Science Is Unlikely to Change the World Without a Heterogeneity Revolution". *Nature Human Behaviour* 5 (8), 980–989.

Bryan, Gharad, Dean Karlan, and Scott Nelson (2010). "Commitment Devices". *Annual Review of Economics* 2 (1), 671–698.

Burger, Nicholas, Gary Charness, and John Lynham (2011). "Field and Online Experiments on Self-Control". *Journal of Economic Behavior & Organization* 77 (3), 393–404.

Byrne, David P., Andrea La Nauze, and Leslie A. Martin (2018). "Tell Me Something I Don't Already Know: Informedness and the Impact of Information Programs". *Review of Economics and Statistics* 100 (3), 510–527.

Cabrera, José María and Alejandro Cid (2017). "Gender Differences to Relative Performance Feedback: A Field Experiment in Education". *mimeo*.

Cadena, Brian C. and Benjamin J. Keys (2015). "Human Capital and the Lifetime Costs of Impatience". *American Economic Journal: Economic Policy* 7 (3), 126–53.

Camerer, Colin F., Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan, Emma Heikenstein, Felix Holzmeister, Taisuke Imai, Siri Isaksson, Gideon Nave, Tomas Pfeiffer, Michael Razen, and Hang Wu (2016). "Evaluating Replicability of Laboratory Experiments in Economics". *Science* 351 (6280), 1433–1436.

Cantoni, Davide, David Y. Yang, Noam Yuchtman, and Y. Jane Zhang (2019). "Protests as Strategic Games: Experimental Evidence from Hong Kong's Antiauthoritarian Movement". *Quarterly Journal of Economics* 134 (2), 1021–1077.

Carrell, Scott E. and James E. West (2010). "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors". *Journal of Political Economy* 118 (3), 409–432.

Casari, Marco (2009). "Pre-Commitment and Flexibility in a Time Decision Experiment". *Journal of Risk and Uncertainty* 38 (2), 117–141.

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

Celik Katreniak, Dagmara (2018). “Dark Side of Incentives: Evidence From a Randomized Control Trial in Uganda”. *mimeo*.

Charness, Gary, David Masclet, and Marie Claire Villeval (2013). “The Dark Side of Competition for Status”. *Management Science* 60 (1), 38–55.

Chase, Jared A., Ramona Houmanfar, Steven C. Hayes, Todd A. Ward, Jennifer Plumb Vilardaga, and Victoria Follette (2013). “Values are Not Just Goals: Online ACT-based Values Training Adds to Goal Setting in Improving Undergraduate College Student Performance”. *Journal of Contextual Behavioral Science* 2 (3-4), 79–84.

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

Cialdini, Robert B. (2003). “Crafting Normative Messages to Protect the Environment”. *Current Directions in Psychological Science* 12 (4), 105–109.

— (2011). “The Focus Theory of Normative Conduct”. *Handbook of Theories of Social Psychology* 2, 295–312.

Cialdini, Robert B., Linda J. Demaine, Brad J. Sagarin, Daniel W. Barrett, Kelton Rhoads, and Patricia L. Winter (2006). “Managing Social Norms for Persuasive Impact”. *Social Influence* 1 (1), 3–15.

Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren (1990). “A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places”. *Journal of Personality and Social Psychology* 58 (6), 1015–1026.

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

Coffman, Lucas, Clayton R. Featherstone, and Judd B. Kessler (2015). “A Model of Information Nudges”. *mimeo*.

Coffman, Lucas C., Clayton R. Featherstone, and Judd B. Kessler (2017). “Can Social Information Affect What Job You Choose and Keep?” *American Economic Journal: Applied Economics* 9 (1), 96–117.

Compte, Olivier and Andrew Postlewaite (2004). “Confidence-Enhanced Performance”. *American Economic Review* 94 (5), 1536–1557.

Corcoran, Katja, Jan Crusius, and Thomas Mussweiler (2011). “Social Comparison: Motives, Standards, and Mechanisms”. *Theories in Social Psychology*. Ed. by Derek Chadee. Wiley-Blackwell.

Croson, Rachel and Jen Yue Shang (2008). "The Impact of Downward Social Information on Contribution Decisions". *Experimental Economics* 11 (3), 221–233.

