Institutional influences on education investment and pro-social behaviour

Author:
Chen, Jie

Publication Date:
2020

DOI:
https://doi.org/10.26190/unsworks/3972

Institutional influences on education investment and pro-social behaviour

Jie Chen

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Economics

UNSW Business School

Supervisors: Gigi Foster and Alberto Motta

May 2020

Thesis/Dissertation Sheet

Surname/Family Name	:	**Chen**
Given Name/s	:	**Jie**
Abbreviation for degree as give in the University calendar	:	**PhD**
Faculty	:	**UNSW Business School**
School	:	**School of Economics**
Thesis Title	:	**Institutional influences on education investment and pro-social behaviour**

Abstract 350 words maximum: (PLEASE TYPE)

This thesis consists of three chapters. It studies, as a broad theme, the effectiveness of several institutional changes on individual decision-making based on experimental evidence. Chapter 1 is self-contained, with results purely based on a laboratory experiment. Chapter 2 and Chapter 3 are based on one field experiment in education. Chapter 2 describes the experimental settings and presents the overall results of the experiment, whereas Chapter 3 extends the analysis and focuses on treatment effects on women and men respectively.

Chapter 1 shows how reward or punishment opportunities change contributions in a public goods game with `privileged' members, where `privilege' indicates that one's per-unit contribution to the public good produces a higher monetary return than is the case for others in the group. The main finding is that reward opportunities strongly increase group contributions in such groups while punishment opportunities do not. Reward also mitigates contribution decay over successive periods and improves social welfare.

Chapter 2 mainly studies how rank incentives (i.e., relative performance information) in a milestone-based online assignment system affect students' academic performance. I find that rank incentives increase the likelihood of a student putting more effort in the online assignment. Rank incentives also have positive effects on low-performing students' exam marks while they have negative effects on high-performing students' exam marks. The positive effects seem driven by increased self-perceived stress, increased effort, and decreased procrastination. The negative effects seem driven by increased self-perceived happiness and re-allocation of effort.

Chapter 3 studies how rank incentives and milestone information (i.e., information with reference to achievement milestones corresponding to different amounts of points earned) affect men's and women's academic performance differently. Women with access to the rank incentives experience a 0.19 SDs mark decrease in the first midterm, compared to women without this access. In the absence of relative performance information, men with access to the milestone information experience a 0.26 SDs mark increase in the final exam, compared to men without the access. The negative effects on women seem driven by their increased stress level, whereas men's improved exam performance seem driven by increased effort.

INCLUSION OF PUBLICATIONS STATEMENT

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in their thesis in lieu of a Chapter if:

- The candidate contributed greater than 50% of the content in the publication and is the "primary author", ie. the candidate was responsible primarily for the planning, execution and preparation of the work for publication
- The candidate has approval to include the publication in their thesis in lieu of a Chapter from their supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Please indicate whether this thesis contains published material or not:

☒ This thesis contains no publications, either published or submitted for publication
(if this box is checked, you may delete all the material on page 2)

☐ Some of the work described in this thesis has been published and it has been documented in the relevant Chapters with acknowledgement
(if this box is checked, you may delete all the material on page 2)

☐ This thesis has publications (either published or submitted for publication) incorporated into it in lieu of a chapter and the details are presented below

CANDIDATE'S DECLARATION

I declare that:

- I have complied with the UNSW Thesis Examination Procedure
- where I have used a publication in lieu of a Chapter, the listed publication(s) below meet(s) the requirements to be included in the thesis.

Candidate's Name	Signature	Date (dd/mm/yy)
JIE CHEN		31/05/2020

Originality Statement

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed ______________________

Date ________________________

Copyright Statement

'I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation. I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstract International (this is applicable to doctoral theses only). I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.'

Signed ________________________

Authenticity Statement

'I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.'

Signed ________________________

Acknowledgement

First and foremost, I would like to thank my supervisors Gigi Foster and Alberto Motta for always being available and for keep pushing me beyond my limits. During the research progress, ups and downs are inevitable. I am hugely grateful to their patience, support, and encouragement through various challenging moments. I thank Alberto for inspiring me to aim high and for trusting me with various research opportunities. I especially thank Gigi for encouraging me to try new things in life beyond research and nudging me to challenge myself. These experiences not only made me a better researcher, but also a richer soul.

My special thanks also goes to Isabella Dobrescu. Her persistence and creativity have expanded my imagination, leading me to increase expectations on my own work. It is a real privilege to work so closely with a senior scholar, in addition to one's supervisors. I am deeply grateful to this experience. As a result, Chapters 2 and 3 of my thesis are planned to be published as joint papers with my supervisors and Isabella.

I thank Johannes Hoelzemann for his selfless help in preparing for the lab experiment for the first chapter of my thesis. I am grateful to the members of my Annual Progress Review panel Sarah Walker, Pauline Grosjean, Federico Masera, Tess Stafford, and Arghya Ghosh for giving me useful feedback at various stages of my program.

I have also been blessed to cross path and work with many great colleagues along the PhD journey: Hang Bahn, Yi Ding, Bianca Bonollo, Xinyi Geng, Dandan Yu, Nadine Yamout, Na Ry Hong, Adam Gorajek, Barton Lee, Shipei Zeng, Fabio Martinenghi, Akashlina Arno, and others. I especially thank Hang Bahn, with whom I studied the PhD courses together and discussed many technical questions along the way. I also thank the School of Economics for the open and lovely atmosphere. I gratefully acknowledge the financial support from the University of New South Wales, UNSW Business School, the School of Economics, and UNSW Bizlab.

Many people have enriched my PhD journey. I thank Paul Andon for coordinating the wonderful PhD Teaching Fellows program, through which I advanced my knowledge and skill set to survive academia. I thank my mentor Benjamin Avanzi for the deep and sometimes philosophical conversations and for helping me think through my career goals. I am especially grateful to the rich support networks provided by UNSW. Through those networks, I get to spoil myself with all kinds of seminars and workshops and learn from others' experiences. These are invaluable..

Above all, I thank my parents and brother for always being there through my ups and downs. I thank my best friend and life partner, my husband Charlie. Thank you for your unconditional support in this challenging journey.

Abstract

This thesis consists of three chapters. It studies, as a broad theme, the effectiveness of several institutional changes on individual decision-making based on experimental evidence. Chapter 1 is self-contained, with results purely based on a laboratory experiment. Chapter 2 and Chapter 3 are based on one field experiment in education. Chapter 2 describes the experimental settings and presents the overall results of the experiment, whereas Chapter 3 extends the analysis and focuses on treatment effects on women and men respectively.

Chapter 1 shows how reward or punishment opportunities change contributions in a public goods game with 'privileged' members, where 'privilege' indicates that one's per-unit contribution to the public good produces a higher monetary return than is the case for others in the group. The main finding is that reward opportunities strongly increase group contributions in such groups while punishment opportunities do not. Reward also mitigates contribution decay over successive periods and improves social welfare.

Chapter 2 mainly studies how rank incentives (i.e., relative performance information) in a milestone-based online assignment system affect students' academic performance. I find that rank incentives increase the likelihood of a student putting more effort in the online assignment. Rank incentives also have positive effects on low-performing students' exam marks while they have negative effects on high-performing students' exam marks. The positive effects seem driven by increased self-perceived stress, increased effort, and decreased procrastination. The negative effects seem driven by increased self-perceived happiness and reallocation of effort.

Chapter 3 studies how rank incentives and milestone information (i.e., information with reference to achievement milestones corresponding to different amounts of

points earned) affect men's and women's academic performance differently. Women with access to the rank incentives experience a 0.19 standard deviations decrease of marks in the first midterm, compared to women without this access. In the absence of relative performance information, men with access to the milestone information experience a 0.26 standard deviations increase of marks in the final exam, compared to men without the access. The negative effects on women seem driven by their increased stress level, whereas men's improved exam performance seems driven by increased effort.

Contents

Acknowledgement **i**

Abstract **iii**

Contents **iv**

Figures **ix**

Tables **xi**

Introduction **1**

1 Carrots and sticks: New Evidence in Privileged Groups **5**

1.1 Introduction 5

1.2 Literature Review 9

1.3 Motivation and Contribution 12

1.3.1 *Homo Oeconomicus* Predictions 15

1.3.2 Predictions with Fehr & Schmidt (1999) Utilities 15

1.4 Experimental Design 17

1.4.1 Further Comments on Experimental Structure 19

1.5 Results 19

1.5.1 Reward Strongly Motivates Contributions 19

1.5.2 Reward Outperforms Punishment in Motivating Contributions 21

1.5.3 Punishment and Reward Behaviour 23

1.5.4 Comparisons Between Privileged Groups and Non-privileged Groups 26

1.5.5 Discussion on Sustained Coordination and Improved Efficiency 29

1.6 Conclusion 32

Appendix A **35**

A.1 Experimental Instructions 35

A.2 Theoretical Predictions With *Homo Oeconomicus* Agents 40

A.3 Theoretical Predictions with Fehr & Schmidt (1999)'s Agents 41

A.4 Other Results 48

2 Rank and Human Capital: Lessons from a Randomized Controlled Trial **59**

2.1 Introduction 59

2.2 Literature Review 62

2.3 Randomized Controlled Trial 65

2.3.1 Environment 65

2.3.2 Treatments 67

2.4 Data and Empirical Analysis 69

2.4.1 Data 69

2.4.2 Empirical Methodology 72

2.5 Results 76

2.5.1 Baseline Course Effect Estimates 76

2.5.2 Heterogeneous Course Effects 79

2.5.3 Beyond Course Effects 81

2.6 The Mechanism 82

2.7 Robustness 86

2.8 Concluding Remarks 88

Appendix B **91**

B.1 Main Results 91

B.2 Other Results 100

3 (High-performing) Women Hate Ranks and Men Like Leagues **121**

3.1 Introduction 121

3.2 Literature Review 125

3.3 Data and Empirical Analysis 128

3.4 Results 130

3.4.1 The Effect of Providing Relative Performance Feedback (Social Goals vs. Goals) 130

3.4.2 The Effect of Setting Social Goals (Social Goals vs. Control) . 132

3.4.3 The Effect of Setting Goals (Goals vs. Control) 133

3.5 Non-linear and Heterogeneous Effects on Course Grades 134
3.5.1 Non-linear Effects on Course Grades 134
3.5.2 Heterogeneous Effects on Course Grades 137
3.6 Mechanisms 140
3.7 Robustness 144
3.8 Conclusions and Discussions 145
3.8.1 Conclusions 145
3.8.2 Discussions: Chapter 2 vs. Chapter 3 147

Appendix C **151**

Conclusion **173**

Bibliography **175**

List of Figures

A.1 Average Individual Contributions - Reward vs. Control 48

A.2 Average Individual Contributions - Reward vs. Punish 48

A.3 Average Group Contributions - Reward vs. Control 49

A.4 Average Group Contributions - Reward vs. Punish 49

B.1 Example of Performance Information as Shown in the Assignment . . 98

B.2 Procrastination and Assignment Progression 99

List of Tables

A.1 Public Goods Game With Sanctioning Opportunities . . . 47

A.2 Experimental Design 48

A.3 Punishment/Reward Depending on Contribution Differences 50

A.4 Punishment/Reward Depending on Differences in First-Stage Earnings 51

A.5 Effect of A High Player's Contributions on Low Players' Contributions 52

A.6 Effect of Low players' Contributions on A High Player's Contributions 53

A.7 Mean Comparison: Baseline vs. Control 54

A.8 Effects of Punishment/Reward on Next Period Contributions 55

A.9 Treatment Effect of Sanctions on Group Contributions 56

A.10 Treatment Effect of Sanctions on High Players' Contributions 57

A.11 Treatment Effect of Sanctions on Low Players' Contributions 58

B.1 Descriptive Statistics 92

B.2 Balance Tests for Treatments and Control Groups in Semester 2 2018 93

B.3 Effect on Course Performance: Social Goals vs. Goals 94

B.4 Beyond Course Effects: Social Goals vs. Goals 95

B.5 Mechanism for Low-performing Students: Social Goals vs. Goals . . 96

B.6 Mechanism for High-performing Students: Social Goals vs. Goals . . 97

B.7 Effects on Assignment Progression, Procrastinators vs. Non-procrastinators 100

B.8 Effects on Assignment Progression 101

B.9 Effects on Other Course Grades 102

B.10 Effects on Other Course Grades, Procrastinators vs. Non-procrastinators 103

B.11 Mechanism for Low-performing Students: Can Overcome Difficulties 104

B.12 Mechanism for Low-performing Students: Course Platform Accesses 105

B.13 Mechanism for High-performing Students: Considers Oneself Happy 106

B.14 Mechanism for High-performing Students: Has 1,000+ Assignment Points 107

B.15 Effects on Assignment Progression, Age 108

B.16 Effects on Assignment Progression, Gender 109

B.17 Effects on Assignment Progression, International status 110

B.18 Effects on Assignment Progression, Degree 111

B.19 Effects on Other Course Grades, Age 112

B.20 Effects on Other Course Grades, Gender 113

B.21 Effects on Other course Grades, International status 114

B.22 Effects on Exam Grades, Degree 115

B.23 Placebo Treatment Effects 116

B.24 Treatment Effects, Ability and Tutorial Fixed Effects 117

B.25 Effects on Other Course Grades, CQR 118

B.26 Full Assignment Completion 119

3.1 Descriptive Statistics 129

C.1 Effects on Exam Grades 151

C.2 Effects on Assignment Progression 152

C.3 Effects on Exam Grades (UQR)- Social Goals vs. Goals 153

C.4 Effects on Exam Grades (UQR)- Social Goals vs. Control 154

C.5 Effects on Exam Grades (UQR)- Goals vs. Control 155

C.6 Mechanism for Females - Social Goals vs. Goals 156

C.7 Mechanism for Females - Social Goals vs. Control 157

C.8 Mechanism for Females - Goals vs. Control 158

C.9 Mechanism for Males - Goals vs. Control 159

C.10 Mechanism for Males - Social Goals vs. Control 160

C.11 Mechanism for Males - Social Goals vs. Goals 161

C.12 Country of Birth: Effects on Exam Grades - Social Goals vs. Goals . 162

C.13 Country of Birth: Effects on Exam Grades - Social Goals vs. Control 163

C.14 Country of Birth: Effects on Exam Grades - Goals vs. Control 164

C.15 Field of Study: Effects on Exam Grades - Social Goals vs. Goals . . . 165

C.16 Field of Study: Effects on Exam Grades - Social Goals vs. Control . . 166

C.17 Field of Study: Effects on Exam Grades - Goals vs. Control 167

C.18 Ability: Effects on Exam Grades 168

C.19 Procrastination: Effects on Exam Grades 169

C.20 Effects on Exam Grades: Control for Tutorial Fixed Effects or Cluster by Tutorials 170

C.21 Effects on Exam Grades: Control for Ability or Procrastination Level 171

Introduction

"Institutions are the rules of the game in a society or, more formally, are the humanly devised constraints that shape human interaction. In consequence they structure incentives in human exchange, whether political, social or economic."[1] In other words, institutions are rules created *by* the human and *for* the human. Historically, the emergence or disappearance of certain institutions has revolutionized people's productivity, income, wealth, living standards, and inevitably their educational outcomes and human capital (Bowles, 2009). Small changes in the content or timing of an institutional design may build up over time, resulting in strong and cumulative differences in performance (Bowles, 2009). Even after its disappearance, an institution continue affecting a population for decades afterward (Acemoglu et al., 2002). Yet, on their own, individual agents often fail to advance institutions towards socially optimal or individually desirable outcomes. Regarding the social aspect, as Bowles (2009) explains, externalities of individual decisions are ubiquitous, significantly hindering an individual's ability to account for the social welfare aspect of his decisions. For example, in writing a paper, one relies on electricity and the internet, not to mention survival goods such as food, water and the like. However, it is almost impossible to evaluate the related harms or benefits, both direct and indirect, to the environment and to the rest of the population. On an individual level, a person often suffers from lack of complete information or lack of sufficient calculation power in his decision making, rendering non-optimal outcomes; worse is the case in which countless daily decisions are made unconsciously or passively. Kahneman (2011) documents extensive evidence in which individual decision making is constrained by inbuilt functionality of the brain or psychological traits such as *cognitive illusions*[2],

[1] North et al. (1990)
[2] (Kahneman, 2011, p. 27)

anchoring effect[3], and *loss aversion*[4]. These facts necessitate devising institutional changes to facilitate better decision makings, better in the sense that an individual would have made such a decision had he "paid full attention and possessed complete information, unlimited cognitive abilities, and complete self-control"[5].

In this thesis, I explore the effect of several institutional changes on either pro-social behaviour or educational outcomes. On the one hand, the concept of pro-social behaviour has been ever-present in economics since Smith (1759), albeit not always referred to using this exact expression. It has received wider attention in recent decades following Benabou and Tirole (2006)'s influential paper which theorizes the interplay between incentives and pro-social behaviour. Despite its popularity, experimental evidence on the effectiveness of incentives (specifically, reward and punishment) remains unsettling. Chapter 1 contributes new experimental evidence to this ongoing discussion, using public goods game with a more realistic design of group structure that allows different returns to contributions among group members. On the other hand, education is one of the key channels through which human capital accumulates and thus greater social welfare can be obtained. Technological advancement has made it possible to implement large-scale field experiments with small institutional changes. Chapters 2 and 3 present the results from a large field experiment among undergraduate university students, with the experimental interventions being two variations on the content and format of performance information provision. More details on the laboratory experiment and the field experiment are given below.

Chapter 1 presents the results of a laboratory experiment investigating the effect of incentives (reward and punishment) on contribution behaviour in a public goods game with privileged members. The existing experimental literature has produced inconclusive evidence on the effectiveness of punishment and reward opportunities in standard public goods games. Standard public goods games often assign group members the same marginal per capita returns on public good production, but in reality group members facing different individual returns often must collaborate to produce a public good. This paper uses a laboratory experiment to investigate public goods contributions in groups that contain a "relatively privileged" participant, in which each participant may also punish or reward other participants. In public

[3](Kahneman, 2011, p. 119)
[4](Kahneman, 2011, p. 283)
[5](Thaler and Sunstein, 2009, p. 5)

good settings, a privileged participant is one who has higher incentives to contribute to the public good than other group members do. In a public-goods setting, will the presence of a privileged participant affect the efficacy of punishment or reward in privileged groups? Findings from the experiment are as follows: 1) Reward increases group contributions while punishment does not; 2) both incentives significantly mitigate contribution decay over successive periods, with reward being more effective than punishment; and 3) the presence of a privileged participant increases the mutual dependence among group members, compared to conventional non-privileged public-goods settings, as measured by regressing the current period contributions of one type of participants on the previous period contributions of the other type of participants. Many real-world groups are composed of groups with members with varying degrees of interest in a common purpose. The present investigation of the efficacy of reward and punishment in groups with a privileged participant can shed light on how to facilitate better cooperation and induce better outcomes in these real-world settings.

Chapter 2 discusses the overall effect of two interventions, namely, relative performance information (i.e., rank incentives, in the form of ranking) and goal-setting (in the form of milestone-referenced league-based absolute performance information), on students' academic performance. Both interventions are implemented via an online assignment system integral to the intervention course. I find 1) weak evidence that access to relative performance information motivates students to achieve higher goals in the online assignment; 2) providing relative performance information has a positive effect on the very low-performers, resulting in exam marks that are approximately 0.27 standard deviations higher than those of students only exposed to the milestone-referenced league-based information; and 3) providing relative performance information negatively affects high-performing students by 0.19 standard deviations. The mechanism for low-performing students involves the ranking's social comparison process, which induces more stress and increased effort, specifically more frequent course material accesses and lower procrastination. In contrast, high-performing students seem happier, with their superior rank inducing a reallocation of effort towards procrastinating less and overachieving in the assignment on which they are ranked. Finally, I show that the relative performance information effects are significantly accentuated for the younger, male, international and business students. By and large, however, the goal-setting treatment does not lead to robust overall treatment effects.

As different behaviour patterns emerge for females and males, Chapter 3 focuses on the treatment effect by gender. The results show that females are susceptible to relative performance information, while males are responsive to the milestone-referenced league-based information. Specifically, females who have access to relative performance information perform 0.19 SDs worse in their first midterm exam than do females without such access. Meanwhile, males who have access to the milestone-referenced league-based information perform 0.26 SDs better in their final exam than males without the access, resulting in significant increases in their average exam grades and overall course grades. Further investigation shows that high-performing females are the most negatively affected among all females. The detrimental effects on females are a result of their increased stress, as they report feeling less able to overcome difficulties. The motivational effects on males are driven by their increased efforts towards the end of the semester: they log into the online assignment more often; they earn and lose more points in the online assignment; and they submit more assignment answers, both correct ones and incorrect ones.

Chapter 1

Carrots and sticks: New Evidence in Privileged Groups[1]

1.1 Introduction

Imagine a group of friends start up a non-profit organisation together, one member being randomly assigned to the role of group representative. Providing effort (e.g., time or money) into the NPO is one real-life example of contributing resources to a public good: all can benefit the most from the group outcome if every individual maximises own contributions, yet everyone can free-ride on others by making lower contributions while sharing in the non-excludable joint benefit. Every member knows that all will be better off if all members fully cooperate by maximising their own contributions. However, not knowing the other members' decisions, the best strategy of a rational member who wants to maximise his earnings would be to free-ride on others without contributing anything. Consequently, contributions tend to suffer under-provision in public goods games and real-world cooperative situations alike. Existing research has found that, in public goods games, members in a group with punishment opportunities (i.e., in which each member can decrease any other

[1]This chapter is my own work, with input from my supervisors in an advisory capacity only. Special thanks goes to Johannes Hoelzemann and Andreas Ortmann for their help and advice in preparing for the experiment. I thank Fanghua Li and two anonymous referees for their comments on drafts of the paper. This work was supported by UNSW Bizlab Higher Degree Research Small Project Grant and UNSW Business School. The experiment was approved by the Human Ethics Research Committee at the University of New South Wales (HC180350).

member's payoff at a cost) contribute significantly more to the public good than do those in a group without such opportunities (Fehr and Gachter, 2000). Reward opportunities are at most as effective as punishment opportunities in motivating contributions, according to a meta-analysis by Balliet et al. (2011). Now suppose that the NPO representative, in addition to sharing the group's joint benefit, has private gains from the operation of the NPO. For example, the representative may have more public exposure, or be more likely to become the CEO, which brings about larger public influence and more opportunities for gains in money and status. These higher private gains mean that the representative has a foreseeable 'privilege' compared to his collaborators, in the way that the term is used in this and prior experimental papers. I implement this type of privilege in the lab experiment by assigning one randomly selected member a higher marginal return (i.e., return for each additional unit of contribution) that is higher than that of the other members. With the public-goods-game structure otherwise unchanged, will reward still motivate more contributions as effectively as punishment, or will one be more effective than the other?

Similar examples can be observed in many real-life scenarios, such as start-up companies with a leading founder and talk shows featuring one host and a back room of producers and organisers. For example, Steve Jobs would probably never have founded Apple had he always been working on his own, yet few people know the names of the other founding members of Apple. Oprah Winfrey needs a team to air her show, yet few know the names of the members of her team. Both Jobs and Winfrey benefitted far more from the (joint) success of the outcome than many others who were working towards the same end. The unequal benefit distribution revealed ex-post in these and other analogous cases are real-world examples of the conditions studied here in the laboratory.

The laboratory experiment consists of a public goods game with a "relatively privileged" group member. A "relatively privileged" member is one who faces a higher marginal per capita return (MPCR), meaning the individual return rate per unit of contribution to the public good, than the other players (Glöckner et al., 2011). Depending on the treatment, a participant may have the opportunity to reward or punish other group members. The term "privilege" corresponds to the advantageous position held by a group representative. In line with existing literature, it is implemented in the experiment by assigning one member a higher MPCR relative to other members. It is distinct from power or leadership[2] as these latter concepts have been understood in experimental public-goods settings studied in past research, in that power grants a member additional choice and leadership allows a member to move before other members, whereas privilege does neither. Privilege, in the current experiment, is merely externally driven and does not change a member's choice set regarding either contributions or sanctioning options. The first research question is whether peer-reward opportunities after a contribution stage alters a player's contributions, as well as overall group contributions. To my best knowledge, no existing literature has explored the effectiveness of reward in groups facing heterogeneous MPCRs. The second research question is whether punishment opportunities or reward opportunities are more effective in motivating group coordination, as measured by total group contribution. Pervasive evidence has shown that punishment opportunities are powerful in mitigating contribution decay in a standard public goods game, while evidence on the effect of reward opportunities is less consistent Fehr and Gachter (2000); Choi and Ahn (2013); Balliet et al. (2011). In contrast, Reuben and Riedl (2009) and Reuben and Riedl (2013) find that punishment opportunities, at best, only weakly mitigates contribution decay in a public goods game with heterogeneous MPCRs. The different level of effectiveness of punishment opportunities on contributions under two public goods game settings (i.e., one with homogeneous MPCRs and one with heterogeneous MPCRs) leads one to wonder whether the effectiveness of reward also changes in a public goods game with heterogeneous MPCRs. Consequently, it would be interesting to investigate the comparative efficacy of punishment and reward opportunities in a public goods game with heterogeneous MPCRs. Additionally, I implement a baseline condition in which standard public goods games with identical MPCRs for each player are played. Reuben and Riedl (2009) find that, with punishment opportunities, contributions in a heterogeneous-MPCR group are smaller than those in a standard public goods game group. I thus question whether heterogeneous-MPCR groups with sanctioning opportunities retain higher contributions than standard public goods game groups without sanctioning opportunities.

The experiment uses a linear public goods game with fixed three-person groups. Within a group, one participant is randomly chosen to have an MPCR of 0.9, while

[2]For example, in Guth et al.'s (2004) experiment, a leader can determine his own contributions first while other members have to decide after him. In the same experiment, power is used to describe an option to exclude members from a group.

the other two participants each have an MPCR of 0.4[3]. The group member with an MPCR of 0.9 is a "relatively privileged" player in the sense of Olson (1965). Olson (1965, p. 50) defines privileged groups as those in which at least one member has an MPCR greater than 1, meaning that she is willing to bear the full cost of providing public good because her marginal return is greater than her marginal cost per unit of contribution. In my experiment, the player with an MPCR of 0.9 is privileged in the sense that he faces a higher marginal rate than his group mates; yet the privilege is only relative in that he is not always willing to bear the full cost of provide public good if neither of his group mates contribute. Formally, a "relatively privileged" player is one who faces an MPCR higher than the other players but smaller than 1. While abundant literature has generated important insights on collective action in standard public goods games, public goods games with heterogeneous MPCRs smaller than 1 have received much less attention. This may be because, as long as MPCRs are all smaller than 1, public goods games with heterogeneous MPCRs have the same Nash equilibrium prediction as standard public goods games. However, heterogeneity in MPCR across players more closely matches real-life cooperative experiences. For expositional convenience, groups with a " privileged player" will henceforth be called "privileged groups". Groups without such a player are referred to as "non-privileged groups". Similarly, a relatively privileged player (one with an MPCR of 0.9) is used interchangeably with a "high player"; a non-privileged player (one with an MPCR of 0.4) is used interchangeably with a "low player" in the following sections.

In countless cases, people are expected to cooperate despite differences in their privilege. Cooperation is a necessary skill in daily life, yet we often face a conflict between maximising individual benefit versus optimizing collective group outcomes. How to promote cooperation in general social settings is an ongoing topic of daily conversation and debate. Experimental economists have actively contributed to this debate by constructing similar but simplified interactions in the lab and purposefully varying one or two elements, hoping to identify key determinants of group cooperation and to come up with ways to encourage better cooperation outside the lab. Social dilemma games are the most commonly used games in lab-based cooperation studies. Public goods games, one of the most commonly used social dilemma games

[3] The choice of the MPCRs follows the settings used in Glöckner et al. (2011). These values make my results directly comparable to those of Glöckner et al. (2011), while the MPCR of 0.9 makes the behaviour of Fehr and Schmidt's (1999) agents verifiable (See Appendix A.3 for further details).

in cooperation studies, have been used extensively in studying the effects of material punishment or reward on cooperation levels. However, most existing studies focus on standard public goods games with homogeneous players characterized by identical marginal return rates per unit of contribution.

Experimentalists have repeatedly found that punishment significantly increases group contributions, while the effect of reward on contributions is inconclusive (Balliet et al., 2011). Even if conclusive conclusions are drawn from standard public goods games, one should be cautious in generalizing the conclusions to non-standard public goods games, for example, those characterized by different individual marginal returns, because different individual marginal returns create different individual incentives, which are found to shift attention and trigger different norm enforcement (Reuben and Riedl, 2013). As non-standard public goods games with privileged players more closely mimic the above mentioned real-life examples without sacrificing simplicity, they are chosen in this study to explore the effectiveness of punishment and reward. The results contribute to the ongoing cooperation debate, as well as offering a new perspective on the public goods games literature in experimental economics. Incentives and cooperation mechanisms in public goods games can also help predict and foster cooperation in real-life research groups.

1.2 Literature Review

A standard public goods game in an N-person group is one in which all N members in a group face the same MPCR. In standard public goods games, while the social optimum is reached when all players contribute the full amount of their endowments to the public good, few instances of full contributions are observed in the lab. This is to be expected for strategic reasons: if one expects others to contribute a non-zero amount, he can take advantage of others' contributions without contributing himself. Over time when many public goods games are played in succession, contributions often decrease to zero. Adding a punishment or reward stage after the public good contribution stage in each round of play has been suggested, and tested in the lab, as a potential way to motivate higher contributions to a public good. In a punishment (reward) stage, any player can decide to decrease (increase) any other player's total payout at a cost.

The existing literature examining public goods games provide various insights into how the underprovision of public good can be mitigated in a standard N-person group. In an experiment focusing on non-privileged groups, Choi and Ahn (2013) find that the total contribution levels in groups with a reward stage and in groups with a punishment stage are statistically indistinguishable, and statistically higher than the total contribution levels observed in groups with neither rewards nor punishments. A meta-analysis of punishment and reward in cooperation games has noted that the effect sizes of the two incentives are similar, while punishment is slightly more effective (Balliet et al., 2011). As one of the several papers that explore reward in a public goods game, Sefton et al. (2007) find reward effectively increases contributions, although it is found less successful than punishment in sustaining high levels of contributions over multiple rounds. Vyrastekova and Van Soest (2008) observe, in a public bads game where players can exploit a public resource for own benefit at the cost of socially optimal outcomes, that the effectiveness of reward relates to the cost-effectiveness ratio (i.e., the unit cost of one reward point to the person doing the rewarding): when the cost effectiveness ratio is lower, reward is utilized more and is more effective. Effectiveness may also be moderated by different experimental designs (Andreoni et al., 2003; Walker and Halloran, 2004).

Yet much less attention has been devoted to the possibility of mitigating under-provision of public goods through the presence of a privileged player. As briefly noted above, a privileged player is one who faces a higher MPCR than other players. Groups with a privileged player are termed in two different ways in existing literature. One strand of literature term these groups "heterogeneous groups", contrasting "homogeneous groups" which stand for groups without a privileged player (Fisher et al., 1995). The other and more recent strand of literature term these groups "privileged groups", contrasting "non-privileged groups" (Reuben and van Winden, 2008; Glöckner et al., 2011). I follow the second strand of literature and term a group with a privileged member a "privileged group". The earliest experiment focusing on privileged players is Fisher et al. (1995). A privileged player is implicitly defined in two different ways in the existing literature. In Reuben and Riedl (2009)'s case, one is privileged only when one's MPCR is greater than 1. In Glöckner et al. (2011), one is privileged as long as one has a higher MPCR than other group members. In both cases, a privileged player is one who has stronger incentives than other group members to contribute to the public good in a group facing a public good contribution problem. Reuben and Riedl (2009)'s player has "absolute privilege", in the sense that their privileged player's total payout from playing the public goods game is never less than his initial contribution to the public good, regardless of what others in his group contribute. By contrast, Glöckner et al. (2011)'s player only has "relative privilege": he merely has a higher marginal return on his public goods contributions compared to others in the group, but he may be paid out less than his initial contribution to the public good if others do not contribute at all. Because a "relatively privileged" player's total payoff may still not fully refund his contributions (depending on all players' contributions), maximising his public good contribution may not be his optimal contribution decision. In many real-world groups, a leading member is essentially a "relatively privileged" player: he expects to get higher returns than the other colleagues but, also faces the possibility of making a net loss. Groups with an "absolutely privileged" player have been observed in lab settings to make higher total contributions to the public good compared to groups without a privileged player, and this is because a privileged player contributes significantly more than other members; the other players contribute at levels similar to what is observed in standard public goods games (Reuben and Riedl, 2009). Whether the presence of a "relatively privileged" player will have a similar effect has not yet been formally tested in the lab.

Even less evidence exists for the effectiveness or otherwise of punishment and reward incentives in privileged groups. For example, in Fisher et al. (1995), players' contributions are found to heavily depend on own MPCR. But they do not give players opportunity to punish or reward each other. In an experiment comparing the effect of punishment on everyone's contributions in privileged groups and non-privileged groups, Reuben and Riedl (2009) find that when all players can choose to punish another member, groups with an "absolutely privileged" player lose their advantage (in terms of total contributions made to the public good) over non-privileged groups. Initially, privileged groups in their experiment make higher total contributions to the public good than non-privileged groups, mainly because of significantly higher contribution levels of the privileged player, whose personal interest is in line with the public interest because they always earn weakly more from contributing to the public good than from refraining from contributing. After punishment is introduced, the "privilege gap" in total contributions disappears, because low contributors in non-privileged groups increase their contributions significantly, while members of privileged groups (including the privileged player) are less responsive to punishment. According to Reuben and Riedl (2009), in privileged groups, privileged players punish less strategically than do members of non-privileged groups. In another experiment including a "relatively privileged" player, Reuben and Riedl (2013) mention that punishment significantly increases the contributions of "relatively privileged" players, but has no effect on other players. In a recent lab-in-the-field experiment, Asiedu and Gross (2017) find that heterogeneous within-group returns, characterized by different MPCRs, increase contributions. Their explanation is that having a privileged player whose payoff structure is known to all group members helps to reduce strategic uncertainty.

The experiment in this paper is designed to explore the effect of reward opportunities on contributions to public good in a laboratory experiment, and to estimate the comparative efficacy of punishment and reward incentives in groups with a "relatively privileged" player.

1.3 Motivation and Contribution

Prior authors have looked at the impact of punishment on public goods contributions in privileged groups containing either an absolutely privileged player or a relatively privileged player. However, no previous design has studied the effect of reward on public goods contributions in privileged groups with a relatively privileged player, much less compared the effectiveness of reward and punishment in relatively privileged groups. This is despite the strong real-world relevance of relatively privileged groups and the fact that theoretically, the stronger likelihood of an interior solution to the contribution decision problem makes a relatively privileged player more interesting to study than an absolutely privileged player. On the one hand, because the marginal return of private goods is strictly larger than the marginal return of public good, a rational relatively privileged player is expected to follow the zero contribution strategy. On the other hand, players have been found to more frequently violate the dominant strategy of zero contribution when the marginal return of public good gets closer to the marginal return of private goods (Palfrey and Prisbrey, 1997). For all MPCRs smaller than 1, the Nash equilibrium for a rational person is to contribute 0 in the first stage, and never punish or reward in the second stage. However, this prediction is only based on the assumption of a first-order public good (Foster et al., 2017). An individual may commit to contribute more, to punish, or to reward, despite others' contribution decisions, when he or she is concerned about a second-order public good (e.g., maintain a social norm, form an institution etc.). Existing literature has repeatedly found violations of the equilibrium prediction: when incentives are available, participants will utilize the incentives to promote greater cooperation, or to elicit own preferences (Herrmann et al., 2008). Thus, I expect to see deviations from Nash equilibrium prediction. Along with that, I pay special attention to the differences between different types of groups, between different types of players, and between different incentives.

When punishment opportunities are available, players in homogeneous groups in which all players face exactly the same experimental conditions, largely agree on punishing low contributors in groups (Fehr and Gachter, 2000; Reuben and Riedl, 2013). Yet in heterogeneous groups where some experimental conditions are different for different players (e.g., MPCR, endowment etc.), disagreements on the desirable contribution rule arise (Reuben and Riedl, 2013). These disagreements vary depending on the source of heterogeneity as well. Some individuals use their punishment opportunities to encourage contributions, which, if successful, would maximise the group outcome. Some others prioritize equal or similar earnings among group members. So far, only Reuben and Riedl (2013) have examined the effect of punishment incentives on relatively privileged groups in a study on the enforcement of contribution norms. No previous literature has studied the effect of reward in such groups, although the effect of reward in general social dilemma games has been found to be statistically indistinguishable from the effect of punishment. What is the effect of reward on relatively privileged groups? Does punishment or reward more effectively increase public goods contributions in relatively privileged groups? Will total public-goods contributions in relatively privileged groups far exceed those in non-privileged groups, even in the presence of punishment? The answers are ambiguous.

The experiment in this paper is designed to answer three main questions. The first question is, in groups facing a public good problem that include a relatively privileged player, whether the peer-reward opportunities increase overall contributions to the public good. The prediction is that, compare to non-privileged groups with no sanctioning opportunities, groups with sanctioning opportunities deliver higher level of contributions and have mitigated decay in contributions over successive periods. The second question is whether punishment or reward is the more effective tool in increasing overall public goods contributions in groups with a relatively privileged player. The prediction is that reward will be at least as effective as punishment, because average group earnings are strictly increasing in reward and decreasing in punishment points. In other words, any reward points will cause a net increase in average group earnings. The third question is whether heterogeneous groups with sanctioning opportunities deliver a higher level of contribution than groups without a privileged player. The prediction is that a higher level of contribution can be maintained in relatively privileged groups with reward opportunities, but not necessarily in groups with punishment opportunities. Based on previous findings in Reuben and Riedl (2013), punishment opportunities only weakly lead to an increase in contributions for players. Given the net positive effect of reward on group earnings by design, privileged groups with reward opportunities are expected to generate significantly higher contributions than those with punishment opportunities, and thus higher than privileged groups without sanctioning opportunities.

The general formula to calculate an individual player's earnings in one round of a public goods game is as follows.

$$\pi_i = e_i - c_i + a_i \times \sum_i c_i \pm 3 \sum_{i \neq j} n_{ji} - \sum_{i \neq j} m_{ij}$$

where π_i denotes the earnings, e_i denotes initial endowments. a_i is the MPCR, which, in the present study will be set to 0.9 for privileged players and 0.4 for non-privileged players. c_i denotes one contribution by individual i, and n_{ji} denotes the incentive points assigned by player j to player i. The direction of incentives to the public good is positive in a reward treatment, and negative in a punishment treatment, which is shown above in the '± 3' term. Correspondingly, m_{ij} denotes the points assigned by player i to player j, which costs player i to implement.

To predict contribution behaviours of players, two theories are adopted here. *Homo oeconomicus* model assumes rational and purely selfish agents. The other one is a model with inequality aversion from Fehr and Schmidt (1999). Both models can tractably derive explicit predictions for a one-shot public goods game.[4] Importantly, this experiment is by no means a rigorous test of the theories. Rather, theories are adopted here to facilitate constructions of theoretical predictions.

[4]In repeated games, other dynamics may be present and are unmodelled here.

1.3.1 *Homo Oeconomicus* Predictions

A player predicted by this theory cares only about his own payoffs. For any $\frac{1}{n} < a < 1$, the equilibrium solution is zero. Furthermore, any individual's contribution is independent of other players' contributions.

With sanctioning opportunities, theoretical predictions do not change. Any individual player will contribute zero to the public good project. Because no selfish player would actually enforce the incentive scheme at own cost.

1.3.2 Predictions with Fehr & Schmidt (1999) Utilities

Assuming there exists subjects who dislike inequitable outcomes, both when they are better off and when they are worse off in monetary payoffs compared to another player. Additionally, subjects dislike material disadvantage more than material advantage. The utility function is as below.

$$U_i(\pi) = \pi_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} max\{\pi_j - \pi_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} max\{\pi_i - \pi_j, 0\}$$

where n denotes number of subjects in a group, α denotes level of disadvantageous inequality aversion, and β denotes level of advantageous inequality aversion ($\alpha > \beta$).

This paper modifies the theoretical model of Fehr and Schmidt (1999) to account for having a relatively privileged player in three steps (see Appendix A.3). First, taking into account Sutter et al. (2010)'s extension of the original inequality aversion model, I reconstruct the conditions for standard public goods game. Second, following the same structure, I calculate changes in utilities in public goods games with a relatively privileged player. Third, theoretical predictions are derived for public goods game with heterogeneous MPCRs. The predictions for a one-shot public goods game are listed below.

- *Prediction 1: Groups with a relatively privileged player will generate fewer zero* *contributions compared to groups without such a player.*[5]

- *Prediction 2: There is a continuum of positive contribution bundles forming Nash Equilibria. If a high player has* $\beta_i > 1/15$*, then he always contributes full endowment, while all low players contribute a positive amount c. Otherwise, the high player will contribute the same amount as everyone else does.*

- *Prediction 3: Reward is more credible than punishment in privileged groups, because the binding conditions for punishment are always more stringent than those for reward.*

- *Prediction 3.1: Reward opportunities more effectively increase group coordination than punishment opportunities in privileged groups.*

Based on these predictions, I investigate public goods contributions in groups that contain a relatively privileged player in an environment in which it is possible for all players to punish or reward any other member. Not only does this experiment fill a gap in the research literature concerned with the effects of the presence of privileged players on behaviour in public goods games, but it responds directly to three real-world empirical observations. First, many groups or teams working towards a common purpose are composed of members with varying degrees of interest in that purpose, be it a small research project or a group fighting against global warming. This reality matches better to a public goods game setup in which players' incentives to contribute to the public good vary, than to the traditional setup in which all players' contribution incentives are identical. Second, when interest does vary, a relatively privileged player arguably appears more often than an absolutely privileged player, as it is often possible for one highly-committed group member to lose some of his contributions to the public good, depending on whether others also contribute. Third, friendly reward and spiteful punishment are both often available to real-world groups. It would be interesting to identify which one better promotes contributions to the public good.

[5] Given the distribution of β in Fehr and Schmidt (1999), the prediction will not change for $a_h \in (0.75, 1)$.

1.4 Experimental Design

The game implemented in this experiment is a linear public goods game played in three-person groups. Participants are randomly assigned into three-person groups at the start of an experimental session. The overall design consists of four distinct conditions: one control condition, two treatment conditions, and one baseline condition. The baseline condition, as a benchmark to a conventional public goods game, runs public goods games with players facing the same MPCR of 0.4 (Fehr and Gachter, 2000). The control condition runs public goods games with one randomly chosen player facing an MPCR of 0.9 and two other players each facing an MPCR of 0.4. Whoever is chosen to have an MPCR of 0.9 in the first period retains the same marginal return rate throughout all periods. This player is the relatively privileged player. In both the baseline condition and the control condition, a player can only make one decision per period; that is, how much he contributes to the group public good. The treatment conditions retain the MPCR design in the control condition and add a sanctioning stage after each contribution decision. The sanctioning option is either reward or punishment depending on the treatment. In all four conditions, 1) a public goods game is repeated for ten periods and groups remain unchanged throughout the ten periods; 2) participants receive feedback, at the end of each period, on each group member's contributions and earnings in that period (see Table A.2 in Appendix A.4 for a tabulated explanation of the experimental design).

After participants enter the lab, they are instructed to read the instructions (see Appendix A.1 for instructions in the punishment condition[6]) carefully. A participant may raise his hand to ask questions, yet communication between participants is not allowed. Before starting the actual game, participants first answer several control questions to demonstrate their understanding of the experimental instructions.[7] Participants then answer three sets of calculation questions (different from the control questions) to earn their 20 starting tokens for each period. In other words, once a participant answers all calculation questions correctly, he is guaranteed to have 20 experimental tokens at the start of each period and has to make a contribution decision given an endowment constraint of 20 tokens in a period. All players make their contribution decisions simultaneously.

[6]Instructions for other conditions are note presented here for brevity.

[7]For a few participants who had doubts or questions, the experimenter explained the details until they fully understand the instruction.

If there is a second stage providing sanctioning opportunities, participants are informed about the contributions to the project made by each of the other two group members. They can then decide whether to decrease, increase, or leave unchanged the amount of tokens that each of the other group members earned in the first stage. The words "punish" and "reward", are deliberately replaced with the more neutral expressions "increase" and "decrease" in the experiment to minimise any emotional arousal effect. If a participant decides to decrease or increase others' first-stage earnings, he is able to decide the amount of decrease or increase for each person separately, by assigning points to each group member. Between 0 and 10 sanctioning points are available to each participant to be allocated as he sees fit across the other two participants. The upper and lower limits are pre-designed in the experiment following existing literature (Fehr and Gachter, 2000; Reuben and Riedl, 2009). Each point a participant assigns towards decreasing or increasing another participant's earnings will cost the assigning participant 1 token, while it decreases or increases another participant's earnings by 3 tokens. A participant cannot see how many each individual group member spent to decrease or increase his or her first-stage earnings, but only the total amount of points assigned by the two other group members combined. At the end of each period, a participant sees how many reward or punishment points are assigned to or by himself. He can also see how his earnings are calculated step by step.

A total of 11 experimental sessions are run, including one pilot session, two baseline sessions, four punishment sessions, and four reward sessions. The experiment is programmed in z-Tree (Fischbacher, 2007). In total, 87 subjects participate in punishment sessions, 81 subjects participate in reward sessions, and 30 subjects participate in baseline sessions.[8] Subjects are recruited via ORSEE (Greiner, 2015). Each session runs a single treatment, consisting of 2×10 periods of a public goods game played in fixed groups. Other than the baseline session, all sessions consist of 10 periods with sanctioning opportunities provided, plus 10 periods without sanctioning opportunities. The periods without sanctioning opportunities form the control observations. As is illustrated in Table A.2, control observations are pooled from different sessions. The order of periods with and without sanctioning opportunities is altered to control for potential order effects. An experimental session on

[8]Five participants (three in punish sessions and two in reward sessions) left halfway in the experiment. Their data are dropped in the analysis.

average lasts 60-90 minutes[9]. Earnings in this experiment are paid according to final earnings in one randomly chosen period out of the 20 experimental periods in total.

1.4.1 Further Comments on Experimental Structure

A public goods game with sanctioning opportunities can be viewed as a combination of a public goods game and a prisoner's dilemma game. When repeated over periods, as it is in this experiment, this game structure is equivalent to a repeated version of two games. Due to the theoretical complexity arising in this setting, there is no clear analytical or theoretical predictions arise with which to frame empirical analysis.[10]

1.5 Results

1.5.1 Reward Strongly Motivates Contributions

Figure A.3 presents average group contributions over 10 periods, contrasting groups in the reward condition with groups in the control condition. Except in the first period, average group contributions in the reward condition always exceed those in the control condition. In fact, the average group contributions in the reward condition are 25.159 (SD 15.935) tokens, whereas the average group contributions in the control condition are only 16.225 (SD 16.235) tokens. A mean difference test assuming unequal variances show that the difference in contributions is strongly significant ($p < 0.00001$). Results remain the same after controlling for order effects. An additional ordinary least square (OLS) regression with robust standard errors returns the same result (Results are not presented here for the sake of brevity).

Average group contributions in the reward condition become statistically significantly different from those in the control condition starting in period 3 (see Table A.9 in Appendix A for mean different tests by each period). The difference remains till the last period of the experiment and grows bigger over successive periods. According to Figure A.3, the growth in the difference over successive periods seems

[9]Two sessions were particularly long because a few participants struggled in the control and calculation questions and spent more than half hour in answering those questions, significantly slowing down a whole session

[10]This section owes special thanks to a discussant Jose Rodrigues-Neto.

to be driven by the contribution decay in the control condition which does not seem to appear in the reward condition. Spearman's ρ tests are conducted to verify this conjecture. Group contributions in the control condition have a ρ of -0.339 ($p < 0.0001$), indicating a significant decreasing trend with a correlation coefficient of -0.339. Group contributions in the reward condition have a ρ of -0.054 ($p = 0.377$), indicating no increasing or decreasing trend. Indeed, reward opportunities nullify the decreasing trend of contributions over successive periods.

Given the large difference in average group contributions between the reward condition and the control condition, one might wonder whether high players, low players, or both types are the main driving force of the difference. I investigate this aspect by looking at contributions by player type. Figure A.1 dissects the groups and presents average contributions for high and low players respectively. A high player in the reward condition on average contributes 10.663 (SD 7.177) tokens, whereas a higher player in the control condition only contributes 7.377 (SD 7.624) tokens. In other words, the reward institution increases a high player's average contributions by 44.5%. Meanwhile, a low player in the reward condition contributes on average 7.248 (SD 6.895) tokens, whereas a low player in the control condition only contributes 4.410 (SD 6.033) tokens on average. A low player contributes 64.4% more when exposed to the reward institution than when in the control condition. Mean difference tests assuming unequal variances confirm that both contribution differences are strongly significant ($p < 0.00001$). Results remain the same after controlling for order effects. All OLS regressions with alternative specifications (using robust standard errors and clustering by groups) retain the same conclusion. Thus, both high players and low players increase their contributions significantly, when exposed to the reward condition than when exposed to the control condition. The magnitude of the relative increase is larger for low players, indicating that reward increases low players coordination more than it does on high players.

Mean difference tests by period for either player type are presented in Panel A of Table A.10 and Table A.11. High players in the reward condition start to contribute significantly more than low players in period 4 and remain so till the end of the experiment. Similarly, the contribution difference of low players between the reward condition and the control is significant from Period 5 onwards. Spearman's ρ tests are also conducted to verify whether the decay mitigation manifests for both types of players. The contributions of high (low) players in the control condition decrease by a ρ of -0.278 (0.297). Both decreasing trends are strongly significant ($p < 0.0001$). Turning to players in the reward condition, statistical evidence cannot reject ($p > 0.59$ for both types of players) the hypothesis that contributions are independent of Period. Put differently, reward opportunities nullify the decreasing trend of contributions for both types of players over successive periods.

Overall, reward opportunities have been found to strongly motivate higher level of contributions, at both the group level and the individual level. The motivational effects mainly manifests toward latter half (Period 6 to 10) of the experiment where strong and significant contribution differences between groups in the reward condition and groups in the control condition are captured. Further investigation shows that the reward opportunities eliminate contribution decay over successive periods. Given that contributions in the control condition decay quickly over successive periods, the elimination of contribution decay is a main driver of the strong overall contribution differences.

1.5.2 Reward Outperforms Punishment in Motivating Contributions

The previous section shows that reward opportunities strongly motivate higher individual and group contributions in a privileged group. Given the mixed and limited evidence on the comparative efficacy of punishment and reward in standard groups facing homogeneous MPCRs (Choi and Ahn, 2013), one wonders whether one sanctioning institution outperforms the other in privileged groups. In this section, I investigate the comparative efficacy of punishment and reward in privileged groups.