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

Damgaard, Mette Trier and Helena Skyt Nielsen (2018). "Nudging in Education". *Economics of Education Review* 64, 313–342.

Danzer, Natalia and Victor Lavy (2018). "Paid Parental Leave and Children's Schooling Outcomes". *Economic Journal* 128 (608), 81–117.

Darolia, Rajeev and Casandra Harper (2018). "Information Use and Attention Deferment in College Student Loan Decisions: Evidence From a Debt Letter Experiment". *Educational Evaluation and Policy Analysis* 40 (1), 129–150.

Dawson, Rachel Fulcher, Melissa S. Kearney, and James X. Sullivan (2020). "Comprehensive Approaches to Increasing Student Completion in Higher Education: A Survey of the Landscape". *NBER Working Paper No. 28046.*

De Paola, Maria and Vincenzo Scoppa (2015). "Procrastination, Academic Success and the Effectiveness of a Remedial Program". *Journal of Economic Behavior & Organization* 115, 217–236.

Dean, Mark, Özgür Kıbrıs, and Yusufcan Masatlioglu (2017). "Limited Attention and Status Quo Bias". *Journal of Economic Theory* 169, 93–127.

Deaton, Angus and Nancy Cartwright (2018). "Understanding and Misunderstanding Randomized Controlled Trials". *Social Science & Medicine* 210, 2–21.

Delavande, Adeline and Basit Zafar (2019). "University Choice: The Role of Expected Earnings, Nonpecuniary Outcomes, and Financial Constraints". *Journal of Political Economy* 127 (5), 2343–2393.

Delfgaauw, Josse, Robert Dur, Joeri Sol, and Willem Verbeke (2013). "Tournament Incentives in the Field: Gender Differences in the Workplace". *Journal of Labor Economics* 31 (2), 305–326.

DellaVigna, Stefano and Elizabeth Linos (2020). "RCTs to Scale: Comprehensive Evidence from Two Nudge Units". *mimeo.*

DellaVigna, Stefano and Ulrike Malmendier (2006). "Paying Not to Go to the Gym". *American Economic Review* 96 (3), 694–719.

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

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

Doyle, William R. (2011). "Effect of Increased Academic Momentum on Transfer Rates: An Application of the Generalized Propensity Score". *Economics of Education Review* 30 (1), 191–200.

Duckworth, Angela Lee and Martin E. P. Seligman (2006). "Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores." *Journal of Educational Psychology* 98 (1), 198.

Duflo, Esther, Rachel Glennerster, and Michael Kremer (2007). "Using Randomization in Development Economics Research: A Toolkit". Ed. by T. Paul Schultz and John A. Strauss. Vol. 4. Handbook of Development Economics. Elsevier, 3895–3962.

Dustmann, Christian, Johannes Ludsteck, and Uta Schönberg (2009). "Revisiting the German Wage Structure". *Quarterly Journal of Economics* 124 (2), 843–881.

Dustmann, Christian and Uta Schönberg (2012). "Expansions in Maternity Leave Coverage and Children's Long-Term Outcomes". *American Economic Journal: Applied Economics* 4 (3), 190–224.

Duvendack, Maren, Richard Palmer-Jones, and W. Robert Reed (2017). "What Is Meant by "Replication" and Why Does It Encounter Resistance in Economics?" *American Economic Review* 107 (5), 46–51.

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

Eil, David and Justin M. Rao (2011). "The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself". *American Economic Journal: Microeconomics* 3 (2), 114–138.

Ellis, Albert and William J. Knaus (1977). *Overcoming Procrastination: How to Think and Act Rationally in Spite of Life's Inevitable Hassles*. New York: Institute for Rational Living.

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

Ericson, Keith Marzilli (2017). "On the Interaction of Memory and Procrastination: Implications for Reminders, Deadlines, and Empirical Estimation". *Journal of the European Economic Association* 15 (3), 692–719.

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

Ertac, Seda (2005). “Social Comparisons and Optimal Information Revelation: Theory and Experiments”. *Job Market Paper, UCLA*.

Escueta, Maya, Andre Joshua Nickow, Philip Oreopoulos, and Vincent Quan (2020). “Upgrading Education with Technology: Insights from Experimental Research”. *Journal of Economic Literature* 58 (4), 897–996.

Feld, Jan, Nicolás Salamanca, and Ulf Zölitz (2020). “Are Professors Worth It? The Value-Added and Costs of Tutorial Instructors”. *Journal of Human Resources* 55 (3), 836–863.

Fellner, Gerlinde, Rupert Sausgruber, and Christian Traxler (2013). “Testing Enforcement Strategies in the Field: Threat, Moral Appeal and Social Information”. *Journal of the European Economic Association* 11 (3), 634–660.

Festinger, Leon (1954). “A Theory of Social Comparison Processes”. *Human Relations* 7 (2), 117–140.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr (2001). “Are People Conditionally Cooperative? Evidence from a Public Goods Experiment”. *Economics Letters* 71 (3), 397–404.

Fischer, Mira and Valentin Wagner (2019). “Effects of Timing and Reference Frame of Feedback: Evidence from a Field Experiment in Secondary Schools”. *mimeo*.

Freedman, David A. (2006). “Statistical Models for Causation: What Inferential Leverage Do They Provide?” *Evaluation Review* 30 (6), 691–713.

French, Robert and Philip Oreopoulos (2017). “Behavioral Barriers Transitioning to College”. *Labour Economics* 47, 48–63.

Frey, Bruno S. and Stephan Meier (2004). “Social Comparisons and Pro-social Behavior: Testing “Conditional Cooperation” in a Field Experiment”. *American Economic Review* 94 (5), 1717–1722.

Galperti, Simone (2015). “Commitment, Flexibility, and Optimal Screening of Time Inconsistency”. *Econometrica* 83 (4), 1425–1465.

Gammarano, Rosina (2018). “Where are the Jobs? Employment Patterns Across Sectors and Occupations”. *ILOSTAT Spotlight on Work Statistics* 2.

Gee, Laura K. (2019). “The More You Know: Information Effects on Job Application Rates in a Large Field Experiment”. *Management Science* 65 (5), 2077–2094.

Gill, David, Zdenka Kissová, Jaesun Lee, and Victoria Prowse (2019). “First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision”. *Management Science* 65 (2), 494–507.

Giné, Xavier, Dean Karlan, and Jonathan Zinman (2010). “Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation”. *American Economic Journal: Applied Economics* 2 (4), 213–235.

Goldin, Claudia and Lawrence F. Katz (2008). *The Race between Education and Technology*. Harvard University Press.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius (2008). “A Room With a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels”. *Journal of Consumer Research* 35 (3), 472–482.

Goos, Maarten, Alan Manning, and Anna Salomons (2014). “Explaining Job Polarization: Routine-Biased Technological Change and Offshoring”. *American Economic Review* 104 (8), 2509–26.

Goulas, Sofoklis and Rigissa Megalokonomou (2021). “Knowing Who You Are: The Effect of Feedback Information on Short and Long Term Outcomes”. *Journal of Economic Behavior & Organization* 183, 589–615.

Gurantz, Oded, Jessica Howell, Michael Hurwitz, Cassandra Larson, Matea Pender, and Brooke White (2021). “A National-Level Informational Experiment to Promote Enrollment in Selective Colleges”. *Journal of Policy Analysis and Management* 40 (2), 453–479.

Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev (2017). “The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance”. *Journal of Public Economics* 148, 14–31.

Hamermesh, Daniel S. (2007). “Replication in Economics”. *Canadian Journal of Economics* 40 (3), 715–733.

Hanushek, Eric A. and Ludger Woessmann (2008). “The Role of Cognitive Skills in Economic Development”. *Journal of Economic Literature* 46 (3), 607–68.

— (2010). “Education and Economic Growth”. *Economics of Education*. Ed. by Dominic J. Brewer, Barry McGaw, and Patrick J. McEwan. Amsterdam: Elsevier, 60–67.

Harrison, Glenn W. and J. A. List (2004). “Field Experiments”. *Journal of Economic Literature* 42 (4), 1009–1055.

Hellmich, Simon Niklas (2019). “Are People Trained in Economics “Different,” and if so, Why? A Literature Review”. *American Economist* 64 (2), 246–268.