Figure A.4 is identical to Figure A.3, except for the additional blue line connecting the squares which represents average group contributions over 10 periods in the punishment condition. Pairwise comparisons between the lines indicate 1) average contributions in the punishment condition is lower than those in the reward condition, and 2) both sanctioning institutions mitigate contribution decay. Recall that the average group contributions in the reward condition are 25.159 (SD 15.935) tokens. The average group contributions in the punishment condition are only 15.379 (SD 13.148) tokens. A mean difference test again rejects ($p < 0.00001$) the hypothesis of equal means (OLS regressions retain the conclusion). Additionally, a mean difference test of average group contributions in the punishment condition versus those in the control condition does not reject the equal mean hypothesis, indicating that punishment opportunities do not significantly increase average group contributions. Interestingly, a Spearman's ρ test captures that contributions in the punishment condition decay with a correlation coefficient of -0.145 ($p < 0.014$) over successive periods, a seemingly mitigated decreasing trend compared to the -0.339 in the control condition. However, a Mann-Whitney test does not reject the hypothesis that the control and punishment samples are from the same population distribution, whereas the test strongly reject the same hypothesis for the reward-versus-punish comparison. In other words, compared to punishment opportunities, reward opportunities significantly motivate higher average group contributions, as well as mitigating contribution decay over successive periods.

To further untangle which type of players drives the significant contribution difference between the reward and the punishment condition at the group level, Figure A.2 presents average individual contributions by high and low players respectively. Unsurprisingly, the patterns are quite similar to those in the reward-vs-control comparisons. Recall a high player in the reward condition on average contributes 10.663 (SD 7.177) tokens. A higher player in the punishment condition only contributes 7.238 (SD 6.866) tokens, an amount almost identical to 7.377 tokens in the control condition. The relative increase when switching from a punish institution to a reward institution is around 47.3%. As for the average contributions (mean 4.069, SD 4.840) of a low player in the punishment condition, the relative increase is 78.1% when exposed to the reward condition. The contribution differences of either type of players between punishment and reward condition are strongly significant ($p < 0.00001$) (All OLS regressions with alternative specifications retain the same conclusions; Pairwise mean difference comparisons by each period can be found in Table A.9 to Table A.11 in the appendix.) Hence, reward opportunities motivate both high players and low players to increase their contributions significantly, than do punishment opportunities. Both types of players increase their contributions by a similar magnitude of 3 tokens, rendering a larger relative increase in low players' contributions. As for the dynamics over successive periods, both types of players show statistically significant trends of contribution decay, at a rate of around -0.11 ($p < 0.05$). Statistical evidence again does not reject that the punishment and the control samples are from the same distribution, but strongly rejects this hypothesis for the punish-versus-reward comparison.

Overall, this section provides further evidence that reward opportunities are effective in motivating higher contributions, not only relative to the control condition but also to the punishment condition. The effectiveness of the reward institution manifests at both the group level and the individual level. Punishment opportunities per se, if anything, only weakly increase contributions in some periods and weakly mitigate contribution decay over successive periods. However, these effects are not supported by statistical evidence, which could potentially be driven by the small sample size.

1.5.3 Punishment and Reward Behaviour

The previous section confirms that reward is more effective than punishment in increasing individual contributions, as well as in mitigating contribution decay over periods. In this section I investigate what is driving the difference between those two conditions.

1.5.3.1 Contributions and Sanctioning Decisions

One reason might be that high and low players follow different strategies in punishing and rewarding others. For example, a high player in a group punishes low contributors and rewards high contributors to increase overall group contributions. On the contrary, low players within a group might be spiteful of a high player's privilege, so that they punish or reward to shrink the earning differences among players. Another possibility is that, upon being punished or rewarded, different types of players respond differently. For example, a high player may decide to revenge in future periods if punished and reciprocate if rewarded, while a low player may be very reluctant to revenge fearing a high player's further revenge back. To explore these possibilities, we can compare the average punishment and reward points received and assigned by each player. In the punishment condition, high players receive more punishment points than low players do (2.14 vs. 1.54, $p < 0.001$). This is despite the fact that high players on average contribute almost twice the amount that low players do. This result is consistent with Reuben and Riedl (2013), where low players appear to enforce a norm that leads to equal earnings. Whether players in this experiment is driven by a similar norm is discussed in Section 1.5.3.2. Although high players appear to assign slightly fewer punishment points to others, the difference is not statistically significant (1.51 vs. 1.73, $p < 0.278$). In the reward condition, high players receive fewer reward points than low players (2.35 vs. 3.19, $p < 0.002$). Meanwhile, high players assign significantly more reward points than low players do (4.13 vs. 2.30, $p < 0.001$).

To analyse the determinants of punishment and reward behaviour, Tobit regressions are fitted, applying a lower limit of 0 and an upper limit of 10 points on the dependent variable (e.g., the punishment points or reward points). The choice of variables to include in the regression follows that in Reuben and Riedl (2009). The dependent variable is the amount of punishment or reward points assigned by player i to player j, and control variables are as follows: the deviation (i.e., number of experimental tokens) of player j's contributions from punisher i's contributions; the deviation of player j's contributions from the third player; whether or not player j is a high type; period; and an interaction term between period and high player. Results are reported in Table A.3. Columns (1) and (3) are for high players; columns (2) and (4) are for low players. Here the deviation is expressed by two dummies. One dummy equals 1 for positive differences, and the other dummy equals 1 for negative differences. Results are similar using actual differences in number of experimental tokens (not reported here for brevity).

A player's contributions do not seem to significantly change the punishment he receives. This seemingly puzzling finding is consistent with Reuben and Riedl (2009)'s finding that players do not punish strategically in privileged groups. However, punishment significantly correlates with the contribution differences between players. Specifically, a positive deviation from the punisher (i.e., where the punished player contributed more than the punisher) is more likely to get punished, while a negative deviation from the third player (i.e., where the punished player contributed less than the third player) is more likely to get punished. This effect is significant and consistent between two types of players. If player j's contributions are higher than player i's contributions, player i, if high, tends to punish player j more and reward less. This is a bit counter-intuitive. However, it is possible that a high player wants to maintain his preferred contribution level. For example, if the high player prefers to contribute little himself, then he might dislike seeing another player contributing more than him. Or if the high player prefers to contribute a lot, then he might dislike seeing low contributors.

Columns (3) and (4) report results for the reward treatment. Overall, if a player contributes more, he gets significantly more rewards. This is strongly significant and consistent for both types. If a player deviates positively from the potential rewarder's contributions, then the player is less likely to get rewarded. Reward decreases over successive periods. Interestingly, a high player appears to receive more reward points.[11]

Overall, reward is used to a larger extent by both types of players as an incentivizing tool, compared to punishment. Although high contributions are more likely to attract more reward points among players, players do not vary punishment points much depending on another player's contributions. Both high and low players like to punish whoever contributes more than self and reward whoever contributes less than self. This is suggesting that players use punishment and reward more as norm enforcement tools, rather than ways to promote better social outcomes.

1.5.3.2 Earning Differences and Sanctioning Decisions

In addition to contributions, another potentially significant determinant of punishment and reward behaviours is players' first-stage earnings. Because players decide how many points to assign to another player immediately after seeing each group member's first-stage earnings. Table A.4 displays the corresponding results. π_i, π_j, and π_k denote the first-stage earnings of the punisher/rewarder, the person being punished/rewarded, and the third player, respectively. All other notations are the same as in Table A.3. A high-type punisher does not seem to care much about the difference between own and the punished person's earnings. Instead, a high player cares more about whether two low players are earning the same amount. Whichever low player is earning more gets more punishment from the high player. Low-type players also increase punishment if the difference between two other players' earnings increases. One puzzling effect is that when the punished player earns more than a low-type punisher, the punisher significantly decreases punishment points as well. Maybe when the low-type punisher perceives that the earnings of another person is beyond the power of his punishment, he simply refrain from acting. In the reward condition, whoever earns more than the rewarder gets more reward. This action is possibly out of the intention to get more reciprocal reward points from others in future periods.

[11] A conversation after one experimental session with a subject reveals an interesting thought process. The subject said that he wanted to show kindness to the high player by giving reward, hoping that the high player would then act more generously in future periods.

Reward increases the level of overall contributions to the public good, and punishment does not; and reward is the more effective tool in increasing overall public goods contributions, in groups with a relatively privileged player. Compared to low players, high players make more contributions, and they are more responsive to both sanctions.

1.5.4 Comparisons Between Privileged Groups and Non-privileged Groups

1.5.4.1 Increased Coordination in Privileged Groups

First I consider whether low players adjust their next-period contributions according to the contributions made by high players. Table A.5 demonstrates that this is indeed the case. In Model 1, contributions of low players are regressed on a high player's previous period contributions, controlling for session and last period effects. Because no high player exists in the baseline condition, a player is randomly selected in a group to compare with the other two players. Overall, low players' contributions in the current period significantly depend on the high player's contribution in the previous period. The effect remains significant for all four conditions.[12] In Model 2, own and the third group member's contributions in the previous period are added. Results show that own contributions in a previous period significantly influence contributions in the next period. Low players' contributions show consistency, as the positive coefficient of $C_{Low,t-1}$ remains positive and significant throughout various specifications. The effect of a high player's contributions remain significant in the reward and control conditions, and marginally significant in the punishment and baseline conditions. In terms of magnitude, in the punishment condition another low player's contributions are stronger drivers of a low player's contributions than the high player's contributions. In model 3, all three contribution variables lagged two periods are added as regressors. In the control and baseline conditions, a one unit increase in the high player's contributions in the previous period significantly increases a low player's contributions by approximately 0.1 unit experimental tokens. A high player's contributions appear to have the largest effect in the reward

[12]Own contributions are also regressed on the average of the other two group members' contributions. Results are similar and hence not reported here, except that the coefficient in the baseline condition becomes more significant ($p < 0.001$). But the R^2s are slightly smaller, indicating less explanatory power of the alternative modelling method.

condition, where the coefficient on low players' contributions is about twice or even three times that of the others three conditions and is statistically significant.

An additional interesting finding is regarding the correlations between two low players. One low player's contributions significantly influence the other low player's contributions in non-incentivized conditions. Specifically, a one token increase in one low player's contribution increases another low player's contribution by 0.18 token in privileged groups. In baseline condition, the effect is similar in magnitude, but only marginally significant.[13] However, when reward or punishment opportunities are available, the interdependence effects between two low players disappear. One low player's contributions do not significantly depend on another other low player's contributions any more. Additionally, the magnitude of another low player's effect is almost twice that of the high player's. This is suggesting that, although a low player's contributions are significantly affected by a high player's contribution decisions, another low player's contributions has a larger impact compared to the high player.

In parallel, I explore the effect of a low player's contributions on the contributions of the high player. Results are reported in Table A.6. A high player is much more responsive to a low player's contribution, than is true in reverse (see Table A.5): if a low player increases his contributions by one unit, a high player will increase his contributions by 0.2 unit in the next period. The analogous coefficient capturing the reciprocal effect of high players' contributions on low players' contributions is 0.09. Comparing Table A.5 and Table A.6, several interesting findings emerge. First, low players' contributions are significantly affected by contributions of a high player in the reward condition, but not in the punishment condition. On the contrary, a high player's contributions are more significantly affected in punishment condition than in reward condition. As expected, both low players affect the high player to a similar extent. When punishment or reward opportunities are available, the effect of another low player's contributions on a high player's contributions is less robust compared to in the control condition, but the effect remains statistically significant. This finding is slightly different from Reuben and Riedl (2009)'s study, where a high player's contributions no longer significantly relate to a low player's contributions once there is punishment. Their explanation is that, without punishment opportunities, high players only cooperate conditionally and use own contributions as a

[13]Designation of high and low players are random in Baseline groups.

disciplinary tool, but once punishment opportunities are available, players can resort to punishment for motivational purposes instead of varying own contributions strategically. Participants in this experiment are less willing to use punishment, as shown in Section 1.5.3 where average punishment points are much smaller than average reward points (and smaller than found in Reuben and Riedl (2009)). Thus players still need to strategically vary their own contributions as a compensatory tool. We observe lower magnitude and a lower significance level of the impact of low players' contributions on high players' contributions in the reward condition, consistent with the previous conjecture: when participants can incentivize each other using reward, they no longer need to resort to strategic contributions. As in Table A.5, "relative privilege" seems to increase mutual dependence among players in groups, serving as a coordination device. Additionally, in Model 3 the reward condition has an adjusted R^2 greater than 0.8, indicating a stronger explanatory power of the model than of three other conditions.

One surprising pattern is that the existence of a high player in a group dramatically changes the contribution levels of low players, compared to the baseline condition. This is despite the fact that low players in privileged groups have exactly the same amount of endowment and same MPCR as baseline players have.[14] Comparing Table A.5 with Table A.6, we see that in all privileged groups, be it incentivized or not, players' contributions significantly depend on contributions of other group members. This is suggesting that a high player's "privilege" is well-perceived by other players who adjust own contributions and coordination accordingly in the next period. By contrast, in non-privileged groups (baseline), players' contributions either do not significantly depend on another member's contributions, or the dependence is only marginally significant. A possible explanation for the significant effect in privileged groups is that low players are more willing to follow a high player's contribution pattern using tit-for-tat strategies. Players in privileged groups show greater mutual dependency in terms of contributing decisions, compared to those in non-privileged groups. In other words, "privilege" serves as a coordination device among group members.

[14] Designation of high and low players are random in Baseline groups.

1.5.4.2 Does Privilege Increase Public Good Provision?

Table A.7 compares contributions by player types in control and baseline condition. The comparison between low players are especially interesting, since low players always have the same starting endowment and the same MPCR despite treatment condition. Table A.7 shows that both high players and low players in privileged groups contribute significantly more than participants in baseline groups. Results are similar when comparing participants in the reward or punishment condition with those in the baseline condition (Results are not presented here for brevity). Previously, Fisher et al. (1995) find weak and insignificant evidence that in privileged groups, high types tend to lower their contributions and low types tend to increase their contributions relative to their equivalents in homogeneous groups. The results from this experiment are in line with their findings on low players, and are statistically significant. Specifically, low players contribute significantly more in the privileged condition than their parellels in the control condition throughout successive periods.

Table A.7 also indicates that, although the difference between high players in privileged and non-privileged conditions are significant over successive periods, it tends to shrink quickly.[15] In the last period, the difference between a high player and a randomly selected player in the baseline condition is no longer significantly different. Another noticeable difference is the large standard deviations in privileged groups, indicating a lack of consensus among players in privileged groups.

1.5.5 Discussion on Sustained Coordination and Improved Efficiency

1.5.5.1 Reward Sustains Coordination

After answering the three research questions, one step further is to identify which design more successfully sustains group coordination. Table A.8 sheds light on this question. An individual player's next-period contribution is significantly altered by reward points in the current period, but is not influenced by punishment points. Based on this evidence, reward more successfully sustains coordination in privileged groups. Meanwhile, we need to bear in mind that players on average assign much

[15]Designation of high and low players are random in Baseline groups.

fewer punishment points than reward points. It is possible that if punishment is utilized to a greater extent, punishment will effectively sustain coordination as well. This question remains to be answered in future researches.

1.5.5.2 Reward Improves Efficiency

Strong evidence in previous sections has shown that reward opportunities motivate higher contributions compared to punishment opportunities or lack of sanctioning opportunities. One may wonder whether the reward institution is efficient. To evaluate efficiency, two methods are adopted here. First, I compare the proportion of nonzero individual contributions across different conditions (Dawes et al., 1986). This is to explore, compared to participants in a different condition, whether participants in one condition are more willing to deviate from the theoretical prediction of zero contribution and cooperate by contributing a positive amount. Second, I compare participants' earnings among different conditions (Reuben and Riedl, 2009). This method captures whether participants in one condition end up being better off than those in another condition.

Using the first method, the outcome of interest is binary: zero for zero contributions and one for nonzero contributions. The binary nature of the variable, which eliminates potential confounding effects of different MPCRs, allows me to include the baseline condition in the comparisons. The proportion of nonzero per-period contributions in the reward condition is 85.6%. The corresponding proportions in the punishment, control, and baseline conditions are 73.3%, 62.7%, and 46.5%, respectively. This observation is in line with prediction 1 in Section A.3. Pairwise mean difference tests assuming unequal variances reject the hypothesis that two means are the same. In other words, the reward condition gives rise to a a significantly ($p < 0.00001$) larger proportion of coordinating participants compared to the punishment condition, whereas the punishment condition results in a significantly larger proportion of coordinating participants compared to the control condition. Regarding the control and the baseline conditions, it is reasonable to assume that the change of MPCRs only affects pre-existing coordinators who would have contributed a positive amount anyway. Because the change of MPCRs from the baseline condition to the control condition does not change the Nash equilibrium prediction, a non-coordinator has no reason to turn to a coordinator. Interestingly, the control condition seems to bring about a higher proportion of coordinators compared to the baseline condition. This observation suggests that other mechanisms might be at play. For example, a larger marginal per group return may increase participants' willingness to coordinate.

Turning to the efficiency measure using participants' earnings, a participant in the reward condition on average earns 31.685 tokens per period. This is strongly larger than both the average per-period earnings of 17.007 tokens in the punishment condition and the average per-period earnings of 23.860 tokens in the control condition. Both differences are statistically significant with $p < 0.0001$. Additionally, a participant's average per-period earnings in the control condition are statistically significantly larger than a participant's earnings in the punishment condition. However, it is worth noting that the first-stage earnings, earnings after a contribution stage and before a sanctioning stage, between participants in the control condition and those in the punishment condition are not statistically different from each other. Recall that the cost-benefit ratio of a one-point sanction is 3 (i.e., it costs a participant one point to increase or decrease another participant's earnings by three points). This lead me to wonder whether the significant difference in participants' earnings between the punishment condition and the control condition is purely due to resource destruction, a inborn nature of punishment. Similarly, it is possible that the significant efficiency improvement in the reward condition, compared to all other conditions, is a result of resource creation which is easy to implement in an experiment but seems less practical in real life. Assuming any material reward one gains comes from some material loss another one suffers, I investigate how the net social welfare changes in the sanctioning conditions.

I recover the social welfare by, 1) for participants in the reward condition, deducting earning increases purely due to the creation effect of the reward factor and 2) for participants in the punishment condition, adding back earning losses purely due to the destruction effect of the punishment factor. Although the magnitudes of the differences are smaller (Results are not presented here for brevity), all conclusions remain the same and all differences remain strongly significant. This suggests that, if resources from another sector of the society can be temporarily reallocated to the collaboration projects in privileged groups, the net social welfare will increase even after paying back the temporary reallocations for rewarding purposes. On the contrary, a punishment institution does not bring about net social welfare improvement, even when treating the resource destruction as a resource reallocation to another sector in the society.

1.6 Conclusion

Using a public goods game, this experiment studies 1) the effectiveness of reward in motivating contributions in privileged groups, and 2) the comparative efficacy of punishment and reward in privileged groups. Overall, reward opportunities strongly increase average group contributions in privileged groups. Reward opportunities also effectively increase both high and low players' contributions, as well as mitigating contribution decay over periods. Punishment opportunities only weakly increase group and individual contributions in some periods. However, the contribution differences between participants exposed to the punishment institution and those not exposed to sanctioning opportunities are not statistically different, indicating punishment opportunities do not effectively increases contributions in privileged groups. Unsurprisingly, reward is found to be much more effective than punishment in motivating contributions, as indicated by mean difference comparisons at both the group level and the individual level.

To the best of my knowledge, I am the first to explore the effectiveness of reward opportunities on contributions in a public goods game in a privileged group (i.e., a group with differential MPCRs for different group members). I find that reward opportunities are strongly effective in motivating public goods contributions in privileged groups, while punishment opportunities are not effective (at most weakly effective) in this situation. The weakened (or lack of) effectiveness of punishment opportunities is in line with the findings in Reuben and Riedl (2009) and Reuben and Riedl (2013). I also find that reward institution is strongly more effective than punishment institution in motivating contributions in privileged groups. This finding fills in a research gap relating to the comparative efficacy of punishment and reward in privileged groups. This finding is in contrast with what is commonly observed in non-privileged groups facing homogeneous MPCRs. In non-privileged groups, reward is found at most as effective as punishment in motivating group coordination (Balliet et al., 2011). This contrasting difference merits future research. For example, future research can explore the implications of the effectiveness of punishment and reward in privileged groups with more than three members.

Turning back to the motivating example introduced at the start of this paper, the group of friends facing heterogeneous returns may well accept the "privilege" of one member, because a "privileged" member has a higher rate of private returns in the start-up NPO and tends therefore to more readily contribute a higher level of effort. As a result of this increased incentive faced by the privileged member, all groups members' effort levels can be driven up, bringing about greater group outcomes, which can benefit every member. Meanwhile, the privileged member himself may also be aware that other members perceive the difference in returns and can increase their contributions in the start-up NPO. When it comes to sanctions, reward is the better option for increasing contributions and mitigating decays. In other words, if a member is willing to devote time and resource at his own cost for the greater benefit of another member, all group members are likely to end up enjoying a better outcome. Recalling the real-world example motivating this experiment of the team containing one member who stands to gain more than the other members if a start-up NPO on which they are all working succeeds, these results imply that when any members of such a team praise or otherwise reward others in the team (at their own personal effort, time, or monetary cost), individual contributions to the team project are significantly increased– and thereby everyone is benefited. Although privilege generally results from the fact that one member stands to gain more from the joint output than his collaborators, it is worth noting that a "privileged" member in a group project may actually be a disadvantaged member in a broader sense. Imagine two professors collaborating on a research project with a PhD student. The PhD student is "privileged" in the sense that she expects higher marginal returns from this project compared to the professors. However, the "privilege" is actually driven by the student's relative disadvantage in her research career. The current experiment assigns "privilege" to a player exogenously.

Future experiments may extend the current research in several dimensions. First, one may vary the starting endowments for group members and investigate the interaction between heterogeneous endowments and heterogeneous marginal returns. Second, one can compare the effect of reward in privileged groups to its effect in standard public-goods-game groups. Third, it would be interesting to look at the robustness of the effectiveness of reward opportunities by varying the cost-benefit ratio of a reward point. Fourth, one could integrate the cost of sanctioning opportunities into a participant's endowment and see whether different patterns emerge. The sanctioning options are intentionally kept simple and constant throughout the experiment, so that changes of conditions are minimised across treatments and I can provide clean evidence to answer the main research question. Future experiments may allow sanctioning options to vary for members with different MPCRs and explore the interplay between those two elements. Last but not least, one may vary the value of MPCRs or the size of a group, elements that have been documented[16] to affect public goods contributions, to see whether similar results hold.

[16]See Kagel and Roth (2016) for a discussion of these elements.

Appendix A

A.1 Experimental Instructions

Experiment Instructions

Welcome to the experiment! If you read these instructions carefully, you can, depending on your decisions, earn a considerable amount of money. The money you earn will be paid to you privately and in cash immediately after the session.

Communication between participants is not allowed. Please use the computer to input your decisions. Please do not start or end any programs, browse the web, or change any settings on the computer.

The experiment consists of two parts. At the start of Part 1, you will be randomly allocated to a group of 3 players (yourself and two others). You will stay in the same group throughout Part 1, and be reassigned to a new group in Part 2. Your Part 2 group will not include either member of your Part 1 group.

Your earnings in this experiment will be paid according to what you earn in one randomly chosen period out of 20 experimental periods in total (10 in Part 1, and 10 in Part 2).

Total Payment
Your payment for participating in this experiment will be calculated as follows:

Your earnings in Australian dollars (AUD) from a randomly selected period
= number of tokens from that period × Exchange rate*
+ 5 AUD show-up fee
= Total payment to take home

*The exchange rate will be calculated to ensure the average payout of participants in this session is between $18 and $20 per hour.

Below are the instructions for Part 1. The instructions for Part 2 will be provided later.

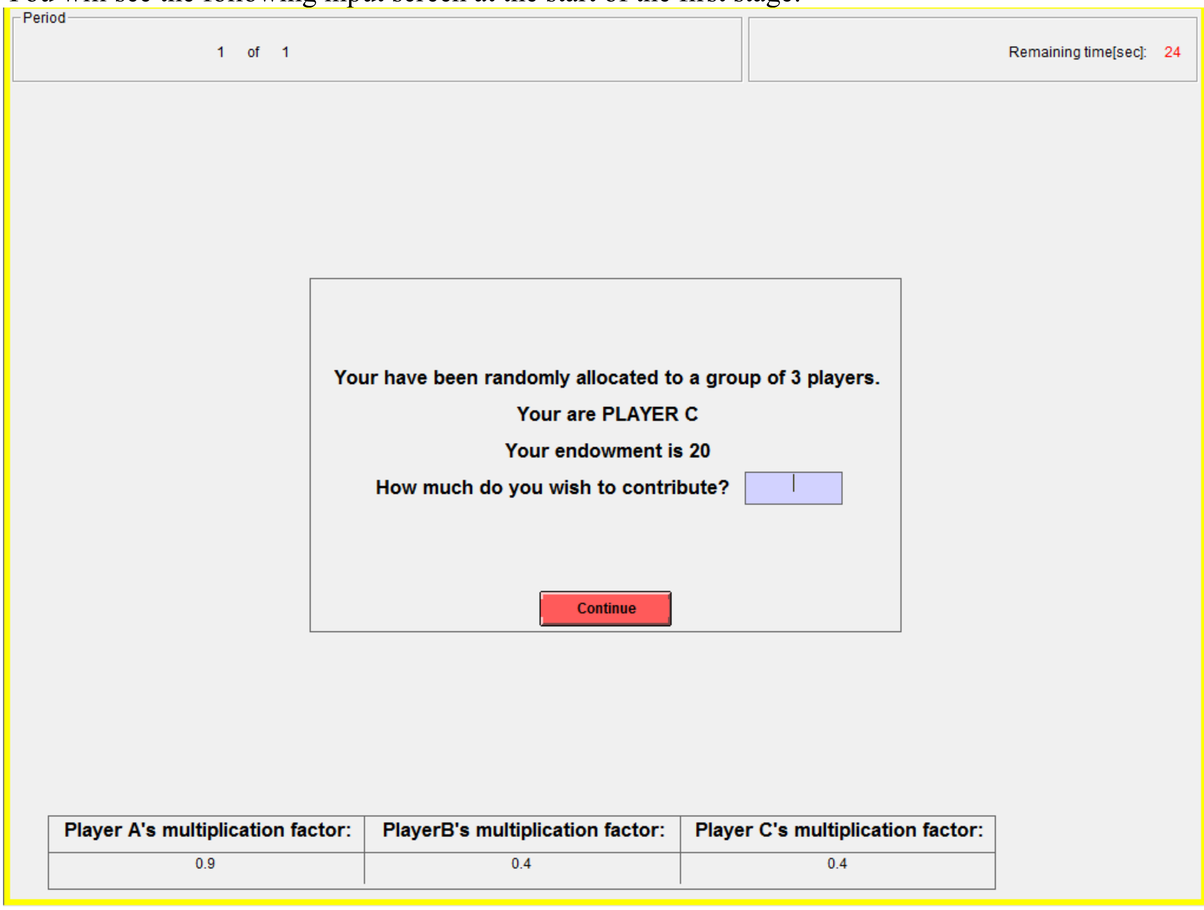
Overview of Part 1
This part of the experiment consists of 10 periods in total. At the start of each part, you will need to answer several control questions. These questions are designed to make sure that you understand the instructions. Then you need to answer three calculation questions to earn the 20 starting tokens for each of the 10 periods.

Each of the 10 periods consists of one stage. In the first stage, you will decide how many tokens you would like to contribute to a group project. First-stage earnings will then be generated, following a process described in more detail below. A more detailed description about Part 1 is provided below.

Detailed information about Part 1
In each period, you need to decide how many of your 20 starting tokens you want to contribute to a group project, described below, and how many to keep for yourself. You may contribute as many of your tokens to the group project as you wish. Each other group member will also choose how many of his or her tokens to contribute to the group project, and how many to keep. You and your group members will make your decisions simultaneously, and nobody will be informed about the decision of the other group members before everyone has made a decision. Your first-stage earnings will be generated following a process described in detail below.

Stage 1

You will see the following input screen at the start of the first stage:



You will indicate how many tokens you want to contribute to the group project by typing a number between 0 (i.e., contributing zero tokens) and 20 (i.e., contributing all 20 tokens) in the input field. In the top right corner of the screen, you can see how many more seconds remain for you to decide on the distribution of your tokens. Please make your decision before the time is up.
Your earnings in tokens, in Stage 1, are calculated as follows.

1) The starting tokens you have kept for yourself.

2) Your income from the group project.

Your income from the group project will be calculated as follows:

[multiplication factor] × [total tokens contributed to the group project]

"Total tokens contributed to the group project" is calculated simply as the sum of all three members' contributions to the group project. ***The multiplication factor used in the payout formula for the group project will be different for different members of your group. In each group, the earnings of one group member will be calculated using a multiplication factor of 0.9, and the earnings of the other two group members will be calculated using a multiplication factor of 0.4. Participants will be randomly assigned to one of these values by the computer (i.e., 0.9 or 0.4, with 1/3 chance of 0.9 and 2/3 chance of 0.4).*** The multiplication factor each person is assigned will be the same for all periods.

The multiplication factors for all members in your group will be displayed at the bottom of the screen throughout all periods. In each period you have 30 seconds to view the earnings screen. If you are finished before the time is up, please press the continue button. Once the viewing time has elapsed, the first stage is over, and the second stage commences.

Your earnings in tokens from the first stage of a period are therefore:

$$20 - \text{your contribution} + \underbrace{(\text{your multiplication factor}) \times (\text{total tokens contributed})}_{\boxed{\text{Your income from the group project}}}$$

Example

The multiplication factor for group member 1 is 0.4, for group member 2 is 0.4, and for group member 3 is 0.9. Suppose that group member 1 contributes X tokens to the group project, group member 2 contributes Y tokens to the group project, and group member 3 contributes Z tokens to the group project.

The earnings in tokens of each of the participants are given by:

20 – own tokens contributed + (multiplication factor) × (total tokens contributed)

For group member 1 this equals: $20 - X + [0.4 \times (X + Y + Z)]$ tokens.

For group member 2 this equals: $20 - Y + [0.4 \times (X + Y + Z)]$ tokens.

For group member 3 this equals: $20 - Z + [0.9 \times (X + Y + Z)]$ tokens.

The calculation of your earnings in each period has now been explained in full.

Please raise your hand if you have any questions about Part 1 of the experiment.

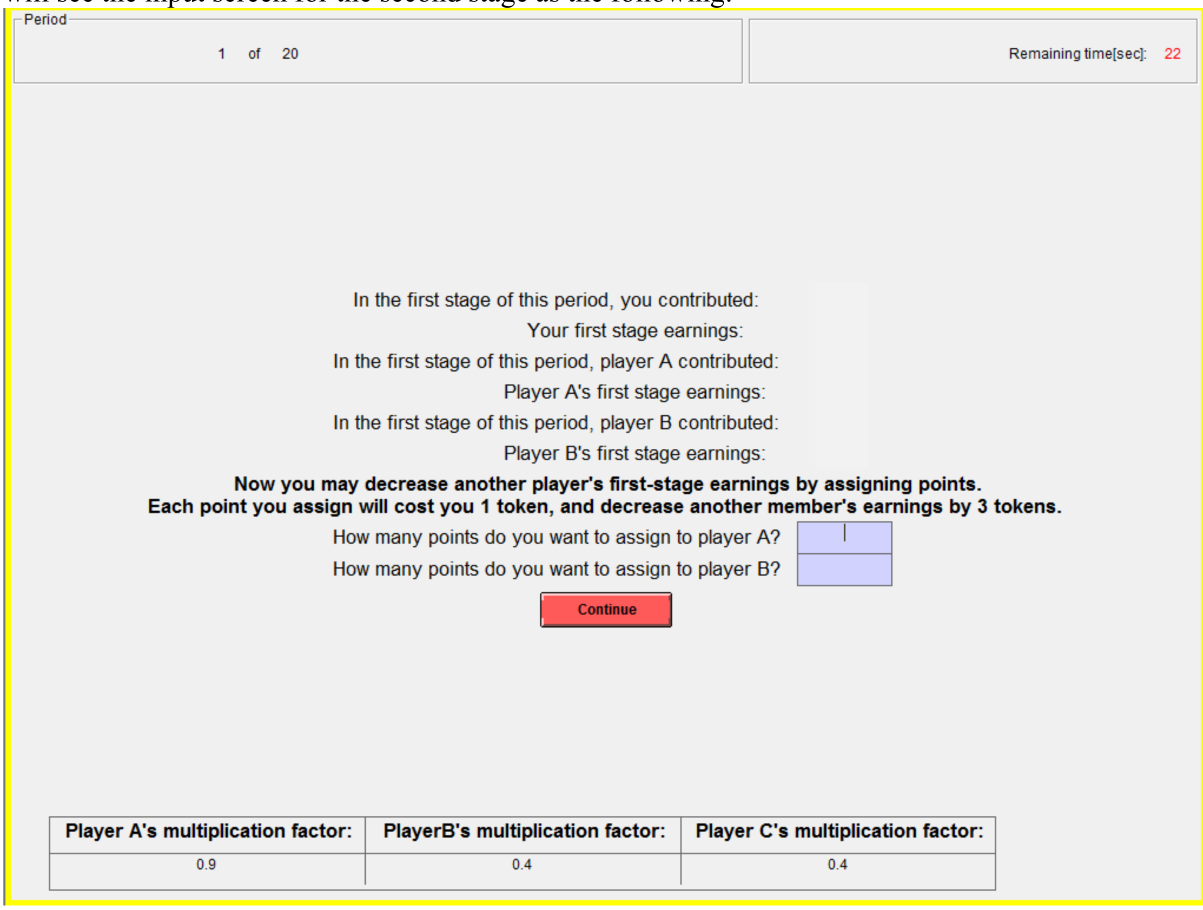
Before starting the experiment, you will need to type in your answer for several questions designed to make sure that you understand the instructions. Then, you can proceed to the calculation questions, through which you earn the 20 starting tokens for each of the 10 periods in Part 1.

Part 2
Part 1 of the experiment will now be repeated, with one difference. You will be assigned into a new group of three players (including yourself and two others), containing neither member of your group from Part 1. As before, you will need to earn the 20 starting tokens for each of the 10 periods by completing three calculation questions. At the start of each period you will have to decide how many of your initial 20 tokens you want to contribute to the group project. **The difference is that now, there will be a second stage.**

In the second stage, you will be informed about the contributions to the project made by each of the other two group members. You can then decide whether to *decrease* the amount of tokens that each of them earned in the first stage. If you decide to *decrease* these earnings, you will be able to decide the amount of *decrease* for each person separately, by assigning points to each group member. **Each point you assign towards *decreasing* others' earnings will cost you 1 token**. Your earnings, too, can be *decreased* by other group members in the same manner. A more detailed description about Stage 2 is provided below.

<u>Stage 2</u>

At the beginning of the second stage, everyone in the group will see how many tokens each of the other group members contributed to the project, as well as everyone's earnings from the first stage. In this stage you are given the opportunity to *decrease* the earnings of each other group member by purchasing points using tokens that you earned in the first stage. **Each point you assign towards *decreasing* another's earnings will cost you 1 token. For each point that you pay towards the *decrease* of another group member, his or her earnings will be *decreased* by 3 tokens.** The other group members can also *decrease* your earnings if they wish to, through the same mechanism. You will see the input screen for the second stage as the following:



All decisions are made simultaneously. That is, nobody will be informed about the decisions of the other group members before everyone has made a decision. If you do not wish to change the first-stage earnings of another group member, then you should assign 0 points to him or her. You are not allowed to assign more than 10 tokens to a group member.

After everyone has made a decision, your total earnings in tokens for the period (including the results of the first and second stages) will be displayed. You will not see how many of his or her tokens each individual group member spent on points to *decrease* your first-stage earnings; you will only be informed of the total amount of points assigned to you by the two other group members combined. *Please note that the final earnings of no group member will be allowed to fall below zero.*

Your earnings in tokens at the second stage of a period are therefore: (Your first stage earnings - *3 ×* points allocated to you) - points you allocated* * If the earnings in tokens at the second stage are negative, then they are replaced with a zero.

Example
Remember that for each point that you pay towards the *decrease* of another group member, his or her earnings will be *decreased* by 3 tokens. Suppose that you are group member 1 and that after the first stage you have earnings of 30 tokens. In the second stage, you decide to assign A points to group member 2 (this *decreases* group member 2's earnings by 3 × A tokens) and B points to group member 3 (this *decreases* group member 3's earnings by 3 × B tokens). After all group members have made their decisions, you learn that the others assigned you a total of C points. In this case, your total earnings in tokens in this period are equal to: *(30 – 3 ×* C*) – (A + B) tokens*.

The calculation of your earnings in each period has now been explained in full.
Please raise your hand if you have any questions about Part 2 of the experiment.

After you finish the 10 periods of Part 2, a short demographic survey will complete this experimental session.

A.2 Theoretical Predictions With *Homo Oeconomicus* Agents

If a player cares only about his own material benefit, then his utility can be represented by the payoff function. [17]

$$\pi_i = e_i - c_i + a_i \times \sum_i c_i \pm 3 \sum_{i \neq j} m_{ji} - \sum_{i \neq j} m_{ij}$$

If we take the first-order derivative of π_i over c_i, for different a_is, we have that

$$\frac{\partial \pi}{\partial c_i} = -1 + a_i \begin{cases} < 0, \text{ if } a_i < 1 \\ \geq 0, \text{ if } a_i \geq 1 \end{cases}$$

Two things can be inferred from the above equations: 1) As long as $a_i < 1$, different a_i will not change the equilibrium prediction, which is zero contribution from all players; 2) For any a_i, adding an incentive stage does not alter the equilibrium prediction. That is, to contribute 0 when $a_i < 1$ and to contribute all when $a_i > 1$ is individually rational regardless of incentives, although contributing all is of the greatest benefit to the whole group in both cases. The group objective is in conflict with personal interest if $a_i < 1$, but is in line with personal interest if $a_i \geq 1$. The equilibrium also predicts that no punishment or reward will be used since no player expects another player to use the punishment or reward incentive at a cost. Yet previous experiments have repeatedly demonstrated that, for identical $a_i < 1$, players will use punishment and reward (Fehr and Gachter, 2000). Two previous experiments by Reuben and colleagues (Reuben and Riedl, 2009, 2013) find that for heterogeneous $a_i < 1$, players will still use punishment when it is available, but the punishment decision is less strategic and less effective in changing other members' contributions.

[17] I acknowledge that, in general, *homo oeconomicus* agents maximise utility and not necessarily monetary payoff. Thus, a Fehr & Schmidt agent may well be a(n) *homo oeconomicus*. However, the argument that utility (rather than monetary payoff) motivates behaviour can be reduced *ad absurdum* to justify any strategic response as being driven by unseen sources of utility. I follow Sutter et al. (2010) in treating Fehr & Schmidt's agents as being distinct from self-oriented *homo oeconomicus* agents, in that they receive utility from sources beyond their own monetary payoffs, such as from (in)equality.

A.3 Theoretical Predictions with Fehr & Schmidt (1999)'s Agents

- *Prediction 1: Groups with a relatively privileged player will deliver more nonzero contributions, compared to standard groups.*[18]

  Given the parameter distributions in Fehr and Schmidt (1999), the probability of observing nonzero contributions can be calculated. For $a = 0.4$ and $n = 3, P(k = 0) = P(\beta_i \geq 0.6)^3 = 0.4^3 = 0.064$. This is saying that on average, almost all individuals fully defect to zero contributions. With heterogeneous MPCRs $a_h = 0.9$ and $a_l = 0.4$, $P(k = 0) = P(\beta_i \geq 0.6)^2 \times P(\beta_j \geq 0.1) = 0.4^2 \times 0.7 = 0.112$. The probability of observing nonzero contributions is higher. This leads us to the first prediction. Given the distribution of β in Fehr and Schmidt (1999), the prediction will not change for $a_h \in (0.75, 1)$. For example, the theory given Fehr and Schmidt's (1999) agents predicts that some alternative combinations of MPCRs (e.g., 0.8 and 0.5, or 0.8 and 0.4) will leave the main results unchanged. For a_h outside of this range, the situation is more complex. An exploration of the predictions and realised outcomes in such settings is left to future research.

- *Prediction 2: There is a continuum of positive contribution bundles forming Nash Equilibria. If a high player has $\beta_i > 1/15$, then he always contributes full endowment, while all low players contribute a positive amount c. Otherwise, he will contribute the same amount as everyone else does.*

  It is straightforward to see that even with Fehr and Schmidt (1999)'s agents, zero contribution is an equilibrium. Next, we want to verify whether positive contributions can be sustained in such groups.

  Assume all players choose to contribute c, we verify whether an individual player of either low or high type has an incentive to deviate from c.

  - If a high type player chooses to contribute c, then his utility is as below

$$U_h(c) = e - c + a_h nc - \frac{\beta_i}{n-1}(a_h - a_l)nc(n-1)$$

[18] All predictions in this paper are specific to the parameter values of this experiment. Predictions may change depending on parameter values.

If a high type player deviates and chooses to contribute a different amount c_h, then

$$U_h(c_h) = e - c + a_h[(n-1)c + c_h] - \frac{\beta_i}{n-1}[c - c_h + (a_h - a_l)((n-1)c + c_h)(n-1)$$

For a player to deviate, we need to have

$$U_h(c_h) > U_h(c)$$

Simplify the above conditions,

$$U_h(c_h) - U_h(c) = (c_h - c)[a_h - 1 + \beta_i(1 + a_h - a_l)]$$

Given the parameter values in this experiment, we get

$$U_h(c_h) - U_h(c) = (c_h - c)[1.5\beta_i - 0.1]$$

Then the following two predictions can be generated

1. If $c_h > c$, then for any $\beta_i > 1/15$ a high type player can gain more than he loses from deviating and contributing more. A high type player's utility is maximised if he contributes all his endowment.
2. If $c_h < c$, then for any $\beta_i < 1/15$ a high type player will have pure loss from deviating and contributing less than other players.

– If a low player chooses to contribute c, then his utility is

$$U_l(c) = \pi - c + a_l nc - \frac{\alpha_i}{n-1}(a_h - a_l)nc$$

If a low chooses to contribute c_l, then we need to discuss separately depending on whether the deviation is positive or negative.

1. If $\beta_i < 1/15$ for a high player, then he contribute the same amount as everyone else does.
   (a) If $c_d > c$, then

$$\begin{aligned} U_l(c_l) = e - c_l + a_l[(n-1)c + c_l] - \frac{\alpha_i}{n-1}(c_l - c)(n-2) \\ - \frac{\alpha_i}{n-1}\{c_l - c + (a_h - a_l)[(n-1)c + c_l]\} \end{aligned}$$

Similarly, for a player to deviate, we need to have

$$U_l(c_l) > U_h(c)$$

Simplify the above conditions,

$$U_l(c_l) - U_l(c) = (c_l - c)(a_l - 1) - \alpha_i(c_l - c) - \frac{\alpha_i}{n-1}(a_h - a_l)(c_l - c)$$

which is always smaller than zero, despite the values of parameters. Hence, low players do not have an incentive to deviate positively.

(b) If $c_d < c$, then

$$U_l(c_l) = e - c_l + a_l[(n-1)c + c_l] - \frac{\beta_i}{n-1}(c - c_l)(n-2) - \frac{\alpha_i}{n-1}\{c_l - c + (a_h - a_l)[(n-1)c + c_l]\}$$

Similarly, for a player to deviate, we need to have

$$U_l(c_l) > U_h(c)$$

Simplify the above conditions and plug in parameter values we have

$$U_l(c_l) - U_l(c) = -0.6 + \frac{\beta_i}{2} - \frac{\alpha_i}{2} \times 1.5$$

This is saying that low players have an incentive to deviate and contribute less than others if $\beta_i - 1.5\alpha_i - 1.2 > 0$. However, given the distribution of αs and βs, this condition can never be satisfied.

Hence, low players do not have an incentive to deviate from c if everyone else is contributing c.

2. If a high player has $\beta_i > 1/15$, then he always contributes full endowment e, while all low players contribute a positive amount c.

(a) If $c_d > c$, then

$$U_l(c) = e - c + a_l[(n-1)c + e] - \frac{\alpha_i}{n-1}\{(c - e + (a_h - a_l)[(n-1)c + e]\}$$

$$\begin{aligned}U_l(c_l) &= e - c_l + a_l[(n-2)c + c_l + e] - \frac{\alpha_i}{n-1}(c_l - c)(n-2) \\ &- \frac{\alpha_i}{n-1}\{c_l - e + (a_h - a_l)[(n-2)c + c_l + e] + a_h c - a_l c_l\}\end{aligned}$$

Simplify the above conditions,

$$U_l(c_l) - U_l(c) = (c_l - c)(a_l - 1) - \alpha_i(c_l - c) - \frac{\alpha_i}{n-1}(a_l c - a_l c_l)$$

This condition can never be satisfied.

(b) If $c_d < c$, then

$$U_l(c) = e - c + a_l[(n-1)c + e] - \frac{\alpha_i}{n-1}\{(c - e + (a_h - a_l)[(n-1)c + e]\}$$

$$\begin{aligned}U_l(c_l) &= e - c_l + a_l[(n-2)c + c_l + e] - \frac{\beta_i}{n-1}(c - c_l)(n-2) \\ &- \frac{\alpha_i}{n-1}\{c_l - e + (a_h - a_l)[(n-2)c + c_l + e] + a_h c - a_l c_l\}\end{aligned}$$

Given parameter values, the above condition is equivalent to saying $\beta_i - 0.6\alpha_i > 1.2$, which cannot be satisfied for any $\alpha \in [0, 1]$ and $\beta \in [0, 1]$. Hence, low players do not have an incentive to deviate from c if a high player is contributing e and all other low type players are contributing c.

- *Prediction 3: Reward is more credible than punishment in privileged groups.*

For punishment strategies, assume every group member chooses the same punishment strategy, assigning k to a deviator. We need to verify whether a player can gain more by deviating from this punishment strategy. As for reward strategy, following the assumptions in Sutter et al. (2010), whoever rewards will reward every group member. In the same logic, a deviating action needs to bring player i greater utility gain than utility loss. In addition to direct changes in monetary payoff, a player may encounter six kinds of inequalities, as listed below.

  - a. Disadvantageous inequality toward other contributors who never punish/reward
  - b. Disadvantageous inequality toward the deviator (punishment condition only)

- c. Disadvantageous inequality toward other conditional collaborators
- d. Advantageous inequality toward other contributors who never punish/reward
- e. Advantageous inequality toward the deviator (punishment condition only)
- f. Advantageous inequality toward other conditional collaborators

Depending on the type of players, in total there can be four scenarios in punishment condition and three scenarios in reward condition.

- Scenario P1: player i is high type and everyone else is low type.
- Scenario P2: player i is low type, the defector is high type, and everyone else is low type.
- Scenario P3: player i is low type, the defector is also low type, one of the non-punishing players is high type
- Scenario P4: player i is low type, the defector is also low type, one of the punishing players is high type
- Scenario R1: player i is high type and everyone else is low type.
- Scenario R2: player i is low type, one of the non-rewarding players is high type
- Scenario R3: player i is low type, one of the rewarding players is high type

Next, the utility changes due to inequalities are listed in Table A.1 for each scenario. To make punishment or reward credible strategies, the following condition needs to be satisfied.

$$U_{ScenarioX,Y} \geqslant U_{ScenarioX,N}$$

Given that $n = 3$, $a_h = 0.9$, $a_l = 0.4$, $L/k = 3$, the predictions for this experiment are solved below.

1. Scenario P1:
   - For $\forall n'$, $-2 - \beta_i \geqslant 0$. This inequality can never be satisfied.
2. Scenario P2 & P3:

- For $n' = 1$, $\alpha_i \geqslant 2$. This inequality can never be satisfied.
- For $n' = 2$, $2\alpha_i - \beta_i \geqslant 2$
- For $n' = 3$, there is no solution. With a group of three members including a deviator, we cannot have three enforcers at the same time.

3. Scenario P4:
   - For $n' = 1$, $\beta_i \leqslant -2$. This inequality can never be satisfied.
   - For $n' = 2$, $\alpha_i \geqslant 2$. This inequality can never be satisfied.
   - For $n' = 3$, there is no solution. With a group of three members including a deviator, we cannot have three enforcers at the same time.
4. Scenario R1:
   - For $\forall n'$, $\beta_i \geqslant \frac{2}{5}$
5. Scenario R2:
   - For $n' = 1$, $\alpha_i \leqslant \frac{-2}{5}$. This inequality can never be satisfied.
   - For $n' = 2$, $\beta_i - \alpha_i \geqslant \frac{4}{5}$
   - For $n' = 3$, $\beta_i \geqslant \frac{2}{5}$
6. Scenario R3:
   - For $n' = 1$, $-2 - \frac{15}{2}\alpha_i \geqslant 0$. This inequality can never be satisfied.
   - For $n' = 2$, $-2 - 5\alpha_i \geqslant 0$. This inequality can never be satisfied.
   - For $n' = 3$, $\beta_i - \alpha_i \geqslant \frac{4}{5}$

In each scenario, the binding conditions for punishment are more stringent than those for reward. Hence, reward is a more credible incentivizing strategy than punishment in privileged groups.

Furthermore, if all members stick to a reward strategy, then every member is more willing to collaborate and contribute in the first stage. On the contrary, members are less likely to be motivated by the punishment opportunities given the difficulty in sustaining punishment.

- *Prediction 3.1: Reward opportunities are more effective than punishment opportunities in privileged groups.*

Table A.1: Public Goods Game With Sanctioning Opportunities

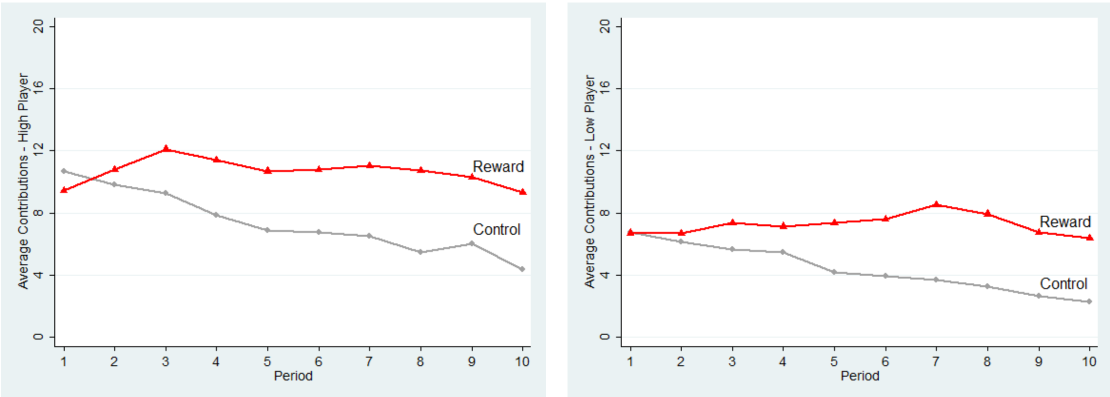
Situation	Payoff change	a	b	c	d	e	f
Panel A: Punishment							
Panel A1: Standard public goods game (Fehr and Schmidt, 1999; Sutter et al., 2010)							
Y	$-k$	$-\frac{\alpha_i}{n-1}(n-n'-1)k$	$-\frac{\alpha_i}{n-1}(c-c_d-n'L+k)$	0	0	0	0
N	0	0	$-\frac{\alpha_i}{n-1}(c-c_d-(n'-1)L)$	0	0	0	$-\frac{\beta_i}{n-1}(n'-1)k$
Panel A2: public goods game with a high type player							
P1, Y	$-k$	0	0	0	$-\frac{\beta_i}{n-1}(c-c_h+(a_h-a_l)c_{sum}-k)(n-n'-1)$	$-\frac{-\beta_i}{n-1}(c_d-c_h+(a_h-a_l)c_{sum}+n'L-k)$	$-\frac{\beta_i}{n-1}(c-c_h+(a_h-a_l)c_{sum})(n'-1)$
P1, N	0	0	0	0	$-\frac{\beta_i}{n-1}(c-c_h+(a_h-a_l)c_{sum})(n-n'-1)$	$-\frac{\beta_i}{n-1}(c_d-c_h+(a_h-a_l)c_{sum}+(n'-1)L)$	$-\frac{\beta_i}{n-1}(c-c_h+(a_h-a_l)c_{sum}+k)(n'-1)$
P2, Y	$-k$	$-\frac{\alpha_i}{n-1}(n-n'-1)k$	$-\frac{\alpha_i}{n-1}(c-c_d-n'L+k+(a_h-a_l)c_{sum})$	0	0	0	0
P2, N	0	0	$-\frac{\alpha_i}{n-1}(c-c_d-(n'-1)L+(a_h-a_l)c_{sum})$	0	0	0	$-\frac{\beta_i}{n-1}(n'-1)k$
P3, Y	$-k$	$-\frac{\alpha_i}{n-1}(n-n'-2)k-\frac{\alpha_i}{n-1}(c-c_h+(a_h-a_l)c_{sum}+k)$	$-\frac{\alpha_i}{n-1}(c-c_d-n'L+k)$	0	0	0	0
P3, N	0	$-\frac{\alpha_i}{n-1}(c-c_h+(a_h-a_l)c_{sum})$	$-\frac{\alpha_i}{n-1}(c-c_d-(n'-1)L)$	0	0	0	$-\frac{\beta_i}{n-1}(n'-1)k$
P4, Y	$-k$	$-\frac{\alpha_i}{n-1}(n-n'-1)k$	$-\frac{\alpha_i}{n-1}(c-c_d-n'L+k)$	$-\frac{\alpha_i}{n-1}(c-c_h+(a_h-a_l)c_{sum})$	0	0	0
P4, N	0	0	$-\frac{\alpha_i}{n-1}(c-c_d-(n'-1)L)$	$-\frac{\alpha_i}{n-1}(c-c_h+(a_h-a_l)c_{sum}-k)$	0	0	$-\frac{\beta_i}{n-1}(n'-2)k$
Panel B: Reward							
Panel B1: Standard public goods game (Sutter et al., 2010)							
Y	$-(n-1)k+(n'-1)L$	$-\frac{\alpha_i}{n-1}(n-n')[L-(n-1)k]$	0	0	0	0	0
N	$0+(n'-1)L$	0	0	0	0	0	$-\frac{\beta_i}{n-1}[L+(n-1)k](n'-1)$
Panel B2: public goods game with a high type player							
R1, Y	$-(n-1)k+(n'-1)L$	0	0	0	$-\frac{\beta_i}{n-1}(n-n')[c-c_h+(a_h+a_l)c_{sum}-L-(n-1)k]$	0	$-\frac{\beta_i}{n-1}(n'-1)(c-c_h+0.5c_{sum})$
R1, N	$(n'-1)L$	0	0	0	$-\frac{\beta_i}{n-1}(c-c_h+(a_h-a_l)c_{sum})(n-n')$	0	$-\frac{\beta_i}{n-1}(c-c_h+(a_h-a_l)c_{sum}+L+(n-1)k)(n'-1)$
R2, Y	$-(n-1)k+(n'-1)L$	$-\frac{\alpha_i}{n-1}(n-n'-1)[L+(n-1)k]-\frac{\alpha_i}{n-1}[c-c_h+L+(n-1)k+(a_h-a_l)x_{sum}]$	0	0	0	0	0
R2, N	$(n'-1)L$	$-\frac{\alpha_i}{n-1}[c-c_h+(a_h-a_l)c_{sum}]$	0	0	0	0	$-\frac{\beta_i}{n-1}(n'-1)[L+(n-1)k]$
R3, Y	$-(n-1)k+(n'-1)L$	$-\frac{\alpha_i}{n-1}(n-n')[L+(n-1)k]$	0	$-\frac{\alpha_i}{n-1}(c-c_h+(a_h-a_l)c_{sum})$	0	0	0
R3, N	$(n'-1)L$	0	0	$-\frac{\alpha_i}{n-1}[c-c_h-L-(n-1)k+(a_h-a_l)c_{sum}]$	0	0	$-\frac{\beta_i}{n-1}(n'-2)[L+(n-1)k]$

Notes: Y indicates the scenario where the conditional collaborator chooses to punish or reward, N indicates the scenario where the conditional collaborator deviate from his or her incentivizing strategy.

A.4 Other Results

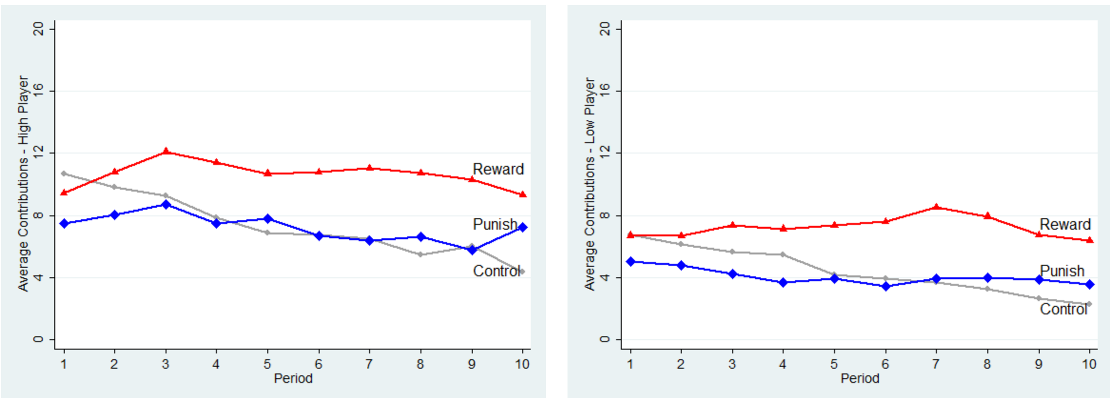
Table A.2: Experimental Design

Group Type	MPCR	Punish/Reward		Sessions
		Period 1-10	Period 11-20	
1	0.9/0.4	Punish	No(Control)	2
2	0.9/0.4	No(Control)	Punish	2
3	0.9/0.4	Reward	No(Control)	2
4	0.9/0.4	No(Control)	Reward	2
5	0.4	No(Baseline)	No(Baseline)	2



(a) High Player (b) Low Player

Figure A.1: Average Individual Contributions - Reward vs. Control



(a) High Player (b) Low Player

Figure A.2: Average Individual Contributions - Reward vs. Punish

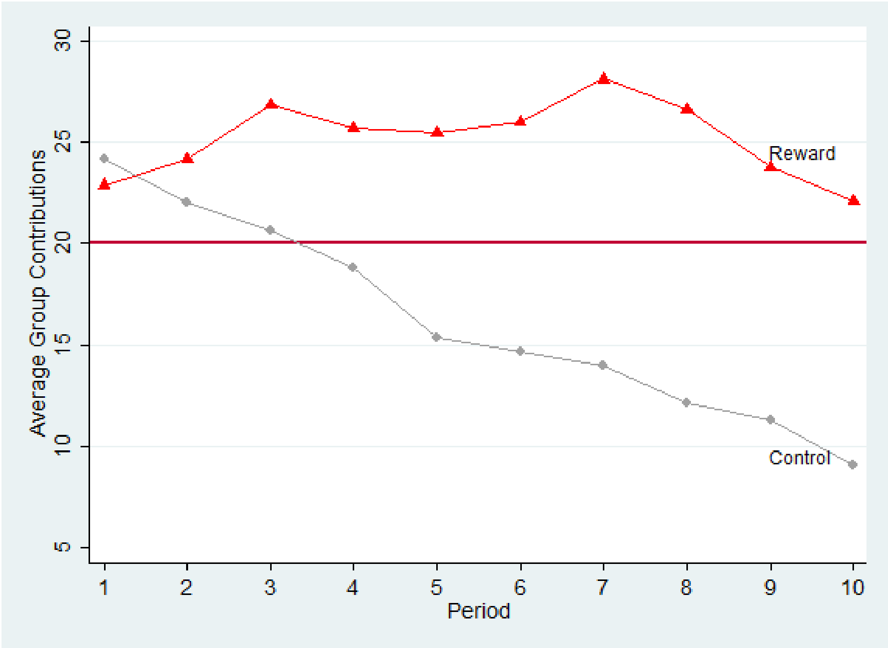


Figure A.3: Average Group Contributions - Reward vs. Control

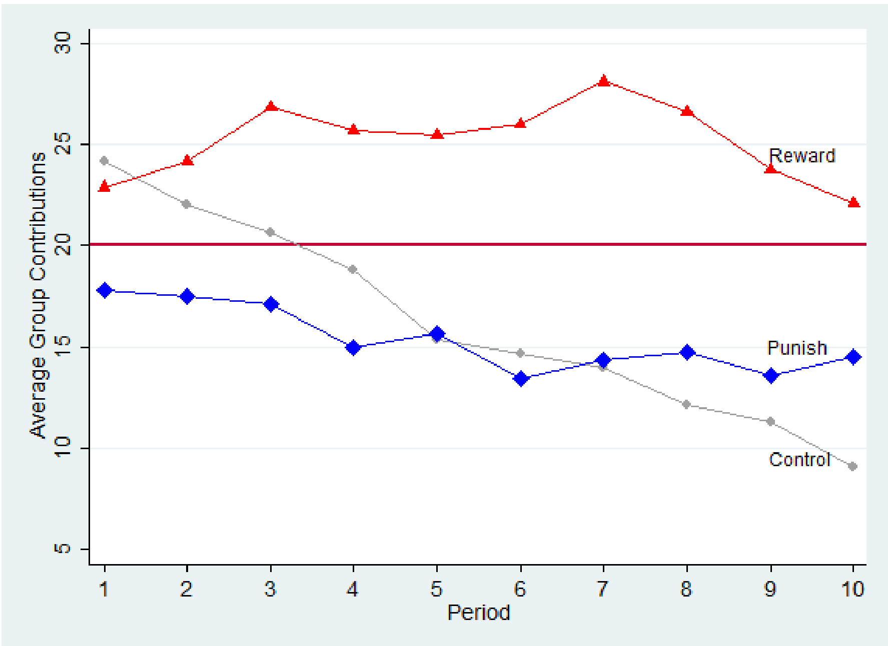


Figure A.4: Average Group Contributions - Reward vs. Punish

Table A.3: Punishment/Reward Depending on Contribution Differences

	Punishment		Reward	
Punisher/Rewarder (i) Type	High (1)	Low (2)	High (3)	Low (4)
Dependent Variable: Points assigned by i to j				
Contributions of j	0.106	-0.035	0.206***	0.239***
	(0.065)	(0.047)	(0.073)	(0.082)
Positive deviation of c_j from c_i	2.936***	2.640***	-2.486*	-3.056***
	(1.115)	(0.475)	(1.356)	(0.790)
Negative deviation of c_j from c_i	0.545	0.898*	0.654	-0.079
	(0.577)	(0.530)	(0.877)	(0.848)
Positive deviation of c_j from c_k	-0.555	-0.321	0.481	1.297*
	(0.706)	(0.571)	(1.062)	(0.687)
Negative deviation of c_j from c_k	1.839***	1.205**	-0.768	0.561
	(0.526)	(0.543)	(0.838)	(0.685)
j is a high player		0.394		1.402*
		(0.845)		(0.743)
Period	-0.052	-0.058	-0.226***	-0.198***
	(0.060)	(0.063)	(0.048)	(0.064)
j is a high player $\times$ Period		-0.022		-0.126
		(0.093)		(0.098)
Constant	-2.225***	-2.827***	-0.662	-0.573
	(0.846)	(0.785)	(1.096)	(1.368)
Observations	520	1,040	540	1,080
Loglikelihood	-708	-1237	-791.9	-1815
Fstat	5.596	7.204	4.984	6.884

Notes: Tobit regression with a lower limit of 0 and an upper limit of 10 on the dependent variable. Standard errors are in parentheses, clustering at the group level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.4: Punishment/Reward Depending on Differences in First-Stage Earnings

	Punishment		Reward	
Punisher/Rewarder (i) Type	High (1)	Low (2)	High (3)	Low (4)
Dependent Variable: Points assigned by i to j				
Positive deviation of π_j from π_i	1.320*	0.807	2.250**	1.932**
	(0.749)	(0.493)	(1.067)	(0.977)
Negative deviation of π_j from π_i	2.114***	2.684***	1.153	-2.444**
	(0.710)	(0.617)	(1.203)	(0.996)
Positive deviation of π_j from π_k	1.494**	1.584***	-0.559	-0.265
	(0.607)	(0.536)	(0.775)	(0.639)
Negative deviation of π_j from π_k	-0.340	0.471	0.015	1.464**
	(0.723)	(0.610)	(0.949)	(0.711)
Contribution of j	0.099	-0.028	0.178*	0.234***
	(0.089)	(0.048)	(0.092)	(0.076)
Period	-0.049	-0.071*	-0.211***	-0.266***
	(0.063)	(0.040)	(0.060)	(0.051)
j is a high player		-0.212		0.443
		(0.594)		(0.679)
Constant	-3.061***	-2.692***	-1.763	-1.963
	(0.897)	(0.802)	(1.562)	(1.399)
Observations	520	1,040	540	1,080
loglikelihood	-716.7	-1237	-805.8	-1801
Fstat	6.491	6.473	4.354	10.63

Notes: Tobit regression with a lower limit of 0 and an upper limit of 10 on the dependent variable. π_i, π_j, and π_k denote the first-stage earnings of the punisher/rewarder, the person being punished/rewarded, and the third player, respectively. Robust standard errors are in parentheses, clustering at the group level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.5: Effect of A High Player's Contributions on Low Players' Contributions

	Model 1				Model 2				Model 3			
Variables	Reward	Punish	Control	Baseline	Reward	Punish	Control	Baseline	Reward	Punish	Control	Baseline
	Dependent Variable : Contributions of Low Players in Period t											
$C_{High,t-1}$	0.435***	0.364***	0.402***	0.144*	0.127**	0.100*	0.086***	0.081	0.197**	0.072*	0.092**	0.086
	(0.094)	(0.059)	(0.063)	(0.069)	(0.035)	(0.046)	(0.021)	(0.040)	(0.063)	(0.035)	(0.031)	(0.045)
$C_{Own,t-1}$					0.757***	0.547***	0.606***	0.397***	0.713***	0.439***	0.537***	0.498***
					(0.061)	(0.066)	(0.045)	(0.073)	(0.041)	(0.084)	(0.050)	(0.067)
$C_{Other,t-1}$					0.036	0.217***	0.179***	0.150**	0.076	0.145	0.178***	0.170*
					(0.041)	(0.058)	(0.042)	(0.048)	(0.053)	(0.095)	(0.050)	(0.062)
$C_{High,t-2}$									-0.083	-0.023	-0.023	-0.011
									(0.057)	(0.047)	(0.028)	(0.041)
$C_{Own,t-2}$									0.069	0.104	0.146**	-0.001
									(0.050)	(0.064)	(0.055)	(0.051)
$C_{Other,t-2}$									-0.050	0.182*	-0.075	-0.022
									(0.043)	(0.069)	(0.046)	(0.048)
Period	0.063	0.005	-0.200	-0.253**	-0.094	0.056	-0.038	-0.018	-0.128	0.120	-0.059	-0.008
	(0.171)	(0.077)	(0.113)	(0.085)	(0.071)	(0.046)	(0.055)	(0.065)	(0.102)	(0.062)	(0.067)	(0.048)
Last Period	-1.085	0.167	-0.393	0.175	-0.018	-0.305	0.061	-0.122	0.100	-0.604	0.067	-0.096
	(0.592)	(0.553)	(0.608)	(0.322)	(0.758)	(0.447)	(0.364)	(0.171)	(0.808)	(0.465)	(0.386)	(0.134)
Constant	2.348	1.245	2.310**	2.783***	0.671	-0.197	0.074	0.417	0.938	-0.684	0.309	0.226
	(1.590)	(0.648)	(0.853)	(0.643)	(0.526)	(0.349)	(0.348)	(0.490)	(0.765)	(0.466)	(0.473)	(0.377)
N(Players)	54	58	112	40	54	58	112	40	54	58	112	40
Observations	486	495	1,008	360	486	495	1,008	360	432	440	896	320
Adjusted R^2	0.197	0.255	0.300	0.102	0.671	0.555	0.663	0.332	0.687	0.593	0.678	0.399

Notes: Designation of high and low players are random in Baseline groups. Model 1 controls for high players' contributions in the previous period ($C_{High,t-1}$) plus session and last period effects. Model 2 additionally controls for own and the third group member's contributions ($C_{Own,t-1}$ and $C_{Other,t-1}$). In Model 3, a two-period lagged variable is added for each individual contribution level. Robust standard errors are clustered at one group level. *** p<0.001, ** p<0.01, * p<0.05.

Table A.6: Effect of Low players' Contributions on A High Player's Contributions

	Model 1				Model 2				Model 3			
Variables	Reward	Punish	Control	Baseline	Reward	Punish	Control	Baseline	Reward	Punish	Control	Baseline
	Dependent Variable : Contributions of High Players in Period t											
$C_{Low,t-1}$	0.451***	0.817***	0.690***	0.208*	0.091*	0.331***	0.204***	0.111	0.104**	0.329**	0.198***	0.047
	(0.103)	(0.106)	(0.072)	(0.085)	(0.033)	(0.080)	(0.028)	(0.079)	(0.036)	(0.107)	(0.043)	(0.075)
$C_{Own,t-1}$					0.814***	0.531***	0.595***	0.487*	0.788***	0.391**	0.502***	0.408*
					(0.056)	(0.072)	(0.045)	(0.176)	(0.074)	(0.116)	(0.076)	(0.175)
$C_{Other,t-1}$					0.091*	0.302***	0.204***	0.111	0.104**	0.287**	0.198***	0.047
					(0.033)	(0.069)	(0.028)	(0.079)	(0.036)	(0.097)	(0.043)	(0.075)
$C_{Low,t-2}$									0.049	0.194*	0.110	0.146
									(0.076)	(0.094)	(0.066)	(0.109)
$C_{Own,t-2}$									-0.030	-0.033	-0.003	-0.075
									(0.029)	(0.086)	(0.045)	(0.073)
$C_{Other,t-2}$									-0.030	-0.008	-0.003	-0.075
									(0.029)	(0.084)	(0.045)	(0.073)
Period	-0.235	-0.219	-0.245	-0.126	-0.230*	-0.124	0.049	0.082	-0.188	-0.188	0.091	0.075
	(0.212)	(0.164)	(0.128)	(0.169)	(0.102)	(0.122)	(0.078)	(0.135)	(0.115)	(0.164)	(0.099)	(0.149)
Last Period	-0.283	1.225	-0.280	0.390	-0.018	1.605	-1.121	-0.504	-0.100	1.788*	-1.123	-0.528
	(1.005)	(0.753)	(0.885)	(0.437)	(0.900)	(0.840)	(0.810)	(0.735)	(0.877)	(0.864)	(0.827)	(0.735)
Constant	8.925***	5.074***	5.296***	2.249**	2.037**	1.373	0.349	-0.020	1.745	1.616	-0.046	0.342
	(1.513)	(1.205)	(1.095)	(0.773)	(0.715)	(0.833)	(0.550)	(0.810)	(0.870)	(1.013)	(0.646)	(0.936)
N(Players)	27	29	56	20	27	29	56	20	27	29	56	20
Observations	486	495	1,008	360	486	495	1,008	360	432	440	896	320
Adjusted R^2	0.188	0.328	0.332	0.037	0.789	0.629	0.648	0.298	0.803	0.637	0.648	0.250

Notes: Model 1 controls for one low player's contributions in the previous period ($C_{Low,t-1}$) plus session and last period effects. Model 2 additionally controls for own and the third group member (the other low player)'s contributions ($C_{Own,t-1}$ and $C_{Other,t-1}$). In Model 3, a two-period lagged variable is added for each individual contributions. Robust standard errors are clustered at group level. *** p<0.001, ** p<0.01, * p<0.05.

Table A.7: Mean Comparison: Baseline vs. Control

	Control		Baseline		Difference	
Period	Mean	SD	Mean	SD	Diff.	SE
Panel A: High Player						
1	10.696	7.178	4.050	5.443	6.646***	1.765
2	9.804	7.445	3.750	5.902	6.054**	1.845
3	9.286	7.981	2.200	2.093	7.086***	1.814
4	7.875	7.749	2.450	4.594	5.425**	1.843
5	6.893	7.473	1.200	2.093	5.693**	1.701
6	6.768	7.685	1.950	3.859	4.818**	1.799
7	6.536	7.226	1.250	2.770	5.286**	1.664
8	5.482	7.076	1.150	3.150	4.332*	1.643
9	6.036	7.573	2.200	5.502	3.836*	1.849
10	4.393	6.959	1.550	4.286	2.843	1.662
Panel B: Low Player						
1	6.732	4.563	4.850	3.305	1.882*	0.787
2	6.116	5.384	3.250	2.734	2.866**	0.891
3	5.661	5.906	2.450	2.851	3.211**	0.973
4	5.473	6.067	1.750	2.016	3.723***	0.980
5	4.250	5.693	1.900	2.658	2.350*	0.936
6	3.946	5.175	1.350	1.916	2.596**	0.840
7	3.723	4.973	1.175	1.821	2.548**	0.806
8	3.348	4.811	1.050	1.679	2.298**	0.778
9	2.643	3.994	0.825	1.487	1.818**	0.648
10	2.348	3.489	0.750	1.373	1.598**	0.568

Notes: Each panel reports mean differences of individual contributions in the Punishment (Panel A) and Reward treatment (Panel B) vs. Baseline, respectively. Baseline treatment does not have a high player, so a player is randomly selected for mean comparison. The last two columns report the difference in means and the corresponding standard error of the difference, respectively. *** p<0.001, ** p<0.01, * p<0.05.

Table A.8: Effects of Punishment/Reward on Next Period Contributions

	Punishment		Reward	
	(1)	(2)	(3)	(4)
Dep. Var.: Individual Contributions in the Next Period				
Total punish Points received by i	0.173	-0.013	0.721***	0.116**
	(0.132)	(0.049)	(0.116)	(0.043)
i is a high player		2.040***		0.873*
		(0.465)		(0.466)
Contribution		0.385***		0.689***
		(0.079)		(0.075)
Total group contributions		0.181***		0.075***
		(0.028)		(0.026)
Period		0.013		-0.046**
		(0.026)		(0.021)
Constant	4.793***	-0.513**	6.253***	0.521
	(0.753)	(0.233)	(0.881)	(0.349)
Observations	756	756	729	729
Adjusted R^2	0.005	0.603	0.183	0.728

Notes: Robust standard errors are in parentheses, clustering at the group level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A.9: Treatment Effect of Sanctions on Group Contributions

	Treatment Group		Control Group		Difference	
Period	Mean	SD	Mean	SD	Diff.	SE
Panel A: Reward versus Control						
1	22.889	11.092	24.161	12.200	-1.272	2.686
2	24.148	11.951	22.036	15.704	2.112	3.114
3	26.815	13.511	20.607	18.118	6.208	3.553
4	25.704	15.334	18.821	18.443	6.882	3.845
5	25.444	17.037	15.393	17.281	10.052*	4.010
6	26.000	17.589	14.661	16.515	11.339**	4.041
7	28.111	18.037	13.982	15.784	14.129**	4.062
8	26.630	17.478	12.179	15.048	14.451***	3.919
9	23.778	18.575	11.321	13.988	12.456**	4.034
10	22.074	18.292	9.089	12.091	12.985**	3.873
Panel B: Punish versus Control						
1	17.828	10.286	24.161	12.200	-6.333*	2.511
2	17.517	12.229	22.036	15.704	-4.519	3.092
3	17.138	13.196	20.607	18.118	-3.469	3.445
4	14.966	13.899	18.821	18.443	-3.856	3.569
5	15.690	14.185	15.393	17.281	0.297	3.503
6	13.448	13.284	14.661	16.515	-1.212	3.310
7	14.345	13.494	13.982	15.784	0.363	3.275
8	14.724	14.370	12.179	15.048	2.546	3.341
9	13.586	12.935	11.321	13.988	2.265	3.044
10	14.552	14.292	9.089	12.091	5.462	3.107
Panel C: Reward versus Punish						
1	22.889	11.092	17.828	10.285	5.061	2.864
2	24.148	11.951	17.517	12.229	6.631*	3.232
3	26.815	13.511	17.138	13.196	9.677**	3.573
4	25.704	15.334	14.966	13.899	10.738**	3.920
5	25.444	17.037	15.690	14.185	9.755*	4.206
6	26.000	17.589	13.448	13.284	12.552**	4.189
7	28.111	18.037	14.345	13.494	13.766**	4.281
8	26.630	17.478	14.724	14.370	11.905**	4.294
9	23.778	18.575	13.586	12.935	10.192*	4.307
10	22.074	18.292	14.552	14.292	7.522	4.409

Notes: Each panel reports mean differences of group contributions in pairwise comparisons between Reward versus Control (Panel A), Punish versus Control (Panel B), and Reward versus Punish (Panel C). T-test assuming unequal variances between two samples is used to compare mean differences. The last two columns report the difference in means and the corresponding standard error of the difference, respectively. *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Table A.10: Treatment Effect of Sanctions on High Players' Contributions

	Treatment		Control		Difference	
Period	Mean	SD	Mean	SD	Diff.	SE
Panel A: Reward versus Control						
1	9.444	7.250	10.696	7.178	-1.252	1.693
2	10.778	6.980	9.804	7.445	0.974	1.672
3	12.111	7.122	9.286	7.981	2.825	1.737
4	11.407	7.012	7.875	7.748	3.532*	1.701
5	10.704	6.696	6.893	7.473	3.811*	1.630
6	10.778	7.239	6.768	7.685	4.010*	1.731
7	11.037	7.298	6.536	7.226	4.501*	1.705
8	10.741	7.578	5.482	7.076	5.259**	1.738
9	10.296	7.630	6.036	7.573	4.261*	1.783
10	9.333	7.691	4.393	6.959	4.940**	1.748
Panel B: Punish versus Control						
1	7.517	6.168	10.696	7.178	-3.179*	1.494
2	8.069	6.740	9.804	7.445	-1.735	1.599
3	8.690	7.117	9.286	7.981	-0.596	1.698
4	7.517	6.937	7.875	7.748	-0.358	1.653
5	7.793	7.257	6.893	7.473	0.900	1.677
6	6.690	6.783	6.768	7.685	-0.078	1.625
7	6.414	6.946	6.536	7.226	-0.122	1.611
8	6.655	7.301	5.482	7.076	1.173	1.653
9	5.793	6.488	6.036	7.573	-0.243	1.573
10	7.241	7.424	4.393	6.959	2.849	1.663
Panel B: Reward versus Punish						
1	9.444	7.250	7.517	6.168	1.927	1.805
2	10.778	6.980	8.069	6.740	2.709	1.836
3	12.111	7.122	8.690	7.117	3.421	1.904
4	11.407	7.012	7.517	6.937	3.890*	1.866
5	10.704	6.696	7.793	7.257	2.911	1.865
6	10.778	7.239	6.690	6.783	4.088*	1.878
7	11.037	7.298	6.414	6.946	4.623*	1.907
8	10.741	7.578	6.655	7.301	4.086*	1.991
9	10.296	7.630	5.793	6.488	4.503*	1.899
10	9.333	7.691	7.241	7.424	2.092	2.023

Notes: Each panel reports mean differences of individual contributions in pairwise comparisons between Reward versus Control (Panel A), Punish versus Control (Panel B), and Reward versus Punish (Panel C). T-test assuming unequal variances between two samples is used to compare mean differences. The last two columns report the difference in means and the corresponding standard error of the difference, respectively. *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Table A.11: Treatment Effect of Sanctions on Low Players' Contributions

	Treatment		Control		Difference	
Period	Mean	SD	Mean	SD	Diff.	SE
Panel A: Reward versus Control						
1	6.722	6.600	6.771	6.222	-0.048	1.078
2	6.685	6.273	6.147	6.230	0.538	1.042
3	7.352	6.334	5.651	6.687	1.700	1.074
4	7.148	6.781	5.477	6.801	1.671	1.129
5	7.370	7.106	4.193	6.280	3.178**	1.139
6	7.611	7.056	3.945	5.903	3.666***	1.114
7	8.537	7.348	3.688	5.614	4.849***	1.135
8	7.944	7.162	3.275	5.363	4.669***	1.102
9	6.741	7.346	2.651	4.822	4.089***	1.101
10	6.370	7.138	2.303	4.438	4.068***	1.060
Panel B: Punish versus Control						
1	5.055	4.916	6.771	6.222	-1.716	0.891
2	4.782	5.216	6.147	6.230	-1.365	0.922
3	4.255	4.915	5.651	6.687	-1.397	0.922
4	3.727	4.633	5.477	6.801	-1.750	0.903
5	3.964	4.468	4.193	6.280	-0.229	0.851
6	3.436	4.315	3.945	5.903	-0.509	0.811
7	3.964	5.221	3.688	5.614	0.276	0.886
8	4.018	4.980	3.275	5.363	0.743	0.845
9	3.891	4.593	2.651	4.822	1.240	0.773
10	3.600	5.202	2.303	4.438	1.297	0.820
Panel B: Reward versus Punish						
1	6.722	6.600	5.055	4.916	1.668	1.116
2	6.685	6.273	4.782	5.216	1.903	1.106
3	7.352	6.334	4.255	4.915	3.097**	1.087
4	7.148	6.781	3.727	4.633	3.421**	1.114
5	7.370	7.106	3.964	4.468	3.407**	1.139
6	7.611	7.056	3.436	4.315	4.175***	1.123
7	8.537	7.348	3.964	5.221	4.573***	1.223
8	7.944	7.162	4.018	4.980	3.926***	1.184
9	6.741	7.346	3.891	4.593	2.850*	1.176
10	6.370	7.138	3.600	5.202	2.770*	1.198

Notes: Each panel reports mean differences of individual contributions in pairwise comparisons between Reward versus Control (Panel A), Punish versus Control (Panel B), and Reward versus Punish (Panel C). T-test assuming unequal variances between two samples is used to compare mean differences. The last two columns report the difference in means and the corresponding standard error of the difference, respectively. *** $p<0.001$, ** $p<0.01$, * $p<0.05$.

Chapter 2

Rank and Human Capital: Lessons from a Randomized Controlled Trial[1]

2.1 Introduction

Social comparisons are ubiquitous in everyday life. From sports competitions and college rankings to competitive cooking shows to followers on Instagram and Twitter, the salience of ranking is pervasive throughout society. These comparisons not only affect individuals' private lives but are also encouraged in their professional ones. Allowing individuals to compare their performance can increase their productivity in both educational (Azmat and Iriberri, 2010; Tran and Zeckhauser, 2012; Katreniakova, 2014) and labour settings (Mas and Moretti, 2009; Blanes i Vidal and Nossol, 2011). However, while they may increase productivity in some instances, social comparisons are not without their drawbacks. If individuals discover that they have overestimated their competitors, they may choose to reduce the level of effort

[1]This chapter was jointly developed with Isabella Dobrescu, Gigi Foster, and Alberto Motta. All four of us were involved in all stages, from formulating the research questions, through designing the experiment, and to writing up the draft. I played a major role in analysing the data and writing up the draft. Special thanks goes to Denzil Fiebig, Valentyn Panchenko, Federico Masera and seminar and conference participants at several institutions for helpful discussions and suggestions. Great appreciation also goes to Chong Eng Tay for outstanding support with the administrative records. The work was approved by the Human Ethics Research Committee at the University of New South Wales (HC180136).

that they expend (Cabrales et al., 2019), while a similar decrease in effort can occur through demoralization if individuals find they have overestimated their relative ability (Barankay, 2011).

These mixed results (Bandiera et al., 2013; Bursztyn and Jensen, 2015; Blader et al., 2016) are likely due to the wide range of incentives potentially at play. These include inherent preferences for high rank (*rank incentives*, henceforth), financial and signaling motives, learning and experimentation processes, peer-peer pressure and changes in beliefs about future compensation. In the absence of a tailored experimental design, all these incentives could be active at once, hindering the identification of their individual contributions.

In this paper, I report results from a semester-long randomized controlled trial (RCT) in the education sector, involving hundreds of university students in an introductory microeconomics course. The students received instantaneous private feedback on their performance on a semester-long online assignment that included graded and voluntary academic questions. The contribution of this paper is multifold.

Firstly, the experimental design allows me to cleanly capture the effect of *rank incentives* per se, joining the handful of RCTs that has done so in the literature (Barankay, 2012; Dobrescu et al., 2019). The relative feedback is anonymous, has no impact on the course grade, and is unlikely to generate any artificial boost in effort owing to student learning or experimentation.[2] Secondly, I am the first to study the impact of rank incentives both at the intensive and extensive margin. As part of their online assignment, students can engage in graded assignment (intensive margin) as well as many voluntary, non-graded assignments (extensive margin). I present the first evidence on how rank incentives affect the way students achieve goals over time and how rank incentives distort the way they allocate effort across a range of different activities within their education production function. Thirdly, I exploit the exceptional panel data on student effort captured in the online learning environment to study the underlying mechanism. This paper is the first to examine all the channels previously explored in the literature (e.g., the effect on prior believes, confidence, happiness, etc) in a unified framework, while also including new channels,

[2]Students can affect their assignment rank almost instantly. Experimenting with the effort level required to change their rank is likely to be a transitory process that does not affect the overall performance by the end of the semester.

such as procrastination, by leveraging a unique data set.

Additionally, the experimental design allows me to clearly disentangle the impact of rank incentives from the goal-setting component. The students are randomized into three groups: Control, Goals, and Social Goals. In the Control group students only see their absolute performance and the cumulative number of points they have at any point during the semester. In the Goals group students could also see which milestone they are currently in and how close they are to the next one. In the Social Goals group students can see their points, how close they are to the next goal as well as their relative performance information within their current milestone. Finally, I use rich administrative data to investigate whether the effects of the treatment carry on to future periods.

I find that rank incentives have a motivational effect that sees students 2.6% more likely to go beyond full completion on their assignment by mid-semester. Since the magnitude by which a student's score is above the minimum point threshold does not affect their grade, this provides a clean understanding of how students react to social comparisons at the extensive margin. I find that this effect is heterogeneous across high and low performing students. For the very low performers this brings in roughly 0.27 standard deviations higher exam marks compared to those only exposed to milestones. Rank incentives induce more stress and increase effort, specifically higher course material accesses and lower procrastination. In contrast, high-performing students seem happier, with their superior rank inducing a reallocation of effort towards procrastinating less and overachieving in the assignment on which they are ranked. Yet, they ultimately score between 0.19 and 0.25 standard deviations lower grades. Finally, I show that the rank incentives effects are significantly accentuated for the younger, male, international and business students. There are no overall long term effects beyond the current semester.

My results bring a slate of interesting policy implications. They suggest that rank incentives can distort student effort in a way that is unproductive, and it clearly identifies which groups are more likely to get negatively impacted. The design in Dobrescu et al. (2019), where rank incentives are only allowed to affect student behaviour at the intensive margin seems preferable, in this respect. The general lesson seems to be that whenever relative performance is disclosed, it should be done within well-defined boundaries, on a finite activity that prevents individuals from overexerting effort.

Unlike the results in Cabrales et al. (2019), which point to the fact that students who underestimate their rank perform worse than those who overestimate it, I show that when the effect of rank incentive is the only incentive at play, the priors of the distribution of ability on students do not necessarily change. Rather, the results suggest that a change in psychological factors in response to the assignments/points, which directly results from rank incentives, leads low-performing students to increase their performance and high-performing students to decrease their performance.

From a welfare analysis point of view, my results suggest that rank incentives operate via an increase in stress that results in greater effort being exerted. It does not seem to improve the accuracy of information available to students, rather it seems to operate via a behavioural channel. It is therefore hard to evaluate whether using rank incentives is welfare improving, even for those who see their exam grade improved as a result of it (Katreniakova, 2014).

2.2 Literature Review

Experimental investigation on relative performance feedback has been a fast-growing research area over the past decade. In an influential paper by Schultz et al. (2007), the authors used a field experiment and documented both the constructive and destructive effects of relative performance feedback. Specifically, feedback on whether one's own energy consumption is above or below the average consumption level causes low energy consumers to increase their consumption (i.e., the so-called "boomerang effect") and high energy consumers to decrease their consumption level. Over time, a particular kind of relative performance feedback, ranking, has received increasing attention (Bandiera et al., 2013; Charness et al., 2011; Cabrales et al., 2019; Dobrescu et al., 2019; Villeval, 2020).

To date, the effect of ranking feedback on performance remains ambiguous (Villeval, 2020). Some find positive effects for the whole treatment group (Azmat and Iriberri, 2010; Bandiera et al., 2013; Charness et al., 2011; Katreniakova, 2014), while others find negative or no effects overall (Cabrales et al., 2019; Eriksson et al., 2009). Various behavioural explanations have been adopted to explain those mix effects.[3]

[3]For a most-updated summary of relative feedback related researches and potential underlying

Many existing studies also identify various heterogeneous treatment effects across the distribution. On the one hand, some studies find that high and low performers are the most significantly motivated. (Gill et al., 2018) find that subjects increase effort after being ranked first or last. On the other hand, some researches capture strong effect on one tail of the distribution, yet a lower or no effect on the other tail. Bandiera et al. (2013) find, through a field experiment in a soft fruit producer, that ranking actually decreases average productivity by 14 percent, with productivity reduction occurring only at the bottom of the productivity distribution. Murphy and Weinhardt (2014) find that high performers gain 0.29 standard deviations more than low performers when ranked. Elsner and Isphording (2017) and Goulas and Megalokonomou (2015) also find that high-performing students benefit more from additional ranking information than do low-performing students. Bhanot (2017) finds that poor performers are significantly more discouraged, than high performers are encouraged, by peer ranking information on water conservation. In fact, they even observe a boomerang effect, such that high-performers become demotivated and consume more water after receiving ranking feedback. This effect disappears when ranking information is framed competitively, linking better performance to a winning ribbon. Bhanot (2017) also captures a "last place effect", in which the worst performers perform even worse upon receiving competitively-framed rank information. Barankay (2011) finds negative overall effect of ranking, with a null effect for subjects ranked among the top 10. However, Barankay (2011) finds no heterogeneous treatment effect across various dimensions, such as age, prior ranking, and location of ranking in the distribution.

Two major reasons could have contributed to the mixed evidence related to ranking feedback. One reason is that most of the ranking feedback is provided periodically in a deferred manner. The timing of the feedback matters (Villeval, 2020). A time lag could weaken, or even revert if provided immediately before a subsequent task, the effectiveness of the feedback (Fischer and Wagner, 2018). The other reason is that most of the ranking feedback is provided offline, which may bear a potential signal or financial value and consequently confound the findings (Dobrescu et al., 2019). Technological advancement has made real-time and online ranking feedback possible. However, research evidence on the effectiveness of real-time online ranking feedback is still limited; much less, if any, investigates local ranking. To the best of my knowledge, I am the first to explore the effect and

theories, please refer to Villeval (2020) mechanism of real-time local ranking feedback on performance.

Another strand of related literature is that studying goal-setting. Goals represent status, in the sense that accomplishing different goals signals different status of task completion level. Existing studies have investigated status with or without monetary incentives. Moldovanu et al. (2007) find that symbolic award, such as a congratulatory card, encourages students to take on significantly more work even though the payment remains the same. Experimental evidence specifically related to goal-setting, from the lab or the field, is a rapidly growing area of literature (Van Lent and Souverijn, 2017). The earliest experimental evidence on goal-setting is from Latham and Locke (1979). They explore a series of field experiments and find that goals significantly increase production level by around 19 percent on average. Goerg and Kube (2012) find, in a randomized field experiment, that both endogenous goals (i.e., goals set by oneself) and exogenous goals (i.e., goals set by others) have positive effect on work performance. McCalley and Midden (2002) find similar effects of endogenous and exogenous goals on energy conservation using a laboratory experiment with simulated washing behaviours. Van Lent and Souverijn (2017) use a large RCT involving 1092 students to study the effect of having students set realistic goals and having them set ambitious goals. They find that setting realistic goals significantly increases course performance yet setting ambitious goals does not generate significant difference.

Topic-wise, a similar paper has been written by Akın and Karagözoğlu (2017). Using a laboratory experiment, they study the effect of absolute performance feedback, self-specified goals, and exogenous joint goals with relative performance feedback on work performance. They find that absolute performance feedback lowers performance compared to no feedback at all, while self-specified non-binding goals (i.e., goals without monetary incentives) have no effect. They also find that exogenous goals combined with relative performance feedback decrease performance by 8 percent. However, they target a different kind of relative performance feedback, namely, information on group average, rather than ranking. Their results also have one major limitation: they do not untangle the pure effect of exogenous goal-setting or that of relative performance feedback. Especially for the treatment that joins goal-setting with relative performance feedback, they are not able to identify whether the effect is due to exogenous goal-setting or due to relative performance feedback, since they vary more than one condition across treatments. For example, when comparing endogenous and exogenous goal-setting, they also vary payoff schemes. Furthermore, they do not provide mechanisms for observed effects. In contrast, this experiment, in which one condition is varied per treatment, provides clean evidence on the effect and mechanism of each manipulation.

Comparing rank incentives and exogenous goal-setting in the same experiment is particularly useful, because either one can be laid on to the other with virtually zero cost. This experimental design allows the possibility of singling out the effect of local ranking from the effect of exogenous goal-setting. I also identify the joint effectiveness of ranking and goal-setting, a clear research gap in the literature. Apart from contributions to the related literature in economics, this research also sheds light on management practices, in which the design of incentive-compatible contracts has always been a challenging issue (see Gibbons (2005) for a review of related discussions). Furthermore, this research fits well into the longstanding and still-flourishing literature in psychology and education, as per the aforementioned literature and references therein.

2.3 Randomized Controlled Trial

2.3.1 Environment

The Randomized Controlled Trial (RCT) was implemented in Semester 2, 2018 among students enrolled in a first-year introductory microeconomics course at an Australian university. More than 1,000 students enrol in this course every semester, over a period of 13 teaching weeks. Each teaching week consisted of a 2-hour lecture delivered by academic staff, and a 1-hour tutorial where TAs (also known as "tutors") discussed topics that students found problematic. Students were randomly assigned into lectures and tutorials at the beginning of the semester and could not subsequently switch classes. Class attendance was not compulsory, with tutorial attendance recorded but not rewarded and lecture attendance neither recorded nor rewarded. All instructors (lecturers and tutors) remained the same throughout the semester, and they all used the same teaching materials, including course textbook, lecture slides, and tutorial questions with standardized solutions prepared by the course coordinator.

Besides the 13 teaching weeks, the semester also included a 1-week mid-semester break and a 2-week study period, after which final exams took place. During the entire semester, students (and instructors) had access to (i) a discussion board environment that was used to post comments related to course content and/or administration, and (ii) an educational software that provided them with an online database of approximately 500 questions linked to the course textbook (online assignment henceforth). The textbook covered all the topics traditionally taught in a standard introductory microeconomics course, from the principle of comparative advantage to externalities and public goods. The questions became available from Week 3 onwards and focused on different economic topics corresponding to each one of the 10 textbook chapters, and were grouped as such. Several types of questions were available for students to attempt (e.g., graph, maths, short answer, multiple choice questions), and the software tracked all correct and incorrect attempts via a points-based score system. A correct answer provided by a student would earn points for that student, whereas an incorrect one would cause points to be taken away.[4] After a question was correctly answered, it could also be re-attempted in 'review' mode, with no further points accrued or lost.

Fully completing the online assignment was worth 20% of the overall course grade and it involved earning minimum 1,000 points. Partial completion of the online assignment was also possible, and yielded a proportional grade[5] (approximated to the first decimal). The remaining 80% of the course grade included (i) two invigilated mid-term exams - one in Week 6 and one in Week 10, each worth 15%, and (ii) an invigilated final exam, worth 50%. All exam papers were set up by the course coordinator who drew the corresponding questions from a pre-existing database of uniformly difficult exercises. Both mid-term exams were administered during tutorials over a 45-minute period, and involved questions that required a combination of maths calculations, graph drawing and written short answers. These exam papers were marked by tutors in a double-blind manner, following a strict, detailed set of guidelines provided by the course coordinator with rigorous consistency checks in place. The final exam was scheduled over a 2-hour window after the study period concluded, involved answering 50 multiple choice questions, and was

[4] Multiple choice questions earned 10 points, while maths/graph and short essay questions earned 15 and 40 points, respectively; an incorrect answer caused a deduction of 50% of the points awarded for a correct answer.

[5] The terms 'mark' and 'grade' are used interchangeably in this thesis, both denoting the percentage scores out of 100 earned by a student.

machine-graded.

2.3.2 Treatments

As mentioned, students completed the online assignment by answering online questions and gaining the associated points. To implement the treatments, the score system of the online assignment was used to create a milestone-referenced league-based hierarchy, and the information students had about their performance in the online assignment was varied across treatments. The online assignment was designed around several milestones. Each milestone required students to accrue a minimum specific amount of points and a range between two adjacent milestones is termed "league". Thus, achieving the milestone earned students a spot in a particular league.

At the start of the semester, all students were allocated to a "Beginner" league and randomly divided into 30-person groups.[6] By correctly answering online questions, students started accumulating points and moving into higher leagues (e.g., Bronze III-II-I, Silver III-II-I, Gold III-II-I, etc.).[7] Incorrect answers, as explained above, reduced their overall points count, eventually moving them into the previous (lower) leagues if their points count fell below specific thresholds. Students could progress to as high as the 23rd league (labelled "Nobel" and worth 8,000 points); whenever they 'moved up' into a new league, they became part of a new group in that higher league. If they regressed to a previous league, they were allocated back to the group they previously belonged to. To test whether this milestone-referenced league-based information system has any behavioural impact, an RCT was implemented that varied the progression information that students had access to in relation to their performance in the online assignment. Specifically, students were asked at the start of the semester if they wanted to add a score system to their online assessment. A total of 890 out of 1,062 students agreed to do so, and were subsequently randomly assigned - based on the last number of their student ID - into one of three groups, namely (i) Social Goals treatment; (ii) Goals treatment; and (iii) Control. Students in the Control group had constant access to information about their absolute assignment performance, as captured by the number of points they accumulated up to that moment in time. Students exposed to the Goals

[6]The size of these groups was selected to match the size of the tutorial classes.

[7]For a complete list of the available leagues, please refer to the footnote of Figure B.1

treatment could see that information and could also see what milestones they had achieved (and thus what league they were currently in) and how many points were required to achieve the next milestones (i.e, move into the next leagues). Finally, students in the Social Goals treatment had access to all of the information provided to those in the other two treatments, plus information about the performance of their peers (including those in the other treatment and in the control groups), as denoted by the amount of points and the performance rank[8] achieved by each other student in the student's present group.

Figure B.1 shows an example of the information received by the students in each of the three groups. A quick glance reveals that in terms of the absolute performance information available to students, the treatment arms were perfectly equivalent. When compared to control, the Goals treatment then added information about progression milestones in isolation, via a left panel that showed the corresponding points thresholds and a main panel that showed one's position with respect to the current milestone. Finally, the Social Goals treatment added to the Goals treatment yet another informational layer related to peer progression, by showing in the main panel also the points count and corresponding rank of everyone else in one's group. Overall, 272 students were assigned to the control group, 330 were assigned to the Goals treatment and 288 were in the Social Goals group.

Neither the highest milestone achieved (i.e., the league) nor one's relative performance (i.e., the rank) had any bearing on grades. Students did not know what information or design was varied across treatments. Instructors were not involved in the RCT and were never aware of the treatment group that a student was assigned to. Neither the research study nor the treatments were ever discussed in class (in lectures or tutorials), while the online course platform - where students could access all the materials and ask any questions related to the course - was separated in three different interfaces, with each treatment group being able to access only its assigned one. While I cannot completely rule out spillover possibilities[9], I believe that the limited interaction opportunities that students had online (via separate discussion boards) or in-class (via optional class attendance) can go a long way to-

[8]The ranking information is the student's rank within a 30-student group. A typical 30-student group is composed of participating students whose assignment points fall into the range of a league. Multiple 30-student groups can exist within a league.

[9]For example, a student in the control condition might learn, from a peer who had access to the league information, which league his points fell into. He might adjust his learning strategies and effort exertion accordingly.

wards minimising them. This in turn allows clean identification of the impact of the league-based score system on students' goal-setting behaviour in the online assessment and on academic performance more broadly. In what follows, I estimate these effects. I also investigate the potential for heterogeneous treatment effects and the mechanism behind the results.

2.4 Data and Empirical Analysis

This section presents an overview of the data, verifies that the sample predetermined characteristics are well-balanced across groups, and outlines the empirical analysis strategy.

2.4.1 Data

I use data from three sources, namely university administrative records, (pre- and post-treatment) course survey data, and educational software logs. University records contain several (i) demographic and enrolment indicators (e.g., age, gender, country of birth, whether a student is enrolled full-time or part-time, the field of her degree), as well as (ii) academic performance indicators (e.g., measures of previous academic ability, grades (marks) on each piece of assessment in the course, Semester 2 2018 GPA excluding the course in which the intervention studied here was implemented, etc.). To supplement this information, I use the education software logs that automatically collected timestamped data on (i) the number of times each student accessed her online course platform, (ii) the number of times each student accessed her online assignment, and (iii) the number of points each student accrued. Finally, all students were requested to answer two almost-identical, paper-based surveys - one during Week 2 tutorials (i.e., before the online assignment was available) and another during Week 13 tutorials (i.e., the last week classes were held). Both surveys aimed to capture students' perceived stress and happiness levels, via questions about self-perceived ability to overcome difficulties and self-perceived happiness.[10]

[10]The questions were: "How often have you felt difficulties were piling up so high that you could not overcome them?" with available answers *Always, Often, Sometimes, Rarely, Never* (Cohen and Williamson, 1988), and "In general, I consider myself" with available answers *Very very happy, Very happy, Little happy, Neutral, Little unhappy, Very unhappy, Very very unhappy* (Lyubomirsky and Lepper, 1999). Each question was used to construct a dummy variable equal to one for above ' sometimes/neutral' levels and 0 otherwise.