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

Himmler, Oliver, Robert Jäckle, and Philipp Weinschenk (2019). “Soft Commitments, Reminders and Academic Performance”. *American Economic Journal: Applied Economics* 11 (2), 114–142.

Hoekstra, Mark (2009). “The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach”. *Review of Economics and Statistics* 91 (4), 717–724.

Holzer, Harry J. and Sandy Baum (2017). *Making College Work: Pathways to Success for Disadvantaged Students*. Washington, DC: Brookings Institution.

Houser, Daniel, Daniel Schunk, Joachim Winter, and Erte Xiao (2018). "Temptation and Commitment in the Laboratory". *Games and Economic Behavior* 107, 329–344.

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

Hoxby, Caroline M. and Sarah Turner (2015). "What High-Achieving Low-Income Students Know about College". *American Economic Review* 105 (5), 514–17.

Hummel, Dennis and Alexander Maedche (2019). "How Effective is Nudging? A Quantitative Review on the Effect Sizes and Limits of Empirical Nudging Studies". *Journal of Behavioral and Experimental Economics* 80, 47–58.

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

Huntington-Klein, Nick and Andrew M. Gill (2019). "An Informational Intervention to Increase Semester Credits in College". *Series of Unsurprising Results in Economics* 1, 1–17.

Ilgen, Daniel R. and Cori Davis (2000). "Bearing Bad News: Reactions to Negative Performance Feedback". *Applied Psychology* 49 (3), 550–565.

Ilgen, Daniel R., Cynthia D. Fisher, and M Susan Taylor (1979). "Consequences of Individual Feedback on Behavior in Organizations." *Journal of Applied Psychology* 64 (4), 349.

Imbens, Guido W. (2010). "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)". *Journal of Economic Literature* 48 (2), 399–423.

Imbens, Guido W. and Donald B. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press.

Jensen, Robert (2010). "The (Perceived) Returns to Education and the Demand for Schooling". *Quarterly Journal of Economics* 125 (2), 515–548.

John, Anett (2020). "When Commitment Fails: Evidence from a Field Experiment". *Management Science* 66 (2), 503–529.

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

Kara, Elif, Mirco Tonin, and Michael Vlassopoulos (2021). "Class Size Effects in Higher Education: Differences across STEM and Non-STEM Fields". *Economics of Education Review* 82, 102104.

Karlan, Dean, Margaret McConnell, Sendhil Mullainathan, and Jonathan Zinman (2016). "Getting to the Top of Mind: How Reminders Increase Saving". *Management Science* 62 (12), 3393–3411.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan (2015). “Self-Control at Work”. *Journal of Political Economy* 123 (6), 1227–1277.

Kent, David M. and Rodney A. Hayward (2007). “Limitations of Applying Summary Results of Clinical Trials to Individual Patients: The Need for Risk Stratification”. *JAMA* 298 (10), 1209–1212.

Kent, David M., David van Klaveren, Jessica K. Paulus, Ralph D’Agostino, Steve Goodman, Rodney Hayward, John P.A. Ioannidis, Bray Patrick-Lake, Sally Morton, Michael Pencina, Gowri Raman, Joseph S. Ross, Harry P. Selker, Ravi Varadhan, Andrew Vickers, John B. Wong, and Ewout W. Steyerberg (2020). “The Predictive Approaches to Treatment Effect Heterogeneity (PATH) Statement: Explanation and Elaboration”. *Annals of Internal Medicine* 172 (1), W1–W25.

Kent, David M., Ewout Steyerberg, and David van Klaveren (2018). “Personalized Evidence Based Medicine: Predictive Approaches to Heterogeneous Treatment Effects”. *BMJ* 363.

Keser, Claudia and Frans Van Winden (2000). “Conditional Cooperation and Voluntary Contributions to Public Goods”. *Scandinavian Journal of Economics* 102 (1), 23–39.

Kim, Kyung Ryung and Eun Hee Seo (2015). “The Relationship Between Procrastination and Academic Performance: A Meta-Analysis”. *Personality and Individual Differences* 82, 26–33.

King, Jacqueline E. (2004). “Missed Opportunities: Students Who Do Not Apply for Financial Aid”. *American Council on Education Issue Brief.*

Koch, Alexander, Julia Nafziger, and Helena Skyt Nielsen (2015). “Behavioral Economics of Education”. *Journal of Economic Behavior & Organization* 115, 3–17.