Table B.1 presents descriptive statistics of the whole sample. Panel A focuses on student and tutor characteristics, while Panel B shows the outcome variables of interest. A quick glance at Panel A reveals that the students in the sample are about 19 years old on average. There are slightly more males (56.5%) than females (43.5%), and most of them (82.2%) are full-time enrolled. The vast majority are pursuing degrees in Commerce & Economics (44.7%) and STEM (41.3%), with only 14.0% pursuing Humanities degrees. Finally, 60.3% of the sample are international students, while the leading geographical regions of origin seem to be Asia (65.4%) and Oceania (29.0%). Prior academic ability is primarily captured by the GPA attained by a student in the semester prior to the semester of the intervention studied here (i.e., Semester 1 2018). Previous semester GPA is non-existent, however, for first-year students entering university in Semester 2 2018 (16.3% of the sample). If, however, Semester 2 2018 first-time enrolees are domestic students (3.4% of the sample[11]), I can proxy their previous ability using the Australian Tertiary Admission Rank score (ATAR henceforth).[12] Unfortunately ATAR is not available for Semester 2 2018 first-time international enrolees, and given that different countries have different academic standards, the university does not maintain a record of their original high school information. As a result, 13.4% of the sample are missing information about prior academic ability. Among those for whom prior ability information is available, I observe levels of roughly 53.3 out of 100. I note that the GPA and ATAR scores are two distinct measures. To tackle this problem, I first predict prior-semester GPA using ATAR for students who have ATAR and then calculate the predicted value of prior-semester GPA based on the fitted equation for all students who only have ATAR. The prior-semester GPA, both original and fitted if an original GPA is missing, forms the eventual prior ability measure used in the regression analyses. Finally, tutorial classes are taught by 15 tutors, of whom 65.2% are males, and only 4.3% are international.

Panel B provides information on performance levels (i.e., course assessment grades)[13] and effort indicators (i.e., course engagement). Final grades in the online assignment appear quite high (9.92 out of 10), with 98.2% of students achieving

[11]Majority (87.4%) of the domestic students enrolled in Semester 1 2018.

[12]ATAR is the primary selection criterion for Australian university undergraduate program admission and represents a student's high-school ranking relative to their peers when completing secondary education. It is computed based on a combination of their score in (i) the final high-school year assessments, and (ii) a national final exam across their best five subjects (or equivalent).

[13]All assessment grades are scaled between zero and 10 to improve comparability.

full marks.[14] The three invigilated exams generated unsurprisingly lower grades, with students scoring 6.3, 5.8 and 6.5 (out of 10) in the Week 6 exam, the Week 10 exam and the final exam, respectively. This brought the weighted average exam grade to 6.2 out of 10 marks, a level slightly lower than the average final course mark achieved in the other courses taken the same semester (7.1 out of 10). In terms of course engagement, on average students logged into the course platform more than 20 times, accessed their assignment about 28 times and accrued roughly 1,253.2 total points (a level equivalent to the achievement of membership in the Bronze III league) by the end of the semester, on average. Silver I was the highest league achieved. Finally, by the end of the semester, about 83.7% (68.8%) of students considered themselves able to overcome difficulties (to be happy), compared with 89.5% (72.0%) at the start of the term.

Table B.2 presents checks on the balance of the sample across the three treatments. The first two columns provide the mean and standard deviations for the Social Goals (Panel B) and Goals (Panel C) groups, while the following two provide the same descriptive statistics for the control group. The final two columns show the difference in means between columns (1) and (3) and the associated standard errors. I find that none of the observable characteristics available in the data are significantly different between treated and control students at conventional levels. Panel A shows the same type of differences, this time between the Social Goals and the Goals groups and reassuringly yields the same conclusion. Finally, an F-test do not reject the hypothesis that students' assignment to treatments is random. Therefore the randomization is confirmed successful.

As in Dobrescu et al. (2019), the baseline specifications control for all predetermined characteristics, except prior ability. The reason is threefold: First, controlling for prior ability would reduce the sample size by 13.4%, increasing the noisiness of the estimates. Second, while this might not seem like a dramatic sample drop, it comes from a particular, non-random student demographic, i.e., all international students who enter university in Semester 2 2018. However, international students can differ from domestic students across various unobservable dimensions (Foster, 2012) and may value their university degree differently.[15] With 94.1% of

[14] The high engagement results from the formative nature of the assignment, which is by design. As long as a student devotes enough effort throughout the semester, she will almost certainly earn full marks on this portion of the course assessment.

[15] Note, for instance, the government requires that tuition fees for international students normally

the missing prior ability data coming from the international subsample, selection becomes an issue. It can also likely considerably affect in particular the specifications that explore the potential for treatment effect heterogeneity in the domestic vs. international subgroups. Third, students were randomized solely based on the last digit of their student ID.[16] Table B.2 shows that the randomization worked well across all available observable characteristics, including all prior ability proxies. It is thus highly unlikely to have failed in one single dimension (i.e., a uniform measure of prior ability). The robustness check specifications that include prior ability yield results consistent with the baseline ones, further alleviating this concern (see Table B.24 in Appendix B).

2.4.2 Empirical Methodology

I am interested in evaluating the impact of the intervention described in the prior section on students' academic performance both in the course where the intervention was implemented, and in the other courses taken simultaneously, as well as in understanding what might be driving the results.

The identification strategy relies on comparing the outcomes of students with similar characteristics and the same tutor, but who were exposed to two different milestone-referenced league-based treatments. Along with that, I also compare students who were exposed to either treatment with students in the control group. I analyse the effect on two types of course outcomes. First, I examine whether the treatments have any influence on students' performance in the online assignment, as indicated by their number of points. Second, I examine the effect of the interventions on academic performance, as measured by students' three exam grades, namely the two midterms and the final exam. To capture overall effects, I also use as outcomes the average exam grade (computed as the weighted mean of the three invigilated exams) and the overall course grade[17]. Using these performance indicators has sev-

cannot be lower than those for domestic students (Ferguson and Sherrell, 2019). International students in fact pay much higher tuition fees than their domestic peers (Hurley and Van Dyke, 2020).

[16]The allocation of student IDs is *as if* random, since the ID numbers simply count upwards based on students' registration time.

[17]Overall course grade is the simple sum of three exam grades, in their original scale (0-15, 0-15, and 0-50, respectively), plus online assignment grade (0-20). Weighted average exam grade is the average of exam grades after they have been scaled to 0-10. For example, a first midterm grade of 15 out of 15 would be weighted to equal $\frac{15}{1.5} = 10$, whereas a final exam grade of 15 out of 50 would be weighted to equal $\frac{15}{5} = 3$. Thus, overall course grade is not to be mixed with weighted

eral advantages. First, none of the instructors was aware of a student's treatment group, which rules out their chance to systematically influence the results of students in different treatments. Second, all exams were invigilated, closed-book, and administered in-class, and thus provide an objective measure of course performance. Third, the final exam was machine-graded, while the midterms were marked by tutors following strict marking guidelines and undergoing several consistency checks. Finally, there was no 'marking on a curve', with no grade adjustment or re-weighting occurring.

Course Effects. To capture the direct effect of the intervention on the assignment progression, I use the milestone structure embedded in the league-based score system. Students were clearly informed that attaining full marks for the assignment meant they had to accrue 1,000 points. Despite knowing it, many accumulated more than the required 1,000 points, effectively going above the call of duty as far as their assignment performance was concerned.[18] To capture this overachievement effect, I look at the extent to which students have passed the subsequent milestones after achieving full marks in their online assignment. Specifically, I capture whether each student has passed the 1,000, 1,300, 1,600, 1,800 and 2,000 points milestones by specific times during the semester (i.e., Week 7, Week 11, Week 13). Recall that earning 1,000 points is equivalent to reaching the first milestone (i.e., Bronze III league) and qualifies a student for full marks of the online assignment, whereas 1,300 points, 1,600 points, and 2,000 points correspond to the cut-off levels of Bronze II, Bronze I, and Silver III leagues respectively.[19] To the last two leagues, I add an intermediary level (1,800 points) that I use to check the robustness of the results related to those (rather few) extremely high-achieving students reaching Silver III. The weeks are chosen as they mark key moments of the course - Week 7 and Week 11 are the weeks right after each mid-term exam, while Week 13 is the last in the semester. By looking at the league transition patterns in time, I am thus able to disentangle whether, when and to what extent students go beyond the call of duty in their online assignment. Additionally, I extend the analysis above to also include course grades as additional outcomes of interest. Specifically, I investigate specifi-

average exam grade.

[18]Table B.1 shows that students ended up in the Bronze III league with 1,253.2 points on average, with the front-runner finishing in the Silver I league (with 2,843 points).

[19]The next available milestone is Silver II with a 2,300 points cut-off; by Week 13, only 2 students achieved it.

cations that involve the grade achieved in each of the three invigilated exams, as well as their weighted average and the overall course grade. By doing so, I check whether the overachieving behaviour recorded in relation to the online assignment has potentially also led to better academic performance.

I employ the following baseline estimating equation

$$Y_{i,t} = \alpha + \beta Treatment_i + \gamma X_i + TutorFE_t + u_{i,t} \tag{2.1}$$

where $Y_{i,t}$ is (i) a dummy variable equal to one if student i taught by tutor t surpasses a certain milestone L, with $L \in \{1,000;\ 1,300;\ 1,600;\ 1,800;\ 2,000\}$ in Week 7, Week 11 and Week 13, respectively, and zero otherwise, or (ii) the standardized[20] grade achieved by student i taught by tutor t in the three invigilated exams (either separately or on average), and overall in the course. $Treatment_i$ equals to (i) one if a student is in the Social Goals group and zero if she in the Goals one; or (ii) one if a student is in the Social Goals (Goals) group and zero if in control. X_i represents student i's characteristics, such as age, gender, dummies for country of birth groups, as well as two dummies equal to one if a student is enrolled full-time or in an Economics degree, respectively. To account for any systematic differences among students' learning experiences in tutorials, I include tutor fixed effects. Robust standard errors are clustered at tutor level to further account for the possibility of common shocks driven by the instructor team (Clustering at the tutorial level to account for common shocks at the class level leaves our results unchanged - see Section 2.7.)

Finally, I explore whether the intervention has an impact on the grades achieved in the other courses taken in the same semester. To do so, I use a (standardized) variable labelled 'GPA, S2-18' that represents Semester 2 2018 GPA adjusted to exclude the grade of the course in which the intervention was implemented. As a result, I am able to test two additional hypotheses: First, given that students might go beyond the call of duty in the course featuring the intervention, it would not be unreasonable to think that they could be allocating less effort into the other courses that they take at the same time. Second, the potential overachieving behaviour I see in relation to their online assignment could become a habit that spills over into

[20]To improve comparability across estimates, all the specifications not involving dummy outcomes use variables of interest that are standardized (to a zero mean and a standard deviation of one).

other courses, leading students to exhibit such behaviour more generally and score better in the other courses too. Furthermore, it is also interesting to investigate the heterogeneity of the effects by gender, international status or field of study, for instance, particularly given the lack of compelling evidence on how such groups might react differently to performance information.

The Mechanism. To uncover the potential mechanisms behind the academic performance results, I take advantage of the survey data available via the course records and the software logs. Specifically, I focus on several self-perception and effort indicators as follows: First, I exploit the survey variables about self-perceived happiness and ability to overcome difficulties, discussed in the prior section.[21] Next, I choose as measures of effort two sets of indicators denoting (i) the level of course engagement (specifically, the number of times a student accessed the online course platform), and (ii) the extent to which overachievement is demonstrated (captured by (a) whether a student achieves milestones beyond the one that, as explained above, is the threshold for achieving full points for the online assignment, and (b) the number of logins to the online assignment during the semester). While the first set of indicator (i.e., the level of course engagement) is a direct course effort proxy, the second set of indicators effectively captures whether the extra (un-incentivized) effort students put in the online assignment affected in some form their (more standard) academic performance. Finally, I also use the number of logins both in the course platform and in the assignment to compute two independent Procrastination Indexes (PI hereafter) as the area under the cumulative curve denoting the number of individual accesses over the semester, multiplied by minus one (Steel et al., 2018). Compared to the simple number of course platform accesses, the PI further quantifies the extent to which students procrastinate. Procrastination stems from delays in beginning or completing an intended course of action, and has been linked to several major problems from national and consumer debt (Sunstein, 2011) to unemployment and job search (Van Hooft et al., 2013) to workplace cyberslacking and presenteeism (O'Neill et al., 2014; Wan et al., 2014). Non-procrastinating students who maintain a sustained 'pace' with the course materials or with their assignment show a progress line that approaches the maximum level by Week 12-13 and then plateaus until the final exam; the area under which the cumulative curve of their logins over time

[21]Recall that these measures were collected both before the intervention was deployed and after it has finished, and the initial levels of both indicators were balanced across three groups. Balance checks are available from the authors upon request.

would therefore be large. Procrastinating students who delay engaging in the course or with the assignment should reach the maximum level only toward the end of the course; the area under which the cumulative curve of their logins over time should be small. Multiplying by minus one makes the index consistent with the direction of other outcomes (under the assumption that less procrastination is better).

2.5 Results

This section presents the impact of the Social Goals vs. Goals treatments on students' goal-setting behaviour and academic performance, while results separately contrasting each treatment with the control are presented in the Appendix.

2.5.1 Baseline Course Effect Estimates

I start by evaluating the direct effect of the intervention on students' performance in the assignment and present results in Panel A of Table B.3. The first row documents the effect of providing students with relative performance information - as opposed to only showing their milestones - on the likelihood of surpassing various points thresholds immediately after the first midterm, while the remaining rows report impact estimates immediately after the Week 10 midterm and at the end of the semester. The estimates unsurprisingly capture the natural progression of students through the milestone structure as the semester advances. I note that by mid-semester, those in the Social Goals group are already 2.6% more likely to have gained full assignment marks than those in the Goals group. By the second time snapshot (Week 11), I see the milestones further shifting again, with the Social Goals group being this time more likely to achieve 60% and 80% more points than necessary for full marks. Notably, the two groups are as likely to have surpassed the full-marks milestone at this point, which seems to suggest that Social Goals students tend to complete the online assignment earlier and then move into higher leagues. This overachieving trend continues until the end of the semester, when the Social Goals group records again a 1.6% higher chance of reaching the 1,800 milestone than their Goals peers. Interestingly, at this point I also observe statistically significant average effects on points accumulated by students in the first league, which might be due to those who left completing the assignment to the last minute (i.e., the procrastinators) doing so now, more successfully in the Social Goals group than in the Goals one. To check whether this is indeed the case, I split the sample in two groups denoted by whether a student has actively engaged with their assignment (and started accumulating points) within the first fortnight or not,[22] and re-run the analysis from Table B.3, Panel A. Table B.7 shows that the Week 13 significant effect on the full-marks milestone is indeed coming solely from the procrastinators, who are 13% more likely to complete the assignment in the Social Goals group than in the Goals one. This is in fact the only significant effect recorded for this subsample, with the non-procrastinator subsample driving all the other significant results in Table B.3, Panel A. Hence, the Social Goals treatment significantly increases the likelihood of achieving full marks in the assignment for both non-procrastinators and procrastinators, earlier in the semester for the former group and later on for the latter one.

Taken together, these estimates point to a consistent tendency of the students exposed to the Social Goals treatment to complete the assignment early (except the procrastinators), and then go above the call of duty and achieve increasingly higher milestones compared to their Goals counterparts. This holds true also when comparing the Social Goals group with the control group (see Table B.8, Panel A), with estimates suggesting higher chances to achieve superior milestones for Social Goals students than for their control peers. I find however no behavioural differences between those in the Goals group and in the control group (see Table B.8, Panel B), which implies that the simple presence of a milestone system does not induce one to go above the call of duty, while providing individual rank information does.

Does this behaviour boost or hinder academic performance in the other course assessments? To answer this question, I run model (2.1) with each of the course grades series as outcome variables. Panel B of Table B.3 presents the overall effects - again contrasting the Social Goals and the Goals groups, while the comparison of each treatment group with the control group is shown as the first row in Table B.9. Interestingly, I only find a small negative effect (equivalent to 0.09 standard deviations (SDs henceforth)) on the grades achieved in the first mid-term. So, compared to the Goals group, the overachieving behaviour displayed in the online assignment early in the semester by the Social Goals group is detrimental to their first invigilated exam score, but has no further impact on the other grades, including on the overall course one.

[22]Roughly 40.3% of the sample reach strictly positive points balances within the first two weeks.

This early negative effect on grades is, however, not present when comparing the treatment groups with the control group (see the first row of each panel in Table B.9). Despite this, it remains a rather interesting result that merits further study. I note that the baseline specification 2.1 assumes a linear impact of an intervention on course performance. It might, however, well be that low and high-achieving students react differently to the treatments, putting more or less weight on the milestone leagues or the rank information. To investigate whether the results vary across the grade distribution, I run unconditional quantile regression models (Firpo et al., 2009) to estimate the effect of the intervention at several percentiles $\theta \in [0,1]$ of the distribution of grades.[23] An attractive feature of an unconditional quantile regression is that its coefficients are directly comparable to standard OLS coefficients (Borah and Basu, 2013). Panel C of Table B.3 shows the coefficients from these quantile regressions (marginal effects) at five academic performance levels, corresponding to the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentile of the grade distribution. Unsurprisingly, no significant effects for median students emerge. When focusing on the tails of the distributions, however, I observe (i) a negative effect on Week 6 midterm marks at the 75^{th} percentile, maintained also for those at the 90^{th} percentile of the distribution, and (ii) a negative effect on Week 10 midterm marks at the 75^{th} percentile of ability counteracted by a positive effect at the 10^{th} percentile. All these effects are not only robust but also sizeable, ranging from roughly -0.25 SDs to 0.27 SDs. For instance, being provided with rank information as opposed to only knowing the relevant milestones leads to 0.22 SDs (0.24 SDs) lower Week 6 exam grades at the 75^{th} (90^{th}) percentile. Furthermore, I find a monotonically decreasing pattern for Week 10 exam effects, with estimates going from 0.27 SDs at the 10^{th} percentile to -0.25 SDs at the 75^{th} one. Most of the results are coming from the non-procrastinator subsample (see Table B.10), with no further significant results present either for the final exam or for the overall course mark. Notably, similar impact patterns emerge when comparing those exposed to the Social Goals treatment with their peer in the control (see UQR estimates in Table B.9), with a 0.42 SDs beneficial Week 6 exam effect at the first decile and a 0.26 SDs detrimental one at the last decile (Panel A). Finally, Panel B in Table B.9 shows no significant difference between those in the Goals group and the control, regardless of the

[23] An unconditional quantile regression model is preferred to the standard conditional quantile model, because the quantile effects estimated by the former model is independent of the set of available conditioning covariates (Borah and Basu, 2013). Robustness checks show that using the latter models yield similar results - see Section 2.7.

assessment.

All in all, these patterns seem to suggest that the presence of a milestone-referenced league-based progression structure does not significantly alter students' academic performance. However, when combined with positive news (i.e., higher rank position), it may demotivate the best students as they can be tempted to 'rest on their laurels'; when combined with negative news (i.e., inferior rank position), it may induce low-performing students to increase their effort and ultimately do better. Doing better (or worse, in relation to better students) refers to performance on in-term assessments, but not in final course marks. Below I will put these conjectures to the test by investigating the mechanism behind these findings.

2.5.2 Heterogeneous Course Effects

I explore the heterogeneity of the impact across several dimensions, including age, gender, whether the student is international or domestic, and whether she is enrolled in a Business degree or not. To do so, I split the full sample using these observable characteristics and re-run the benchmark specification 2.1 separately for different subsamples. Results are reported in the appendix, Tables B.15 - B.22.

We start from Table B.15 and Tables B.19 that look at the impacts by age. I first note that most of the results related to students going above the call of duty come from the younger subsample (see Table B.15, Panel A). Indeed, younger Social Goals students are consistently more likely (by 1.7% to 6.1%) to achieve higher milestones than their Goals peers. The only statistically significant effect for the older subsample is related to achieving the first milestone (and full assignment marks) at the end of the semester. Similar to before, this likely reflects the last-minute cramming of the procrastinators, done (7.5%) more successfully by those in the Social Goals group than in the Goals one, with ultimately no significant academic performance effects. As for grades, Table B.19, Panel A interestingly shows that those aged 19 or less are also the source of all the negative performance effects related to adding rank information to one's milestones. Specifically, I find estimates of -0.35 SDs (-0.29 SDs) for Week 6 (Week 10) midterms at the 75^{th} percentile, and even -0.32 SDs (-0.30 SDs) at the highest performance decile in the final exam (overall course). The strong negative effects on younger cohorts' exam marks could be because younger students see their course mark as a stronger indicator of their academic potential than older students.

Turning to gender, I note no clear behaviour difference between males and females in relation to the overachieving results. I do, however, find (i) a positive effect on Week 10 mid-term marks for males at both the lowest decile and lowest quartile of ability (see Table B.20, Panel A), and (ii) a substantial negative effect on Week 6 mid-term marks for the highest (at 75^{th} and 90^{th} percentile) performing females (see Table B.20, Panel B). Subsequent investigations reveal that this effect is driven by the non-procrastinator subsample,[24] which is unsurprising if one considers this finding in the context of the (weak) goal-setting result that sees female students in the Social Goals group 3.6% more likely than their Goals peers to go 80% above the call of duty in their assignment by the end of the semester.

The third heterogeneity dimension I investigate involves international status. Splitting the sample by whether one is an international student or not reveals strong results for (i) the international subsample both early and very late in the semester, and (ii) the domestic subsample for the period in between (see Table B.17). Indeed, domestic students in the Social Goals treatment are 6.2%, 5.1% and 2.5% more likely to go 30%, 60% and 80%, respectively, above the threshold corresponding to full assignment mark than their peers in the Goals treatment by Week 11. The international subsample in the Social Goals treatment is 4.1% and 8.9% more likely to get to full marks early (Week 7) and late (Week 13), respectively. Interestingly, the subsample of international students is also the main driving force of the mixed - positive (0.49 SDs for Week 10 exam) for low-achievers and negative (0.25 SDs for Week 6 exam) for high-achievers - results on course grades (see Table B.21). In contrast, domestic high-performing (75^{th} percentile) Social Goals students score 0.29 SDs lower Week 6 exam grades than the latter group, an effect driven particularly by the procrastinators.

Finally, I partition the sample by whether a student is pursuing a Business (i.e., Commerce and/or Economics) degree or not. Table B.18 shows that, in contrast with STEM or Humanities students, Business students exposed to the Social Goals treatment are roughly 3.2% more likely to achieve higher milestones than their Goals counterparts but they also generally do so later, after Week 11. In terms of grades, high performing Business students in the Social Goals treatment also score lower

[24]Results are not presented here for brevity.

marks in the two midterms than their Goals peers (by 0.32 SDs and 0.43 SDs, respectively at the third performance quartile), but manage to catch up towards the end of the term such that I find no general detrimental effect of providing rank information on the overall course grade (see Table B.22). No behavioural or performance effects are present for STEM or Humanities students.

To sum up, I find that the younger cohort are more responsive to the relative performance information than the older cohort, whereas men are more responsive than women to the information. Both international students and domestic students overachieve in the online assignment; international students seem overachieving only the first milestone, whereas domestic students seem overachieving several successive higher milestones. As for major, only Business students overachieve milestones and students majoring in other areas do not. When it comes to exam performance, high-performing students (those at the 75^{th} and 90^{th} percentile) that are younger, women, domestic, or studying Business-related degrees are significantly negatively affected by the relative performance information. On the contrary, low-performing students (those at the 10^{th} and 25^{th} percentile) that are older, men, or international students are significantly positively affected in their exam marks by the relative performance information.

2.5.3 Beyond Course Effects

Finally, I use the university's rich administrative data to investigate whether the treatments have any spillover effects onto students' behaviour or outcomes in other courses. To do so, I use the information related to students' academic performance in the other courses taken in Semester 2 2018 and re-run the baseline model (Equation 2.1) in Chapter 2 and the unconditional quantile regressions (Firpo et al., 2009). Table B.4 presents the results, for the full sample (in specifications (1)-(2)), as well as for several subsamples related to the heterogeneity dimensions I investigated (in specifications (3)-(10)). In all models, the outcome variable is the standardized Semester 2 2018 GPA excluding the mark in the course in which the intervention was implemented.

I find that, on average, being treated did not crowd out effort in other courses. (Specification (8) in Table B.4, Panel A is not robust to controlling for student characteristics and tutor fixed effects.) This holds true also for males and females when examnied separately, as well as for the international and domestic subsamples in isolation. Interestingly, however, the very high-performing (i.e., highest decile) students who are either at most 19 years old or enrolled in Business degrees score about 0.33 SDs and 0.34 SDs lower in the other courses taken the same semester, respectively. One potential reason for this could be related to impact of the intervention on the beliefs of these two groups about the inherent value of ability versus effort. 'High praise' for their ability (provided through the intervention) would have potentially enhanced their perceived competence in the course (Koestner et al., 1989), making them also more prone to self-handicap by subsequently withdrawing effort from academic activities, particularly those at which they felt less successful (Hirt and McCrea, 2009).

2.6 The Mechanism

The results so far have shown that providing students with relative performance information, compared to only revealing the milestone structure and their position within it, (i) drives students to achieve milestones well above the ones relevant for their course grades, and (ii) has a beneficial impact for low-performing students in their second midterm and a detrimental one on high-achievers in both midterms, with no further final exam and overall grade effects. What is the mechanism behind these results, however? To answer this question, I maintain the main focus on the comparison between Social Goals and Goals treatments and take advantage of the course survey and software logs data. These two sources contain rich information on self-perception and effort, respectively. For self-perception, I use two indicators that proxy one's level of stress and happiness, captured by whether they consider themselves able to overcome difficulties and whether they consider themselves happy. For effort, I use two measures of course engagement, namely (i) the number of course platform accesses, and (ii) whether one goes above the call of duty in the assignment (i.e., achieves more than the points required for full marks in the assessment). In the following mechanism analysis, I purposefully split the sample by post-intervention ability (hence by the outcome), rather than by pre-intervention ability. This is to induct backwards the most likely mechanisms relating to the formation of the subgroups.

Results contrasting the Social Goals and the Goals groups are presented in Table

B.5 - B.6 and Figure B.2, while the ones related to the treatments versus control are relegated in the Appendix (Tables B.11 - B.14). Since I found the pattern of the effects to be quite different for low-performing students compared to the high-achievers, I present estimates for several subsamples split by various levels of ability (as denoted by different course grades) and discuss the results pertaining to the high- and low-performing students separately. For instance, Panel A in Table B.5 presents estimates on the impact of disclosing rank information (in addition to the milestones) on the perception of one's ability to overcome difficulties for different subsamples identified by the level of academic performance (showed on the rows - e.g., below 10^{th} percentile, above 25^{th} percentile, etc.) in various course assessments (showed on the columns - e.g., first midterm, second midterm, etc.). By looking closely at the same groups of students, based on the intervention-induced ability distribution, the mechanism analysis explores possible drivers behind the academic performance patterns.

We start from the findings presented in Table B.5 that are particularly relevant - and showed in greater detail - for the low-performing students. Panel A presents the estimates related to the self-perception indicator, while Panel B shows the ones associated with the total number of course platform accesses. Remember that the only salient baseline result on grades involved a positive effect of rank information for very low-performance Social Goals students, compared to the Goals group, in the second mid-term exam. Interestingly, specifications (3)-(4) in Panel A show that indeed those scoring below-median in Week 10 midterm are also 12.5% more likely to feel stressed and consider themselves unable to overcome difficulties (Cohen and Williamson, 1988). This is a sizeable and robust result that carries largely also to the lowest 10^{th} or 25^{th} percentiles although the corresponding estimates are not significant, likely due to the considerable drop in sample size. The same pattern but in reverse is showed in Panel B, with below-median Week 10 exam students in the Social Goals group engaging with the course platform 0.35 SDs more compared to their Goals counterparts. Just like for the self-perception indicator, this direction of the result is maintained also in the lowest ability subsamples, with those scoring below the 25^{th} percentile engaging online 0.44 SDs more (a result significant at 15%). Finally, and perhaps most interestingly, a quick glance at the top row of plots in Figure B.2 reveals a marked overall tendency for the Social Goals students to procrastinate less than the Goals one when it comes to engaging with the course materials in general. Note that weekly differences in PI indexes in Figure B.2 are indicated by larger markers if significant.[25] If one considers the entire sample, the difference in course engagement PI appears statistically significant only towards the end of the semester (i.e, from Week 10 onwards). Does this difference aggregate, however, different patterns for high- or low-performing students? To answer this question, I zoom in on the procrastination behaviour related to course platform accesses of those at the bottom and top quartile of academic performance. I see indeed a much lower tendency to procrastinate for low-achievers in the Social Group, as indicated by their significantly lower PI values compared to their Goals peers, starting as early as Week 5. In contrast, providing rank information to high-achievers does not alter their rate of procrastination regardless of the group they belong to. I find generally the same overall results pattern when comparing the Social Goals and Control groups, with no meaningful further differences between the Goals group and Control (see Tables B.11 - B.12).

I interestingly note that good news 'travel faster' than bad news, however. Given that the assignment started in Week 3, it appears that low-performing students take about six weeks to show a positive turnaround in academic performance. The negative effects on high-performing students, however, materialize considerably earlier, with both the first and second mid-term grades of high-achievers being negatively impacted by the good news on their high rank in the assignment. Indeed, Panel C in Table B.3 showed (i) 0.22 SDs and 0.25 SDs lower grades in the Social Goals group compared the Goals ones at the 75^{th} percentile for Week 6 and Week 10 mid-terms, respectively; and even (ii) 0.24 SDs lower grades at the 90^{th} percentile for Week 6 exam. When looking at the happiness level for the associated subsamples, I note that those with top quartile Week 10 grades exposed to the Social Goals treatment also have 19% higher chances than their Goals peers to self-report as happy (see Table B.6, Panel A). This holds true, albeit in a slightly weaker form (at 15% significance), also for those scoring above median in the same assessment, and it appears quite robustly also for those scoring above median in their first mid-term. Finally, despite not being significant, the same direction of the effect is found for the top 10% students in both assessments. Taken together, these estimates suggest that providing good students with their rank information increases their life enjoyment (Lyubomirsky and Lepper, 1999) but lowers their grades. To check if this is related

[25]Statistical significance is established based on standard means tests; a specification including dummies for treatment and for performance quartile, as well as their interaction is also employed as a robustness check and confirms the findings.

in any way to good students being somewhat lulled into a 'false sense of security' by making particularly good progress in the assignment, I look at their likelihood of overachieving by gaining more than the necessary points for full assignment marks. I do so for the subsample of students who score above-median in their exams and show results in Table B.6, Panel B.[26] I immediately note the advantage that good students in the Social Goals treatment exhibit with respect to their Goals peers from very early on. Indeed, among those scoring above-median in Week 6, those exposed to their rank information were 4.6% more likely to continue accumulating points above the maximum required in the course a result significant at 15%. Considering ability proxied by Week 10 score, however, shows robust estimates of 2.2%. By Week 11, the good Social Goals students (with above-median final exam grades) were 5.1% more likely to score well above the required level for full assignment marks. Unsurprisingly, these results are also in line with those related to the assignment PI. Not only better Social Goals students go above the call of duty more than their Goals peers, but they also do so earlier - i.e., procrastinate less engaging with their assignment even before mid-semester. Indeed, a quick glance at the last two rows of plots in Figure B.2 shows that in both cases in which rank information affects high-performing Social Goals students grades (i.e., in their Week 6 and Week 10 exams), I also find this group procrastinating less on assignment engagement than the Goals one, starting from Week 5 at the latest. There are no further significant findings for the bottom performance quartile in either Week 6 or Week 10 rows. No substantial robust statistically significant results are found when comparing either treatment group with the control group (see Tables B.13 - B.14).

To sum up, the findings suggest that low-performing students are feeling more stressed as a result of knowing how they rank compared to their peers rather than just knowing their general milestone position, they work harder (by engaging more with the course and procrastinating less) and score significantly higher in their second exam assessment. In contrast, high-performing students exposed to good news in relation to their assignment tend to continue to outdo their peers in this assessment, grow happier and end up scoring lower grades in both in-semester exams. Notably, none of these effects - positive or negative - survive the end of term, as I ultimately find the intervention to have no impact on either the final exam or the

[26]Restricting the sample to those scoring above 75^{th} percentile yields unsurprisingly weaker results due to $2/3^{rd}$ drop in sample, but they maintain the same direction as the results in Table B.6, Panel B.

overall course grade at any level of academic prowess.

2.7 Robustness

In this section, I discuss potential confounding factors and perform various checks that might bias the analysis, methodological or otherwise, to verify the robustness of the results.

First, I carry out a randomized treatments-type of robustness test. Specifically, I randomly regroup the 890 students in the sample into three alternative, placebo groups and re-run the baseline specifications in Table B.3. Since the placebo treatment groups are absolutely different compared to the actual treatment groups, I should expect no treatment effects. If the placebo treatment effects end up being significant determinants of the assignment progression or academic performance, this would imply that students might have reacted to confounding factors (not perfectly coinciding with the real treatments) and gotten a performance boost. Table B.23 shows no placebo treatment effects. All specifications produce estimates that are close to zero and not significant; there is one exception but but it lacks robustness. I can thus conclude that the results are unlikely to be driven by any concurrent factors other than the real treatment assignment.

Second, I also experiment with two alternative specifications that control for (i) tutorial fixed effects instead of tutor fixed effects, and (ii) prior academic ability. Tutorial fixed effects accounts for any systematic differences between students' learning experiences in tutorials, while including prior ability will net out previous academic performance. As mentioned, I proxy a student's prior ability with her previous GPA score, or with one's ATAR score if GPA data is unavailable. This still leaves 13.4% missing values for prior ability, of which 94% is related to the international students subsample. To tackle the missing prior ability issue, I employ a Multiple Imputation using Chained Equations (MICE) methodology. MICE represents the standard route to estimating models with missing covariate data under a missing-at-random assumption (i.e., if other covariates - but not the problematic covariate itself - in the dataset can be used to predict missingness on a given covariate). In this case, the international status of a student can largely predict the missingness of the prior ability measure. As a result, I include international status both as an auxiliary variable in the imputation models and further control for it in the standard OLS models. Furthermore, prior ability is imputed via truncated OLS methods censored between zero and 100. The number of imputed datasets is equal to the proportion of missing observations of the imputed variable (i.e., I create 14 imputed datasets, rounding up the missing proportion). Table B.24 shows that the findings remain the same.

Third, I also check whether the results in Panel C of Table B.3 are robust in standard conditional quantile regression specifications. Doing so shows that the baseline results generally hold, although the magnitudes vary slightly (see Table B.25). Indeed, I still observe positive effects for the Week 10 exam grade of low-performing students, and negative effects for high-performing students in Week 6 exam grades, as well as on the Week 10 exam and average exam grades when clustering by tutorial. Unlike the UQR results, however, the Social Goals treatment, when compared to the Goals one, appears to also decrease the overall course grade for the top decile high-performers (and even top quartile when clustering by tutorial). Hence, the treatment effects on high-performing students might be potentially underestimated.

Fourth, clustering standard errors at tutorial level to further account for the possibility of common class shocks and using logit models in all specifications involving indicator outcomes also leaves the results unchanged. (Results are not presented here.)

Finally, I also explore if the significant mechanism results in Tables B.5 - B.6 found for specific subsamples are indeed statistically different compared to the non-results present for other samples. To do so, I adjust the model 2.1 to include dummy variables that capture whether one belongs or not to a particular ability level (as denoted by the standard levels showed in Tables B.5 - B.6) and their interaction with the treatment dummies. I find that the findings indeed hold. (Results are not presented here.)

I also note that a major concern in long-term experimental studies is related to attrition. Participants may drop out during the intervention period for various reasons, which could greatly bias results. In this case, however, attrition is zero because treatments were implemented by design right after the census date and so, only permanent course enrolees participated in the study. Additionally, although a drop out option was readily available, no participant actually exercised it. Attrition is thus not an issue in this setting.

Another valid concern is whether a significant difference exists in the proportion of students who complete (at least) 100% of the online assignment. Indeed, students who only partially complete the whole assignment (as measured by scoring above the full assignment marks requirement or not) are less exposed to the treatment than those who complete it. Column (3) in Table B.26 reports the number of students who complete the assignment (i.e., 100% completion rate). Percentages are very similar across treatments (from 97.2% to 99.3%).

2.8 Concluding Remarks

Using exploratory evidence from a large field experiment, I find that students who have access to relative performance information are 2.6% more likely to gain full assignment marks in Week 7 than those without the access. By Week 11, students with the relative performance information access are more likely to achieve 60% and 80% more points than necessary for full marks, and retain this overachieving tendency till Week 13, end of the teaching weeks. These overachieving patterns in the online assignment interestingly translate into a negative 0.09 SD (equivalent to a decrease of 2 marks out of 100) effect on the grades achieved in the first midterm. In other words, the overachieving behaviour seems to harm students' first midterm performance, although no further impact is captured on other course grades. By contrast, no overachieving behaviour in the online assignment is captured for students who have access only to the milestones information. Based on the literature (e.g., Collins and Gan (2013)), I suspect treatment effects might vary along the grade distributions and run additional unconditional quantile regressions. On the one hand, I find that the relative performance information has a negative impact on the two midterms of the highest performance quartile, rendering significant decreases in average exam grades by around 0.19 SDs (equivalent to a decrease of 3 marks out of 100). I also find a positive effect (0.27 SDs, equivalent to an increase of 5.6 marks out of 100) of the performance information on the lowest decile students' second midterm. On the other hand, I find that the relative performance information joint with the milestones information has similar effects on students' exam grades: with access to these two types of information, the lowest decile students experience a performance boost in their first midterm by 0.34 SDs (equivalent to an increase of

7 marks out of 100), whereas the highest performance quartile and decile students suffer a setback of around 0.26 SDs in their first midterm, giving rise to a 0.20 SD (equivalent to 3 marks out of 100) decrease in their average exam grades. All these results survive robustness checks.

As it turns out, high-performing and low-performing students seem driven by different mechanisms, respectively. Low-performing students report feeling more stressed and are found to access the course platform more often, compared to their controlled peers. High-performing students report feeling happier and overachieve leagues (i.e., milestones) in the online assignment. Following Dobrescu et al. (2019), I investigate heterogeneous effects across several dimensions: age, gender, international status, and field of study. I find that the relative performance information mostly motivate overachieving behaviour among the younger students rather than the old ones and among men rather than women. Interestingly, international students are motivated to surpass the full-mark milestone, while domestic students are motivated to further surpass higher milestones above the full-mark threshold. This seems to be suggesting that the relative performance information triggers more effort exertion, pushing students at least one milestone up from where they would have been without the information. I also find that only business-majored students are overachieving milestones. Turning to course grades, the most severely negatively affected groups are always the high-performing students, within certain subgroups: the younger cohort, females, international students, and business-majored students. Meanwhile, low-performing international students and male students are the most positively affected subgroups.

Overall, results from this experiment provide suggestive evidence of the effectiveness of performance information on students' goal-achieving behaviour and academic performance. Although no overall treatment effect is captured for either final exam or overall course grade, I note several pieces of evidence implying potential treatment effects on subsamples. When splitting the sample by procrastinators vs. non-procrastinators, strong and robust negative median effects of the relative performance information are captured for non-procrastinators' first midterm grades, and even stronger effects are captured for high-performing non-procrastinators at the 75^{th} percentile. I also note the non-robust but large positive effects (0.41 SD) on low-performing non-procrastinators' second midterm. When assuming linear treatment effects, these positive and negative effects may largely cancel each other out, rendering insignificant overall effects. I thus suspect that strong treatment effects might exist within subsamples. Additionally, given that existing literature (e.g., Goerg and Kube (2012); McCalley and Midden (2002)) has evidence that exogenous goals significantly alter people's behaviour, yet essentially little treatment effects are captured in this experiment so far with respect to the milestones treatment. I further doubt the overall lack of overall treatment effect of the milestones information may be due to the cancelling-out or diluting noises. I formally explore these conjectures in Chapter 3.

Appendix B

B.1 Main Results

Table B.1: Descriptive Statistics

Variable	Mean	SD	Min.	Max.	N
Panel A.1: Student Level Characteristics					
Age	19.730	1.541	17	31	890
Male	0.565	0.496	0	1	890
Full Time	0.822	0.382	0	1	890
Undertaking an Economics Degree	0.045	0.207	0	1	890
International Status	0.603	0.489	0	1	890
COB: Australia	0.280	0.449	0	1	890
COB: Other Oceania	0.010	0.100	0	1	890
COB: Europe	0.027	0.162	0	1	890
COB: Asia	0.654	0.476	0	1	890
COB: America	0.010	0.100	0	1	890
COB: Africa and Middle East	0.019	0.137	0	1	890
Prior Ability: Prev. GPA/ATAR Score	53.295	28.457	0	99.800	771
GPA Previous Semester	51.441	28.061	0	94.250	745
ATAR (1st Semester Domestic)	92.233	5.906	75.850	99.800	30
Panel A.2: Tutor Level Characteristics					
Male Tutor	0.652	0.482	0	1	46
Tutor International Status	0.043	0.206	0	1	46
COB: Australia	0.891	0.315	0	1	46
COB: Asia	0.065	0.250	0	1	46
COB: America	0.043	0.206	0	1	46
Panel B: Performance, Effort & Self-perception Indicators					
Online Assignment Grade	9.917	0.769	0	10	890
Week 6 Exam Grade	6.300	1.975	0	10	884
Week 10 Exam Grade	5.855	2.089	0	10	872
Final Exam Grade	6.531	1.454	2	9.8	890
(Weighted) Average Exam Grade	6.233	1.560	1.289	9.522	866
Overall Course Grade	7.074	1.184	2.2774	9.575	890
Adjusted GPA Semester 2 2018	6.684	1.223	0	9.433	887
Number of Course Platform Accesses	20.573	22.260	1	197	810
Number of Assignment Accesses	27.763	33.242	1	357	890
Number of Assignment Points	1,253.174	432.867	36.200	2843	886
(Post) Whether Can Overcome Difficulties	0.837	0.369	0	1	639
(Post) Whether Considers Oneself Happy	0.688	0.464	0	1	641

Notes: The classification of the country of birth (*COB*) follows the Standard Australian Classification of Countries, 2011. The Oceania group includes Oceania countries other than Australia. *ATAR* (Australian Tertiary Admission Rank) score denotes a student's ranking relative to his/her peers when completing secondary education. *GPA Previous Semester* is Semester 1 2018 GPA for students enrolled at the university before the intervention semester (i.e., Semester 2 2018). *ATAR (1st Semester Domestic)* is the ATAR score of domestic students first enrolled at the university in Semester 2 2018. *Adjusted GPA Semester 2 2018* is Semester 2 2018 GPA adjusted to exclude the intervention course. All course grades and GPA variables in Panel B are scaled between zero and 10. *Number of Course Platform Accesses* denotes a student's total number of logins into the course platform during the intervention semester. *Number of Assignment Points* captures a student's progress in completing the online assignment (with 1,000 points denoting full marks). The stress and happiness (survey) indicators are the ones collected post-treatment, at the end of the intervention semester.

Table B.2: Balance Tests for Treatments and Control Groups in Semester 2 2018

	Treatment Group		Control Group		Difference	
Variable	Mean	SD	Mean	SD	Diff.	SE
Panel A: Social Goals vs. Goals Group						
Age	19.628	(1.408)	19.818	(1.649)	0.190	(0.124)
Male	0.590	(0.493)	0.558	(0.497)	-0.033	(0.040)
Full-time	0.983	(0.131)	0.991	(0.095)	0.008	(0.009)
Undertaking an Economics Degree	0.042	(0.200)	0.045	(0.209)	0.004	(0.017)
International Status	0.580	(0.494)	0.618	(0.487)	0.038	(0.040)
COB: Australia	0.288	(0.454)	0.267	(0.443)	-0.022	(0.036)
COB: Other Oceania	0.010	(0.102)	0.015	(0.122)	0.005	(0.009)
COB: Europe	0.021	(0.143)	0.033	(0.180)	0.013	(0.013)
COB: Asia	0.649	(0.478)	0.664	(0.473)	0.014	(0.038)
COB: America	0.014	(0.117)	0.006	(0.078)	-0.008	(0.008)
COB: Africa and Middle East	0.017	(0.131)	0.015	(0.122)	-0.002	(0.010)
Prior Ability: Prev. GPA/ATAR Score	53.766	(28.348)	52.275	(28.794)	-1.491	(2.471)
Panel B: Social Goals vs. Control Group						
Age	19.628	(1.408)	19.732	(1.538)	0.103	(0.125)
Male	0.590	(0.493)	0.548	(0.499)	-0.042	(0.042)
Full-time	0.983	(0.131)	0.993	(0.086)	0.010	(0.009)
Undertaking an Economics Degree	0.042	(0.200)	0.048	(0.214)	0.006	(0.017)
International Status	0.580	(0.494)	0.610	(0.489)	0.030	(0.042)
COB: Australia	0.288	(0.454)	0.287	(0.453)	-0.001	(0.038)
COB: Other Oceania	0.010	(0.102)	0.004	(0.061)	-0.007	(0.007)
COB: Europe	0.021	(0.143)	0.026	(0.159)	0.005	(0.013)
COB: Asia	0.649	(0.478)	0.647	(0.479)	-0.002	(0.040)
COB: America	0.014	(0.117)	0.011	(0.105)	-0.003	(0.009)
COB: Africa and Middle East	0.017	(0.131)	0.026	(0.159)	0.008	(0.012)
Prior Ability: Prev. GPA/ATAR Score	53.766	(28.348)	54.072	(28.233)	0.306	(2.584)
Panel C: Goals vs. Control Group						
Age	19.818	(1.649)	19.732	(1.538)	-0.087	(0.131)
Male	0.558	(0.497)	0.548	(0.499)	-0.010	(0.041)
Full-time	0.991	(0.095)	0.993	(0.086)	0.002	(0.007)
Undertaking an Economics Degree	0.045	(0.209)	0.048	(0.214)	0.002	(0.017)
International Status	0.618	(0.487)	0.610	(0.489)	-0.008	(0.040)
COB: Australia	0.267	(0.443)	0.287	(0.453)	0.020	(0.037)
COB: Other Oceania	0.015	(0.122)	0.004	(0.061)	-0.011	(0.008)
COB: Europe	0.033	(0.180)	0.026	(0.159)	-0.008	(0.014)
COB: Asia	0.664	(0.473)	0.647	(0.479)	-0.017	(0.039)
COB: America	0.006	(0.078)	0.011	(0.105)	0.005	(0.007)
COB: Africa and Middle East	0.015	(0.122)	0.026	(0.159)	0.011	(0.011)
Prior Ability: Prev. GPA/ATAR Score	52.275	(28.794)	54.072	(28.233)	1.797	(2.513)

Notes: Each panel reports differences in pre-determined characteristics for students in the Social Goals vs. Goals group (Panel A), as well as in the Social Goals (Panel B) and Goals (Panel C) groups vs. control. The last two columns report the difference in means and the associated standard error of the difference, respectively.

Table B.3: Effect on Course Performance: Social Goals vs. Goals

Panel A: Number of Assignment Points

	1,000+ points		1,300+ points		1,600+ points		1,800+ points		2,000+ points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Week 7	**0.029***	**0.026***	0.007	0.007	0.007	*0.007#*	0.000	0.000	0.000	0.000
	(0.014)	(0.014)	(0.007)	(0.007)	(0.005)	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	615	615	615	615	615	615	615	615	615	615
Week 11	0.028	0.030	0.018	0.018	**0.025***	**0.024***	**0.014****	**0.013****	0.007	0.006
	(0.032)	(0.031)	(0.018)	(0.017)	(0.014)	(0.013)	(0.006)	(0.006)	(0.005)	(0.004)
Obs	615	615	615	615	615	615	615	615	615	615
Week 13	**0.071***	**0.074***	0.022	0.023	0.020	0.020	**0.015***	**0.016****	0.007	0.007
	(0.040)	(0.036)	(0.030)	(0.028)	(0.020)	(0.019)	(0.007)	(0.007)	(0.007)	(0.006)
Obs	615	615	615	615	615	615	615	615	615	615

Panel B: Other Course Grades

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall Course	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Grade	**-0.087***	**-0.094***	0.012	0.003	-0.020	-0.025	-0.033	-0.038	-0.041	-0.053
	(0.048)	(0.045)	(0.089)	(0.088)	(0.077)	(0.077)	(0.062)	(0.061)	(0.067)	(0.068)
Obs	613	613	607	607	618	618	602	602	618	618

Panel C: Other Course Grades, Non-linearities

P10	0.030	0.029	**0.298****	**0.274****	0.140	0.104	0.203	0.173	0.131	0.088
	(0.140)	(0.152)	(0.125)	(0.130)	(0.121)	(0.119)	(0.133)	(0.144)	(0.119)	(0.122)
Obs	613	613	607	607	618	618	602	602	618	618
P25	-0.026	-0.038	0.177	0.137	0.017	0.014	0.044	0.014	0.073	0.058
	(0.120)	(0.122)	(0.139)	(0.137)	(0.121)	(0.121)	(0.121)	(0.131)	(0.099)	(0.106)
Obs	613	613	607	607	618	618	602	602	618	618
P50	0.047	-0.062	-0.071	-0.087	-0.113	-0.109	-0.028	-0.023	-0.058	-0.076
	(0.095)	(0.101)	(0.110)	(0.120)	(0.107)	(0.112)	(0.136)	(0.111)	(0.105)	(0.113)
Obs	613	613	607	607	618	618	602	602	618	618
P75	**-0.227****	**-0.225****	**-0.252****	**-0.251****	0.014	0.021	**-0.199***	**-0.186***	-0.060	-0.059
	(0.096)	(0.106)	(0.105)	(0.100)	(0.124)	(0.126)	(0.103)	(0.103)	(0.107)	(0.097)
Obs	613	613	607	607	618	618	602	602	618	618
P90	**-0.251****	**-0.237****	-0.110	-0.096	-0.072	-0.094	-0.153	-0.140	-0.108	-0.107
	(0.101)	(0.096)	(0.088)	(0.083)	(0.117)	(0.118)	(0.094)	(0.102)	(0.111)	(0.104)
Obs	613	613	607	607	618	618	602	602	618	618
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: ***Panel A:*** Each row presents estimates from separate OLS models. The dependent variable in column (1)-(2) is a binary indicator equal to one if a student's online assignment points by a certain week are above 1,000 and zero otherwise. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for points above 1,300, 1,600, 1,800, and 2,000. ***Panel B-C:*** Each column presents estimates from separate OLS (Panel B) and Unconditional Quantile Regression UQR (Panel C) models. P_i in Panel C represents the UQR result at ith percentile. The dependent variable in column (1)-(2) is the standardized first mid-term grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second mid-term, final exam, weighted average exam, and overall course grades. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses in Panel A and Panel B. Bootstrapped standard errors with 200 replications are reported in parentheses in Panel C. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table B.4: Beyond Course Effects: Social Goals vs. Goals

	All		Age<=19	Age>19	Females	Males	Dom.	Int.	Bus.	Non-Bus.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Adj. GPA Semester 2 2018										
GPA, S2-18	0.031	0.013	-0.093	0.108	0.067	0.015	-0.203	**0.166****	-0.053	0.055
	(0.075)	(0.071)	(0.101)	(0.101)	(0.120)	(0.077)	(0.119)	(0.075)	(0.122)	(0.143)
Obs	615	615	325	290	264	351	245	370	277	338
Panel B: Adj. GPA Semester 2 2018, Non-linearities										
P10	0.168	0.108	0.077	0.091	0.112	0.028	-0.295	0.147	-0.087	0.177
	(0.162)	(0.169)	(0.208)	(0.229)	(0.224)	(0.204)	(0.216)	(0.217)	(0.268)	(0.206)
Obs	615	615	325	290	351	264	245	370	277	338
P25	0.040	0.021	0.009	0.031	0.060	0.128	-0.099	0.160	-0.088	0.114
	(0.100)	(0.107)	(0.172)	(0.123)	(0.136)	(0.169)	(0.181)	(0.145)	(0.150)	(0.154)
Obs	615	615	325	290	351	264	245	370	277	338
P50	-0.007	-0.019	-0.145	0.017	-0.010	0.041	-0.218	0.173	-0.064	-0.025
	(0.085)	(0.089)	(0.116)	(0.130)	(0.113)	(0.142)	(0.144)	(0.117)	(0.120)	(0.139)
Obs	615	615	325	290	351	264	245	370	277	338
P75	0.001	0.004	-0.138	0.144	-0.074	0.135	-0.152	0.051	0.013	-0.086
	(0.093)	(0.090)	(0.118)	(0.121)	(0.117)	(0.164)	(0.149)	(0.101)	(0.151)	(0.141)
Obs	615	615	325	290	351	264	245	370	277	338
P90	-0.162	-0.144	**-0.334****	0.030	-0.193	-0.142	-0.209	0.069	**-0.342****	0.002
	(0.105)	(0.108)	(0.133)	(0.169)	(0.158)	(0.133)	(0.158)	(0.160)	(0.166)	(0.181)
Obs	615	615	325	290	351	264	245	370	277	338
Student Char	X	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tutor FE	X	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each row presents estimates from separate OLS (Panel A) and UQR (Panel B) models. P_i represents UQR results at ith percentile. The dependent variable is the standardized GPA in the intervention semester (Semester 2 2018) adjusted to exclude the intervention course. Columns (1)-(2) show results for the full sample, while columns (3)-(4), (5)-(6), (7)-(8) and (9)-(10) focus on subsamples split by age, gender, international status and degree field, respectively. Column (1) only include treatment indicators; subsequent columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses in Panel A. Bootstrapped standard errors with 200 replications are reported in parentheses in Panel B. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively. Panel A, specification (8) is not robust.

Table B.5: Mechanism for Low-performing Students: Social Goals vs. Goals

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Can Overcome Difficulties										
Below 10	-0.000	-0.062	-0.237	-0.034	0.073	0.031	0.000	0.107	0.089	-0.019
	(0.106)	(0.204)	(0.185)	(0.265)	(0.096)	(0.204)	(0.141)	(0.389)	(0.106)	(0.165)
Obs	40	40	34	34	43	43	32	32	33	33
Below 25	-0.066	-0.061	-0.059	0.001	0.022	0.051	-0.105	-0.072	-0.066	-0.074
	(0.060)	(0.083)	(0.070)	(0.048)	(0.062)	(0.074)	(0.080)	(0.119)	(0.075)	(0.095)
Obs	97	97	94	94	121	121	87	87	93	93
Below 50	-0.047	-0.064	**-0.127****	**-0.125****	*-0.079#*	*-0.071#*	**-0.106***	**-0.134****	-0.083	*-0.087#*
	(0.052)	(0.052)	(0.046)	(0.044)	(0.047)	(0.045)	(0.054)	(0.054)	(0.056)	(0.056)
Obs	225	225	214	214	229	229	202	202	214	214
Above 75	-0.112	-0.094	0.009	0.033	-0.008	-0.006	-0.014	-0.013	-0.002	-0.014
	(0.089)	(0.110)	(0.051)	(0.068)	(0.058)	(0.074)	(0.072)	(0.096)	(0.072)	(0.087)
Obs	116	116	118	118	113	113	117	117	120	120
Above 90	**-0.348****	-0.182	**0.189****	0.218	0.021	0.021	0.013	-0.063	-0.029	-0.114
	(0.132)	(0.227)	(0.082)	(0.212)	(0.077)	(0.128)	(0.073)	(0.078)	(0.099)	(0.162)
Obs	49	49	37	37	46	46	45	45	50	50
Panel B: Number of Course Platform Accesses										
Below 10	-0.119	-0.062	-0.016	0.181	-0.315	-0.094	**-0.291****	-0.176	-0.108	0.019
	(0.140)	(0.222)	(0.096)	(0.208)	(0.262)	(0.542)	(0.133)	(0.269)	(0.169)	(0.380)
Obs	56	56	52	52	61	61	52	52	52	52
Below 25	0.035	0.009	*0.346#*	*0.437#*	**-0.307***	-0.187	-0.026	0.050	0.049	0.098
	(0.125)	(0.149)	(0.218)	(0.270)	(0.154)	(0.164)	(0.146)	(0.182)	(0.156)	(0.192)
Obs	135	135	140	140	167	167	131	131	132	132
Below 50	0.171	0.145	**0.421***	**0.355***	-0.076	-0.049	*0.213#*	*0.248#*	**0.207***	**0.243***
	(0.119)	(0.114)	(0.219)	(0.197)	(0.136)	(0.144)	(0.134)	(0.142)	(0.115)	(0.121)
Obs	294	294	285	285	301	301	272	272	281	281
Above 75	0.027	0.136	-0.134	-0.101	0.207	0.323	0.067	0.251	0.117	0.197
	(0.156)	(0.148)	(0.122)	(0.132)	(0.239)	(0.298)	(0.216)	(0.288)	(0.109)	(0.181)
Obs	118	118	138	138	128	128	128	128	135	135
Above 90	**-0.417*****	**-0.446***	-0.156	-0.250	-0.341	-0.580	-0.285	-0.274	-0.236	-0.031
	(0.130)	(0.217)	(0.290)	(0.400)	(0.324)	(0.726)	(0.195)	(0.201)	(0.185)	(0.247)
Obs	48	48	43	43	45	45	48	48	53	53
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models run on the subsample of students with grades (showed on the columns) above or below certain percentiles (showed on the rows). [For instance, *Below 10* in columns (1)-(2) refers to the subsample of students with first mid-term grades below the 10^{th} percentile.] ***Panel A:*** The dependent variable is a binary variable equal to one if one considers themselves able to overcome difficulties and zero otherwise. ***Panel B:*** The dependent variable is the standardized number of times one accessed the course platform during the intervention semester. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

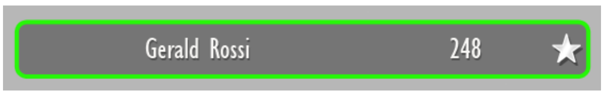
Table B.6: Mechanism for High-performing Students: Social Goals vs. Goals

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Considers Oneself Happy										
Below 10	0.050	-0.173	0.071	0.212	-0.251	-0.309	0.017	-0.013	-0.100	**-0.539****
	(0.129)	(0.285)	(0.112)	(0.236)	(0.186)	(0.337)	(0.153)	(0.432)	(0.198)	(0.209)
Obs	40	40	34	34	43	43	32	32	33	33
Below 50	0.012	-0.001	0.086	0.126	0.008	0.064	0.068	0.091	-0.001	0.093
	(0.090)	(0.118)	(0.065)	(0.122)	(0.064)	(0.062)	(0.095)	(0.108)	(0.087)	(0.090)
Obs	97	97	94	94	121	121	87	87	93	93
Above 50	**0.122****	**0.124****	**0.089***	*0.080#*	0.037	0.025	0.071	0.066	0.075	0.069
	(0.052)	(0.047)	(0.046)	(0.048)	(0.053)	(0.047)	(0.050)	(0.056)	(0.045)	(0.048)
Obs	231	231	235	235	229	229	245	245	244	244
Above 75	0.071	0.075	**0.206*****	**0.190*****	0.018	0.032	0.052	-0.009	0.021	0.007
	(0.070)	(0.085)	(0.041)	(0.058)	(0.052)	(0.039)	(0.060)	(0.076)	(0.062)	(0.055)
Obs	116	116	119	119	114	114	118	118	121	121
Above 90	0.013	-0.089	0.177	0.219	-0.062	-0.081	-0.071	**-0.196****	-0.037	-0.138
	(0.111)	(0.237)	(0.130)	(0.205)	(0.108)	(0.142)	(0.119)	(0.089)	(0.125)	(0.122)
Obs	49	49	37	37	46	46	45	45	50	50
Panel B: Has 1,000+ Assignment Points (for 'Above 50' subsample)										
Week 7	**0.055***	*0.046#*	**0.023***	**0.022***	0.016	0.017	0.000	0.000	0.000	0.000
	(0.028)	(0.028)	(0.011)	(0.011)	(0.010)	(0.012)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	293	293	293	293	288	288	305	305	308	308
Week 11	0.038	0.044	0.046	0.041	**0.056***	**0.051***	**0.029****	**0.027***	0.014	0.013
	(0.051)	(0.058)	(0.033)	(0.029)	(0.030)	(0.026)	(0.012)	(0.013)	(0.009)	(0.009)
Obs	293	293	293	293	288	288	305	305	308	308
Week 13	0.065	0.077	0.040	0.037	0.052	*0.047#*	**0.031****	**0.033****	0.015	0.015
	(0.070)	(0.073)	(0.052)	(0.049)	(0.035)	(0.030)	(0.014)	(0.014)	(0.013)	(0.012)
Obs	293	293	293	293	288	288	305	305	308	308
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: ***Panel A:*** Each row presents estimates from separate OLS models run on the subsample of students with various grades (showed on the columns) above or below certain percentiles (showed on the rows). [For instance, *Below 10* in columns (1)-(2) refers to the subsample of students with first mid-term grades below the 10^{th} percentile.] The dependent variable is a binary indicator equal to one if one considers themselves happy and zero otherwise. ***Panel B:*** Each row presents estimates from separate OLS models run on the subsample of students with above-median grades (showed on the columns) scoring above the 1,000 points threshold by a certain week (showed on the rows). [For instance, *Week 7* in columns (1)-(2) of Panel B refers to the subsample of students with above-median first mid-term grades that accrue (or not) at least 1,000 points by Week 7.] Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. $^{\#}$, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

A: Social Goals Treatment

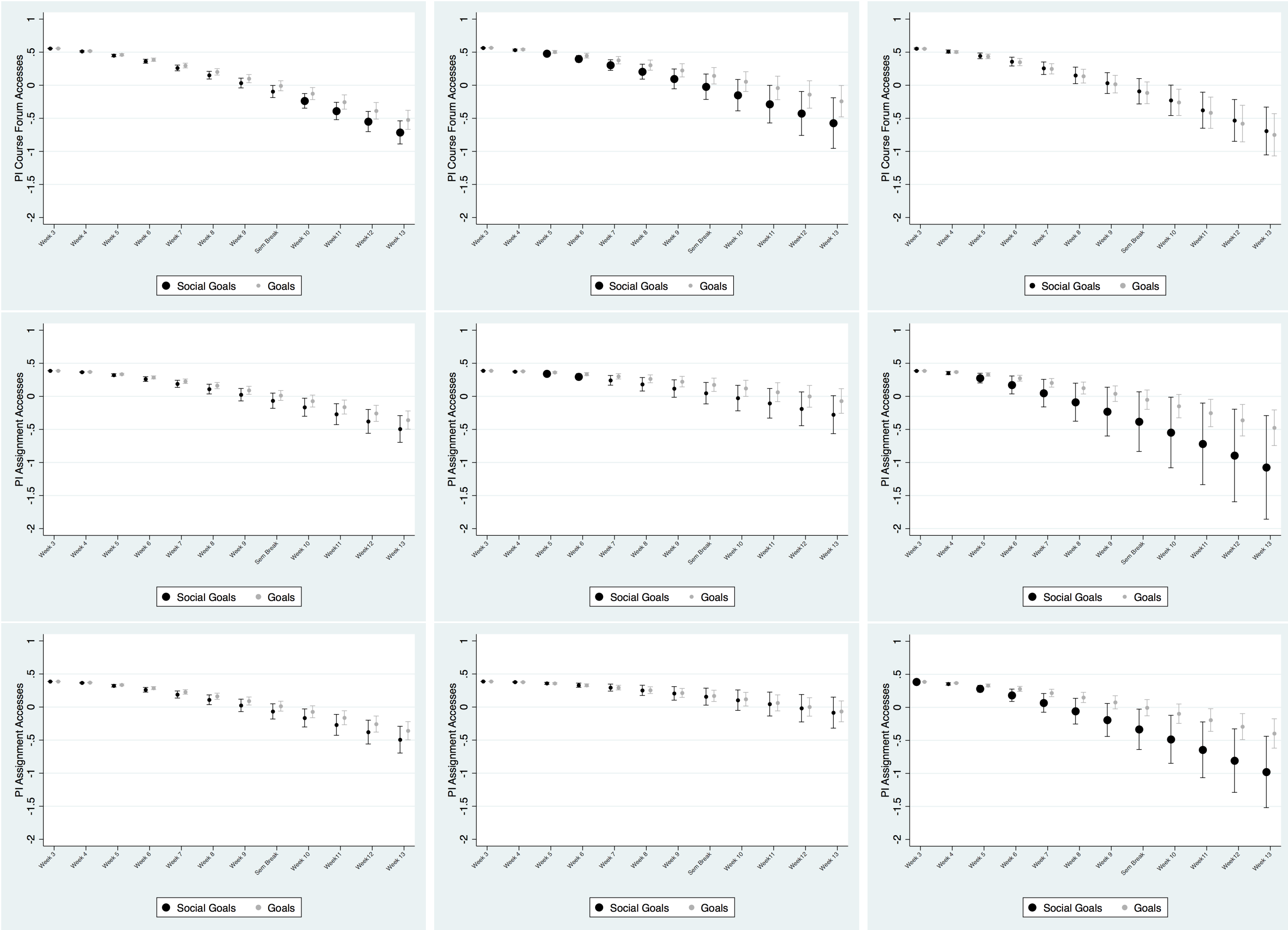
B: Goals Treatment



C: Control

Notes: All names are fictional to guarantee anonymity. A complete list of the available leagues corresponding to points are as follows: Beginner [0,1000), Bronze III [1000, 1300), Bronze II [1300, 1600), Bronze I [1600, 2000), Silver III [2000, 2300), Silver II [2300, 2600), Silver I [2600, 3000), Gold III [3000, 3300), Gold II [3300, 3600), Gold I [3600, 4000), Platinum III [4000, 4300), Platinum II [4300, 4600), Platinum I [4600, 5000), Diamond III [5000, 5300), Diamond II [5300, 5600), Diamond I [5600, 6000), Scholar III [6000, 6300), Scholar II [6300, 6600), Scholar I [6600, 7000), Professor III [7000, 7300), Professor II [7300, 7600), Professor I [7600, 8000), and Nobel for 8000 and above.

Figure B.1: Example of Performance Information as Shown in the Assignment



Notes: The figure presents the procrastination index (PI) on (i) the course platform accesses based on Week 6 performance (top row plots), and (ii) the assignment accesses based on Week 6 performance (middle row plots) and Week 10 performance (bottom row plots). First column plots the PI for the full sample, while the subsequent two columns refer to the bottom and top 25^{th} percentile of performance, respectively. The only exception is the bottom right plot that considers the top 25^{th} percentile below the 10^{th} percentile, to account for the lack of grade effect at the 10^{th} percentile. Social Goals PI indices found significantly different from their Goals counterparts are indicated by larger markers.

Figure B.2: Procrastination and Assignment Progression

B.2 Other Results

Table B.7: Effects on Assignment Progression, Procrastinators vs. Non-procrastinators

	1,000+ points		1,300+ points		1,600+ points		1,800+ points		2,000+ points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Procrastinators										
Week 7	0.015	0.012	0.007	0.006	0.000	0.000	0.000	0.000	0.000	0.000
	(0.010)	(0.008)	(0.007)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	290	290	290	290	290	290	290	290	290	290
Week 11	0.049	0.037	0.015	0.012	0.007	0.006	0.007	0.006	0.007	0.006
	(0.036)	(0.034)	(0.014)	(0.011)	(0.007)	(0.006)	(0.007)	(0.006)	(0.007)	(0.006)
Obs	290	290	290	290	290	290	290	290	290	290
Week 13	**0.132****	**0.131****	0.031	*0.029#*	0.001	0.002	0.007	0.006	0.007	0.006
	(0.055)	(0.050)	(0.021)	(0.018)	(0.010)	(0.007)	(0.007)	(0.006)	(0.007)	(0.006)
Obs	290	290	290	290	290	290	290	290	290	290
Panel B: Non-procrastinators										
Week 7	**0.044***	**0.049****	0.008	0.007	*0.013#*	**0.012***	0.000	0.000	0.000	0.000
	(0.025)	(0.022)	(0.011)	(0.009)	(0.009)	(0.007)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	325	325	325	325	325	325	325	325	325	325
Week 11	0.012	0.024	0.022	0.025	*0.041#*	*0.039#*	**0.020***	*0.018#*	0.007	0.005
	(0.054)	(0.051)	(0.032)	(0.033)	(0.025)	(0.024)	(0.011)	(0.011)	(0.007)	(0.005)
Obs	325	325	325	325	325	325	325	325	325	325
Week 13	0.022	0.043	0.016	0.025	0.039	0.039	*0.022#*	*0.024#*	0.008	0.006
	(0.053)	(0.051)	(0.048)	(0.047)	(0.035)	(0.034)	(0.013)	(0.014)	(0.011)	(0.010)
Obs	325	325	325	325	325	325	325	325	325	325
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models. The dependent variable in column (1)-(2) is a binary indicator equal to one if a student's online assignment points by a certain week are above 1,000 and zero otherwise. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for points above 1,300, 1,600, 1,800, and 2,000. Procrastinator status is assigned based on whether one has actively engaged with their assignment (and started accumulating points) within the first fortnight or not. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.8: Effects on Assignment Progression

	1,000+ points		1,300+ points		1,600+ points		1,800+ points		2,000+ points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Control										
Week 7	*0.019#*	0.010	**0.010***	0.007#	0.007	0.005	0.000	0.000	0.000	0.000
	(0.011)	(0.013)	(0.005)	(0.005)	(0.005)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	558	558	558	558	558	558	558	558	558	558
Week 11	0.014	-0.000	0.023	0.020	0.021	**0.019***	0.010	0.011	0.007	0.006
	(0.036)	(0.038)	(0.019)	(0.018)	(0.014)	(0.011)	(0.008)	(0.008)	(0.005)	(0.004)
Obs	558	558	558	558	558	558	558	558	558	558
Week 13	0.056	0.042	0.022	0.015	0.005	0.001	-0.001	0.001	-0.001	-0.000
	(0.042)	(0.045)	(0.031)	(0.032)	(0.016)	(0.015)	(0.009)	(0.010)	(0.007)	(0.008)
Obs	558	558	558	558	558	558	558	558	558	558
Panel B: Goals vs. Control										
Week 7	-0.010	-0.013	0.003	0.004	0.000	0.000	0.000	0.000	0.000	0.000
	(0.016)	(0.018)	(0.003)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	601	601	601	601	601	601	601	601	601	601
Week 11	-0.014	-0.030	0.005	0.003	-0.004	-0.005	-0.004	-0.004	0.000	0.000
	(0.030)	(0.030)	(0.008)	(0.009)	(0.004)	(0.004)	(0.004)	(0.004)	(0.000)	(0.000)
Obs	601	601	601	601	601	601	601	601	601	601
Week 13	-0.015	-0.032	-0.000	-0.006	-0.015	-0.018	*-0.016#*	*-0.017#*	-0.008	-0.009
	(0.030)	(0.030)	(0.016)	(0.016)	(0.017)	(0.017)	(0.010)	(0.010)	(0.007)	(0.007)
Obs	601	601	601	601	601	601	601	601	601	601
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models. The dependent variable in column (1)-(2) is a binary indicator equal to one if a student's online assignment points by a certain week are above 1,000 and zero otherwise. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for points above 1,300, 1,600, 1,800, and 2,000. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.9: Effects on Other Course Grades

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Control										
Grade	0.023	0.018	0.046	0.026	-0.028	-0.040	0.024	0.012	-0.014	-0.028
	(0.085)	(0.091)	(0.067)	(0.063)	(0.070)	(0.071)	(0.061)	(0.059)	(0.062)	(0.061)
Obs	556	556	545	545	560	560	541	541	560	560
P10	**0.344***	**0.424***	0.247	0.234	0.111	0.154	0.123	0.146	0.121	0.149
	(0.200)	(0.220)	(0.153)	(0.144)	(0.135)	(0.131)	(0.160)	(0.168)	(0.151)	(0.162)
Obs	556	556	545	545	560	560	541	541	560	560
P25	0.095	0.120	0.116	0.066	0.031	0.043	0.092	0.071	0.091	0.097
	(0.119)	(0.120)	(0.135)	(0.137)	(0.107)	(0.103)	(0.124)	(0.127)	(0.102)	(0.117)
Obs	556	556	545	545	560	560	541	541	560	560
P50	-0.019	-0.053	0.083	0.064	-0.144	-0.172	0.098	0.060	-0.040	-0.084
	(0.124)	(0.129)	(0.129)	(0.121)	(0.131)	(0.131)	(0.138)	(0.116)	(0.112)	(0.119)
Obs	556	556	545	545	560	560	541	541	560	560
P75	**-0.201****	**-0.249****	-0.119	-0.147	-0.050	-0.087	*-0.173#*	**-0.203***	-0.111	-0.155
	(0.097)	(0.115)	(0.103)	(0.111)	(0.105)	(0.124)	(0.110)	(0.118)	(0.104)	(0.106)
Obs	556	556	545	545	560	560	541	541	560	560
P90	**-0.242****	**-0.262****	-0.134	-0.152	-0.015	-0.069	**-0.195***	**-0.231****	-0.100	-0.148
	(0.112)	(0.112)	(0.105)	(0.100)	(0.131)	(0.124)	(0.116)	(0.110)	(0.105)	(0.118)
Obs	556	556	545	545	560	560	541	541	560	560
Panel B: Goals vs. Control										
Grade	0.110	0.100	0.035	0.027	-0.008	-0.006	0.057	0.050	0.027	0.019
	(0.069)	(0.073)	(0.082)	(0.075)	(0.052)	(0.058)	(0.063)	(0.057)	(0.050)	(0.046)
Obs	599	599	592	592	602	602	589	589	602	602
P10	0.244	0.287	-0.091	-0.084	-0.042	-0.008	-0.023	-0.030	0.050	0.054
	(0.215)	(0.190)	(0.159)	(0.176)	(0.131)	(0.135)	(0.148)	(0.150)	(0.129)	(0.135)
Obs	599	599	592	592	602	602	589	589	602	602
P25	0.086	0.070	-0.067	-0.061	-0.054	-0.055	0.013	0.018	0.022	0.032
	(0.112)	(0.116)	(0.146)	(0.146)	(0.110)	(0.126)	(0.139)	(0.133)	(0.104)	(0.110)
Obs	599	599	592	592	602	602	589	589	602	602
P50	0.038	0.019	0.117	0.096	-0.021	-0.062	0.085	0.049	0.059	0.033
	(0.130)	(0.134)	(0.126)	(0.125)	(0.114)	(0.123)	(0.114)	(0.121)	(0.123)	(0.141)
Obs	599	599	592	592	602	602	589	589	602	602
P75	0.013	-0.006	0.146	0.125	-0.061	-0.040	0.089	0.106	-0.055	-0.056
	(0.101)	(0.102)	(0.098)	(0.103)	(0.109)	(0.126)	(0.108)	(0.122)	(0.107)	(0.111)
Obs	599	599	592	592	602	602	589	589	602	602
P90	-0.065	-0.076	-0.010	-0.017	0.058	0.091	-0.017	-0.029	0.041	0.034
	(0.099)	(0.091)	(0.079)	(0.094)	(0.129)	(0.116)	(0.101)	(0.096)	(0.099)	(0.112)
Obs	599	599	592	592	602	602	589	589	602	602
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each column presents estimates from separate OLS (1st row in each panel) and UQR models. The dependent variables in columns (1)-(2), (3)-(4), (5)-(6), (7)-(8), and (9)-(10) are the standardized grades as indicated. Odd numbered columns only include treatment indicators; even ones additionally control for age, gender, dummies for countries of birth groups, whether a student is enrolled full-time, whether a student is enrolled in an Economics degree and tutor fixed effects. Robust standard errors clustered by tutor (1st row in each panel) or bootstrapped with 200 replications are in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.10: Effects on Other Course Grades, Procrastinators vs. Non-procrastinators

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Procrastinators										
P10	0.147	0.086	**0.324***	*0.296#*	0.118	0.085	**0.373***	0.293	0.123	0.050
	(0.158)	(0.194)	(0.180)	(0.192)	(0.146)	(0.162)	(0.192)	(0.235)	(0.171)	(0.205)
Obs	288	288	282	282	290	290	280	280	290	290
P25	0.140	0.085	0.251	0.206	0.206	0.175	0.225	0.174	0.073	0.035
	(0.201)	(0.220)	(0.181)	(0.176)	(0.147)	(0.170)	(0.149)	(0.170)	(0.147)	(0.145)
Obs	288	288	282	282	290	290	280	280	290	290
P50	0.134	0.181	-0.081	0.002	0.034	0.070	0.000	0.039	0.094	0.108
	(0.140)	(0.156)	(0.176)	(0.153)	(0.167)	(0.184)	(0.143)	(0.166)	(0.134)	(0.140)
Obs	288	288	282	282	290	290	280	280	290	290
P75	-0.072	-0.011	-0.005	0.065	-0.041	0.019	-0.024	0.061	0.002	0.085
	(0.180)	(0.194)	(0.178)	(0.179)	(0.175)	(0.180)	(0.184)	(0.184)	(0.162)	(0.156)
Obs	288	288	282	282	290	290	280	280	290	290
P90	-0.314	-0.144	0.091	0.222	0.181	0.271	-0.078	0.149	0.027	0.152
	(0.206)	(0.220)	(0.163)	(0.160)	(0.251)	(0.263)	(0.215)	(0.213)	(0.236)	(0.218)
Obs	288	288	282	282	290	290	280	280	290	290
Panel B: Non-procrastinators										
P10	-0.045	-0.074	**0.410***	*0.353#*	-0.012	0.019	0.049	0.008	-0.087	-0.115
	(0.188)	(0.200)	(0.238)	(0.227)	(0.180)	(0.178)	(0.218)	(0.211)	(0.199)	(0.181)
Obs	325	325	325	325	328	328	322	322	328	328
P25	-0.207	**-0.316****	0.007	-0.011	-0.010	-0.031	-0.064	-0.133	0.072	0.007
	(0.142)	(0.142)	(0.194)	(0.196)	(0.161)	(0.183)	(0.182)	(0.171)	(0.154)	(0.164)
Obs	325	325	325	325	328	328	322	322	328	328
P50	**-0.275****	**-0.299****	-0.132	-0.138	-0.026	-0.028	-0.057	-0.073	-0.148	-0.155
	(0.136)	(0.146)	(0.133)	(0.162)	(0.157)	(0.170)	(0.132)	(0.147)	(0.136)	(0.138)
Obs	325	325	325	325	328	328	322	322	328	328
P75	-0.198	-0.180	**-0.284****	**-0.317*****	-0.093	-0.089	**-0.325*****	**-0.352*****	-0.113	-0.121
	(0.128)	(0.130)	(0.115)	(0.119)	(0.144)	(0.137)	(0.099)	(0.109)	(0.138)	(0.136)
Obs	325	325	325	325	328	328	322	322	328	328
P90	0.000	-0.015	-0.172	-0.198	-0.073	-0.091	-0.132	-0.134	-0.137	-0.128
	(0.159)	(0.150)	(0.108)	(0.122)	(0.137)	(0.136)	(0.118)	(0.126)	(0.118)	(0.124)
Obs	325	325	325	325	328	328	322	322	328	328
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each column presents estimates from separate UQR models. P_i represents the UQR result at ith percentile. The dependent variable in column (1)-(2) is the standardized first mid-term grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second mid-term, final exam, weighted average exam, and overall course grades. Procrastinator status is assigned based on whether one has actively engaged with their assignment (and started accumulating points) within the first fortnight or not. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Bootstrapped standard errors with 200 replications are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.11: Mechanism for Low-performing Students: Can Overcome Difficulties

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Control										
Below 10	0.063	0.149	**-0.455****	*-0.307*$^{\#}$	0.083	0.295	0.058	0.283	0.133	0.160
	(0.139)	(0.142)	(0.172)	(0.199)	(0.157)	(0.275)	(0.202)	(0.165)	(0.163)	(0.136)
Obs	39	39	24	24	34	34	25	25	25	25
Below 25	-0.007	0.101	-0.043	-0.097	-0.103	-0.093	-0.104	-0.083	-0.088	-0.056
	(0.071)	(0.085)	(0.113)	(0.147)	(0.077)	(0.101)	(0.093)	(0.110)	(0.084)	(0.104)
Obs	93	93	78	78	105	105	81	81	88	88
Below 50	-0.075	-0.055	-0.092	-0.101	**-0.099***	-0.078	*-0.113*$^{\#}$	**-0.118***	-0.102	-0.095
	(0.051)	(0.051)	(0.062)	(0.059)	(0.055)	(0.057)	(0.065)	(0.066)	(0.067)	(0.064)
Obs	197	197	191	191	198	198	171	171	181	181
Above 75	**-0.165****	**-0.165***	-0.058	-0.083	-0.031	-0.036	-0.063	-0.070	-0.014	-0.008
	(0.074)	(0.092)	(0.066)	(0.067)	(0.056)	(0.065)	(0.055)	(0.060)	(0.059)	(0.061)
Obs	94	94	86	86	98	98	93	93	104	104
Above 90	**-0.302****	-0.203	-0.016	-0.013	0.014	0.091	-0.075	0.021	-0.102	-0.031
	(0.136)	(0.157)	(0.096)	(0.172)	(0.091)	(0.199)	(0.106)	(0.224)	(0.109)	(0.174)
Obs	46	46	32	32	34	34	36	36	37	37
Panel B: Goals vs. Control										
Below 10	0.063	-0.049	**-0.217*****	-0.284	0.010	0.019	0.058	0.054	0.044	-0.087
	(0.130)	(0.230)	(0.072)	(0.206)	(0.128)	(0.281)	(0.127)	(0.250)	(0.136)	(0.210)
Obs	39	39	36	36	41	41	33	33	33	33
Below 25	0.059	0.037	0.017	-0.026	**-0.125*****	**-0.128****	0.001	-0.044	-0.022	-0.053
	(0.075)	(0.096)	(0.075)	(0.087)	(0.041)	(0.054)	(0.072)	(0.083)	(0.062)	(0.084)
Obs	94	94	102	102	112	112	88	88	97	97
Below 50	-0.027	-0.014	0.035	0.031	-0.020	-0.005	-0.007	-0.013	-0.019	-0.022
	(0.024)	(0.029)	(0.055)	(0.055)	(0.037)	(0.047)	(0.045)	(0.048)	(0.038)	(0.039)
Obs	204	204	213	213	203	203	195	195	195	195
Above 75	-0.053	-0.039	-0.067	-0.060	-0.023	-0.030	-0.049	-0.046	-0.012	-0.034
	(0.070)	(0.088)	(0.057)	(0.073)	(0.062)	(0.086)	(0.061)	(0.073)	(0.072)	(0.094)
Obs	120	120	114	114	109	109	124	124	118	118
Above 90	0.046	0.074	**-0.205***	-0.306	-0.007	-0.025	-0.088	**-0.158***	-0.073	-0.108
	(0.082)	(0.148)	(0.103)	(0.247)	(0.090)	(0.218)	(0.070)	(0.086)	(0.076)	(0.115)
Obs	59	59	41	41	42	42	49	49	49	49
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models run on the subsample of students with grades (showed on the columns) above or below certain percentiles (showed on the rows). [For instance, ***Below 10*** in columns (1)-(2) refers to the subsample of students with first mid-term grades below the 10^{th} percentile.] The dependent variable is a binary indicator equal to one if one considers themselves able to overcome difficulties and zero otherwise. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. $^{\#}$, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.12: Mechanism for Low-performing Students: Course Platform Accesses

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Control										
Below 10	-0.260	-0.387	-0.081	-0.022	-0.149	-0.475	-0.194	-0.197	-0.126	-0.074
	(0.292)	(0.514)	(0.115)	(0.166)	(0.243)	(0.388)	(0.138)	(0.279)	(0.182)	(0.272)
Obs	58	58	42	42	51	51	40	40	45	45
Below 25	-0.038	-0.181	0.295	0.174	-0.101	0.008	0.045	0.059	0.180	0.221
	(0.161)	(0.266)	(0.221)	(0.214)	(0.134)	(0.147)	(0.166)	(0.155)	(0.158)	(0.170)
Obs	134	134	127	127	151	151	122	122	129	129
Below 50	0.170	0.118	**0.413***	**0.300***	0.024	0.071	**0.306***	**0.306****	**0.266***	*0.250*#
	(0.106)	(0.141)	(0.202)	(0.147)	(0.112)	(0.104)	(0.148)	(0.143)	(0.133)	(0.152)
Obs	270	270	267	267	269	269	249	249	258	258
Above 75	0.144	0.152	-0.073	-0.019	**0.547*****	**0.559****	0.134	0.219	0.157	0.202
	(0.200)	(0.219)	(0.144)	(0.233)	(0.152)	(0.190)	(0.220)	(0.283)	(0.179)	(0.214)
Obs	106	106	108	108	122	122	110	110	125	125
Above 90	-0.230	-0.008	0.113	-0.149	0.251	-0.015	-0.145	-0.243	-0.053	-0.004
	(0.298)	(0.422)	(0.239)	(0.426)	(0.162)	(0.225)	(0.100)	(0.180)	(0.108)	(0.173)
Obs	48	48	40	40	38	38	45	45	45	45
Panel B: Goals vs. Control										
Below 10	-0.141	-0.326	-0.065	-0.092	0.166	0.192	0.097	0.251	-0.017	0.066
	(0.298)	(0.629)	(0.076)	(0.117)	(0.229)	(0.208)	(0.154)	(0.238)	(0.177)	(0.204)
Obs	56	56	58	58	60	60	50	50	51	51
Below 25	-0.074	-0.084	-0.050	-0.052	0.207	0.165	0.070	0.178	0.131	0.184
	(0.164)	(0.232)	(0.066)	(0.074)	(0.130)	(0.126)	(0.137)	(0.145)	(0.108)	(0.117)
Obs	135	135	143	143	160	160	127	127	135	135
Below 50	-0.001	-0.033	-0.008	-0.029	0.100	0.064	0.093	0.107	0.059	0.049
	(0.107)	(0.128)	(0.055)	(0.053)	(0.128)	(0.133)	(0.100)	(0.093)	(0.108)	(0.112)
Obs	272	272	276	276	278	278	265	265	265	265
Above 75	0.116	0.077	0.060	0.048	**0.340***	*0.284*#	0.067	0.065	0.041	-0.005
	(0.136)	(0.167)	(0.113)	(0.102)	(0.173)	(0.177)	(0.121)	(0.135)	(0.146)	(0.167)
Obs	142	142	146	146	130	130	152	152	146	146
Above 90	0.187	0.089	0.269	0.548	**0.592***	0.576	0.140	**0.257****	0.183	0.126
	(0.229)	(0.400)	(0.201)	(0.393)	(0.309)	(0.492)	(0.178)	(0.112)	(0.199)	(0.184)
Obs	64	64	55	55	45	45	61	61	58	58
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models run on the subsample of students with grades (showed on the columns) above or below certain percentiles (showed on the rows). [For instance, *Below 10* in columns (1)-(2) refers to the subsample of students with first mid-term grades below the 10^{th} percentile.] The dependent variable is the standardized number of times one accessed the course platform during the intervention semester. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.13: Mechanism for High-performing Students: Considers Oneself Happy

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Control										
Below 10	-0.234	-0.206	-0.133	-0.567	-0.299	-0.053	-0.276	-0.356	-0.200	-0.315
	(0.177)	(0.268)	(0.150)	(0.399)	(0.186)	(0.338)	(0.199)	(0.337)	(0.217)	(0.393)
Obs	39	39	24	24	34	34	25	25	30	30
Below 50	0.036	0.046	0.004	0.021	0.028	0.053	-0.019	0.002	0.006	0.045
	(0.075)	(0.071)	(0.061)	(0.056)	(0.066)	(0.067)	(0.080)	(0.078)	(0.079)	(0.080)
Obs	197	197	191	191	198	198	171	171	181	181
Above 50	0.091	0.080	0.110	0.081	0.101	0.079	0.121	0.095	0.118*	0.102
	(0.087)	(0.100)	(0.078)	(0.090)	(0.066)	(0.082)	(0.091)	(0.109)	(0.067)	(0.081)
Obs	196	196	193	193	197	197	211	211	214	214
Above 75	0.053	0.108	0.106	0.067	0.087	0.053	0.061	0.001	0.088	0.020
	(0.088)	(0.115)	(0.118)	(0.144)	(0.102)	(0.142)	(0.094)	(0.124)	(0.115)	(0.140)
Obs	95	95	88	88	100	100	95	95	106	106
Above 90	0.151	0.198	-0.056	-0.092	-0.023	0.021	-0.122	-0.044	0.000	0.061
	(0.147)	(0.227)	(0.111)	(0.295)	(0.148)	(0.275)	(0.157)	(0.340)	(0.133)	(0.249)
Obs	46	46	33	33	35	35	37	37	38	38
Panel B: Goals vs. Control										
Below 10	**-0.284***	**-0.397***	-0.204	-0.261	-0.048	-0.357	**-0.292***	*-0.385#*	-0.100	-0.052
	(0.145)	(0.192)	(0.166)	(0.244)	(0.150)	(0.225)	(0.141)	(0.225)	(0.190)	(0.411)
Obs	39	39	36	36	41	41	33	33	33	33
Below 50	0.022	0.015	-0.021	0.012	-0.071	-0.031	-0.060	-0.022	-0.049	0.016
	(0.045)	(0.047)	(0.061)	(0.062)	(0.055)	(0.050)	(0.047)	(0.047)	(0.058)	(0.052)
Obs	204	204	213	213	203	203	195	195	195	195
Above 50	-0.031	-0.005	0.021	0.028	0.064	0.082	0.050	0.057	0.043	0.052
	(0.072)	(0.070)	(0.070)	(0.072)	(0.055)	(0.064)	(0.066)	(0.063)	(0.055)	(0.056)
Obs	223	223	208	208	226	226	224	224	234	234
Above 75	-0.018	0.029	-0.101	-0.081	0.069	0.107	0.009	0.048	0.067	0.103
	(0.094)	(0.097)	(0.106)	(0.110)	(0.091)	(0.115)	(0.074)	(0.084)	(0.087)	(0.092)
Obs	121	121	115	115	110	110	125	125	119	119
Above 90	0.138	0.022	-0.233	-0.324	0.039	-0.030	-0.051	-0.129	0.037	-0.001
	(0.123)	(0.153)	(0.146)	(0.185)	(0.108)	(0.142)	(0.108)	(0.147)	(0.119)	(0.150)
Obs	59	59	42	42	43	43	50	50	50	50
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models run on the subsample of students with grades (showed on the columns) above or below certain percentiles (showed on the rows). [For instance, *Below 10* in columns (1)-(2) refers to the subsample of students with first mid-term grades below the 10^{th} percentile.] The dependent variable is a binary indicator equal to one if one considers themselves happy and zero otherwise. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. $^{\#}$, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.14: Mechanism for High-performing Students: Has 1,000+ Assignment Points

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Control (for the 'Above 50' subsample)										
Week 7	0.031	0.014	**0.023***	**0.020***	0.016	0.012	0.000	0.000	0.000	0.000
	(0.025)	(0.029)	(0.011)	(0.011)	(0.010)	(0.011)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	262	262	252	252	262	262	267	267	276	276
Week 11	-0.000	-0.019	0.050	0.040	*0.047#*	**0.040***	0.021	0.024	0.014	0.014
	(0.067)	(0.073)	(0.038)	(0.031)	(0.029)	(0.020)	(0.016)	(0.017)	(0.009)	(0.010)
Obs	262	262	252	252	262	262	267	267	276	276
Week 13	0.038	0.019	0.046	0.032	0.017	0.005	-0.004	0.007	-0.001	0.003
	(0.074)	(0.079)	(0.055)	(0.050)	(0.034)	(0.028)	(0.017)	(0.017)	(0.015)	(0.015)
Obs	262	262	252	252	262	262	267	267	276	276
Panel B: Goals vs. Control (for the 'Above 50' subsample)										
Week 7	-0.024	*-0.038#*	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	(0.023)	(0.024)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	293	293	281	281	292	292	292	292	304	304
Week 11	-0.038	-0.070	0.004	-0.006	-0.009	-0.011	-0.008	-0.007	0.000	0.000
	(0.056)	(0.061)	(0.015)	(0.018)	(0.008)	(0.010)	(0.008)	(0.007)	(0.000)	(0.000)
Obs	293	293	281	281	292	292	292	292	304	304
Week 13	-0.026	-0.062	0.006	-0.008	-0.035	-0.039	**-0.035***	**-0.034***	-0.016	*-0.019#*
	(0.049)	(0.048)	(0.030)	(0.029)	(0.026)	(0.027)	(0.018)	(0.019)	(0.013)	(0.013)
Obs	293	293	281	281	292	292	292	292	304	304
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models run on the subsample of students with above-median grades (showed on the columns) scoring above the 1,000 points threshold by a certain week (showed on the rows). [For instance, *Week 7* in columns (1)-(2) of Panel B refers to the subsample of students with above-median first mid-term grades that accrue (or not) at least 1,000 assignment points by Week 7.] Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.15: Effects on Assignment Progression, Age

	1,000+ points		1,300+ points		1,600+ points		1,800+ points		2,000+ points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Age <=19										
Week 7	0.034	0.033	0.007	0.006	0.006	0.006	0.000	0.000	0.000	0.000
	(0.024)	(0.023)	(0.011)	(0.010)	(0.007)	(0.007)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	324	324	324	324	324	324	324	324	324	324
Week 11	0.026	0.031	0.027	0.027	**0.026***	**0.024***	**0.019***	**0.017***	0.013	0.011
	(0.043)	(0.047)	(0.018)	(0.018)	(0.013)	(0.012)	(0.010)	(0.009)	(0.009)	(0.008)
Obs	324	324	324	324	324	324	324	324	324	324
Week 13	0.058	*0.075*$^{\#}$	**0.061***	**0.060***	*0.027*$^{\#}$	0.022	**0.026****	**0.025****	0.013	0.011
	(0.049)	(0.049)	(0.033)	(0.031)	(0.016)	(0.015)	(0.010)	(0.011)	(0.009)	(0.008)
Obs	324	324	324	324	324	324	324	324	324	324
Panel B: Age>19										
Week 7	**0.024***	0.011	0.008	0.005	0.008	0.005	0.000	0.000	0.000	0.000
	(0.013)	(0.013)	(0.007)	(0.005)	(0.007)	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	291	291	291	291	291	291	291	291	291	291
Week 11	0.029	0.021	0.008	0.004	0.023	0.019	0.008	0.006	0.000	0.000
	(0.037)	(0.030)	(0.024)	(0.022)	(0.016)	(0.014)	(0.008)	(0.007)	(0.000)	(0.000)
Obs	291	291	291	291	291	291	291	291	291	291
Week 13	*0.085*$^{\#}$	**0.075***	-0.022	-0.021	0.013	0.017	0.003	0.005	0.001	0.002
	(0.048)	(0.038)	(0.035)	(0.030)	(0.028)	(0.023)	(0.014)	(0.013)	(0.010)	(0.008)
Obs	291	291	291	291	291	291	291	291	291	291
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models. The dependent variable in column (1)-(2) is a binary indicator equal to one if a student's online assignment points by a certain week are above 1,000 and zero otherwise. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for points above 1,300, 1,600, 1,800, and 2,000. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. $^{\#}$, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.16: Effects on Assignment Progression, Gender

	1,000+ points		1,300+ points		1,600+ points		1,800+ points		2,000+ points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Males										
Week 7	**0.038***	0.024	**0.018***	0.016#	0.012	0.011	0.000	0.000	0.000	0.000
	(0.019)	(0.019)	(0.009)	(0.010)	(0.008)	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	351	351	351	351	351	351	351	351	351	351
Week 11	0.007	-0.010	0.015	0.005	0.024	0.019	0.012	0.013	0.012	0.013
	(0.030)	(0.024)	(0.022)	(0.019)	(0.019)	(0.017)	(0.008)	(0.009)	(0.008)	(0.009)
Obs	351	351	351	351	351	351	351	351	351	351
Week 13	*0.075#*	0.053	0.023	0.008	0.025	0.016	0.006	0.007	0.012	0.013
	(0.044)	(0.040)	(0.028)	(0.026)	(0.023)	(0.022)	(0.010)	(0.011)	(0.008)	(0.009)
Obs	351	351	351	351	351	351	351	351	351	351
Panel B: Females										
Week 7	0.017	0.015	-0.007	-0.007	0.000	0.000	0.000	0.000	0.000	0.000
	(0.026)	(0.029)	(0.007)	(0.007)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	264	264	264	264	264	264	264	264	264	264
Week 11	0.055	0.066	0.022	0.033	*0.025#*	*0.031#*	0.017	0.021	0.000	0.000
	(0.059)	(0.058)	(0.027)	(0.028)	(0.015)	(0.018)	(0.012)	(0.015)	(0.000)	(0.000)
Obs	264	264	264	264	264	264	264	264	264	264
Week 13	0.069	0.085	0.021	0.041	0.013	0.025	*0.027#*	**0.036***	0.002	0.006
	(0.075)	(0.074)	(0.048)	(0.050)	(0.027)	(0.026)	(0.018)	(0.019)	(0.011)	(0.012)
Obs	264	264	264	264	264	264	264	264	264	264
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models. The dependent variable in column (1)-(2) is a binary indicator equal to one if a student's online assignment points by a certain week are above 1,000 and zero otherwise. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for points above 1,300, 1,600, 1,800, and 2,000. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.17: Effects on Assignment Progression, International status

	1,000+ points		1,300+ points		1,600+ points		1,800+ points		2,000+ points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: International Students										
Week 7	**0.042****	**0.041****	0.001	0.001	0.006	0.006	0.000	0.000	0.000	0.000
	(0.015)	(0.016)	(0.008)	(0.008)	(0.006)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	369	369	369	369	369	369	369	369	369	369
Week 11	0.030	0.044	-0.010	-0.010	0.007	0.006	0.006	0.006	0.006	0.006
	(0.040)	(0.045)	(0.017)	(0.017)	(0.010)	(0.012)	(0.006)	(0.006)	(0.006)	(0.006)
Obs	369	369	369	369	369	369	369	369	369	369
Week 13	**0.075***	**0.089****	0.001	-0.002	-0.001	-0.001	0.001	0.002	0.001	0.002
	(0.037)	(0.034)	(0.028)	(0.026)	(0.019)	(0.018)	(0.008)	(0.008)	(0.008)	(0.008)
Obs	369	369	369	369	369	369	369	369	369	369
Panel B: Domestic Students										
Week 7	0.010	0.008	*0.017*$^{\#}$	**0.015***	0.008	0.007	0.000	0.000	0.000	0.000
	(0.030)	(0.033)	(0.010)	(0.008)	(0.008)	(0.006)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	246	246	246	246	246	246	246	246	246	246
Week 11	0.022	0.021	**0.058***	**0.062***	**0.050***	**0.051***	**0.025***	**0.025***	0.008	0.007
	(0.059)	(0.060)	(0.033)	(0.032)	(0.025)	(0.024)	(0.013)	(0.014)	(0.008)	(0.007)
Obs	246	246	246	246	246	246	246	246	246	246
Week 13	0.064	0.083	0.051	0.060	**0.050***	**0.054***	**0.033****	**0.039****	0.017	0.017
	(0.062)	(0.063)	(0.044)	(0.044)	(0.028)	(0.028)	(0.013)	(0.015)	(0.012)	(0.013)
Obs	246	246	246	246	246	246	246	246	246	246
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models. The dependent variable in column (1)-(2) is a binary indicator equal to one if a student's online assignment points by a certain week are above 1,000 and zero otherwise. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for points above 1,300, 1,600, 1,800, and 2,000. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. $^{\#}$, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.18: Effects on Assignment Progression, Degree

	1,000+ points		1,300+ points		1,600+ points		1,800+ points		2,000+ points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Business Degree										
Week 7	0.024	0.018	0.009	0.009	0.008	0.005	0.000	0.000	0.000	0.000
	(0.033)	(0.036)	(0.013)	(0.014)	(0.008)	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	276	276	276	276	276	276	276	276	276	276
Week 11	0.037	0.032	0.020	0.017	**0.032****	**0.032***	**0.032****	**0.032***	0.016	0.015
	(0.054)	(0.056)	(0.022)	(0.019)	(0.014)	(0.015)	(0.014)	(0.015)	(0.011)	(0.012)
Obs	276	276	276	276	276	276	276	276	276	276
Week 13	0.079	**0.093***	0.027	0.018	**0.033***	**0.033***	**0.032****	**0.032***	**0.024***	*0.022#*
	(0.052)	(0.049)	(0.035)	(0.033)	(0.017)	(0.018)	(0.014)	(0.015)	(0.013)	(0.013)
Obs	276	276	276	276	276	276	276	276	276	276
Panel B: Non-business Degree										
Week 7	0.033	0.017	0.006	0.001	0.006	0.001	0.000	0.000	0.000	0.000
	(0.024)	(0.022)	(0.006)	(0.002)	(0.006)	(0.002)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	339	339	339	339	339	339	339	339	339	339
Week 11	0.020	-0.008	0.016	0.011	0.019	0.016	0.000	0.000	0.000	0.000
	(0.051)	(0.046)	(0.028)	(0.023)	(0.018)	(0.012)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	339	339	339	339	339	339	339	339	339	339
Week 13	0.067	0.030	0.018	0.012	0.009	0.006	0.001	0.005	-0.006	-0.005
	(0.051)	(0.048)	(0.038)	(0.031)	(0.028)	(0.022)	(0.009)	(0.009)	(0.006)	(0.005)
Obs	339	339	339	339	339	339	339	339	339	339
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models. The dependent variable in column (1)-(2) is a binary indicator equal to one if a student's online assignment points by a certain week are above 1,000 and zero otherwise. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for points above 1,300, 1,600, 1,800, and 2,000. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.19: Effects on Other Course Grades, Age