König, Tobias, Sebastian Schweighofer-Kodritsch, and Georg Weizsäcker (2019). “Beliefs as a Means of Self-Control? Evidence from a Dynamic Student Survey”. *WZB Discussion Paper SP II 2019–204.*

Köszegi, Botond (2006). “Ego Utility, Overconfidence, and Task Choice”. *Journal of the European Economic Association* 4 (4), 673–707.

Kraft, Matthew A. (2020). “Interpreting Effect Sizes of Education Interventions”. *Educational Researcher* 49 (4), 241–253.

Kuhnen, Camelia M. and Agnieszka Tymula (2012). “Feedback, Self-Esteem, and Performance in Organizations”. *Management Science* 58 (1), 94–113.

Kunz, Johannes S. and Kevin E. Staub (2020). “Early Subjective Completion Beliefs and the Demand for Post-Secondary Education”. *Journal of Economic Behavior & Organization* 177, 34–55.

Laibson, David (1997). “Golden Eggs and Hyperbolic Discounting”. *Quarterly Journal of Economics* 112 (2), 443–478.

Lavecchia, Adam M., Heidi Liu, and Philip Oreopoulos (2016). "Behavioral Economics of Education: Progress and Possibilities". *Handbook of the Economics of Education Vol. 5*. Ed. by Stephen Machin Eric A. Hanushek and Ludger Woessmann. Elsevier, 1–74.

Leaver, Sean (2016). "Behavioural Education Economics". *Routledge Handbook of Behavioral Economics*. Ed. by Roger Frantz, Shu-Heng Chen, Kurt Dopfer, Floris Heukelom, and Shabnam Mousavi. Routledge, 379–397.

Lee, David S. and Thomas Lemieux (2010). "Regression Discontinuity Designs in Economics". *Journal of Economic Literature* 48 (2), 281–355.

Leeper, Thomas J. (2019). "Where Have the Respondents Gone? Perhaps We Ate Them All". *Public Opinion Quarterly* 83 (S1), 280–288.

Lergetporer, Philipp, Katharina Werner, and Ludger Woessmann (2021). "Does Ignorance of Economic Returns and Costs Explain the Educational Aspiration Gap? Representative Evidence from Adults and Adolescents". *Economica* 88 (351), 624–670.

Levitt, Steven D and John A. List (2009). "Field Experiments in Economics: The Past, the Present, and the Future". *European Economic Review* 53 (1), 1–18.

MacDonell, Karen, Sylvie Naar-King, Heather Huszti, and Marvin Belzer (2013). "Barriers to Medication Adherence in Behaviorally and Perinatally Infected Youth Living with HIV". *AIDS and Behavior* 17 (1), 86–93.

Maniadis, Zacharias, Fabio Tufano, and John A. List (2014). "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects". *American Economic Review* 104 (1), 277–90.

Martin, Richard and John Randal (2008). "How is Donation Behaviour Affected by the Donations of Others?" *Journal of Economic Behavior & Organization* 67 (1), 228–238.

Martorell, Paco and Isaac McFarlin Jr (2011). "Help or Hindrance? The Effects of College Remediation on Academic and Labor Market Outcomes". *Review of Economics and Statistics* 93 (2), 436–454.

McGuigan, Martin, Sandra McNally, and Gill Wyness (2016). "Student Awareness of Costs and Benefits of Educational Decisions: Effects of an Information Campaign". *Journal of Human Capital* 10 (4), 482–519.

Mischel, Walter, Yuichi Shoda, and Monica I. Rodriguez (1989). "Delay of Gratification in Children". *Science* 244 (4907), 933–938.

Möbius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat (2014). "Managing Self-Confidence". *mimeo*.

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

Morisano, Dominique, Jacob B. Hirsh, Jordan B. Peterson, Robert O. Pihl, and Bruce M. Shore (2010). "Setting, Elaborating, and Reflecting on Personal Goals Improves Academic Performance." *Journal of Applied Psychology* 95 (2), 255.

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

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

Niederle, Muriel and Lise Vesterlund (2011). "Gender and Competition". *Annual Review of Economics* 3 (1), 601–630.