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Age<=19										
P10	-0.006	-0.015	*0.386*$^{\#}$	**0.361***	0.064	0.041	0.120	0.116	-0.036	-0.033
	(0.217)	(0.227)	(0.247)	(0.212)	(0.149)	(0.138)	(0.202)	(0.202)	(0.169)	(0.167)
Obs	322	322	317	317	325	325	314	314	325	325
P25	-0.058	-0.058	0.001	-0.013	-0.078	-0.069	-0.176	-0.185	-0.062	-0.046
	(0.153)	(0.163)	(0.203)	(0.231)	(0.140)	(0.144)	(0.193)	(0.211)	(0.158)	(0.158)
Obs	322	322	317	317	325	325	314	314	325	325
P50	-0.082	-0.105	-0.145	-0.193	-0.153	-0.201	-0.159	-0.192	-0.116	-0.165
	(0.132)	(0.145)	(0.145)	(0.164)	(0.172)	(0.155)	(0.161)	(0.151)	(0.140)	(0.147)
Obs	322	322	317	317	325	325	314	314	325	325
P75	**-0.328****	**-0.348*****	**-0.247****	**-0.291****	-0.109	-0.141	**-0.309****	**-0.319****	-0.101	-0.135
	(0.134)	(0.123)	(0.125)	(0.129)	(0.195)	(0.207)	(0.124)	(0.141)	(0.156)	(0.163)
Obs	322	322	317	317	325	325	314	314	325	325
P90	**-0.408****	**-0.416*****	**-0.188***	**-0.195***	**-0.309***	**-0.325****	**-0.303****	**-0.304****	**-0.288****	**-0.298****
	(0.162)	(0.160)	(0.109)	(0.107)	(0.162)	(0.164)	(0.132)	(0.138)	(0.130)	(0.142)
Obs	322	322	317	317	325	325	314	314	325	325
Panel B: Age>19										
P10	0.085	0.111	*0.272*$^{\#}$	**0.293***	0.217	0.236	0.189	0.130	0.145	0.097
	(0.213)	(0.222)	(0.175)	(0.176)	(0.174)	(0.172)	(0.186)	(0.204)	(0.201)	(0.194)
Obs	291	291	290	290	293	293	288	288	293	293
P25	0.039	0.035	0.190	0.176	0.204	0.218	0.133	0.114	0.173	0.135
	(0.173)	(0.183)	(0.172)	(0.215)	(0.176)	(0.171)	(0.159)	(0.169)	(0.143)	(0.142)
Obs	291	291	290	290	293	293	288	288	293	293
P50	-0.122	-0.130	0.082	0.135	-0.106	-0.069	0.096	0.132	0.030	0.034
	(0.154)	(0.157)	(0.166)	(0.175)	(0.164)	(0.177)	(0.154)	(0.167)	(0.164)	(0.185)
Obs	291	291	290	290	293	293	288	288	293	293
P75	-0.133	-0.127	-0.138	-0.041	0.095	0.127	0.000	0.069	-0.013	0.033
	(0.160)	(0.170)	(0.160)	(0.178)	(0.162)	(0.172)	(0.192)	(0.194)	(0.169)	(0.147)
Obs	291	291	290	290	293	293	288	288	293	293
P90	-0.015	0.026	-0.018	0.062	0.197	0.188	0.062	0.135	-0.023	-0.012
	(0.146)	(0.147)	(0.132)	(0.142)	(0.181)	(0.190)	(0.152)	(0.168)	(0.161)	(0.183)
Obs	291	291	290	290	293	293	288	288	293	293
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each column presents estimates from separate UQR models. P_i represents the UQR result at ith percentile. The dependent variable in column (1)-(2) is the standardized first mid-term grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second mid-term, final exam, weighted average exam, and overall course grades. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Bootstrapped standard errors with 200 replications are reported in parentheses. $^{\#}$, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.20: Effects on Other Course Grades, Gender

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Males										
P10	0.049	-0.009	**0.362****	**0.283***	0.048	-0.025	0.265	0.232	0.110	0.007
	(0.170)	(0.190)	(0.165)	(0.160)	(0.158)	(0.167)	(0.201)	(0.198)	(0.175)	(0.191)
Obs	350	350	346	346	354	354	342	342	354	354
P25	0.027	0.021	**0.378***	*0.318#*	0.034	-0.019	0.243	0.202	0.142	0.105
	(0.165)	(0.186)	(0.193)	(0.199)	(0.153)	(0.169)	(0.160)	(0.172)	(0.138)	(0.144)
Obs	350	350	346	346	354	354	342	342	354	354
P50	0.015	0.035	0.003	-0.011	-0.156	-0.193	0.048	0.016	0.016	-0.017
	(0.133)	(0.114)	(0.164)	(0.159)	(0.157)	(0.164)	(0.157)	(0.151)	(0.159)	(0.156)
Obs	350	350	346	346	354	354	342	342	354	354
P75	-0.143	-0.125	-0.164	-0.135	0.179	0.183	-0.088	-0.054	-0.023	-0.036
	(0.148)	(0.136)	(0.132)	(0.122)	(0.166)	(0.172)	(0.144)	(0.164)	(0.123)	(0.140)
Obs	350	350	346	346	354	354	342	342	354	354
P90	-0.153	-0.106	-0.090	-0.029	0.066	0.005	-0.022	0.004	0.010	0.010
	(0.131)	(0.144)	(0.138)	(0.136)	(0.149)	(0.150)	(0.132)	(0.155)	(0.135)	(0.130)
Obs	350	350	346	346	354	354	342	342	354	354
Panel B: Females										
P10	-0.147	-0.018	0.315	0.434	0.147	0.241	0.129	0.236	0.191	0.258
	(0.273)	(0.309)	(0.258)	(0.269)	(0.165)	(0.179)	(0.199)	(0.249)	(0.175)	(0.187)
Obs	263	263	261	261	264	264	260	260	264	264
P25	-0.120	-0.017	-0.137	-0.060	-0.019	0.061	-0.198	-0.136	-0.023	0.024
	(0.168)	(0.167)	(0.202)	(0.220)	(0.160)	(0.178)	(0.192)	(0.229)	(0.167)	(0.169)
Obs	263	263	261	261	264	264	260	260	264	264
P50	-0.253	-0.202	-0.167	-0.095	-0.145	-0.095	-0.133	-0.013	-0.236	-0.128
	(0.157)	(0.166)	(0.184)	(0.167)	(0.159)	(0.179)	(0.185)	(0.193)	(0.169)	(0.160)
Obs	263	263	261	261	264	264	260	260	264	264
P75	**-0.442*****	**-0.404****	-0.247	-0.265	-0.109	-0.046	**-0.337****	**-0.319***	-0.144	-0.095
	(0.157)	(0.177)	(0.157)	(0.168)	(0.196)	(0.225)	(0.169)	(0.190)	(0.163)	(0.152)
Obs	263	263	261	261	264	264	260	260	264	264
P90	**-0.398*****	**-0.408****	-0.187	-0.198	**-0.374***	-0.302	**-0.359*****	**-0.353****	**-0.450*****	**-0.419****
	(0.152)	(0.168)	(0.123)	(0.133)	(0.215)	(0.224)	(0.122)	(0.146)	(0.149)	(0.162)
Obs	263	263	261	261	264	264	260	260	264	264
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each column presents estimates from separate UQR models. P_i represents the UQR result at ith percentile. The dependent variable in column (1)-(2) is the standardized first mid-term grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second mid-term, final exam, weighted average exam, and overall course grades. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Bootstrapped standard errors with 200 replications are reported in parentheses. $^{\#}$, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.21: Effects on Other course Grades, International status

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: International										
P10	0.056	0.083	**0.490*****	**0.448****	-0.024	-0.032	0.237	0.180	0.126	0.075
	(0.168)	(0.157)	(0.169)	(0.177)	(0.205)	(0.193)	(0.177)	(0.193)	(0.165)	(0.175)
Obs	370	370	368	368	371	371	367	367	371	371
P25	-0.032	0.006	0.126	0.084	0.113	0.143	0.128	0.142	0.120	0.164
	(0.156)	(0.150)	(0.181)	(0.173)	(0.130)	(0.140)	(0.156)	(0.153)	(0.119)	(0.121)
Obs	370	370	368	368	371	371	367	367	371	371
P50	-0.043	-0.046	0.052	0.107	-0.033	0.027	-0.006	0.047	-0.020	0.032
	(0.117)	(0.137)	(0.154)	(0.148)	(0.138)	(0.139)	(0.147)	(0.169)	(0.131)	(0.126)
Obs	370	370	368	368	371	371	367	367	371	371
P75	**-0.251***	**-0.250***	-0.172	-0.156	-0.024	0.003	-0.170	-0.145	-0.073	-0.040
	(0.139)	(0.139)	(0.146)	(0.147)	(0.137)	(0.143)	(0.156)	(0.136)	(0.145)	(0.147)
Obs	370	370	368	368	371	371	367	367	371	371
P90	**-0.232***	-0.200	**-0.277****	**-0.224***	0.060	0.114	-0.199	-0.128	-0.177	-0.119
	(0.140)	(0.155)	(0.117)	(0.124)	(0.168)	(0.143)	(0.135)	(0.136)	(0.133)	(0.133)
Obs	370	370	368	368	371	371	367	367	371	371
Panel B: Domestic										
P10	-0.098	-0.199	0.165	0.094	-0.015	-0.009	-0.185	-0.308	-0.180	-0.270
	(0.298)	(0.311)	(0.258)	(0.231)	(0.174)	(0.175)	(0.225)	(0.243)	(0.206)	(0.206)
Obs	243	243	239	239	247	247	235	235	247	247
P25	0.022	-0.046	-0.267	-0.279	-0.064	-0.079	0.028	-0.021	0.051	-0.004
	(0.184)	(0.174)	(0.225)	(0.234)	(0.213)	(0.223)	(0.222)	(0.262)	(0.192)	(0.206)
Obs	243	243	239	239	247	247	235	235	247	247
P50	-0.046	-0.105	-0.249	**-0.267***	-0.251	-0.240	-0.164	-0.158	-0.199	-0.200
	(0.162)	(0.164)	(0.165)	(0.158)	(0.216)	(0.215)	(0.169)	(0.185)	(0.163)	(0.169)
Obs	243	243	239	239	247	247	235	235	247	247
P75	**-0.259***	**-0.294***	-0.028	-0.022	-0.122	-0.131	**-0.238***	*-0.261#*	-0.238	-0.245
	(0.144)	(0.165)	(0.128)	(0.127)	(0.177)	(0.182)	(0.144)	(0.163)	(0.167)	(0.196)
Obs	243	243	239	239	247	247	235	235	247	247
P90	-0.043	-0.070	-0.001	0.011	0.017	0.008	-0.156	-0.178	-0.075	-0.110
	(0.178)	(0.175)	(0.121)	(0.127)	(0.171)	(0.197)	(0.154)	(0.183)	(0.173)	(0.188)
Obs	243	243	239	239	247	247	235	235	247	247
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each column presents estimates from separate UQR models. P_i represents the UQR result at ith percentile. The dependent variable in column (1)-(2) is the standardized first mid-term grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second mid-term, final exam, weighted average exam, and overall course grades. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Bootstrapped standard errors with 200 replications are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.22: Effects on Exam Grades, Degree

	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Business Degree										
P10	0.180	0.170	0.314	0.280	0.032	-0.047	0.266	0.234	0.086	-0.015
	(0.224)	(0.245)	(0.229)	(0.238)	(0.149)	(0.156)	(0.251)	(0.214)	(0.170)	(0.184)
Obs	276	276	271	271	277	277	270	270	277	277
P25	0.033	-0.002	0.064	-0.023	-0.003	-0.043	0.001	-0.064	0.089	0.073
	(0.148)	(0.164)	(0.251)	(0.244)	(0.162)	(0.173)	(0.150)	(0.170)	(0.144)	(0.135)
Obs	276	276	271	271	277	277	270	270	277	277
P50	-0.102	-0.182	0.013	-0.012	-0.199	**-0.268***	-0.134	-0.160	-0.126	-0.208
	(0.145)	(0.151)	(0.147)	(0.164)	(0.150)	(0.157)	(0.157)	(0.181)	(0.139)	(0.167)
Obs	276	276	271	271	277	277	270	270	277	277
P75	**-0.285****	**-0.319****	**-0.359*****	**-0.431*****	-0.062	-0.152	**-0.374****	**-0.457*****	-0.142	-0.232
	(0.129)	(0.130)	(0.131)	(0.131)	(0.164)	(0.169)	(0.145)	(0.138)	(0.147)	(0.154)
Obs	276	276	271	271	277	277	270	270	277	277
P90	**-0.286***	**-0.346****	**-0.291****	**-0.338****	-0.175	-0.186	**-0.211***	*-0.233#*	-0.162	-0.176
	(0.146)	(0.133)	(0.135)	(0.153)	(0.206)	(0.195)	(0.123)	(0.146)	(0.147)	(0.176)
Obs	276	276	271	271	277	277	270	270	277	277
Panel B: Non-business Degree										
P10	0.047	0.070	0.228	0.264	0.086	0.153	0.089	0.146	0.070	0.098
	(0.197)	(0.226)	(0.190)	(0.198)	(0.170)	(0.176)	(0.165)	(0.197)	(0.160)	(0.179)
Obs	337	337	336	336	341	341	332	332	341	341
P25	-0.075	-0.086	0.127	0.046	0.093	0.145	0.162	0.176	0.088	0.119
	(0.157)	(0.174)	(0.196)	(0.216)	(0.155)	(0.159)	(0.164)	(0.187)	(0.140)	(0.159)
Obs	337	337	336	336	341	341	332	332	341	341
P50	-0.121	-0.185	-0.047	-0.080	-0.009	-0.068	-0.038	-0.085	0.082	0.030
	(0.140)	(0.149)	(0.182)	(0.159)	(0.207)	(0.173)	(0.181)	(0.172)	(0.162)	(0.159)
Obs	337	337	336	336	341	341	332	332	341	341
P75	-0.156	-0.160	-0.131	-0.112	0.076	0.010	-0.059	-0.054	0.009	-0.063
	(0.163)	(0.162)	(0.146)	(0.157)	(0.162)	(0.176)	(0.178)	(0.169)	(0.149)	(0.173)
Obs	337	337	336	336	341	341	332	332	341	341
P90	-0.220	-0.151	0.041	0.117	-0.011	-0.067	-0.101	-0.079	-0.084	-0.122
	(0.148)	(0.162)	(0.116)	(0.124)	(0.148)	(0.180)	(0.141)	(0.167)	(0.141)	(0.166)
Obs	337	337	336	336	341	341	332	332	341	341
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each column presents estimates from separate UQR models. P_i represents the UQR result at ith percentile. The dependent variable in column (1)-(2) is the standardized first mid-term grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second mid-term, final exam, weighted average exam, and overall course grades. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Bootstrapped standard errors with 200 replications are reported in parentheses. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.23: Placebo Treatment Effects

Panel A: Number of Assignment Points	1,000+ points		1,300+ points		1,600+ points		1,800+ points		2,000+ points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Week 7	0.003	-0.002	-0.000	0.001	0.003	0.004	0.000	0.000	0.000	0.000
	(0.014)	(0.017)	(0.005)	(0.005)	(0.003)	(0.004)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	591	591	591	591	591	591	591	591	591	591
Week 11	0.016	0.012	-0.003	-0.008	0.007	0.005	0.007	0.007	0.003	0.004
	(0.026)	(0.032)	(0.012)	(0.013)	(0.005)	(0.005)	(0.007)	(0.008)	(0.003)	(0.004)
Obs	591	591	591	591	591	591	591	591	591	591
Week 13	-0.008	-0.009	-0.014	-0.023	0.003	-0.001	-0.010	-0.013	-0.007	-0.009
	(0.031)	(0.033)	(0.016)	(0.016)	(0.013)	(0.013)	(0.011)	(0.012)	(0.008)	(0.009)
Obs	591	591	591	591	591	591	591	591	591	591
Panel B: Other Course Grades	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall Course	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Grade	-0.023	0.041	-0.035	0.031	0.008	0.057	-0.024	0.046	-0.010	0.050
	(0.070)	(0.097)	(0.075)	(0.074)	(0.072)	(0.086)	(0.069)	(0.090)	(0.065)	(0.082)
Obs	588	588	580	580	593	593	575	575	593	593
Panel C: Other Course Grades, Non-linearities										
P10	0.111	0.271	**-0.276****	-0.218	0.080	0.199	-0.030	0.077	0.020	0.127
	(0.168)	(0.213)	(0.126)	(0.134)	(0.124)	(0.134)	(0.157)	(0.155)	(0.145)	(0.143)
Obs	588	588	580	580	593	593	575	575	593	593
P25	0.066	0.135	-0.191	-0.114	0.014	0.056	-0.032	0.065	0.033	0.105
	(0.120)	(0.121)	(0.136)	(0.144)	(0.123)	(0.130)	(0.117)	(0.120)	(0.107)	(0.114)
Obs	588	588	580	580	593	593	575	575	593	593
P50	-0.019	0.033	0.082	0.161	-0.014	0.010	0.035	0.084	0.005	0.022
	(0.119)	(0.111)	(0.115)	(0.119)	(0.123)	(0.116)	(0.118)	(0.129)	(0.119)	(0.111)
Obs	588	588	580	580	593	593	575	575	593	593
P75	-0.096	-0.043	0.060	0.126	-0.023	0.003	-0.031	0.003	-0.048	-0.031
	(0.096)	(0.110)	(0.100)	(0.095)	(0.112)	(0.109)	(0.099)	(0.102)	(0.103)	(0.102)
Obs	588	588	580	580	593	593	575	575	593	593
P90	0.011	0.023	-0.029	0.017	-0.047	-0.000	-0.007	0.031	0.012	0.055
	(0.103)	(0.109)	(0.092)	(0.085)	(0.114)	(0.130)	(0.099)	(0.097)	(0.110)	(0.098)
Obs	588	588	580	580	593	593	575	575	593	593
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: ***Panel A:*** Each row presents estimates from separate OLS models. The dependent variable in column (1)-(2) is a binary indicator equal to one if a student's online assignment points by a certain week are above 1,000 and zero otherwise. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for points above 1,300, 1,600, 1,800, and 2,000. ***Panel B-C:*** Each column presents estimates from separate OLS (Panel B) and UQR (Panel C) models. P_i represents the UQR result at ith percentile. The dependent variable in column (1)-(2) is the standardized first mid-term grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second mid-term, final exam, weighted average exam, and overall course grades. Data is generated as follows: First, I create a random variable; next I use it to sort the dataset and then I assign observations to placebo treatment groups. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses for Panel A and Panel B. Bootstrapped standard errors with 200 replications are reported in parentheses in Panel C. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table B.24: Treatment Effects, Ability and Tutorial Fixed Effects

Panel A: Number of Assignment Points										
	1,000+ points		1,300+ points		1,600+ points		1,800+ points		2,000+ points	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Week 7	**0.027***	**0.026***	0.005	0.007	*0.006*$^{\#}$	*0.007*$^{\#}$	0.000	0.000	0.000	0.000
	(0.015)	(0.014)	(0.006)	(0.007)	(0.004)	(0.005)	(0.000)	(0.000)	(0.000)	(0.000)
Obs	615	615	615	615	615	615	615	615	615	615
Week 11	0.027	0.028	0.017	0.018	*0.022*$^{\#}$	**0.024***	**0.009***	**0.013****	0.004	0.006
	(0.034)	(0.032)	(0.016)	(0.017)	(0.013)	(0.013)	(0.005)	(0.006)	(0.003)	(0.004)
Obs	615	615	615	615	615	615	615	615	615	615
Week 13	**0.069***	**0.074***	0.026	0.022	0.023	0.020	**0.013****	**0.015****	0.006	0.007
	(0.035)	(0.037)	(0.024)	(0.029)	(0.018)	(0.019)	(0.005)	(0.007)	(0.005)	(0.006)
Obs	615	615	615	615	615	615	615	615	615	615
Panel B: Other Course Grades										
	Week 6 Exam		Week 10 Exam		Final Exam		Weighted Av. Exam		Overall Course	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Grade	**-0.121****	**-0.096****	0.014	-0.002	-0.022	-0.026	-0.043	-0.042	-0.059	-0.055
	(0.047)	(0.044)	(0.082)	(0.087)	(0.080)	(0.074)	(0.064)	(0.059)	(0.072)	(0.066)
Obs	613	613	607	607	618	618	602	602	618	618
Panel C: Other Course Grades, Non-linearities										
P10	0.022	0.032	**0.252***	**0.270****	0.098	0.109	0.163	0.169	0.091	0.091
	(0.164)	(0.158)	(0.143)	(0.136)	(0.122)	(0.118)	(0.140)	(0.135)	(0.138)	(0.126)
Obs	613	613	607	607	618	618	602	602	618	618
P25	-0.108	-0.043	0.132	0.128	0.024	0.014	-0.014	0.009	0.046	0.056
	(0.133)	(0.122)	(0.151)	(0.138)	(0.119)	(0.115)	(0.127)	(0.133)	(0.114)	(0.105)
Obs	613	613	607	607	618	618	602	602	618	618
P50	-0.085	-0.063	-0.065	-0.094	-0.103	-0.114	-0.032	-0.030	-0.083	-0.079
	(0.105)	(0.100)	(0.120)	(0.114)	(0.132)	(0.114)	(0.122)	(0.121)	(0.121)	(0.111)
Obs	613	613	607	607	618	618	602	602	618	618
P75	**-0.211****	**-0.228****	**-0.198****	**-0.255*****	0.025	0.015	-0.153$^{\#}$	**-0.191***	-0.030	-0.061
	(0.098)	(0.104)	(0.094)	(0.098)	(0.127)	(0.123)	(0.104)	(0.110)	(0.104)	(0.103)
Obs	613	613	607	607	618	618	602	602	618	618
P90	**-0.248*****	**-0.244****	-0.079	-0.098	-0.094	-0.099	-0.131	*-0.144*$^{\#}$	-0.111	-0.108
	(0.093)	(0.101)	(0.085)	(0.087)	(0.116)	(0.116)	(0.104)	(0.095)	(0.107)	(0.106)
Obs	613	613	607	607	618	618	602	602	618	618
Student Ch.	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓
Tutorial FE	✓	X	✓	X	✓	X	✓	X	✓	X
Prior Ability	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: ***Panel A:*** Each row presents estimates from separate OLS models. The dependent variable in column (1)-(2) is a binary indicator equal to one if a student's online assignment points by a certain week are above 1,000 and zero otherwise. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for points above 1,300, 1,600, 1,800, and 2,000. ***Panel B-C:*** Each column presents estimates from separate OLS (Panel B) and UQR (Panel C) models. P_i in Panel C represents the UQR result at ith percentile. The dependent variable in column (1)-(2) is the standardized first mid-term grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second mid-term, final exam, weighted average exam, and overall course grades. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Robust standard errors clustered by tutor are reported in parentheses for Panel A and Panel B . Bootstrapped standard errors with 200 replications are reported in parentheses in Panel C. $^{\#}$, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table B.25: Effects on Other Course Grades, CQR

	Week 6 Exam				Week 10 Exam				Final Exam				Weighted Av. Exam				Overall			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)
P10	0.000	-0.028	0.000	-0.028	0.314**	0.420**	0.314***	0.420***	0.269	0.058	0.269	0.058	0.132	0.114	0.132***	0.114***	0.079	-0.020	0.079***	-0.020
	(0.144)	(0.110)	(0.034)	(.)	(0.155)	(0.180)	(0.007)	(0.046)	(0.181)	(0.159)	(.)	(.)	(0.166)	(0.172)	(0.019)	(0.024)	(0.182)	(0.179)	(0.030)	(0.050)
Obs	613	613	613	613	607	607	607	607	618	618	618	618	602	602	602	602	618	618	618	618
P25	0.000	-0.124	0.000	-0.124**	0.235*	0.054	0.235***	0.054	0.000	0.013	0.000	0.013	0.049	0.160	0.049***	0.160***	0.079	0.110	0.079***	0.110
	(0.081)	(0.095)	(0.030)	(0.061)	(0.127)	(0.157)	(0.061)	(.)	(0.103)	(0.133)	(0.068)	(0.044)	(0.097)	(0.125)	(0.001)	(0.021)	(0.094)	(0.095)	(0.014)	(.)
Obs	613	613	613	613	607	607	607	607	618	618	618	618	602	602	602	602	618	618	618	618
P50	-0.083	-0.041	-0.083***	-0.041***	-0.078	-0.078	-0.078***	-0.078***	-0.135	-0.135	-0.135***	-0.135**	-0.042	-0.010	-0.042***	-0.010***	-0.118**	-0.010	-0.118***	-0.010
	(0.069)	(0.071)	(0.013)	(0.001)	(0.097)	(0.080)	(0.012)	(0.007)	(0.091)	(0.093)	(0.027)	(0.066)	(0.086)	(0.114)	(0.001)	(0.001)	(0.059)	(0.051)	(0.018)	(0.009)
Obs	613	613	613	613	607	607	607	607	618	618	618	618	602	602	602	602	618	618	618	618
P75	-0.249***	-0.249**	-0.249***	-0.249***	-0.157	-0.071	-0.157***	-0.071***	0.000	0.000	0.000	0.000	-0.201***	-0.197**	-0.201***	-0.197***	-0.098	-0.105	-0.098***	-0.105***
	(0.069)	(0.121)	(0.008)	(0.024)	(0.108)	(0.121)	(0.040)	(0.002)	(0.108)	(0.105)	(0.017)	(0.012)	(0.070)	(0.092)	(0.000)	(0.017)	(0.098)	(0.082)	(0.022)	(0.010)
Obs	613	613	613	613	607	607	607	607	618	618	618	618	602	602	602	602	618	618	618	618
P90	-0.249**	-0.249***	-0.249***	-0.249***	-0.078	-0.169*	-0.078***	-0.169	0.000	-0.124	0.000	-0.124***	-0.146	-0.292***	-0.146***	-0.292***	-0.197**	-0.197**	-0.197***	-0.197***
	(0.104)	(0.085)	(0.040)	(0.025)	(0.078)	(0.089)	(0.002)	(.)	(0.065)	(0.133)	(0.043)	(0.029)	(0.102)	(0.087)	(0.020)	(0.024)	(0.097)	(0.088)	(0.001)	(0.025)
Obs	613	613	613	613	607	607	607	607	618	618	618	618	602	602	602	602	618	618	618	618
Student Ch.	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor clustering	✓	✓	X	X	✓	✓	X	X	✓	✓	X	X	✓	✓	X	X	✓	✓	X	X
Tutorial clustering	X	X	✓	✓	X	X	✓	✓	X	X	✓	✓	X	X	✓	✓	X	X	✓	✓

Notes: Each column presents estimates from separate Conditional Quantile Regression (CQR) models. P_i represents the CQR result at ith percentile. The dependent variable in column (1)-(2) is the standardized first mid-term grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second mid-term, final exam, weighted average exam, and overall course grades. Odd numbered columns only include treatment indicators; even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled full-time, and a dummy denoting whether a student is enrolled in an Economics degree) and tutor fixed effects. Bootstrapped standard errors with 200 replications are reported in parentheses in. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table B.26: Full Assignment Completion

	Sample Size (1)	100% Completions (#) (2)	100% Completion Rate (%) (3)
Social Goals	288	280	97.222
Goals	330	324	98.182
Control	272	270	99.265

Notes: Each column shows the number of students belonging to each treatment group, the number of students achieving 100% in the online assignment by the end of the semester, and the corresponding 100% assignment completion rate.

Chapter 3

(High-performing) Women Hate Ranks and Men Like Leagues: Evidence From a Field Experiment[1]

3.1 Introduction

Gender differences in performances is a topic of general social concern. Despite worldwide efforts in promoting equal opportunities to people of all genders, gender differences prevail in the labor market. The International Labor Organisation (ILO, 2019) reports a 19% mean hourly gender pay gap worldwide last year. In a 2017 survey covering 142 countries and more than 100,000 adults, ILO joint with Gallup finds that 70 percent of women prefer to work at paid jobs instead of staying at home (ILO-Gallup, 2017). However, only 45.3% of women are found to have a job in 2018 (Beghini et al., 2019). The contrast between the proportion of women reporting a preference to work and the actual employment rate demonstrates that the gender gap is still a serious problem. Women are even less represented in leadership roles

[1]This chapter was jointly developed with Isabella Dobrescu, Gigi Foster, and Alberto Motta. All four of us were involved in all stages from formulating the research questions, through designing the experiment, to writing up the draft. I played a major role in analysing the data and writing up the draft.

where merely 27.1% of leaders or managers are female and this ratio has barely changed over the past 27 years (Beghini et al., 2019). Some literature has found that maintaining a gender diversity in top management teams relates to better financial performance or helps to resolve stakeholder conflicts (Opstrup and Villadsen, 2015; Adams et al., 2011).[2]

Education is one of the key channels through which human capital accumulates and manifests itself (Bertocchi and Bozzano, 2019). This makes education a powerful way to prevent and mitigate gender gaps in the labor market. Despite its existence, gender inequality in education has been alleviated quantitatively around the world (World Bank, 2019). According to the World Bank, females had a net[3] secondary enrolment rate of 66% in 2017, identical to that of males. As for tertiary education, the gross enrolment ratio of females is 40%, slightly higher than the 36% for males (World Bank, 2019). While the quantitative mitigation of the gender inequality gap is noted, the qualitative gender gap is still of great concern. For example, only fewer than one-third of the female student population in higher education choose STEM fields of study, and merely a quarter of students in tertiary engineering and related fields are female.[4] Lacking women in the related education pipeline, technology or engineering entities can suffer from a lack of women's perspective which hinders their long-run development (Goy et al., 2018). Moreover, researches have shown that females' performance worsens as competition becomes more intense (Niederle and Vesterlund, 2011). When competition is mild, females are as likely as similar-ability males to be admitted to selective universities, yet females are substantially less likely to be admitted if competition in the selection process becomes more intense (Jurajda and Münich, 2011). The persistent gender gap in the global labor market and in education, as well as the commonly observed difference in each gender's attitudes towards competition, motivates this chapter's focus on the mechanism underpinning the gender differences found in Chapter 2.

In Chapter 2, I find that the relative performance information provided in the intervention implemented in a large first-year university course mostly motivated men, but not women, to overachieve. The reason for this could be that relative

[2]In a more recent paper, Adams (2016) notes that more evidence needs to be collected to better understand the benefits of diversifying top management teams.

[3]The World Bank calculates net enrolment ratio as one that only includes children of official school age, excluding over- or under-age children.

[4]Global Education Monitoring Report Team [541] (2020)

performance information significantly affects only men, not women. However, while there was no overachievement effect for women, women's final marks in the course seems to have been negatively affected, while men's exhibit a positive effect. This pattern by gender calls for deeper investigation.

Using the same experimental data as used in the previous chapter, in this chapter I investigate the effect of relative performance information, information about leagues and milestones, on the academic performance of men and women respectively. As it turns out, females and males are strongly different in their response to the interventions. On the one hand, females are significantly negatively affected (0.19 SDs lower first midterm grade) by the relative performance information, with high-performing females being the most negatively affected (0.28 SDs lower overall course grade) group. High-performing females' exam grades and course grades are significantly lower than their female counterparts who do not have access to relative performance information. On the other hand, males are significantly positively affected (0.26 SDs higher overall course grade) by the league information and the increase appears homogeneous along the grade distributions. I explore additional measures recorded by the software logs to further untangle the underlying mechanisms for each gender. High-performing females who have access to relative performance information report feeling less able to overcome difficulties, compared to their female counterparts without the access. Males shown this information exhibit an increase in effort in the online assignment[5], which may in turn be responsible for the uptick in their marks.

These results are robust to all the alternative specifications utilized in the previous chapter: i) controlling for tutorial fixed effects instead of tutor fixed effects; ii) clustering standard errors by tutorials instead of tutors; and iii) controlling for imputed ability. Additionally, given the difference between procrastinators and non-procrastinators that emerge in the previous chapter, I construct an alternative procrastination variable (i.e., procrastination level), treating procrastination as an continuous variable (a detailed explanation is in Section 3.3). The variable remains exogenous to the treatment interventions, in the sense that the procrastination level is captured before a student is treated, since a student cannot be treated until she logs into the online software. I find that procrastination levels strongly negatively

[5]'Online assignment' in this chapter refers to the same educational software, as mentioned in Chapter 2, that provided students with an online database of questions linked to the course textbook.

correlate with students' course grades. In other words, the later a student starts accessing the online assignment, the worse his exam grades are, which is unsurprising. However, all results on the impact of the treatment intervention remain robust.

I also investigate heterogeneous effects within each gender across several dimensions of longstanding research interests, namely country of birth, field of study, and ability. Additionally, I investigate whether effect heterogeneity within each gender is related to procrastination level. For country of birth, I compare Chinese students (47% of the whole sample) with native Australian students (28% of the sample), the two largest ethnic groups[6] in this course and in other Australian universities in general (Foster, 2012). I find that Australian males who have access to relative performance information achieve significantly higher second midterm marks (0.21 SDs), whereas Chinese males who have access to relative performance information achieve significantly lower marks (0.34 SDs) in their final exam, compared to their same-ethnicity peers who are exposed only to the milestone-referenced league information. Turning to major, when access to relative performance information is layered with access to leagues information, business-majored males earn significantly higher (0.30 SDs) average exam grades compared to their same-major male counterparts in the control condition. Interestingly, no statistically significant effect on overall course grades is found for females in any subgroup defined by country of birth or field of study. After controlling for ability, females who have access to both the relative performance information and the league information perform worse (0.30 SDs) in the first midterm than females without access to either type of information. Males with access to the league information, regardless of availability of the relative performance information, perform substantially better (0.67-0.84 SDs, equivalent to an increase of around 10 marks out of 100) in four out of the five course outcomes modelled than males in the control condition, after ability is controlled. Interestingly, if procrastination level is controlled instead, only females who have access to both the relative performance information and the league information suffer a negative impact on their performance compared to their female counterparts in the other conditions.

In what follows, I discuss these gender-specific results in greater detail with statistical evidence and explore the underlying mechanisms for the main results.

[6]They sum up to 75% of the sample. The third largest country of birth in the sample is Indonesia with merely 40 students.

Finally, I compare the findings in this chapter with the findings in the previous chapter and discuss how the gender-specific results resonate with the overall results.

3.2 Literature Review

Bettering understanding of what motivates gender differences is a primary step towards eliminating gender gap. Researches have found that males and females select differently into competitive environments and perform differently when exposed to competitions. In their seminal paper, Niederle and Vesterlund (2007) explore the female and the male's willingness to opt in competitions in a laboratory experiment. Controlling for attitudes and beliefs toward competitions, they find that low-performing males over compete while high-performing females under compete. Moreover, same-ability men select into competition twice that of women, which renders sub-optimal experimental earnings for both genders. These sub-optimal decisions are mainly driven by subjects' confidence levels and attitudes toward competitions: males tend to be overconfident while females prefer less competition. A battery of experiments with slight modifications have followed their design and find similar results (Wozniak et al., 2014; Balafoutas and Sutter, 2010). Niederle and Vesterlund (2011) provide a comprehensive literature review regarding gender differences in competition, both in terms of choice of competitive versus non-competitive environment and in terms of performance under competitive environment. By and large, high-ability females are found to more frequently opt out of competitions, while males do not. If subjects are confronted with a competitive environment without a choice, males and females still perform differently within the same environment. Gneezy et al. (2003) find, in a laboratory experiment, that as the competitiveness level of an environment increases, males increase their performance significantly, while female do not show any significant changes. These different responses to competitiveness enlarge the gender gap in performance, even though all subjects start at the same point. Interestingly, Wozniak et al. (2014) find that relative performance information eliminates differences in competitive choices between female and male, as more high-performing females and less low-performing males choose competitive compensation schemes. Given that relative performance information has been found to eliminate gender differences in competitive choices, it is quite possible that relative performance information also affect each gender's performance differently given a competitive environment, a research gap this chapter sheds light on.

Compared to gender differences under competition (or more specifically, with relative performance information), goal-setting is an under-investigated area in economics (Smithers, 2016), less literature looks into differential gender responses to goal-setting. One major difference between relative performance information and goal-setting is that goal-setting eliminates the peer competition component; instead, one competes against himself along a list of goals. All goals in this experiment are non-binding, in the sense that goals are detached from monetary incentives or coercion (Smithers, 2016). Smithers (2015) investigates goal-setting in a laboratory experiment and find that non-binding goals significantly increase the speed and accuracy of efforts in a timed addition task. More importantly, Smithers (2015) finds persistent evidence that men respond more strongly than women to goal-setting. In a large field experiment involving several thousands university students, Clark et al. (2016) finds that task-based goals give rise to better academic outcomes, while performance-based goals are less effective. Interestingly, given task-based goals, they also find that males are more responsive than females, in terms of both effort and academic performance. Using marathon data, Burdina et al. (2017) finds that relatively attainable goals improve performance while difficult goals hinder performance. Similar patterns are also present in a study by Fan and Gómez-Miñambres (2019) on workers' achievement in response to non-binding goals set by managers and in Harding and Hsiaw (2014)'s study on energy conservation behaviour of residents. In other words, the effect of goals pertains to their attainability. Besides the limited and unsettling empirical evidence, several existing economic theories provide theoretical grounds why goal-setting can be effective in motivating better performance. For example, an individual may have an intrinsic motivation towards achieving higher goals (Benabou and Tirole, 2006; Gómez-Miñambres, 2012). Alternatively, goals, especially when exogenously set by course authorities, could provide a reference point upon which individuals benchmark their self-expectations (Heath et al., 1999). Another possible channel is that goals attenuate self-control bias, which in turn translate into better performance (Hsiaw, 2013). However, most of the theories remain untested in the field. To the best of my knowledge, my experiment provide the first clean field-experimental evidence capturing differential gender responses to goal-setting, which is further backed up by corresponding mechanisms.

In addition to the overall treatment effect of either relative performance information or leagues information on each gender, several other dimensions are worth exploring to shed light on broader areas of research interests. One interesting aspect is whether same-gender students with different country of birth respond differently to the treatments, especially in this specific experimental context in an Australian university. Based on administrative data from business schools of two Australian universities, Foster (2012) finds strong and persistent evidence that international students and non-English speaking students perform worse than other students in undergraduate classes. Foster (2012) also notes that large scale quantitative evidence are quite limited regarding international students' academic performance in Australian context.[7] The second aspect of interest is whether field of study also plays a role in mediating treatment effects on same-gender students. As is captured in Chapter 2, when exposed to the relative performance information. Business students appear more responsive than STEM or Humanities students. It would be interesting to see whether this difference pertains to each gender.

Other than country of birth and field of study, one's previous ability and procrastination level are also two challenging areas of enduring research effort. Prior researches have found that students' prior academic performance is a strong predictor of their later academic performance, as well as job market performance (Wise, 1975; Caudill and Gropper, 1991; French et al., 2015). It would be interesting to see whether my interventions interact with prior ability. As for procrastination, it occurs when a person postpones a pending task knowing that the task has to be done, which results in "counterproductive and needless delay" (Bisin and Hyndman, 2020; Schraw et al., 2007). Economists describe procrastination as a result of present-bias where agents delay tasks that they themselves would rather do earlier (O'Donoghue and Rabin, 1999). Solomon et al. (2007) document that almost half of university students consider themselves heavy procrastinators. In a more recent research, around 90 percent of university students are found to regularly procrastinate in completing their academic tasks (Steel, 2007). As social network software and mobile phones get popularized, students are more easily distracted by various entertainments and postpone academic tasks. Economic theories predict that self-set goals, albeit non-binding, can mitigate time inconsistency and consequently self-control problems (Hsiaw, 2013). As the semester-long intervention window nat-

[7]International Student Data 2019 shows that enrolments of international students in Australian Higher Education programs maintain a steady growth rate of well above 10% from 2016 to 2019, according to Australian Education International (AEI, 2019). A closer investigation shows that international students from mainland China on average increases by 14.3% over the past four years, comprising more than 35% of the international students group. Chinese international students remain the largest group of international students till 2019 (AEI, 2019).

urally form an objective measure of students' procrastination tendency, I use this measure to study whether procrastination predicts academic outcomes and whether it interacts with the interventions for each gender.

3.3 Data and Empirical Analysis

Data Description. Table 3.1 presents a description of the main variables of interest, in addition to the variables described in Chapter 2. A list of additional effort indicators is utilised here, proxying online assignment engagement. Same to Chapter 2, online assignment here specifically refers to the educational software in which the interventions were implemented. An average student accessed the online assignment around two times per week, with a standard deviation of about 6 times.[8] As explained before, after logging into the online assignment system, a student may attempt several types of questions (e.g., graph, maths, short answer, multiple-choice questions), and different questions bare different points. The system tracks the number of all correct and incorrect submission attempts over a week. The number of submission attempts captures a student's attempt in the online assignment. An average student submitted around 23 questions per week, 14 of them being correct ones and 9 being incorrect ones. The week in which a student first started to accumulate non-zero assignment points is used to construct a measure of his procrastination level. For example, if a student started accruing points in Week 3 (the week in which the treatment intervention was deployed), then his procrastination level was 0; if he started earning points in Week 4, then his procrastination level was 1; and so on. As can be seen, the average student's procrastination level was close to 4, meaning he started accessing the online assignment in Week 7. Interestingly, extreme procrastinators started only in the very last week of the semester.

Course Effects. To evaluate the gender-specific impact of the interventions on students' academic performance, an empirical model is estimated which allows computation of the gender-specific treatment effect and enables a comparison of the treatment effect on men and that on women. Same as in Chapter 2, both relative performance information and the league information are available in a Social Goals condition, whereas only the league information is available in a Goals condition. In a

[8]Impressively, the most active students accessed the assignment platform up to 159 times per week, equivalent to 23 times per day and 7 days a week.

Table 3.1: Descriptive Statistics

Variable	Mean	SD	Min.	Max.	N
Number of Assignment Accesses	1.983	5.736	0	159	12460[a]
Number of All Assignment Submissions	22.982	60.745	0	879	12404
Number of Correct Assignment Submissions	13.963	36.203	0	552	12404
Number of Incorrect Assignment Submissions	9.019	25.824	0	437	12404
Procrastination Level (Starting Week)	3.926	3.801	0	13	886

Notes: All effort indicators are measured weekly, except for *Procrastination Level (Starting Week)*. The total observation count of 12404 is the product of 886 observations per week times 14 intervention weeks. For a more detailed description of all the control variables, please refer to Table B.1 in Chapter 2.
a: Four students actively accessed and attempted the online assignment, but their effort indicators were not successfully retrieved from the system due to technical problems.

control condition, neither piece of information is available. The baseline estimating equation is presented below.

$$Y_{i,t} = \alpha + \beta_1 \text{Treatment}_i + \beta_2 \text{Male}_i + \beta_3 \text{Treatment}_i \times \text{Male}_i + \gamma X_i + TutorFE_t + u_{i,t} \quad (3.1)$$

where $Y_{i,t}$ represents different course grades for student i in tutorial t. $Treatment_i$ equals (i) one if a student is in the Social Goals treatment and zero if in the Goals treatment; or (ii) one if a student is in either treatment and zero if in the Control condition. $Male_i$ equals one for males and zero for females. X_i represents student characteristics such as age, a dummy for full-time enrolment mode, a dummy for undertaking an Economics degree, and dummies for students' ethnicity. $TutorFE_{i,t}$ stands for tutor fixed effects, which enable the partialling-out of any time-invariant tutor-specific disturbances. Robust standard errors are clustered by tutor. In the above specification, female students in the control condition that is being compared to the treatment indicated by the treatment dummy form the baseline group. β_1 captures the treatment effect on female students, while $\beta_1+\beta_3$ captures the treatment effect on male students. β_2 captures systematic gender differences and β_3 captures the component of the treatment effect that is male-specific.

Mechanisms. To untangle the potential mechanisms driving the academic performance results for each gender, several variables are exploited. First, the survey variable on students' reported ability to overcome difficulties is used as a proxy of students' perceived stress level.[9] Second, the number of times a student logs in to the online assignment, as well as the number of correct, incorrect, and total questions submitted are used to explain a student's quantity of effort. Recall from Chapter 2 that 98.2% of the students achieved full grades in the online assignment. These effort indicators thus allow investigation of whether the un-incentivized effort students put into the online assignment affects their (more standard) academic performance.

3.4 Results

In this section I investigate whether females and males perform differently in course exam grades, as well as in online assignment progression. Pairwise treatment comparisons are presented in each subsection, where, in line with Chapter 2, results comparing students in the Social Goals treatment to those in the Goals treatment are discussed first, followed by the results of comparing Social Goals to Control, and Goals to Control. In each subsection, analyses of the treatment effect on exam grades is the first focus, and analyses of the treatment effect on online assignment progression is the second focus, after which the potential interrelations between exam grades and online assignment progression are also discussed. Each component of the discussion starts with analyses of treatment effects on females or males respectively and ends with comparisons between females and males.

3.4.1 The Effect of Providing Relative Performance Feedback (Social Goals vs. Goals)

Comparing the outcomes of students in the Social Goals treatment to those of students in the Goals treatment isolates the effect of providing students with milestone-referenced performance information relative to their peers (rather than merely in absolute terms). Panel A of Table C.1 presents the effect of providing relative performance information (compared to providing only absolute milestone-referenced performance information) on three exam grades (the first midterm exam, the second midterm exam, and the final exam), their weighted average, and the overall course grade. The econometric model in the previous section allows direct

[9] Refer to Section 2.4.1 in Chapter 2 for details about the survey questions.

interpretations of coefficients: 1) coefficients in the *Treatment (β_1)* row capture the treatment effect on females; 2) coefficients in the *Male (β_2)* row capture systematic differences in outcomes between men and women, regardless of treatment; and 3) coefficients in the *Treatment* $\times$ *Male (β_3)* row capture any additional male-specific treatment effect. To capture the total treatment effect on males, a linear combination of $\beta_1 + \beta_3$ is constructed after each regression, with its standard error reported in brackets. As can be seen from columns (1) and (2), women in the Social Goals treatment perform 0.19 SDs worse in their first midterm than women in the Goals condition. This grade drop is statistically significant and robust to controlling for a full set of student characteristics and tutor fixed effects. By contrast, no statistically significant treatment effect on men *per se* is captured in row *Treatment Effect on Male*, although men in the Social Goals condition earn significantly higher final exam marks than women in the Goals condition, as indicated by β_3s in columns (5) and (6). The lack of treatment effect on men is due to the negative, albeit non-significant, overall treatment effect cancelling out with the positive gender-specific treatment effect. Notably, men are estimated to perform 0.15 SDs better than women in the final exam even with access only to absolute, but not relative, milestone-referenced performance information.

Panel A of Table C.2 presents the effect of access to relative performance information on online assignment progression in Week 7 (Panel A.1), Week 11 (Panel A.2), and Week 13 (Panel A.3). Specifically, online assignment progression over two milestones, corresponding to total accumulative points of 1,000 and 1,300, is explored.[10] By and large, access to relative performance information does not bring about significant changes in online assignment progression for either females or males, despite the fact that the online assignment is where the intervention takes place. However, the intervention effect is strong enough to affect students' exam grades, the more standard academic outcomes. Overall, access to relative performance information worsens women's first midterm exam grades by 0.19 SDs, compared to their female counterparts with access only to milestone-referenced league information, whereas it boosts men's final exam grades by 0.17 SDs against treated women.

[10]Regression results of online assignment progression over all five milestones are also explored and available upon request from the author.

3.4.2 The Effect of Setting Social Goals (Social Goals vs. Control)

As explained previously, the Social Goals treatment displays both a student's real-time league status and his peers' performance information, in addition to the student's own cumulative points information which is the only available information in the control condition. Panel B of Table C.1 presents the treatment effect of Social Goals on all five course grades, defined as the effect of access to milestone-referenced relative performance information compared to access only to non-milestone-referenced absolute information about one's own performance. Looking at the *Treatment* row and the *Treatment Effect on Male ($\beta_1 + \beta_3$)* row, neither women nor men's exam performance seems to be affected by the treatment, as opposed to their same-gender counterparts. Similar as before, men in the control condition already perform 0.26 SDs better than women in the final exam, in the absence of Social Goals. Working from the results in the *Treatment $\times$ Male (β_3)* row, where the differential treatment effect between men and women is captured, a male student is estimated to perform 0.10 SDs better in his second midterm exam and 0.22 SDs better in his final exam than a treated female. These effects further robustly increase his weighted average exam grade by 0.15 SDs, which drives up his overall course grade in a similar magnitude of 0.14 SDs (although not robust to controlling full covariates).

Panel B of Table C.2 presents the effect of the Social Goals treatment on online assignment progression. Specifically, columns (1) and (2) of Panel B.1 show that men in the Social Goals treatment are 5% more likely to surpass the 1,000 milestone than their male counterparts in the control condition in Week 7. They are also around 2% more likely to surpass the next milestone which corresponds to accumulating 30% more online points in the baseline regression without controlling for other covariates. No such effect is captured for females. As a result, one would expect that the statistically significant treatment effects on males' overachievement in the online assignment come from an additional performance boost for treated males against treated females. This is confirmed in the *Treatment $\times$ Male (β_3)* row. A treated man is 3% more likely to surpass the 1,000 milestone and 2% more likely to surpass the 1,300 milestone than a treated woman. Interestingly, the results in columns (1) and (2) across the whole of Panel B show that men do not differ systematically from women in the absence of the treatment in Week 7, although men seem less likely than women to surpass the 1,000 milestone in Week 13. Panel B.2 and Panel B.3 show that the progression gaps disappear in later weeks of the semester, likely to due to the procrastinating students catching up. However, that by no means indicates the original overachieving males stop progressing forward in the online assignment. In fact, they keep progressing beyond and above the call of duty. In Week 11, the treated males are 5% more likely to surpass the 1,600 milestone and 2% more likely to accumulate more than 1,800 points, compared to treated females (Results are not presented here for brevity).

3.4.3 The Effect of Setting Goals (Goals vs. Control)

The Goals treatment displays a student's real-time league status information, together with the absolute cumulative points information which is also available in the control condition. Panel C of Table C.1 presents the effect of access to the milestone-referenced own performance information - compared to non-milestone-referenced absolute performance information - for all five course grades. The *Treatment (β_1)* row shows no treatment effect on females against controlled females, similar to the other two pairwise treatment comparisons results. In stark contrast, treated males achieve strongly higher exam grades than their male counterparts in the control condition. Specifically, the *Treatment Effect on Male ($\beta_1 + \beta_3$)* row shows that treated males on average perform around 0.25 SDs better in their first midterm exam and 0.33 SDs better in their final exam. These increases further improve the treated males' weighted average exam grade by 0.26 SDs, which drives up their overall course grade in the same magnitude of 0.20 SDs, or a larger magnitude of 0.26 SDs in the alternative specification controlling for a full set of student characteristics and tutor fixed effects. The *Male (β_2)* row shows that males again have higher final exam grades than females without the treatment, coinciding with observations from Panel A and Panel B. Turning to the *Treatment $\times$ Male (β_3)* row, a treated male student is estimated to perform 0.26 SDs better in his final exam than a treated female. These effects further increase his weighted average exam grade by 0.16 SDs, which drives up his overall course grade by 0.13 SDs (or 0.18 SDs in the alternative specification with full controls).

Panel C of Table C.2 presents the effect of the Goals treatment on online assignment progression in Week 7 (Panel C.1), Week 11 (Panel C.2), and Week 13 (Panel C.3). On the whole, provision of the league (i.e., milestone-referenced absolute performance) information does not bring about significant or robust changes in online assignment progression for either females or males, similar to the situation in Section 3.4.1. Given the significant improvements in males' exam performance[11] and in overall course grades, the null results in online assignment progression suggest that the Goals treatment may alter students' exam performance through channels other than directly increasing their online assignment progression. This conjecture is investigated in the Mechanism section. For the same reason, the goal-achieving behaviour - achieving points in the online assignment - is not further discussed in the remaining sections. All analyses henceforth exclusively focus on effects on course grades and their corresponding mechanisms.

3.5 Non-linear and Heterogeneous Effects on Course Grades

3.5.1 Non-linear Effects on Course Grades

To further understand whether the treatment effects by gender are different at different points on the grade distributions, unconditional quantile regression models (Firpo et al., 2009) are run to estimate the effect of different interventions at several percentiles $\theta \in [0, 1]$ of the distribution of grades. Table C.3 to Table C.5 present the coefficients from these quantile regressions (marginal effects) at five academic performance levels, corresponding to the 10^{th}, 25^{th}, 50^{th}, 75^{th} and 90^{th} percentiles of the grade distribution.

Table C.3 compares the Social Goals condition with the Goals condition, capturing the effect of relative performance feedback on grades. Surprisingly, no significant effects for median female students are obtained, seemingly contrary to the result in Section 3.4.1 indicating that treated females are estimated to perform 0.19 SDs worse than their same-gender counterparts in the first midterm. A closer investigation in Table C.3, working from Panels D and E, shows that females at the 75^{th} percentile perform 0.36 SDs worse in their first midterm, while females at the 90^{th} percentile

[11] I note that the effects are statistically significant for the first midterm, final exam, average exam, and overall course grades, but not for the second midterm. However, the coefficients for the second midterm are positive and of similar magnitude. The lack of statistical significance seems a sample size issue due to splitting the sample by gender.

perform 0.42 SDs worse, than their female counterparts who have no access to relative performance information. In other words, having access to relative performance information strongly and robustly depressed the performance of high-performing females in their first midterm, which results in an overall negative treatment effect as captured in Section 3.4.1. Meanwhile, similar negative performance effects on treated females are evident in their other grades as well. Compared to their female counterparts in the Goals condition, treated females at the 75^{th} percentile perform 0.38 SDs worse in their second midterm, while those at the 90^{th} percentile perform 0.29 SDs worse in their final exam. These negative effects in individual exams accumulate to a significant and robust decrease of 0.44 SDs in the weighted average exam grade for females at the 75^{th} percentile, as well as a strong and robust decrease of 0.32 SDs for females at the 90^{th} percentile. Notably, the negative effect for the 90^{th} percentile females is so strong that it leads to a 0.28 SDs decrease in their overall course grade. In stark contrast, barely any effect is seen for low-performing females or men at most points on the performance distribution, for most outcomes. There is weak evidence showing that the male-specific treatment effect is positive in the first midterm and the final exam for males at the 90^{th} percentile (the *Treatment* $\times$ *Male (β_3)* row in columns (2) and (5) of Panel E). Additionally, treated males at the 25^{th} percentile appear to perform 0.45 SDs higher than treated females in their second midterm, based on columns (3) and (4) in the *Treatment* $\times$ *Male (β_3)* row in Panel B. These limited male-specific treatment effects unsurprisingly do not lead to any significant treatment effects on males. Overall, these patterns captured by the unconditional quantile regressions point to the conclusion that the overall negative effect on treated females' first midterm, as captured in Section 3.4.1, is driven by the particularly strong negative effect on high-performing females.

Table C.4 presents the results comparing the Social Goals condition with the control condition. On the one hand, no significant differential treatment effects for median males against median females are captured in the final exam or the weighted average exam. The *Treatment* $\times$ *Male (β_3)* row shows that the treatment additionally drives up median males' first midterm exam by 0.44 SDs, compared to the treatment effect on median females. Both the lack of significant effects in the two exam marks (i.e., final exam marks and average exam marks) and the appearance of significant median effect in the first midterm marks suggest one thing: the previously-captured positive male-specific treatment effects are an overall effect on men at most points on the performance distribution, rather than being driven by non-linear effects occurring at specific percentiles of the grade distributions. On the other hand, we see that high-performing females are negatively affected even when they are compared to females in the control condition, confirming the notion that relative performance information can hurt (Bouton and Kirchsteiger, 2015). Columns (1) and (2) in the first row of Panel D show that females at the 75^{th} percentile perform 0.37 SDs worse in their first midterm than their female counterparts in the control condition, resulting in a decrease of the same magnitude in their weighted average exam, which further decreases their overall course grade by 0.28 SDs. The story is similar for females at the 90^{th} percentile. Columns (1) and (2) in the first row of Panel D show that these females perform 0.35 SDs worse in their first midterm than their female counterparts in the control condition, resulting in a decrease of a larger magnitude of 0.45 SDs in their weighted average exam, although these negative effects only lead to a non-robust decrease of 0.28 SDs in their overall course grades. Again, virtually no effect is captured for low-performing females or men at most points on the performance distribution.

No non-linear effect is captured when comparing the Goals condition with the Control condition, based on Table C.5. This is particularly interesting given that significant and robust positive treatment effects on males are shown in Panel C of Table C.1. The lack of non-linear effect further confirms that the previously-captured positive male-specific treatment effects are an overall effect on men at most points on the performance distribution, rather than effects occurring at specific percentiles of the grade distributions. Overall, all the patterns captured by the unconditional quantile regressions point to two distinct paths for females and males. For females, the overall negative effect of the relative performance feedback on treated females' first midterm, as captured in Table C.1, is driven by the particularly strong negative effect on high-performing females. For males, the overall positive effect of the Goals treatment is indeed reflective of a treatment effect that is reasonably homogeneous across the distribution of male performance. In other words, to untangle the underlying mechanisms of the overall treatment effects, slightly different paths should be taken for each gender. To understand what is driving the negative effects on females, one should focus on high-performing females. To understand the driving force of the positive effects on males, one should take the male group as a whole. The *Mechanisms* section is guided by these differences.

3.5.2 Heterogeneous Effects on Course Grades

In this subsection, heterogeneous effects on course grades are investigated across four different dimensions: country of birth (CoB), major, ability, and procrastination level.

Table C.12 to Table C.14 present results that illuminate whether treatment effects differ between two countries of birth, namely Australia and China, within one gender. Panel A in each table presents analyses on females, while Panel B in each table presents analyses on males. Skimming through Panel A of all three tables, neither are there any treatment effects on Australian women or Chinese women, nor are there any systematic differences between two ethnic groups. In stark contrast, columns (3) and (4) of Panel B in Table C.12 show that having access to relative performance information strongly increases Australian men's second midterm grades by 0.22 SDs, compared to Australian men without access to this information. Turning to Panel B of Table C.13, columns (3) and (4) confirm that Australian men achieve higher second midterm grades even when compared to Australian men in the control condition. The statistical significance disappears in Panel B of Table C.14, indicating that Australian men respond most strongly to the relative performance information, rather than the milestone-referenced (league-based) absolute performance information by itself. Meanwhile, for Chinese men, the relative performance information decreases their final exam grades by 0.34 SDs, than do their same-ethnicity peers in the Goals condition. The decrease in Chinese men's final exam grades is so strong that it results in a roughly 0.26 SDs decrease, although not robust to the alternative specification with full controls, in the average exam and overall course grades. The statistical significance of the Chinese-specific treatment effects disappears in Table C.13 and Table C.14, indicating that Chinese men are also susceptible to the relative performance information. Additionally, it is noted that Chinese men perform systematically worse than Australian men in all exams by 0.30 to 0.51 SDs, based on the *Chinese (β_2)* row in Panel B of all three tables. This is in line with previous findings on the low performance of international students on Australian tertiary education (Foster, 2012).

Table C.15 to Table C.17 show whether treatment effects differ between students majoring in business-related degrees and those majoring in other degrees, within one gender. Panel A in each table presents analyses on females, while Panel B in each table presents analyses on males. Interestingly, regardless of gender, students majoring in non-business degrees do not achieve significantly different exam grades in any pairwise treatment comparison, as shown in the *Treatment (β_1)* rows in every table. However, Panel B of Table C.16 shows that, when exposed to the relative performance information joint with the league (i.e., milestone-referenced absolute performance) information, business-majored men achieve 0.44 SDs higher average exam grades compared to their same-gender counterparts with non-business majors in the control condition. This is understandable: the intervention course is compulsory for business students and thus they would want to earn as higher grades as they can; for non-business students, the course is probably selective or for general education purpose and thus many would be content with a passing grade. In other words, the Social Goals intervention have provided extra motivation for men. However, only business-majored men care enough about the grades to take action, while the remaining men do not.

Turning to Table C.18, students' previous semester GPA (or ATAR if GPA is unavailable) are interacted with each treatment variable respectively. This construction is to investigate whether treatment effects change with ability level. Panel A, B, and C present pairwise comparisons of Social Goals vs. Goals, Social Goals vs. Control, and Goals vs. Control. Each panel presents the results for females and males separately. In all three pairwise comparisons, women's ability either does not relate to or negatively relate to their course grades, while men's ability positively relate to their course grades. After controlling for ability, the relative performance information, as presented in Panel A, does not alter either women's or men's course grades. Columns (1) and (2) of Panel B.1 show that females in the Social Goals condition on average perform 0.32 SDs lower in their first midterm than females in the control condition, regardless of ability. Panel B.2 shows that, compared to males in the control condition, males in the Social Goals condition on average perform 0.93 SDs higher in their first midterm, 0.76 SDs higher in their average exam, and 0.72 SDs higher in their overall course grades. However, these average treatment effects tend to decrease by 0.01 SDs for each grade point increase in previous GPA. With the mean GPA of 53.30, the treatment effect on an average male student's first midterm is around 0.31 SDs, a substantial improvement. The story in Panel C.2 is similar. Compared to males in the control condition, males in the Goals condition on average perform 1.00 SDs higher in their first midterm, 0.72 SDs higher in their final exam, 0.82 SDs higher in their average exam grades, and 0.77 SDs higher in their overall course grades. These average treatment effects again tend to decrease by 0.01-0.02 SDs for each grade point increase in previous GPA. Similarly, given the mean GPA of 53.30, the treatment effect on an average male student's first midterm is around 0.58 SDs, a even more substantial improvement. Even for the extremely high-performing students, for example, one with 100 points in previous GPA, the treatment effect is 0.11 SDs. Overall, after controlling for previous ability, males who have access to the league (i.e., milestone-referenced absolute performance) information earn significantly higher exam grades compared to males who do not have access to the league information.

Table C.19 interacts each treatment variable with *Procrastination Level*. The procrastination level is constructed based on the week in which a student start to earn non-zero points. For example, if a student starts earning points in Week 3 (the intervention deployment week), then his procrastination level is 0; if he starts earning points in Week 4, then his procrastination level is 1; and so on. By its nature, procrastination is an exogenous variable, as one can only be treated after one starts completing the online assignment and thus earning non-zero points. Unsurprisingly, the higher the procrastination level, the lower the course grades. This correlation is unambiguous, persistent, and independent of gender (although the magnitude of the correlation is substantially larger for females than is for males). In columns (1) and (2) of Panel A.1, the average student bears a 0.13 SDs decrease in his first midterm for each additionally procrastinated week. As a side-note, this correlation calls upon the necessity to nudge students to start learning early in a semester. Interestingly, after controlling for procrastination level, treated males no longer perform significantly differently from control males in each pairwise treatment comparison, neither are there any differential treatment effects relating to procrastination level. On the contrary, columns (1) and (2) in Panel A.1 show that the negative effect of relative performance information on treated females' first midterm largely remains (-0.34 SDs), as opposed to their controlled female counterparts. The negative effect manifests in the average exam (-0.36 SDs, not robust to controlling a full set of covariates), as well as in the overall course grades (-0.28 SDs, not robust). Additionally, the negative effect is mitigated by the procrastination level, as shown in the *Procrastination*$\times$*Treatment* row. For each procrastinated week, a female student is less harmed by 0.06 SDs in her first midterm. Put differently, had a woman started accessing the intervened online assignment later in the semester, she might have been less negatively affected. Women are similarly negatively affected when comparing the Social Goals condition with the Control condition. Although the moderating effect of procrastination level disappears, treated females perform substantially worse than controlled females in their first midterm and average exam.

In summary, the heterogeneous effects reveal a richer picture across several dimensions of longstanding research interests. As for country of birth, Australian males who have access to relative performance information achieve significantly higher second midterm grades compared to those who do not have access. However, Chinese males who have access to relative performance information are significantly negatively affected in their final exam. Turning to major, when access to relative performance information is layered with access to leagues information, business-majored males earn significantly higher average exam grades compared to their same-major male counterparts. Female subgroups remain dormant in the lens of both country of birth and field of study. After controlling for ability, females in the Social Goals condition perform worse in the first midterm than females in the Control condition, while males in both the Social Goals condition and the Goals condition perform substantially better in four out of five course grades than males in the Control condition. Interestingly, if procrastination level is controlled instead, males appear dormant, while females in the Social Goals condition remain disadvantaged, when compared to either females in the in the Goals condition or females in the in the Control condition.

3.6 Mechanisms

Thus far, two distinct patterns of the treatment effects on course grades have emerged for females and males respectively. Females are particularly responsive to the relative performance information: those who can access the information have an average first midterm grade 0.19 SDs lower than those who do not have access. The negative effects turn out to be driven by high-performing females at the 75^{th} and 90^{th} percentile of the grade distributions, who perform as large as 0.44 SDs worse in exams. Meanwhile, males are particularly responsive to leagues information (i.e., the Goals treatment): those who can access the league (i.e., milestone-referenced absolute performance) information on average perform better in their final exam, weighted average exam grade, as well as their overall course grade, than those who do not have access. The treatment effects on males are average treatment effects which do not differ across percentiles of a distribution. Thus, the underlying mechanisms for females and males are investigated separately in this section.

Table C.6 regresses students' perceived ability to overcome difficulties on the treatment variable, capturing the effect of relative performance information on one behavioural proxy. Both females and males are split into two groups: those whose grades are below median grades within the same gender group are "low-performing students" and those whose grades are above the corresponding median grades are "high-performing students". The exam grades according to which the groups are split are presented in the column titles of the table. Columns (1) to (5) in Panel A.2 show that high-performing females tend to consider themselves significantly less able to overcome difficulties.[12] Specifically, high-performing females in the first midterm are 15.7% less likely to report feeling able to overcome difficulties, those in the second midterm are 11.5% less likely, and those in the final exam are 16.2% less likely. As expected, the probability of high-performing females in the average exam reporting able to overcome difficulties are also statistically significantly lower by 14.5%, while those according to the overall course grade are lower by 11.8%. Unsurprisingly, no such difference is captured for either low-performing females or the male group. As one may recall from Chapter 2, when low-performing students in the Social Goals condition feel less able to overcome difficulties and work harder by accessing the course platform more often, they end up performing better in their exams than their Goals counterparts. To control for the potential functioning channel of course platform accesses, Table C.6 also includes course platform accesses in each panel. As can be seen, high-performing females who feel less able to overcome difficulties are not estimated to access the course platform significantly more often. Instead, low-performing males in the second midterm view textbooks 0.54 SDs more often, corresponding to the positive differential treatment effect presented in Panel B of Table C.3. Put differently, mere worse feeling about one's ability to overcome difficulties does not necessarily bring about better performance, one also needs to increase efforts to improve the situation. In fact, if one self-pity so much without putting in effort, as is captured for high-performing females here, one's performance may be devastated.

[12] The alternative specification controlling for student characteristics and tutor fixed effects are qualitatively similar.

The story is similar when students in the Social Goals condition are compared with those in the control condition. Panel A.2 in Table C.7 shows that high-performing females (according to first midterm, second midterm, and average exam) report significantly lower probability of overcoming difficulties than their control counterparts. These are in line with Table C.4 where all three exam grades are negative. The lack of significance in Column (5) of Table C.7 is unsurprising, given that the negative effects for the 75^{th} and 90^{th} percentile high-performing females are small in magnitude and only marginally significant. As expected, Table C.8 which compares the Goals condition with the control condition does not show any systematically different behaviours for any subgroups. This again confirms that females, especially high-performing females, are particularly responsive to the relative performance information, while they appear dormant to the milestone-referenced league-based absolute performance information. Overall, the significant detrimental effects of relative performance feedback on high-performing females' course grades are a result of them feeling less able to overcome difficulties, which is not accompanied by sufficiently increased effort.

Turning to males, the analysis will start by comparing males in Goals vs. males in Control, as males are most responsive to the milestone-referenced league-based absolute performance information. Then Social Goals are compared with the control group, followed by the comparison between Social Goals vs. Goals. Males with leagues information access are estimated to put in more effort in the online assignment towards the end of the semester, according to a list of behavioural proxies presented in Table C.9. Each row presents the OLS results for one proxy recorded weekly from Week 9 to Week 13, plus regressions on the total of that proxy. Week 9 is the week right before the second midterm. Week 10 is the week in which the second midterm takes place, while Week 11 to Week 13 are the last three semester weeks. The weeks are selected for three reasons.[13] First, Week 10 to Week 13 are all the semester weeks between the second midterm and the final exam. If anything were happening, they would have happened in this period, otherwise significant treatment effects would have been captured in the first or second midterm. Meanwhile, it is possible that efforts accumulate over the semester but only manifest themselves in the final exam or the average exam. This is the reason why the total of a proxy is included in columns (11) and (12). Additionally, Week 9 enables direct comparisons

[13]Results in all weeks are investigated but not presented here for the sake of space. No robust effects present in other weeks.

of students' behaviours before the second midterm and before the final exam (i.e., after the second midterm). Thus, with the selected weeks, any effort changes that may potentially contribute to grade changes are captured.

The *Number of Assignment Accesses* row shows that treated males log into the online assignment platform significantly, albeit non-robustly, more often in Week 9, 10, 11, and 13. In parallel, the *Correct Online Assignment Submissions* row shows that treated males submit 0.22 SDs more correct answers in Week 10 and Week 11 than controlled males. The story is similar for *All Online Assignment Submissions* and for *Incorrect Online Assignment Submissions*, where treated males submit around 0.24 to 0.29 SDs more answers than controlled males in Week 10 and Week 11. Going back to Week 9, columns (1) and (2) show no robust evidence of treated males putting in more effort than their controlled counterparts, which further confirms that the significant differences in Week 10 and Week 11 drive up treated males' final exam grades, as well as average and overall course grade. At last, the lack of significant differences in Week 12 and Week 13 can be attributed to two possible factors. First, students strategically reallocate efforts in the treated course. Instead of last-minute cramming commonly observed in education settings (Fischer, 2001), the treated students are ahead of time. They review and consolidate their learning right after the second midterm, well before the final exam. Second and in the same vein, students redistribute efforts to other courses as the semester exam period approaches. In other words, after studying for the treated course intensively in Week 10 and 11, students focus more on the other courses in Week 12 and 13. In sum, males in the Goals treatment achieve higher final exam marks than their controlled counterparts, because they increase their efforts in the online assignment by submitting more questions, both correct ones and incorrect ones. They commit more trial and error on average, resulting in significant increases in their average exam grade and overall course grade.

Table C.10 presents the same result specifications as above, yet comparing males in the Social Goals treatment with males in the control condition. Results are similar but weaker. It is unsurprising given that only differential treatment effect is observed between males and females in the Social Goals condition and no significant overall treatment effect is captured for either females or males. Table C.11 presents the same results specifications comparing males in the Social Goals treatment with males in the Goals condition. No robust evidence is captured. These two tables are worth mentioning as they complete the story by showing how the mechanisms weaken and vanish as the overall treatment effects on males changes from significant and robust, to positive but statistically insignificant, and to slightly negative as shown in the *Treatment Effect on Male* row in Table C.1. Compared to their controlled counterparts, males in the Social Goals condition also submit 0.14 SDs more online assignment answers overall and have 0.17 SDs more incorrect attempts in Week 11. They log into the online assignment platform 0.22 SDs more often in Week 12 compared to males in the control condition. By and large, the behavioural patterns observed in Table C.9 persist in Table C.10, but are considerably weaker in the latter case. This seems to be the reason why no significant overall treatment effects on males is captured when comparing males in the Social Goals treatment with males in the control condition. Same logic holds when comparing males in the Social Goals condition with males in the Goals condition.

To sum up, females and males are nothing but different in their responses to the interventions. On the one hand, females are significantly negatively affected by the relative performance information, with high-performing females being the most negatively affected group. High-performing females' exam grades and course grades are significantly lower than their female counterparts who do not have access to relative performance information. The reason is that relative performance information makes high-performing females feel less able to overcome difficulties and yet there is no evidence of them increasing efforts. On the other hand, men at most points on the performance distribution are significantly motivated by the league (i.e., milestone-referenced absolute performance) information and the increase appears homogeneous along the grade distributions. The overall grade increases are a result of increased effort.

3.7 Robustness

To verify whether the main effects on exam grades persist for alternative specifications, several robustness checks are conducted.

First, two alternative specifications are explored: 1) control for tutorial fixed effects instead of tutor fixed effects to account for any systematic differences among tutorials; 2) cluster standard errors by tutorial instead of tutor to further account for the possible common class shocks. Results are reported in Table C.20. The first row in columns (1) and (2) in Panel A shows that the negative effect on females' first midterm is robust to both alternative specifications. Columns ((5) and (6) in the last row of Panel C confirm that the positive effect on males' final exam is also robust to both alternative specifications.

The second robustness check tries to account for prior ability, using the same MICE method as explained in Chapter 2. Odd numbered columns in Table C.21 show that all results remain quantitatively similar and statistically significant. Lastly, robustness is checked against inclusion of procrastination level. As observed in Section 3.5.2, procrastination level correlates so strongly to exam grades that it wipes up any differences in course grades within males. One may again be concerned with omitted variable bias. To address this concern, the baseline OLS regressions are rerun including the procrastination level as an additional control. Results are reported in even numbered rows in Table C.21. Although procrastination remain strongly significant in all regressions, all results remain quantitatively similar and statistically significant.

3.8 Conclusions and Discussions

3.8.1 Conclusions

Building on the suggestive evidence in Chapter 2, this chapter further explores the effect of relative performance information, leagues information, or them combined, on academic performance for female students and male students respectively. As indicated, females and males are startlingly different in their responses to the interventions. On the one hand, females are significantly negatively affected (0.19 SDs lower first midterm grade, equivalent to a decrease of 4 marks out of 100) by the relative performance information, with high-performing females being the most negatively affected (0.28 SDs lower overall course grade, equivalent to a decrease of 4 marks out of 100) group. The overall negative effect, albeit manifesting only in the first midterm grade, is consistent with the existing findings that females perform worse in a competitive environment (Niederle and Vesterlund, 2007; Jurajda and Münich, 2011). High-performing females' exam grades and course grades are significantly lower than their female counterparts who do not have access to relative performance information. The reason is that the relative performance information makes high-performing females feel less able to overcome difficulties and yet there is no evidence of them increasing efforts. These impressive negative effects for females on the right tail of the grade distribution is in line with Niederle and Vesterlund (2010)'s finding that competition especially hurts high-performing females possibly relating to their low-confidence level. The fact that students are all anonymized in my experiment provides further evidence that competition hurts high-performing females, even without revealing gender composition, a key factor at play mentioned by Niederle and Vesterlund (2010). On the other hand, men at most points on the performance distribution is significantly motivated (0.26 SDs higher overall course grade, equivalent to an increase of 3 marks out of 100) by the milestone-based league-referenced absolute performance information and the increase is homogeneous along the grade distribution. The overall grade increases are a result of increased effort in the online assignment. The fact that males are more encouraged by goal-setting (i.e., leagues) is in line with the literature (Smithers, 2015; Clark et al., 2016). Meanwhile, the fact that males exposed to Social Goals do not perform statistically different from males exposed to Goals indicates that, in the presence of leagues, ranking information provides limited additional motivational boost to males.

These results are robust to alternative specifications controlling for tutorial fixed effects instead of tutor fixed effects or clustering standard errors by tutorials instead of tutors. Taking advantage of the semester-long intervention window, a student's procrastination level is captured by his starting week in the online assignment. The beauty of this procrastination measure is two-fold: First, it is objective, as it is recovered from students' actual behaviour and not self-reported. Second, it is exogenous to the interventions, in the sense that the procrastination level is captured before a student is treated since one cannot be treated unless he starts logging in the online assignment. Unsurprisingly, procrastination level strongly and negatively relates to students' course grades. This strong and negative correlation implies that students could potentially be better off if they are nudged for an early start in the semester. Meanwhile, previous GPA records enable an additional control of students' ability. Even after controlling for ability or procrastination level, all the results remain strong, statistically significant, and robust.

Heterogeneous effects are also investigated to uncover a richer and more nuanced picture across several dimensions of longstanding research interests, namely country of birth, field of study, ability, and procrastination level. As for country of birth, Australian males who have access to relative performance information achieve significantly higher second midterm grades compared to those who do not have access. However, Chinese males who have access to relative performance information are significantly negatively affected in their final exam. Turning to major, when access to relative performance information is layered with access to leagues information, business-majored males earn significantly higher average exam grades compared to their same-major male counterparts. Female subgroups remain dormant in the lens of both country of birth and field of study. After controlling for ability, females who have access to both the relative performance information and the league information perform worse in the first midterm than females without the access. Males with access to the league information, regardless of availability of the relative performance information, perform substantially better in four out of five course grades than males without the league information, after ability is controlled. Interestingly, if procrastination level is controlled instead, treated males appear dormant relative to controlled males, while females who have access to both the relative performance information and the league information remain negatively affected compared to their female counterparts in the other conditions.

In a nutshell, this chapter has two main take-away messages. First, one should avoid telling high-performing females how they perform relative to their peers, as they are likely to be stressed out and devastated. Second, one should let males compete in leagues (even if virtual) whenever possible, as they care about their leagues and will work harder to achieve higher leagues.

3.8.2 Discussions: Chapter 2 vs. Chapter 3

Comparing this chapter with Chapter 2, several interesting patterns are noted. First, the overall negative effect of relative performance information on Week 6 exam grades is mainly driven by the strong negative effects on high-performing females. Although the negative effects on high-performing females' other course grades are strong and statistically significant, they seem to be diluted by the mixed effects on the rest of the sample, rendering nil overall effect on other course grades. Second and in the same vein, the league information deems effective in improving males' course grades, but the effects do not manifest in the whole sample most likely because the positive effects on males cancel out with the negative effects on females.

Third, looking at the treatment effects within each gender in Table C.1 again, the relative performance information and the league information are like at two ends of a seesaw. The relative performance information brings about mostly negative coefficients of treatment effect for both genders, although statistical significance emerges only for females. The league information brings about positive coefficients for both genders, although statistical significance emerges only for males. When both types of information are provided, the treatment coefficients become noisy, negative for females and mostly positive for males. In other words, despite the aforementioned opposite gender-specific treatment effects of relative performance information and leagues information, the signs of the coefficients suggest consistent overall treatment effects. That is, relative performance information decreases course grades, whereas leagues information increases grades.

If we further revisit Tables C.3 to C.5, the coefficients are negative and sizable which captures the effect of relative performance information for high-performing males, whereas the corresponding coefficients are jumpy and negligible which captures the effect of leagues information. These observations imply that relative performance information brings about negative and non-linear treatment effects on course grades, while leagues information brings about positive and average treatment effects. This is unsurprising. Given that relative performance information makes one's relative position salient, one takes account of this information in benchmarking his ability and exerting effort (Cabrales et al., 2019). The league information is unlikely to make ability salient, given that milestones are exogenously set. The underlying stories merit further investigation and more clear-cut overall evidence which is beyond the capability of this project.

Looking at the mechanisms in Chapter 2 and Chapter 3, all the positive treatment effects on the sample or a subsample, either of relative performance information or of leagues information, are unanimously driven by increased effort in course platform accesses or online assignment submissions, which is as expected. Meanwhile, the driving force of the negative treatment effects on high-performers are slightly different. High-performers based on the whole sample are found to report feeling happier, while high-performing females are found to feel less able to overcome difficulties. A closer investigation shows that the coefficients reflecting high-performing females' perceived happiness are all positive, albeit statistically insignificant, possibly due to the shrunk sample size within each gender. This is suggesting that one's perceived stress does not necessarily reflect his perceived happiness. Yet, an increase in either stress level or perceived happiness, can lead to worse academic performance.

Appendix C

Table C.1: Effects on Exam Grades

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Goals										
Treatment (β_1)	-0.198*	-0.187**	-0.117	-0.110	-0.092	-0.063	-0.157	-0.140	-0.140	-0.120
	(0.100)	(0.084)	(0.127)	(0.127)	(0.107)	(0.116)	(0.094)	(0.093)	(0.093)	(0.097)
Male (β_2)	-0.017	0.038	-0.073	-0.093	0.154*	0.166**	0.006	0.023	0.056	0.077
	(0.103)	(0.095)	(0.081)	(0.092)	(0.085)	(0.078)	(0.081)	(0.087)	(0.070)	(0.068)
Treatment × Male (β_3)	-0.026	0.014	0.033	-0.005	0.176**	0.170**	0.062	0.062	0.080	0.074
	(0.097)	(0.095)	(0.092)	(0.084)	(0.063)	(0.071)	(0.075)	(0.081)	(0.064)	(0.071)
Treatment Effect on Male ($\beta_1 + \beta_3$)	-0.223	-0.173	-0.084	-0.115	0.084	0.107	-0.095	-0.078	-0.060	-0.046
	[0.191]	[0.183]	[0.143]	[0.197]	[0.150]	[0.124]	[0.135]	[0.163]	[0.143]	[0.155]
Obs	613	613	607	607	618	618	602	602	618	618
Panel B: Social Goals vs. Control										
Treatment (β_1)	-0.090	-0.094	-0.022	-0.045	-0.043	-0.030	-0.057	-0.066	-0.066	-0.065
	(0.089)	(0.094)	(0.113)	(0.116)	(0.121)	(0.128)	(0.096)	(0.100)	(0.100)	(0.105)
Male (β_2)	-0.022	-0.009	0.037	0.014	0.260*	0.265*	0.084	0.080	0.144	0.145
	(0.083)	(0.081)	(0.097)	(0.090)	(0.123)	(0.131)	(0.083)	(0.083)	(0.086)	(0.092)
Treatment × Male (β_3)	0.082	0.098	0.128*	0.096	0.225**	0.217*	0.162**	0.154*	0.155*	0.145
	(0.084)	(0.096)	(0.066)	(0.057)	(0.095)	(0.103)	(0.071)	(0.077)	(0.080)	(0.086)
Treatment Effect on Male ($\beta_1 + \beta_3$)	-0.007	0.004	0.106	0.051	0.182	0.187	0.105	0.088	0.089	0.080
	[0.157]	[0.177]	[0.163]	[0.125]	[0.209]	[0.211]	[0.154]	[0.148]	[0.166]	[0.168]
Obs	556	556	545	545	560	560	541	541	560	560
Panel C: Goals vs. Control										
Treatment (β_1)	0.108	0.107	0.095	0.088	0.049	0.074	0.100	0.103	0.075	0.084
	(0.085)	(0.089)	(0.064)	(0.063)	(0.067)	(0.069)	(0.061)	(0.059)	(0.054)	(0.055)
Male (β_2)	-0.022	0.050	0.037	0.075	0.260*	0.335**	0.084	0.155**	0.144	0.218***
	(0.083)	(0.083)	(0.097)	(0.092)	(0.123)	(0.114)	(0.083)	(0.064)	(0.086)	(0.070)
Treatment × Male (β_3)	0.091	0.145	0.022	0.049	0.203**	0.260***	0.106	0.159*	0.130*	0.182**
	(0.075)	(0.086)	(0.091)	(0.099)	(0.089)	(0.079)	(0.085)	(0.085)	(0.073)	(0.067)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.199	0.251*	0.117	0.137	0.252*	0.334**	0.206	0.262**	0.205*	0.265**
	[0.122]	[0.133]	[0.134]	[0.144]	[0.190]	[0.166]	[0.123]	[0.113]	[0.131]	[0.096]
N	599	599	592	592	602	602	589	589	602	602
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade. The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grades. All dependent variables are standardized. Odd numbered columns only include a treatment indicator, gender, and their interaction. Even numbered columns additionally control for student characteristics (age, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.2: Effects on Assignment Progression

	1,000+ points		1,300+ points		1,000+ points		1,300+ points		1,000+ points		1,300+ points	
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Panel A: Social Goals vs. Goals				*Panel B: Social Goals vs. Control*				*Panel C: Goals vs. Control*			
	Panel A.1: Week 7				*Panel B.1: Week 7*				*Panel C.1: Week 7*			
Treatment (β_1)	0.017	0.019	-0.007	-0.006	0.026	0.021	0.000	-0.002	0.010	0.014	0.007	0.008
	(0.026)	(0.027)	(0.007)	(0.008)	(0.026)	(0.026)	(0.000)	(0.001)	(0.022)	(0.023)	(0.007)	(0.008)
Male (β_2)	-0.007	-0.006	-0.007	-0.004	0.029	0.033	0.000	-0.000	0.029	0.037*	0.000***	0.001
	(0.020)	(0.021)	(0.007)	(0.006)	(0.020)	(0.021)	(0.000)	(0.003)	(0.020)	(0.019)	(0.000)	(0.001)
Treatment × Male (β_3)	0.026	0.019	0.011	0.013	0.035**	0.028*	0.018*	0.014	0.003	0.002	0.000***	0.002
	(0.030)	(0.033)	(0.012)	(0.013)	(0.013)	(0.014)	(0.009)	(0.009)	(0.018)	(0.022)	(0.000)	(0.002)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.042	0.038	0.004	0.006	0.062**	0.049*	0.018*	0.012	0.013	0.015	0.007	0.010
	[0.046]	[0.049]	[0.018]	[0.018]	[0.024]	[0.025]	[0.009]	[0.011]	[0.034]	[0.035]	[0.000]	[0.003]
	Panel A.2: Week 11				*Panel B.2: Week 11*				*Panel C.2: Week 11*			
Treatment (β_1)	0.055	0.068	0.022	0.031	-0.025	-0.040	0.026	0.031	-0.080	-0.083	0.004	0.006
	(0.059)	(0.061)	(0.027)	(0.029)	(0.075)	(0.074)	(0.025)	(0.025)	(0.054)	(0.053)	(0.013)	(0.012)
Male (β_2)	0.024	0.041	0.012	0.014	-0.096*	-0.086	0.011	0.009	-0.096*	-0.071	0.011	0.018
	(0.043)	(0.045)	(0.021)	(0.021)	(0.053)	(0.056)	(0.018)	(0.019)	(0.053)	(0.051)	(0.018)	(0.019)
Treatment × Male (β_3)	0.026	0.036	0.027	0.022	-0.054	-0.062	0.032	0.020	-0.056	-0.056	0.017	0.018
	(0.054)	(0.055)	(0.029)	(0.027)	(0.050)	(0.053)	(0.025)	(0.020)	(0.053)	(0.051)	(0.018)	(0.018)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.081	0.104	0.049	0.053	-0.079	-0.102	0.058	0.051	-0.136	-0.139	0.021	0.024
	[0.106]	[0.098]	[0.046]	[0.044]	[0.098]	[0.104]	[0.043]	[0.031]	[0.097]	[0.099]	[0.023]	[0.033]
	Panel A.3: Week 13				*Panel B.3: Week 13*				*Panel C.3: Week 13*			
Treatment (β_1)	0.069	0.088	0.021	0.038	0.012	-0.002	0.003	0.014	-0.057	-0.064	-0.018	-0.019
	(0.075)	(0.075)	(0.048)	(0.051)	(0.078)	(0.079)	(0.050)	(0.052)	(0.048)	(0.046)	(0.027)	(0.025)
Male (β_2)	-0.044	-0.017	-0.000	0.007	-0.122**	-0.108*	-0.033	-0.021	-0.122**	-0.081	-0.033	-0.023
	(0.056)	(0.054)	(0.034)	(0.040)	(0.054)	(0.055)	(0.032)	(0.032)	(0.054)	(0.051)	(0.032)	(0.033)
Treatment × Male (β_3)	0.026	0.043	0.023	0.019	-0.031	-0.036	0.005	-0.005	-0.102	-0.086	-0.019	-0.017
	(0.058)	(0.054)	(0.045)	(0.046)	(0.057)	(0.053)	(0.043)	(0.043)	(0.066)	(0.065)	(0.037)	(0.040)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.095	0.130	0.045	0.057	-0.019	-0.039	0.008	0.009	-0.159	-0.151	-0.037	-0.036
	[0.126]	[0.121]	[0.088]	[0.091]	[0.102]	[0.098]	[0.088]	[0.071]	[0.101]	[0.111]	[0.066]	[0.071]
Obs	614	614	614	614	557	557	557	557	601	601	601	601
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is a binary variable which equals one if a student's online assignment points in a certain week are above 1,000 and zero otherwise. The dependent variable in columns (3)-(4) refer to the same outcome formulation for online assignment points above 1,300. Odd numbered columns only include a treatment indicator, gender, and their interaction. Even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.3: Effects on Exam Grades (UQR)- Social Goals vs. Goals

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: P10										
Treatment (β_1)	-0.134	-0.065	0.259	0.284	0.172	0.200	0.135	0.155	0.118	0.125
	(0.244)	(0.245)	(0.170)	(0.205)	(0.202)	(0.205)	(0.249)	(0.211)	(0.188)	(0.212)
Male(β_2)	0.050	0.159	-0.140	-0.100	0.089	0.167	-0.057	0.057	0.067	0.148
	(0.222)	(0.229)	(0.187)	(0.229)	(0.183)	(0.192)	(0.211)	(0.226)	(0.176)	(0.206)
Treatment× Male (β_3)	0.278	0.166	0.074	-0.018	-0.059	-0.167	0.120	0.032	0.018	-0.065
	(0.333)	(0.319)	(0.233)	(0.289)	(0.265)	(0.249)	(0.321)	(0.291)	(0.243)	(0.264)
Treatment Effect on Male ($\beta_1+\beta_3$)	0.144	0.101	0.333	0.266	0.113	0.032	0.255	0.187	0.136	0.060
	[0.405]	[0.409]	[0.299]	[0.323]	[0.336]	[0.291]	[0.338]	[0.353]	[0.283]	[0.322]
Panel B: P25										
Treatment (β_1)	-0.101	-0.062	-0.108	-0.123	-0.013	0.030	-0.205	-0.212	-0.032	-0.031
	(0.182)	(0.189)	(0.206)	(0.208)	(0.175)	(0.176)	(0.203)	(0.194)	(0.157)	(0.163)
Male(β_2)	0.022	0.108	-0.347*	-0.350*	0.104	0.150	-0.196	-0.154	0.014	0.042
	(0.157)	(0.177)	(0.183)	(0.206)	(0.157)	(0.162)	(0.174)	(0.159)	(0.156)	(0.156)
Treatment× Male (β_3)	0.126	0.043	0.505*	0.455*	0.044	-0.027	0.437*	0.398	0.177	0.155
	(0.241)	(0.231)	(0.263)	(0.271)	(0.224)	(0.234)	(0.247)	(0.253)	(0.202)	(0.224)
Treatment Effect on Male ($\beta_1+\beta_3$)	0.025	-0.019	0.397	0.331	0.032	0.003	0.232	0.186	0.145	0.124
	[0.296]	[0.310]	[0.355]	[0.365]	[0.273]	[0.301]	[0.298]	[0.329]	[0.253]	[0.286]
Panel C: P50										
Treatment (β_1)	-0.221	-0.227	-0.172	-0.156	-0.113	-0.066	-0.158	-0.109	-0.239	-0.198
	(0.149)	(0.152)	(0.165)	(0.151)	(0.176)	(0.197)	(0.172)	(0.170)	(0.164)	(0.179)
Male(β_2)	0.004	0.021	-0.073	-0.063	0.306*	0.319*	-0.046	-0.036	-0.011	0.002
	(0.134)	(0.141)	(0.147)	(0.142)	(0.179)	(0.191)	(0.153)	(0.147)	(0.150)	(0.161)
Treatment× Male (β_3)	0.296	0.290	0.175	0.120	-0.016	-0.075	0.226	0.151	0.306	0.213
	(0.202)	(0.199)	(0.228)	(0.208)	(0.238)	(0.264)	(0.222)	(0.224)	(0.228)	(0.224)
Treatment Effect on Male ($\beta_1+\beta_3$)	0.075	0.063	0.003	-0.036	-0.130	-0.141	0.068	0.042	0.067	0.015
	[0.248]	[0.236]	[0.270]	[0.279]	[0.288]	[0.312]	[0.277]	[0.288]	[0.257]	[0.292]
Panel D: P75										
Treatment (β_1)	-0.340**	-0.363**	-0.360**	-0.381**	-0.176	-0.144	-0.427***	-0.443**	-0.151	-0.140
	(0.156)	(0.167)	(0.152)	(0.159)	(0.177)	(0.185)	(0.150)	(0.179)	(0.133)	(0.139)
Male(β_2)	-0.036	-0.038	0.036	-0.031	0.122	0.106	-0.115	-0.151	0.106	0.097
	(0.168)	(0.169)	(0.133)	(0.128)	(0.177)	(0.192)	(0.159)	(0.167)	(0.142)	(0.134)
Treatment× Male (β_3)	0.195	0.242	0.182	0.228	0.315	0.288	0.397*	0.453*	0.149	0.140
	(0.206)	(0.220)	(0.181)	(0.188)	(0.269)	(0.273)	(0.216)	(0.238)	(0.195)	(0.187)
Treatment Effect on Male ($\beta_1+\beta_3$)	-0.145	-0.121	-0.178	-0.154	0.139	0.144	-0.030	0.010	-0.002	0.000
	[0.261]	[0.263]	[0.243]	[0.244]	[0.290]	[0.337]	[0.252]	[0.282]	[0.234]	[0.262]
Panel E: P90										
Treatment (β_1)	-0.380***	-0.422***	-0.146	-0.160	-0.287**	-0.289**	-0.321**	-0.321**	-0.287**	-0.284**
	(0.144)	(0.154)	(0.111)	(0.117)	(0.138)	(0.147)	(0.128)	(0.126)	(0.137)	(0.138)
Male(β_2)	-0.084	-0.086	0.133	0.054	0.040	0.046	0.060	0.039	0.019	0.013
	(0.159)	(0.161)	(0.123)	(0.125)	(0.155)	(0.159)	(0.144)	(0.147)	(0.161)	(0.168)
Treatment× Male (β_3)	0.224	0.324*	0.054	0.113	0.362*	0.341	0.287	0.319	0.302	0.309
	(0.197)	(0.190)	(0.166)	(0.179)	(0.203)	(0.220)	(0.201)	(0.212)	(0.206)	(0.206)
Treatment Effect on Male ($\beta_1+\beta_3$)	-0.156	-0.098	-0.091	-0.048	0.075	0.051	-0.034	-0.002	0.015	0.025
	[0.244]	[0.242]	[0.239]	[0.229]	[0.284]	[0.280]	[0.268]	[0.241]	[0.276]	[0.257]
Obs	613	613	607	607	618	618	602	602	618	618
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate unconditional quantile regressions (UQR). All dependent variables are standardized. P_i represents UQR results at ith percentile. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns only include a treatment indicator, gender, and their interaction. Even numbered columns additionally control for student characteristics (age, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Bootstrapped standard errors with 200 replications are reported in parentheses. The linear combination of $\beta_1+\beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.4: Effects on Exam Grades (UQR)- Social Goals vs. Control

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: P10										
Treatment (β_1)	0.232	0.278	0.301	0.294	0.267	0.286	0.159	0.165	0.231	0.244
	(0.348)	(0.334)	(0.210)	(0.210)	(0.238)	(0.248)	(0.264)	(0.268)	(0.227)	(0.257)
Male (β_2)	0.174	0.129	0.025	0.024	0.320	0.270	0.229	0.243	0.262	0.204
	(0.331)	(0.310)	(0.207)	(0.233)	(0.214)	(0.230)	(0.227)	(0.242)	(0.184)	(0.238)
Treatment × Male (β_3)	0.180	0.260	-0.095	-0.108	-0.287	-0.235	-0.075	-0.033	-0.206	-0.168
	(0.419)	(0.414)	(0.289)	(0.285)	(0.296)	(0.311)	(0.307)	(0.325)	(0.259)	(0.313)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.412	0.538	0.206	0.186	-0.020	0.051	0.084	0.131	0.025	0.076
	[0.545]	[0.517]	[0.357]	[0.368]	[0.365]	[0.387]	[0.382]	[0.405]	[0.317]	[0.393]
Panel B: P25										
Treatment (β_1)	0.029	0.054	0.022	-0.025	0.136	0.161	0.023	-0.009	0.063	0.071
	(0.171)	(0.175)	(0.182)	(0.189)	(0.174)	(0.176)	(0.170)	(0.193)	(0.157)	(0.166)
Male (β_2)	0.009	0.005	-0.030	-0.027	0.342**	0.357**	0.101	0.087	0.142	0.130
	(0.168)	(0.175)	(0.168)	(0.166)	(0.157)	(0.168)	(0.174)	(0.182)	(0.167)	(0.155)
Treatment × Male (β_3)	0.113	0.117	0.164	0.164	-0.201	-0.209	0.114	0.145	0.037	0.045
	(0.221)	(0.227)	(0.245)	(0.207)	(0.210)	(0.231)	(0.228)	(0.244)	(0.221)	(0.233)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.142	0.172	0.185	0.139	-0.066	-0.048	0.136	0.135	0.100	0.116
	[0.277]	[0.286]	[0.305]	[0.281]	[0.273]	[0.286]	[0.285]	[0.311]	[0.271]	[0.280]
Panel C: P50										
Treatment (β_1)	-0.267	-0.301	-0.010	-0.034	-0.243	-0.189	-0.027	-0.044	-0.151	-0.155
	(0.197)	(0.192)	(0.187)	(0.171)	(0.195)	(0.200)	(0.177)	(0.174)	(0.195)	(0.175)
Male (β_2)	-0.092	-0.094	-0.015	-0.058	0.156	0.230	0.051	0.093	0.115	0.136
	(0.159)	(0.171)	(0.171)	(0.167)	(0.185)	(0.200)	(0.161)	(0.167)	(0.187)	(0.179)
Treatment × Male (β_3)	0.429*	0.443*	0.160	0.176	0.156	0.029	0.211	0.188	0.181	0.125
	(0.231)	(0.249)	(0.241)	(0.217)	(0.267)	(0.273)	(0.253)	(0.245)	(0.247)	(0.226)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.161	0.142	0.149	0.142	-0.086	-0.160	0.185	0.144	0.030	-0.029
	[0.303]	[0.314]	[0.305]	[0.276]	[0.331]	[0.338]	[0.308]	[0.297]	[0.315]	[0.288]
Panel D: P75										
Treatment (β_1)	-0.331**	-0.372**	-0.085	-0.108	-0.253	-0.253	-0.343*	-0.367**	-0.276*	-0.282*
	(0.166)	(0.172)	(0.142)	(0.148)	(0.163)	(0.169)	(0.178)	(0.181)	(0.145)	(0.163)
Male (β_2)	-0.065	-0.003	0.166	0.134	0.059	0.077	0.004	0.034	-0.004	0.024
	(0.154)	(0.187)	(0.140)	(0.148)	(0.182)	(0.169)	(0.148)	(0.168)	(0.146)	(0.162)
Treatment × Male (β_3)	0.226	0.219	-0.071	-0.070	0.341	0.295	0.292	0.294	0.279	0.225
	(0.203)	(0.220)	(0.188)	(0.206)	(0.233)	(0.228)	(0.220)	(0.235)	(0.185)	(0.202)
Treatment Effect on Male ($\beta_1 + \beta_3$)	-0.105	-0.153	-0.156	-0.178	0.088	0.042	-0.051	-0.073	0.004	-0.057
	[0.255]	[0.289]	[0.235]	[0.254]	[0.295]	[0.284]	[0.265]	[0.296]	[0.235]	[0.260]
Panel E: P90										
Treatment (β_1)	-0.325**	-0.348**	-0.210	-0.241	-0.125	-0.140	-0.424***	-0.452***	-0.230	-0.277*
	(0.149)	(0.160)	(0.143)	(0.167)	(0.156)	(0.154)	(0.151)	(0.151)	(0.145)	(0.155)
Male (β_2)	0.006	0.086	0.088	0.071	0.239	0.250	-0.000	0.035	0.153	0.158
	(0.168)	(0.194)	(0.157)	(0.160)	(0.163)	(0.170)	(0.193)	(0.185)	(0.184)	(0.177)
Treatment × Male (β_3)	0.142	0.154	0.122	0.159	0.169	0.127	0.394	0.397*	0.210	0.230
	(0.212)	(0.231)	(0.219)	(0.230)	(0.224)	(0.225)	(0.244)	(0.231)	(0.214)	(0.220)
Treatment Effect on Male ($\beta_1 + \beta_3$)	-0.183	-0.194	-0.087	-0.082	0.044	-0.013	-0.030	-0.055	-0.020	-0.047
	[0.271]	[0.301]	[0.262]	[0.280]	[0.273]	[0.282]	[0.287]	[0.296]	[0.258]	[0.270]
Obs	556	556	545	545	560	560	541	541	560	560
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate unconditional quantile regressions (UQR). All dependent variables are standardized. P_i represents UQR results at ith percentile. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns only include a treatment indicator, gender, and their interaction. Even numbered columns additionally control for student characteristics (age, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Bootstrapped standard errors with 200 replications are reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.5: Effects on Exam Grades (UQR)- Goals vs. Control