O'Brien, William K. (2002). "Applying the Transtheoretical Model to Academic Procrastination". *Doctoral Dissertation.*

O'Donoghue, Ted and Matthew Rabin (1999). "Doing it Now or Later". *American Economic Review*, 103–124.

OECD (2019). "Education at a Glance 2019: OECD Indicators".

— (2021). "Education at a Glance 2021: OECD Indicators".

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

Oreopoulos, Philip (2021). "What Limits College Success? A Review and Further Analysis of Holzer and Baum's Making College Work". *Journal of Economic Literature* 59 (2), 546–73.

Oreopoulos, Philip and Ryan Dunn (2013). "Information and College Access: Evidence from a Randomized Field Experimentt". *Scandinavian Journal of Economics* 115 (1), 3–26.

Oreopoulos, Philip, Richard W. Patterson, Uros Petronijevic, and Nolan G. Pope (forthcoming). "Low-Touch Attempts to Improve Time Management among Traditional and Online College Students". *Journal of Human Resources.*

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

— (2018). "Student Coaching: How Far Can Technology Go?" *Journal of Human Resources* 53 (2), 299–329.

— (2019). "The Remarkable Unresponsiveness of College Students to Nudging And What We Can Learn from It". *NBER Working Papers No. 26059.*

Oreopoulos, Philip and Kjell G. Salvanes (2011). "Priceless: The Nonpecuniary Benefits of Schooling". *Journal of Economic Perspectives* 25 (1), 159–184.

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

Pearce, Jone L. and Lyman W. Porter (1986). "Employee Responses to Formal Performance Appraisal Feedback." *Journal of Applied Psychology* 71 (2), 211.

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

Psacharopoulos, George and Harry Anthony Patrinos (2018). "Returns to Investment in Education: A Decennial Review of the Global Literature". *Education Economics* 26 (5), 445–458.

Ravallion, Martin (2020). "Should the Randomistas (Continue to) Rule?" *NBER Working Paper No. 27554*.

Reuben, Ernesto, Paola Sapienza, and Luigi Zingales (2015). "Procrastination and Impatience". *Journal of Behavioral and Experimental Economics* 58, 63–76.

Robinson, Carly D., Gonzalo A. Pons, Angela L. Duckworth, and Todd Rogers (2018). "Some Middle School Students Want Behavior Commitment Devices (but Take-Up Does Not Affect Their Behavior)". *Frontiers in Psychology* 9, 206.

Royer, Heather, Mark Stehr, and Justin Sydnor (2015). "Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company". *American Economic Journal: Applied Economics* 7 (3), 51–84.

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

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

Semb, George, D. Marvin Glick, and Robert E. Spencer (1979). "Student Withdrawals and Delayed Work Patterns in Self-Paced Psychology Courses". *Teaching of Psychology* 6 (1), 23–25.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Shang, Jen and Rachel Croson (2009). "A Field Experiment in Charitable Contribution: The Impact of Social Information on the Voluntary Provision of Public Goods". *Economic Journal* 119 (540), 1422–1439.

Solomon, Laura J. and Esther D. Rothblum (1984). "Academic Procrastination: Frequency and Cognitive-Behavioral Correlates." *Journal of Counseling Psychology* 31 (4), 503.

Spybrook, Jessaca, Howard Bloom, Richard Congdon, Carolyn Hill, Andres Martinez, and Stephen Raudenbush (2011). “Optimal Design Plus Empirical Evidence: Documentation for the “Optimal Design” Software”. *Optimal Design Plus Version 3.0.*

Statistisches Bundesamt (2018). “Hochschulen auf einen Blick: Ausgabe 2018”.

— (2020a). “Bildung und Kultur: Prüfungen an Hochschulen 2019”.

— (2020b). “Bildung und Kultur: Studierende an Hochschulen Wintersemester 2019/2020”.

Steel, Piers (2007). “The Nature of Procrastination: A Meta-Analytic and Theoretical Review of Quintessential Self-Regulatory Failure”. *Psychological Bulletin* 133 (1), 65–94.

Stinebrickner, Ralph and Todd R. Stinebrickner (2014a). “A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout”. *Review of Economic Studies* 81 (1), 426–472.

— (2014b). “Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model”. *Journal of Labor Economics* 32 (3), 601–644.

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

Suls, Jerry and Ladd Wheeler (2000). *Handbook of Social Comparison: Theory and Research.* The Plenum Series in Social/Clinical Psychology. New York: Kluwer Academic/Plenum Publishers.

Taubinsky, Dmitry (2014). “From Intentions to Actions: A Model and Experimental Evidence of Inattentive Choice”. *Harvard University Working Paper.*

Taylor, Shelley E., Heidi A. Wayment, and Mary Carillo (1996). “Social Comparison, Self-Regulation, and Motivation”. *Handbook of Motivation and Cognition, Vol. 3. The Interpersonal Context.* Ed. by R. M. Sorrentino and E.T. Higgins. Guilford Press.

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

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

Villeval, Marie Claire (2020). “Performance Feedback and Peer Effects”. *Handbook of Labor, Human Resources and Population Economics.* Ed. by Klaus F. Zimmermann. Cham: Springer, 1–38.

Vyankandondera, Joseph, Kirstin Mitchell, Brenda Asiimwe-Kateera, Kimberly Boer, Philippe Mutwa, Jean-Paul Balinda, Masja van Straten, Peter Reiss, and Janneke van de Wijgert (2013). “Antiretroviral Therapy Drug Adherence in Rwanda: Perspectives from Patients and Healthcare Workers Using a Mixed-Methods Approach”. *AIDS Care* 25 (12), 1504–1512.

Weiss, Michael J., Alyssa Ratledge, Colleen Sommo, and Himani Gupta (2019). "Supporting Community College Students from Start to Degree Completion: Long-Term Evidence from a Randomized Trial of CUNY's ASAP". *American Economic Journal: Applied Economics* 11 (3), 253–97.

Wiswall, Matthew and Basit Zafar (2015a). "Determinants of College Major Choice: Identification using an Information Experiment". *Review of Economic Studies* 82 (2), 791–824.

— (2015b). "How Do College Students Respond to Public Information about Earnings?" *Journal of Human Capital* 9 (2), 117–169.

Wong, Wei-Kang (2008). "How Much Time-Inconsistency Is There and Does It Matter? Evidence on Self-Awareness, Size, and Effects". *Journal of Economic Behavior & Organization* 68 (3-4), 645–656.

Declaration for Admission to the Doctoral Examination

I confirm

1. that the dissertation “Behavioral Interventions and Students’ Success at University - Evidence from Randomized Field Experiments” that I submitted was produced independently without assistance from external parties, and not contrary to high scientific standards and integrity,

2. that I have adhered to the examination regulations, including upholding a high degree of scientific integrity, which includes the strict and proper use of citations so that the inclusion of other ideas in the dissertation are clearly distinguished,

3. that in the process of completing this doctoral thesis, no intermediaries were compensated to assist me neither with the admissions or preparation processes, and in this process,

   - No remuneration or equivalent compensation were provided
   - No services were engaged that may contradict the purpose of producing a doctoral thesis

4. that I have not submitted this dissertation or parts of this dissertation elsewhere.

I am aware that false claims (and the discovery of those false claims now, and in the future) with regards to the declaration for admission to the doctoral examination can lead to the invalidation or revoking of the doctoral degree.

Signature: ____________________________

Date: ____________________________

Author Contributions

The main part of the thesis builds on four research papers, three of which are joint work with co-authors. My contributions to the three co-authored papers are as follows:

- Chapter 3 is based on joint work with Oliver Himmler and Robert Jäckle. I made substantial and significant contributions to the preparation of the experimental data, was the leading researcher in the data analysis, and substantially contributed to the conception and writing of the first and final draft.

- Chapter 4 is based on joint work with Oliver Himmler and Robert Jäckle. I prepared the experimental data, was the leading researcher in the data analysis, and substantially contributed to the conception and writing of the first and final draft.

- Chapter 5 is based on joint work with Oliver Himmler, Robert Jäckle, and Philipp Weinschenk. I made substantial and significant contributions to the preparation of the experimental data, was the leading researcher in the data analysis, prepared the first draft of the paper, and substantially contributed to the conception and writing of the final draft.

Signature: ______________________________

Date: ______________________________