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: P10										
Treatment (β_1)	0.312	0.315	-0.134	-0.142	0.065	0.069	0.169	0.128	0.106	0.083
	(0.297)	(0.288)	(0.256)	(0.268)	(0.247)	(0.211)	(0.237)	(0.248)	(0.213)	(0.201)
Male(β_2)	0.173	0.206	-0.049	-0.037	0.288	0.333*	0.337	0.405*	0.192	0.267
	(0.344)	(0.307)	(0.248)	(0.266)	(0.210)	(0.194)	(0.217)	(0.244)	(0.215)	(0.214)
Treatment× Male (β_3)	-0.125	-0.051	0.079	0.108	-0.198	-0.142	-0.351	-0.295	-0.104	-0.054
	(0.401)	(0.386)	(0.347)	(0.371)	(0.282)	(0.272)	(0.293)	(0.322)	(0.273)	(0.252)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.188	0.264	-0.056	-0.034	-0.133	-0.074	-0.182	-0.167	0.002	0.029
	[0.465]	[0.452]	[0.437]	[0.385]	[0.358]	[0.337]	[0.357]	[0.411]	[0.336]	[0.347]
Panel B: P25										
Treatment (β_1)	0.148	0.137	0.067	0.077	0.071	0.096	0.184	0.178	0.064	0.086
	(0.172)	(0.172)	(0.206)	(0.244)	(0.197)	(0.182)	(0.185)	(0.196)	(0.162)	(0.170)
Male(β_2)	0.009	0.089	-0.111	-0.029	0.346*	0.498**	0.160	0.207	0.099	0.158
	(0.172)	(0.178)	(0.212)	(0.234)	(0.184)	(0.195)	(0.179)	(0.173)	(0.149)	(0.172)
Treatment× Male (β_3)	-0.112	-0.125	-0.238	-0.258	-0.230	-0.279	-0.309	-0.298	-0.077	-0.099
	(0.249)	(0.224)	(0.285)	(0.320)	(0.244)	(0.229)	(0.239)	(0.243)	(0.209)	(0.227)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.036	0.013	-0.171	-0.181	-0.159	-0.184	-0.125	-0.120	-0.013	-0.013
	[0.286]	[0.302]	[0.357]	[0.373]	[0.311]	[0.330]	[0.337]	[0.326]	[0.264]	[0.273]
Panel C: P50										
Treatment (β_1)	-0.022	-0.036	0.171	0.152	-0.116	-0.079	0.135	0.135	0.131	0.159
	(0.201)	(0.213)	(0.204)	(0.174)	(0.177)	(0.178)	(0.184)	(0.179)	(0.175)	(0.181)
Male(β_2)	-0.103	-0.031	0.020	0.101	0.151	0.294	0.053	0.173	0.180	0.288
	(0.195)	(0.198)	(0.192)	(0.182)	(0.175)	(0.200)	(0.166)	(0.182)	(0.161)	(0.195)
Treatment× Male (β_3)	0.109	0.101	-0.098	-0.104	0.168	0.031	-0.090	-0.161	-0.132	-0.233
	(0.260)	(0.268)	(0.248)	(0.232)	(0.246)	(0.253)	(0.249)	(0.238)	(0.225)	(0.248)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.087	0.065	0.073	0.048	0.052	-0.048	0.045	-0.026	-0.001	-0.074
	[0.342]	[0.348]	[0.289]	[0.316]	[0.320]	[0.314]	[0.312]	[0.328]	[0.301]	[0.282]
Panel D: P75										
Treatment (β_1)	0.027	0.056	0.186	0.198	-0.091	-0.044	0.063	0.080	-0.165	-0.140
	(0.148)	(0.151)	(0.149)	(0.142)	(0.157)	(0.157)	(0.170)	(0.169)	(0.153)	(0.144)
Male(β_2)	0.046	0.149	0.117	0.165	0.058	0.106	-0.084	-0.059	-0.039	0.010
	(0.148)	(0.152)	(0.144)	(0.157)	(0.162)	(0.192)	(0.172)	(0.170)	(0.164)	(0.167)
Treatment× Male (β_3)	-0.025	-0.114	-0.075	-0.135	0.052	0.008	0.050	0.047	0.199	0.155
	(0.191)	(0.221)	(0.203)	(0.209)	(0.214)	(0.220)	(0.243)	(0.238)	(0.210)	(0.201)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.002	-0.058	0.111	0.062	-0.039	-0.036	0.113	0.127	0.033	0.015
	[0.270]	[0.298]	[0.233]	[0.261]	[0.284]	[0.311]	[0.281]	[0.302]	[0.262]	[0.277]
Panel E: P90										
Treatment (β_1)	-0.040	-0.025	-0.041	-0.050	0.166	0.206	0.021	0.034	0.143	0.149
	(0.133)	(0.148)	(0.129)	(0.118)	(0.152)	(0.164)	(0.127)	(0.131)	(0.155)	(0.159)
Male(β_2)	0.046	0.130	0.077	0.066	0.239	0.259	0.098	0.136	0.161	0.194
	(0.137)	(0.148)	(0.142)	(0.150)	(0.180)	(0.171)	(0.147)	(0.147)	(0.149)	(0.162)
Treatment× Male (β_3)	-0.044	-0.094	0.054	0.062	-0.198	-0.212	-0.069	-0.117	-0.186	-0.212
	(0.180)	(0.199)	(0.174)	(0.175)	(0.236)	(0.243)	(0.189)	(0.198)	(0.213)	(0.227)
Treatment Effect on Male ($\beta_1 + \beta_3$)	-0.085	-0.119	0.013	0.011	-0.032	-0.006	-0.048	-0.083	-0.043	-0.063
	[0.253]	[0.231]	[0.233]	[0.242]	[0.278]	[0.299]	[0.238]	[0.255]	[0.258]	[0.261]
Obs	599	599	592	592	602	602	589	589	602	602
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate unconditional quantile regressions (UQR). All dependent variables are standardized. P_i represents UQR results at ith percentile. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns only include treatment indicator, gender, and their interaction. Even numbered columns additionally control for student characteristics (age, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Bootstrapped standard errors with 200 replications are reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.6: Mechanism for Females - Social Goals vs. Goals

	Week 6 Exam (1)	Week 10 Exam (2)	Final Exam (3)	Average Exam (4)	Overall (5)	Week 6 Exam (1)	Week 10 Exam (2)	Final Exam (3)	Average Exam (4)	Overall (5)
	Panel A: Female					*Panel B: Male*				
	Panel A.1: Low-performing students					*Panel B.1: Low-performing students*				
Can Overcome Difficulties	-0.078	-0.144	-0.072	-0.117	-0.110	-0.018	-0.119	-0.041	-0.111	-0.051
	(0.094)	(0.099)	(0.086)	(0.079)	(0.085)	(0.064)	(0.100)	(0.068)	(0.079)	(0.081)
Obs	99	95	102	93	96	120	114	122	109	118
Course Platform Accesses	0.276	0.393**	-0.072	0.252	0.302	0.060	0.541*	-0.081	0.132	0.151
	(0.207)	(0.179)	(0.210)	(0.204)	(0.199)	(0.144)	(0.304)	(0.132)	(0.162)	(0.141)
Obs	134	124	140	125	127	151	153	161	147	155
	Panel A.2: High-performing students					*Panel B.2: High-performing students*				
Can Overcome Difficulties	-0.157**	-0.115**	-0.162**	-0.145***	-0.118**	-0.035	0.019	-0.010	0.004	-0.013
	(0.061)	(0.049)	(0.064)	(0.047)	(0.053)	(0.043)	(0.045)	(0.048)	(0.047)	(0.045)
Obs	102	104	100	105	106	134	135	133	139	137
Course Platform Accesses	0.175	0.067	0.249	0.260*	0.209	0.043	0.007	0.162	0.135	0.075
	(0.134)	(0.128)	(0.221)	(0.140)	(0.185)	(0.176)	(0.113)	(0.189)	(0.189)	(0.176)
Obs	122	130	117	128	130	150	143	143	146	149
Student Characteristics	X	X	X	X	X	X	X	X	X	X
Tutor FE	X	X	X	X	X	X	X	X	X	X

Notes: Each row presents estimates from separate OLS models run on the subsample of same-gender students with grades (showed on the columns) above the respective medians. The dependent variable is in the left-most row. [For instance, columns (1)-(2) in Panel A refers to the subsample of female students with first midterm grades above the median, whereas columns (1)-(2) in Panel B refers to the subsample of male students with first midterm grades above the median.] *Can Overcome Difficulties* is a binary variable which equals one if one considers himself able to overcome difficulties and zero otherwise. *Course Platform Accesses* is standardized. All columns only include a treatment indicator. Standard errors are clustered by tutor and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.7: Mechanism for Females - Social Goals vs. Control

	Week 6 Exam (1)	Week 10 Exam (2)	Final Exam (3)	Average Exam (4)	Overall (5)	Week 6 Exam (1)	Week 10 Exam (2)	Final Exam (3)	Average Exam (4)	Overall (5)
	Panel A: Female					*Panel B: Male*				
	Panel A.1: Low-performing students					*Panel B.1: Low-performing students*				
Can Overcome Difficulties	-0.016	-0.036	-0.049	-0.033	-0.056	-0.118*	-0.138	-0.119	-0.184*	-0.161
	(0.101)	(0.098)	(0.084)	(0.100)	(0.096)	(0.066)	(0.116)	(0.074)	(0.090)	(0.097)
Obs	88	90	89	81	88	105	97	103	89	95
Course Platform Accesses	0.153	0.301	0.116	0.332	0.278	0.152	0.516	-0.090	0.257	0.241*
	(0.255)	(0.178)	(0.195)	(0.246)	(0.251)	(0.137)	(0.311)	(0.140)	(0.151)	(0.135)
Obs	118	117	118	113	119	148	144	147	135	143
	Panel A.2: High-performing students					*Panel B.2: High-performing students*				
Can Overcome Difficulties	-0.114*	-0.129***	-0.087	-0.122**	-0.072	-0.034	-0.045	-0.033	-0.003	-0.006
	(0.054)	(0.040)	(0.063)	(0.050)	(0.048)	(0.078)	(0.057)	(0.069)	(0.085)	(0.077)
Obs	91	87	91	95	92	107	108	110	115	118
Course Platform Accesses	0.178	0.041	0.185	0.177	0.145	0.223	-0.064	0.343*	0.166	0.140
	(0.148)	(0.171)	(0.146)	(0.166)	(0.166)	(0.156)	(0.212)	(0.177)	(0.156)	(0.154)
Obs	108	106	109	109	108	133	129	136	136	140
Student Characteristics	X	X	X	X	X	X	X	X	X	X
Tutor FE	X	X	X	X	X	X	X	X	X	X

Notes: Each row presents estimates from separate OLS models run on the subsample of same-gender students with grades (showed on the columns) above the respective medians. The dependent variable is in the left-most row. [For instance, columns (1)-(2) in Panel A refers to the subsample of female students with first midterm grades above the median, whereas columns (1)-(2) in Panel B refers to the subsample of male students with first midterm grades above the median.] *Can Overcome Difficulties* is a binary variable which equals one if one considers himself able to overcome difficulties and zero otherwise. *Course Platform Accesses* is standardized. All columns only include a treatment indicator. Standard errors are clustered by tutor and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.8: Mechanism for Females - Goals vs. Control

	Week 6 Exam (1)	Week 10 Exam (2)	Final Exam (3)	Average Exam (4)	Overall (5)	Week 6 Exam (1)	Week 10 Exam (2)	Final Exam (3)	Average Exam (4)	Overall (5)
	Panel A: Female					*Panel B: Male*				
	Panel A.1: Low-performing students					*Panel B.1: Low-performing students*				
Can Overcome Difficulties	0.062	0.108	0.022	0.083	0.054	-0.100**	-0.019	-0.078	-0.074	-0.110*
	(0.049)	(0.079)	(0.082)	(0.070)	(0.072)	(0.038)	(0.064)	(0.058)	(0.059)	(0.059)
Obs	87	89	87	84	84	115	117	113	110	111
Course Platform Accesses	-0.123	-0.092	0.188	0.080	-0.024	0.092	-0.025	-0.009	0.125	0.090
	(0.181)	(0.061)	(0.208)	(0.178)	(0.219)	(0.107)	(0.053)	(0.165)	(0.137)	(0.119)
Obs	118	121	128	120	120	149	145	148	144	146
	Panel A.2: High-performing students					*Panel B.2: High-performing students*				
Can Overcome Difficulties	0.043	-0.013	0.076	0.022	0.046	0.000	-0.064	-0.023	-0.007	0.007
	(0.044)	(0.049)	(0.049)	(0.044)	(0.048)	(0.082)	(0.066)	(0.070)	(0.091)	(0.085)
Obs	103	101	105	104	108	121	113	123	120	125
Course Platform Accesses	0.003	-0.026	-0.064	-0.083	-0.064	0.180*	-0.070	0.181	0.031	0.065
	(0.181)	(0.139)	(0.155)	(0.174)	(0.171)	(0.097)	(0.192)	(0.199)	(0.166)	(0.144)
Obs	136	132	128	131	136	143	142	145	142	147
Student Characteristics	X	X	X	X	X	X	X	X	X	X
Tutor FE	X	X	X	X	X	X	X	X	X	X

Notes: Each row presents estimates from separate OLS models run on the subsample of same-gender students with grades (showed on the columns) above the respective medians. The dependent variable is in the left-most row. [For instance, columns (1)-(2) in Panel A refers to the subsample of female students with first midterm grades above the median, whereas columns (1)-(2) in Panel B refers to the subsample of male students with first midterm grades above the median.] *Can Overcome Difficulties* is a binary variable which equals one if one considers himself able to overcome difficulties and zero otherwise. *Course Platform Accesses* is standardized. All columns only include a treatment indicator. Standard errors are clustered by tutor and reported in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.9: Mechanism for Males - Goals vs. Control

	Week 9		Week 10		Week 11		Week 12		Week 13		Total	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of Assignment Accesses	*0.131*#	0.113	*0.150*#	0.123	0.190**	0.142	0.128	0.183	0.195*	0.164	0.030	-0.046
	(0.077)	(0.075)	(0.094)	(0.130)	(0.086)	(0.094)	(0.133)	(0.154)	(0.100)	(0.120)	(0.119)	(0.131)
Obs	333	333	333	333	333	333	333	333	333	333	333	333
All Online Assignment Submissions	0.220	*0.239*#	0.241*	0.262*	0.256**	0.249**	-0.007	0.066	-0.089	-0.099	0.132	0.132
	(0.163)	(0.139)	(0.120)	(0.130)	(0.094)	(0.091)	(0.115)	(0.135)	(0.097)	(0.116)	(0.124)	(0.128)
Obs	332	332	332	332	332	332	332	332	332	332	332	332
Correct Online Assignment Submissions	0.219	0.251*	0.218*	0.227*	0.228**	0.220**	-0.022	0.051	-0.099	-0.119	0.078	0.071
	(0.155)	(0.139)	(0.117)	(0.126)	(0.097)	(0.098)	(0.109)	(0.126)	(0.088)	(0.104)	(0.125)	(0.126)
Obs	332	332	332	332	332	332	332	332	332	332	332	332
Incorrect Online Assignment Submissions	0.204	0.203	0.253*	0.289**	0.273***	0.267***	0.012	0.081	-0.070	-0.065	*0.188*#	0.196
	(0.169)	(0.137)	(0.119)	(0.132)	(0.084)	(0.081)	(0.124)	(0.147)	(0.109)	(0.130)	(0.123)	(0.130)
Obs	332	332	332	332	332	332	332	332	332	332	332	332
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models on the male subsample. The dependent variables are listed in the left-most column. Columns (1)-(2) present the variables recorded at the end of Week 9. Columns (3)-(4), (5)-(6), (7)-(8), (9)-(10), and (11)-(12) present the variables recorded in Week 10, 11, 12, 13, and over the whole semester, respectively. All dependent variables are standardized. Odd numbered columns only include a treatment indicator. Even numbered columns additionally control for student characteristics (age, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. #, *, **, and *** denote significance at the 15%. 10%, 5%, and 1% level, respectively.

Table C.10: Mechanism for Males - Social Goals vs. Control

	Week 9		Week 10		Week 11		Week 12		Week 13		Total	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of Assignment Accesses	0.181*	0.147*	0.118	0.090	0.116	0.074	0.186**	0.218**	0.110	0.111	0.151	0.059
	(0.085)	(0.071)	(0.122)	(0.153)	(0.090)	(0.088)	(0.082)	(0.092)	(0.082)	(0.095)	(0.170)	(0.160)
Obs	319	319	319	319	319	319	319	319	319	319	319	319
All Online Assignment Submissions	0.121	*0.172*#	0.059	0.030	0.136*	0.139*	0.117	0.151	-0.057	-0.045	0.046	0.008
	(0.089)	(0.113)	(0.091)	(0.111)	(0.071)	(0.072)	(0.109)	(0.109)	(0.088)	(0.095)	(0.138)	(0.132)
Obs	316	316	316	316	316	316	316	316	316	316	316	316
Correct Online Assignment Submissions	0.096	0.149	0.036	0.001	*0.111*#	0.107	0.120	0.151	-0.075	-0.065	-0.017	-0.065
	(0.088)	(0.110)	(0.093)	(0.113)	(0.071)	(0.072)	(0.111)	(0.106)	(0.090)	(0.097)	(0.138)	(0.130)
Obs	316	316	316	316	316	316	316	316	316	316	316	316
Incorrect Online Assignment Submissions	*0.148*#	*0.191*#	0.086	0.067	0.160**	0.171**	0.106	0.143	-0.029	-0.016	0.120	0.097
	(0.090)	(0.113)	(0.084)	(0.104)	(0.070)	(0.073)	(0.107)	(0.113)	(0.085)	(0.095)	(0.129)	(0.127)
Obs	316	316	316	316	316	316	316	316	316	316	316	316
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models on the male subsample. The dependent variables are listed in the left-most column. Columns (1)-(2) present the variables recorded at the end of Week 9. Columns (3)-(4), (5)-(6), (7)-(8), (9)-(10), and (11)-(12) present the variables recorded in Week 10, 11, 12, 13, and over the whole semester, respectively. All dependent variables are standardized. Odd numbered columns only include a treatment indicator. Even numbered columns additionally control for student characteristics (age, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. #, *, **, and *** denote significance at the 15%. 10%, 5%, and 1% level, respectively.

Table C.11: Mechanism for Males - Social Goals vs. Goals

	Week 9		Week 10		Week 11		Week 12		Week 13		Total	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Number of Assignment Accesses	0.050	0.035	-0.032	-0.031	-0.074	-0.067	0.058	0.007	-0.085	-0.100	0.121	0.056
	(0.072)	(0.067)	(0.122)	(0.126)	(0.070)	(0.084)	(0.087)	(0.095)	(0.130)	(0.151)	(0.118)	(0.127)
Obs	354	354	354	354	354	354	354	354	354	354	354	354
All Online Assignment Submissions	-0.099	-0.111	-0.182	-0.143	-0.120	-0.095	0.123	0.041	0.033	0.019	-0.086	-0.139*
	(0.185)	(0.212)	(0.156)	(0.153)	(0.104)	(0.117)	(0.125)	(0.141)	(0.097)	(0.114)	(0.070)	(0.065)
Obs	350	350	350	350	350	350	350	350	350	350	350	350
Correct Online Assignment Submissions	-0.124	-0.140	-0.182	-0.145	-0.117	-0.091	0.142	0.068	0.025	0.016	-0.095	-0.151**
	(0.176)	(0.204)	(0.150)	(0.151)	(0.101)	(0.112)	(0.114)	(0.130)	(0.092)	(0.106)	(0.068)	(0.062)
Obs	350	350	350	350	350	350	350	350	350	350	350	350
Incorrect Online Assignment Submissions	-0.056	-0.062	-0.167	-0.128	-0.113	-0.092	0.094	0.008	0.041	0.022	-0.068	-0.113
	(0.186)	(0.214)	(0.153)	(0.146)	(0.102)	(0.114)	(0.136)	(0.151)	(0.100)	(0.120)	(0.078)	(0.075)
Obs	350	350	350	350	350	350	350	350	350	350	350	350
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each row presents estimates from separate OLS models on the male subsample. The dependent variables are listed in the left-most column. Columns (1)-(2) present the variables recorded at the end of Week 9. Columns (3)-(4), (5)-(6), (7)-(8), (9)-(10), and (11)-(12) present the variables recorded in Week 10, 11, 12, 13, and over the whole semester, respectively. All dependent variables are standardized. Odd numbered columns only include a treatment indicator. Even numbered columns additionally control for student characteristics (age, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. #, *, **, and *** denote significance at the 15%. 10%, 5%, and 1% level, respectively.

Table C.12: Country of Birth: Effects on Exam Grades - Social Goals vs. Goals

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Females										
Treatment (β_1)	-0.016	0.001	-0.073	-0.112	0.255	0.334	0.045	0.068	0.123	0.164
	(0.236)	(0.239)	(0.270)	(0.304)	(0.239)	(0.266)	(0.260)	(0.282)	(0.238)	(0.264)
Chinese (β_2)	0.184	0.197	-0.338	-0.493*	0.009	-0.005	-0.065	-0.124	-0.044	-0.087
	(0.197)	(0.211)	(0.247)	(0.244)	(0.228)	(0.268)	(0.229)	(0.245)	(0.205)	(0.232)
Treatment × Chinese (β_3)	0.109	0.128	-0.303	-0.391	-0.017	0.031	-0.088	-0.097	-0.063	-0.054
	(0.206)	(0.244)	(0.219)	(0.235)	(0.213)	(0.267)	(0.225)	(0.260)	(0.199)	(0.244)
Treatment Effect on Chinese Students ($\beta_1 + \beta_3$)	0.093	0.129	-0.375	-0.503	0.238	0.365	-0.043	-0.029	0.060	0.110
	[0.388]	[0.430]	[0.444]	[0.444]	[0.429]	[0.478]	[0.420]	[0.494]	[0.394]	[0.446]
Obs	197	197	198	198	198	198	197	197	198	198
Panel B: Males										
Treatment (β_1)	0.057	0.058	0.209*	0.216*	0.047	0.072	0.141	0.167	0.097	0.111
	(0.109)	(0.112)	(0.116)	(0.121)	(0.108)	(0.108)	(0.099)	(0.108)	(0.071)	(0.079)
Chinese (β_2)	-0.302***	-0.335***	-0.343*	-0.415**	-0.421***	-0.302**	-0.421***	-0.420***	-0.382***	-0.310**
	(0.085)	(0.111)	(0.162)	(0.171)	(0.104)	(0.133)	(0.109)	(0.126)	(0.087)	(0.110)
Treatment × Chinese (β_3)	-0.292**	-0.305**	-0.376***	-0.392***	-0.438***	-0.414***	-0.429***	-0.430***	-0.396***	-0.380***
	(0.099)	(0.110)	(0.116)	(0.106)	(0.107)	(0.115)	(0.085)	(0.097)	(0.092)	(0.106)
Treatment Effect on Chinese Students ($\beta_1 + \beta_3$)	-0.235	-0.247	-0.167	-0.176	-0.391**	-0.342*	-0.288*	-0.263	-0.299*	-0.269
	[0.150]	[0.190]	[0.203]	[0.242]	[0.171]	[0.166]	[0.170]	[0.187]	[0.156]	[0.185]
Obs	259	259	254	254	261	261	252	252	261	261
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns only include a treatment indicator, a dummy denoting whether a student is Australian-born or Chinese-born, and their interaction. Even numbered columns additionally control for student characteristics (age, gender, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.13: Country of Birth: Effects on Exam Grades - Social Goals vs. Control

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Females										
Treatment (β_1)	-0.158	-0.081	0.264	0.273	0.234	0.220	0.102	0.127	0.160	0.170
	(0.295)	(0.332)	(0.316)	(0.337)	(0.264)	(0.285)	(0.289)	(0.335)	(0.240)	(0.282)
Chinese (β_2)	-0.057	0.103	0.005	0.013	-0.091	-0.078	-0.075	-0.003	-0.070	-0.020
	(0.262)	(0.328)	(0.258)	(0.287)	(0.260)	(0.276)	(0.255)	(0.313)	(0.229)	(0.273)
Chinese$\times$ Treatment (β_3)	-0.032	0.045	0.034	0.062	-0.037	-0.023	-0.031	0.012	-0.026	0.006
	(0.213)	(0.224)	(0.147)	(0.156)	(0.209)	(0.207)	(0.164)	(0.184)	(0.156)	(0.170)
Treatment Effect on Chinese Students ($\beta_1 + \beta_3$)	-0.190	-0.036	0.297	0.335	0.197	0.198	0.071	0.139	0.134	0.176
	[0.468]	[0.542]	[0.405]	[0.425]	[0.449]	[0.447]	[0.412]	[0.471]	[0.374]	[0.430]
Obs	181	181	180	180	182	182	179	179	182	182
Panel B: Males										
Treatment (β_1)	0.076	0.108	0.343*	0.242*	0.110	0.103	0.211	0.202	0.186	0.145
	(0.147)	(0.176)	(0.168)	(0.130)	(0.129)	(0.167)	(0.124)	(0.120)	(0.130)	(0.149)
Chinese (β_2)	-0.509**	-0.388*	-0.316	-0.285	-0.368**	-0.311*	-0.497**	-0.380**	-0.374**	-0.317
	(0.184)	(0.196)	(0.225)	(0.175)	(0.139)	(0.163)	(0.179)	(0.167)	(0.174)	(0.181)
Chinese$\times$ Treatment (β_3)	-0.273	-0.205	-0.242	-0.203	-0.374*	-0.380	-0.359*	-0.296	-0.307	-0.294
	(0.212)	(0.239)	(0.161)	(0.157)	(0.192)	(0.223)	(0.189)	(0.204)	(0.193)	(0.221)
Treatment Effect on Chinese Students ($\beta_1 + \beta_3$)	-0.197	-0.097	0.101	0.039	-0.264	-0.277	-0.148	-0.094	-0.121	-0.149
	[0.347]	[0.414]	[0.354]	[0.299]	[0.300]	[0.359]	[0.351]	[0.315]	[0.319]	[0.359]
Obs	236	236	230	230	238	238	228	228	238	238
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns only include a treatment indicator, a dummy denoting whether a student is Australian-born or Chinese-born, and their interaction. Even numbered columns additionally control for student characteristics (age, gender, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.14: Country of Birth: Effects on Exam Grades - Goals vs. Control

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Females										
Treatment (β_1)	-0.141	-0.152	0.336	0.406	-0.020	-0.014	0.057	0.075	0.037	0.057
	(0.261)	(0.303)	(0.283)	(0.304)	(0.354)	(0.402)	(0.298)	(0.347)	(0.289)	(0.342)
Chinese (β_2)	-0.057	0.025	0.005	0.004	-0.091	-0.000	-0.075	-0.023	-0.070	0.011
	(0.262)	(0.341)	(0.258)	(0.302)	(0.260)	(0.290)	(0.254)	(0.323)	(0.229)	(0.281)
Chinese$\times$ Treatment (β_3)	0.042	0.089	-0.002	-0.069	-0.012	-0.006	-0.008	-0.029	-0.007	-0.003
	(0.205)	(0.259)	(0.247)	(0.273)	(0.188)	(0.217)	(0.197)	(0.248)	(0.166)	(0.211)
Treatment Effect on Chinese Students ($\beta_1 + \beta_3$)	-0.099	-0.063	0.334	0.336	-0.032	-0.020	0.048	0.047	0.030	0.054
	[0.444]	[0.572]	[0.470]	[0.556]	[0.519]	[0.494]	[0.436]	[0.538]	[0.424]	[0.477]
Obs	202	202	202	202	204	204	200	200	204	204
Panel B: Males										
Treatment (β_1)	0.019	0.043	0.134	0.147	0.064	0.066	0.070	0.110	0.089	0.071
	(0.179)	(0.199)	(0.203)	(0.221)	(0.157)	(0.182)	(0.182)	(0.217)	(0.168)	(0.190)
Chinese (β_2)	-0.509**	-0.400*	-0.316	-0.249	-0.368**	-0.270	-0.497**	-0.358	-0.374**	-0.292
	(0.184)	(0.193)	(0.225)	(0.231)	(0.139)	(0.171)	(0.178)	(0.208)	(0.174)	(0.192)
Chinese$\times$ Treatment (β_3)	-0.283	-0.272	-0.209	-0.210	-0.357**	-0.241	-0.351*	-0.284	-0.293	-0.236
	(0.183)	(0.230)	(0.239)	(0.278)	(0.165)	(0.219)	(0.195)	(0.244)	(0.178)	(0.223)
Treatment Effect on Chinese Students ($\beta_1 + \beta_3$)	-0.264	-0.229	-0.075	-0.063	-0.293	-0.175	-0.281	-0.175	-0.204	-0.165
	[0.340]	[0.412]	[0.413]	[0.470]	[0.276]	[0.372]	[0.352]	[0.426]	[0.334]	[0.399]
Obs	253	253	248	248	253	253	248	248	253	253
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns only include a treatment indicator, a dummy denoting whether a student is Australian-born or Chinese-born, and their interaction. Even numbered columns additionally control for student characteristics (age, gender, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.15: Field of Study: Effects on Exam Grades - Social Goals vs. Goals

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Females										
Treatment (β_1)	-0.113	-0.035	0.046	0.123	0.234	0.321	0.046	0.142	0.104	0.194
	(0.189)	(0.189)	(0.212)	(0.238)	(0.250)	(0.301)	(0.212)	(0.241)	(0.221)	(0.261)
Business (β_2)	0.412*	0.408*	0.594***	0.657***	0.514***	0.567**	0.601***	0.653***	0.543***	0.593***
	(0.213)	(0.195)	(0.132)	(0.164)	(0.168)	(0.207)	(0.173)	(0.193)	(0.161)	(0.185)
Business× Treatment (β_3)	0.217	0.279	0.448**	0.563***	0.209	0.299	0.372**	0.487**	0.277*	0.382**
	(0.186)	(0.177)	(0.152)	(0.167)	(0.143)	(0.181)	(0.154)	(0.172)	(0.134)	(0.154)
Treatment Effect on Business Students ($\beta_1 + \beta_3$)	0.105	0.244	0.494	0.687	0.443	0.621	0.417	0.629	0.381	0.576
	[0.383]	[0.350]	[0.256]	[0.372]	[0.343]	[0.420]	[0.317]	[0.373]	[0.286]	[0.371]
Obs	263	263	261	261	264	264	260	260	264	264
Panel B: Males										
Treatment (β_1)	-0.072	-0.098	0.084	0.064	-0.081	-0.160	-0.002	-0.041	-0.066	-0.146
	(0.113)	(0.120)	(0.124)	(0.126)	(0.130)	(0.139)	(0.110)	(0.117)	(0.108)	(0.126)
Business (β_2)	-0.001	0.068	0.020	0.105	-0.156	-0.077	-0.041	0.054	-0.104	-0.040
	(0.166)	(0.208)	(0.185)	(0.201)	(0.133)	(0.193)	(0.155)	(0.207)	(0.141)	(0.206)
Business× Treatment (β_3)	0.085	0.192	0.157	0.220	0.046	0.129	0.109	0.214	0.075	0.144
	(0.115)	(0.130)	(0.122)	(0.128)	(0.131)	(0.148)	(0.118)	(0.134)	(0.127)	(0.144)
Treatment Effect on Business Students ($\beta_1 + \beta_3$)	0.013	0.094	0.241	0.285	-0.035	-0.031	0.107	0.173	0.008	-0.002
	[0.244]	[0.204]	[0.236]	[0.251]	[0.213]	[0.283]	[0.204]	[0.280]	[0.189]	[0.228]
Obs	350	350	346	346	354	354	342	342	354	354
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns only include a treatment indicator, a dummy denoting whether a student is pursuing a business degree, and their interaction. Even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.16: Field of Study: Effects on Exam Grades - Social Goals vs. Control

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Females										
Treatment (β_1)	-0.118	-0.116	-0.095	-0.048	0.084	0.069	-0.060	-0.043	-0.022	-0.016
	(0.168)	(0.186)	(0.151)	(0.176)	(0.200)	(0.221)	(0.157)	(0.180)	(0.161)	(0.177)
Business (β_2)	0.249	0.220	0.223	0.286	0.206	0.261	0.286	0.324	0.235	0.286
	(0.202)	(0.228)	(0.202)	(0.215)	(0.188)	(0.206)	(0.205)	(0.229)	(0.179)	(0.203)
Business× Treatment (β_3)	0.213	0.216	0.306*	0.392**	0.059	0.100	0.266	0.331*	0.151	0.206
	(0.176)	(0.210)	(0.160)	(0.142)	(0.163)	(0.179)	(0.153)	(0.158)	(0.133)	(0.143)
Treatment Effect on Business Students ($\beta_1 + \beta_3$)	0.095	0.099	0.211	0.344	0.142	0.168	0.206	0.288	0.129	0.190
	[0.326]	[0.370]	[0.269]	[0.273]	[0.316]	[0.345]	[0.339]	[0.280]	[0.246]	[0.324]
Obs	240	240	237	237	241	241	236	236	241	241
Panel B: Males										
Treatment (β_1)	0.127	0.104	0.123	0.042	-0.119	-0.192	0.097	0.040	-0.022	-0.105
	(0.165)	(0.180)	(0.077)	(0.079)	(0.125)	(0.146)	(0.112)	(0.124)	(0.111)	(0.127)
Business (β_2)	0.242	0.452**	0.172	0.176	-0.094	0.023	0.181	0.343	0.063	0.169
	(0.160)	(0.176)	(0.152)	(0.181)	(0.156)	(0.172)	(0.174)	(0.199)	(0.149)	(0.168)
Business× Treatment (β_3)	0.284*	0.512***	0.196	0.250	0.008	0.172	0.207**	0.400***	0.119	0.278***
	(0.154)	(0.144)	(0.124)	(0.156)	(0.097)	(0.108)	(0.091)	(0.107)	(0.069)	(0.085)
Treatment Effect on Business Students ($\beta_1 + \beta_3$)	0.411	0.616**	0.319**	0.292	-0.111	-0.020	0.304*	0.440***	0.097	0.173
	[0.288]	[0.279]	[0.164]	[0.231]	[0.165]	[0.224]	[0.216]	[0.133]	[0.118]	[0.118]
Obs	316	316	308	308	319	319	305	305	319	319
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns only include a treatment indicator, a dummy denoting whether a student is pursuing a business degree, and their interaction. Even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.17: Field of Study: Effects on Exam Grades - Goals vs. Control

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Females										
Treatment (β_1)	-0.005	-0.048	-0.142	-0.134	-0.150	-0.156	-0.106	-0.121	-0.126	-0.140
	(0.192)	(0.200)	(0.181)	(0.192)	(0.151)	(0.183)	(0.174)	(0.179)	(0.148)	(0.166)
Business (β_2)	0.249	0.141	0.223	0.215	0.206	0.172	0.286	0.230	0.235	0.184
	(0.201)	(0.183)	(0.202)	(0.217)	(0.187)	(0.221)	(0.205)	(0.208)	(0.179)	(0.192)
Business× Treatment (β_3)	0.407**	0.308	0.452**	0.448**	0.364**	0.328*	0.495**	0.447**	0.416**	0.367**
	(0.178)	(0.200)	(0.181)	(0.183)	(0.150)	(0.157)	(0.174)	(0.180)	(0.154)	(0.154)
Treatment Effect on Business Students ($\beta_1 + \beta_3$)	0.402	0.260	0.311	0.314	0.213	0.172	0.389	0.326	0.290	0.228
	[0.349]	[0.344]	[0.337]	[0.392]	[0.250]	[0.271]	[0.302]	[0.375]	[0.255]	[0.264]
Obs	267	267	266	266	269	269	264	264	269	269
Panel B: Males										
Treatment (β_1)	0.199	0.183	0.039	0.009	-0.038	-0.019	0.099	0.085	0.044	0.039
	(0.130)	(0.134)	(0.106)	(0.115)	(0.122)	(0.128)	(0.103)	(0.103)	(0.094)	(0.092)
Business (β_2)	0.242	0.492**	0.172	0.256	-0.094	0.030	0.181	0.364	0.063	0.215
	(0.160)	(0.189)	(0.152)	(0.152)	(0.156)	(0.205)	(0.174)	(0.207)	(0.149)	(0.190)
Business× Treatment (β_3)	0.198	0.360	0.058	0.106	-0.195	-0.086	0.058	0.180	-0.060	0.043
	(0.181)	(0.223)	(0.218)	(0.217)	(0.148)	(0.171)	(0.194)	(0.226)	(0.167)	(0.193)
Treatment Effect on Business Students ($\beta_1 + \beta_3$)	0.397	0.543*	0.097	0.115	-0.233	-0.105	0.157	0.265	-0.016	0.082
	[0.268]	[0.302]	[0.288]	[0.282]	[0.268]	[0.340]	[0.330]	[0.284]	[0.283]	[0.236]
Obs	332	332	326	326	333	333	325	325	333	333
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns only include a treatment indicator, a dummy denoting whether a student is pursuing a business degree, and their interaction. Even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.18: Ability: Effects on Exam Grades

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Goals										
Panel A.1: Females										
Ability	-0.002	-0.000	0.003	0.002	-0.000	-0.002	0.000	0.000	-0.000	-0.001
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.004)
Treatment	-0.112	-0.093	0.036	0.015	-0.176	-0.201	-0.085	-0.094	-0.123	-0.136
	(0.233)	(0.229)	(0.170)	(0.176)	(0.154)	(0.207)	(0.155)	(0.162)	(0.137)	(0.163)
Ability× Treatment	-0.001	-0.001	-0.004	-0.003	0.003	0.004	-0.001	-0.000	0.000	0.001
	(0.005)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
Obs	219	219	218	218	220	220	217	217	220	220
Panel A.2: Males										
Ability	0.004	0.003	0.004	0.004	0.009***	0.007**	0.006**	0.005	0.008***	0.006**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Treatment	-0.279	-0.178	-0.094	-0.044	0.016	-0.036	-0.176	-0.137	-0.089	-0.105
	(0.264)	(0.271)	(0.220)	(0.295)	(0.264)	(0.307)	(0.268)	(0.321)	(0.269)	(0.317)
Ability× Treatment	0.005	0.003	0.003	0.002	-0.001	-0.001	0.004	0.003	0.002	0.001
	(0.005)	(0.005)	(0.003)	(0.004)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
Obs	315	315	311	311	319	319	307	307	319	319
Panel B: Social Goals vs. Control										
Panel B.1: Females										
Ability	-0.008**	-0.008*	-0.003	-0.001	-0.001	-0.002	-0.005	-0.005	-0.003	-0.004
	(0.003)	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Treatment	-0.348**	-0.316*	-0.190	-0.116	-0.304	-0.210	-0.326	-0.258	-0.309	-0.227
	(0.153)	(0.168)	(0.203)	(0.185)	(0.211)	(0.203)	(0.188)	(0.163)	(0.181)	(0.160)
Ability× Treatment	0.005	0.004	0.002	0.000	0.003	0.002	0.004	0.002	0.003	0.002
	(0.003)	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)
Obs	193	193	191	191	194	194	190	190	194	194
Panel B.2: Males										
Ability	0.019***	0.018***	0.010**	0.010*	0.020***	0.018***	0.019***	0.017***	0.020***	0.018***
	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.006)
Treatment	0.847**	0.928**	0.360	0.469	0.751**	0.661*	0.720*	0.765*	0.735*	0.719*
	(0.342)	(0.359)	(0.386)	(0.403)	(0.318)	(0.344)	(0.352)	(0.384)	(0.355)	(0.383)
Ability× Treatment	-0.011**	-0.012**	-0.003	-0.005	-0.012**	-0.011*	-0.009	-0.010	-0.011*	-0.011*
	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
Obs	283	283	275	275	286	286	272	272	286	286
Panel C: Goals vs. Control										
Panel C.1: Females										
Ability	-0.008**	-0.004	-0.003	0.002	-0.001	0.002	-0.005	0.000	-0.003	0.001
	(0.003)	(0.005)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)
Treatment	-0.236	-0.139	-0.226	-0.072	-0.128	-0.055	-0.241	-0.097	-0.186	-0.091
	(0.230)	(0.270)	(0.182)	(0.212)	(0.214)	(0.246)	(0.194)	(0.245)	(0.185)	(0.220)
Ability× Treatment	0.006	0.003	0.006	0.002	0.001	-0.001	0.005	0.002	0.003	0.001
	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.005)	(0.004)	(0.005)
Obs	224	224	223	223	226	226	221	221	226	226
Panel C.2: Males										
Ability	0.019***	0.017***	0.010**	0.011*	0.020***	0.018***	0.019***	0.018***	0.020***	0.019***
	(0.004)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.006)	(0.004)	(0.006)
Treatment	1.125***	1.000***	0.454	0.393	0.735**	0.721**	0.896**	0.820**	0.824**	0.767**
	(0.307)	(0.335)	(0.367)	(0.394)	(0.324)	(0.326)	(0.325)	(0.355)	(0.338)	(0.354)
Ability× Treatment	-0.015***	-0.014***	-0.006	-0.005	-0.011**	-0.011**	-0.013***	-0.012**	-0.012**	-0.011**
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)	(0.005)
Obs	296	296	290	290	297	297	289	289	297	297
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns include a treatment indicator, previous GPA, and their interaction. Even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.19: Procrastination: Effects on Exam Grades

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Goals										
Panel A.1: Females										
Procrastination Level	-0.127***	-0.128***	-0.090***	-0.100***	-0.086***	-0.091***	-0.119***	-0.127***	-0.103***	-0.110***
	(0.019)	(0.019)	(0.017)	(0.023)	(0.016)	(0.022)	(0.015)	(0.021)	(0.013)	(0.019)
Treatment	-0.429***	-0.344**	-0.289*	-0.211	-0.162	-0.091	-0.360***	-0.270	-0.282**	-0.205
	(0.115)	(0.131)	(0.140)	(0.166)	(0.152)	(0.191)	(0.114)	(0.157)	(0.131)	(0.169)
Procrastination× Treatment	0.063**	0.061**	0.045	0.040	0.016	0.013	0.051	0.049	0.037	0.035
	(0.026)	(0.028)	(0.029)	(0.032)	(0.037)	(0.042)	(0.031)	(0.037)	(0.032)	(0.038)
Obs	263	263	261	261	264	264	260	260	264	264
Panel A.2: Males										
Procrastination Level	-0.043***	-0.043***	-0.073***	-0.075***	-0.050**	-0.048**	-0.068***	-0.068***	-0.060***	-0.060***
	(0.014)	(0.014)	(0.015)	(0.018)	(0.017)	(0.018)	(0.013)	(0.016)	(0.015)	(0.017)
Treatment	-0.007	0.005	0.079	0.074	0.046	0.041	0.037	0.047	0.047	0.040
	(0.119)	(0.129)	(0.104)	(0.107)	(0.135)	(0.134)	(0.096)	(0.107)	(0.098)	(0.106)
Procrastination× Treatment	0.000	-0.000	0.008	0.007	-0.002	-0.011	0.006	0.002	-0.000	-0.006
	(0.021)	(0.023)	(0.025)	(0.024)	(0.029)	(0.028)	(0.022)	(0.023)	(0.024)	(0.024)
Obs	346	346	342	342	350	350	338	338	350	350
Panel B: Social Goals vs. Control										
Panel B.1: Females										
Procrastination Level	-0.148***	-0.157***	-0.138***	-0.146***	-0.112***	-0.125***	-0.159***	-0.170***	-0.135***	-0.148***
	(0.041)	(0.044)	(0.027)	(0.033)	(0.032)	(0.035)	(0.036)	(0.041)	(0.032)	(0.036)
Treatment	-0.393***	-0.406**	-0.362**	-0.313	-0.203	-0.231	-0.402**	-0.399*	-0.317	-0.332
	(0.122)	(0.140)	(0.164)	(0.197)	(0.215)	(0.222)	(0.176)	(0.201)	(0.186)	(0.204)
Procrastination× Treatment	0.084	0.091	0.093*	0.083	0.042	0.043	0.091	0.092	0.069	0.071
	(0.047)	(0.054)	(0.045)	(0.055)	(0.054)	(0.057)	(0.053)	(0.060)	(0.051)	(0.057)
Obs	240	240	237	237	241	241	236	236	241	241
Panel B.2: Males										
Procrastination Level	-0.084***	-0.082***	-0.073**	-0.061**	-0.035	-0.036	-0.078***	-0.073**	-0.063**	-0.060**
	(0.021)	(0.023)	(0.026)	(0.028)	(0.023)	(0.025)	(0.024)	(0.026)	(0.023)	(0.025)
Treatment	-0.091	-0.040	0.033	0.084	0.040	0.081	-0.015	0.048	0.009	0.053
	(0.144)	(0.154)	(0.181)	(0.186)	(0.142)	(0.160)	(0.150)	(0.165)	(0.132)	(0.149)
Procrastination× Treatment	0.041	0.034	0.008	-0.005	-0.017	-0.030	0.016	0.003	0.002	-0.010
	(0.027)	(0.029)	(0.028)	(0.030)	(0.030)	(0.032)	(0.027)	(0.030)	(0.027)	(0.029)
Obs	313	313	305	305	316	316	302	302	316	316
Panel C: Goals vs. Control										
Panel C.1: Females										
Procrastination Level	-0.148***	-0.150***	-0.138***	-0.135***	-0.112***	-0.119***	-0.159***	-0.159***	-0.135***	-0.139***
	(0.041)	(0.038)	(0.027)	(0.028)	(0.032)	(0.031)	(0.036)	(0.035)	(0.032)	(0.031)
Treatment	0.036	-0.001	-0.074	-0.059	-0.040	-0.105	-0.042	-0.067	-0.035	-0.081
	(0.163)	(0.171)	(0.081)	(0.112)	(0.123)	(0.104)	(0.117)	(0.130)	(0.108)	(0.110)
Procrastination× Treatment	0.021	0.027	0.048	0.043	0.026	0.041	0.040	0.044	0.032	0.041
	(0.046)	(0.045)	(0.032)	(0.036)	(0.031)	(0.029)	(0.038)	(0.039)	(0.032)	(0.032)
Obs	267	267	266	266	269	269	264	264	269	269
Panel C.2: Males										
Procrastination Level	-0.084***	-0.080***	-0.073**	-0.066**	-0.035	-0.040	-0.078***	-0.075**	-0.063**	-0.063**
	(0.021)	(0.023)	(0.026)	(0.029)	(0.023)	(0.025)	(0.024)	(0.026)	(0.023)	(0.025)
Treatment	-0.084	-0.052	-0.046	-0.022	-0.006	-0.001	-0.052	-0.032	-0.038	-0.023
	(0.141)	(0.148)	(0.213)	(0.217)	(0.164)	(0.154)	(0.178)	(0.175)	(0.159)	(0.151)
Procrastination× Treatment	0.040	0.032	0.001	-0.007	-0.015	-0.013	0.011	0.005	0.003	-0.000
	(0.028)	(0.032)	(0.035)	(0.037)	(0.030)	(0.031)	(0.030)	(0.032)	(0.030)	(0.031)
Obs	331	331	325	325	332	332	324	324	332	332
Student Characteristics	X	✓	X	✓	X	✓	X	✓	X	✓
Tutor FE	X	✓	X	✓	X	✓	X	✓	X	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. Odd numbered columns include a treatment indicator, procrastination level, and their interaction. Even numbered columns additionally control for student characteristics (age, gender, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode) and tutor fixed effects. Standard errors are clustered by tutor and reported in parentheses. The linear combination of $\beta_1+\beta_3$ is reported and the corresponding standard error is in brackets. *, **, and *** denote significance at the 10%, 5%, and 1% level, respectively.

Table C.20: Effects on Exam Grades: Control for Tutorial Fixed Effects or Cluster by Tutorials

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Goals										
Treatment (β_1)	-0.195*	*-0.187*#	-0.071	-0.110	-0.060	-0.063	-0.124	-0.140	-0.119	-0.120
	(0.093)	(0.119)	(0.136)	(0.123)	(0.118)	(0.112)	(0.108)	(0.121)	(0.110)	(0.108)
Male (β_2)	0.089	0.038	-0.078	-0.093	0.176*	*0.166*#	0.057	0.023	0.094	0.077
	(0.115)	(0.100)	(0.093)	(0.119)	(0.083)	(0.103)	(0.097)	(0.113)	(0.075)	(0.100)
Male× Treatment (β_3)	0.023	0.014	-0.000	-0.005	0.182**	0.170*	0.076	0.062	0.081	0.074
	(0.111)	(0.096)	(0.082)	(0.094)	(0.070)	(0.097)	(0.094)	(0.097)	(0.074)	(0.090)
Obs	613	613	607	607	618	618	602	602	618	618
Treatment Effect on Male ($\beta_1 + \beta_3$)	-0.172	-0.173	-0.071	-0.115	0.123	0.107	-0.048	-0.078	-0.039	-0.046
	[0.194]	[0.175]	[0.157]	[0.188]	[0.161]	[0.177]	[0.179]	[0.189]	[0.163]	[0.169]
Panel B: Social Goals vs. Control										
Treatment (β_1)	-0.142	-0.094	-0.050	-0.045	-0.027	-0.030	-0.086	-0.066	-0.083	-0.065
	(0.101)	(0.123)	(0.126)	(0.134)	(0.146)	(0.131)	(0.115)	(0.136)	(0.126)	(0.126)
Male (β_2)	0.029	-0.009	0.035	0.014	0.286*	0.265*	0.110	0.080	*0.174*#	0.145
	(0.092)	(0.129)	(0.092)	(0.131)	(0.138)	(0.144)	(0.091)	(0.138)	(0.101)	(0.129)
Male× Treatment (β_3)	0.113	0.098	0.099	0.096	0.239**	0.217*	0.167*	0.154	0.163*	0.145
	(0.095)	(0.123)	(0.070)	(0.108)	(0.111)	(0.120)	(0.083)	(0.126)	(0.092)	(0.115)
Obs	556	556	545	545	560	560	541	541	560	560
Treatment Effect on Male ($\beta_1 + \beta_3$)	-0.029	0.004	0.049	0.051	0.212	0.187	0.081	0.088	0.080	0.080
	[0.141]	[0.228]	[0.141]	[0.214]	[0.238]	[0.248]	[0.176]	[0.247]	[0.183]	[0.220]
Panel C: Goals vs. Control										
Treatment (β_1)	0.092	0.107	*0.093*#	0.088	0.047	0.074	0.087	0.103	0.069	0.084
	(0.088)	(0.132)	(0.058)	(0.129)	(0.075)	(0.118)	(0.063)	(0.135)	(0.061)	(0.118)
Male (β_2)	0.081	0.050	0.089	0.075	0.322**	0.335**	0.168**	0.155	0.223**	0.218*
	(0.097)	(0.124)	(0.095)	(0.127)	(0.120)	(0.132)	(0.072)	(0.127)	(0.078)	(0.117)
Male× Treatment (β_3)	0.173**	0.145	0.049	0.049	0.249**	0.260**	0.163*	0.159	0.185**	0.182*
	(0.080)	(0.115)	(0.086)	(0.116)	(0.097)	(0.099)	(0.087)	(0.115)	(0.077)	(0.096)
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.265**	0.251	0.142	0.137	0.295*	0.334*	0.250*	0.262	0.253**	0.265
	[0.123]	[0.221]	[0.121]	[0.213]	[0.193]	[0.210]	[0.134]	[0.218]	[0.135]	[0.193]
Obs	599	599	592	592	602	602	589	589	602	602
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Fixed Effects	Tutorial	Tutor	Tutorial	Tutor	Tutorial	Tutor	Tutorial	Tutor	Tutorial	Tutor
Clustering	Tutor	Tutorial	Tutor	Tutorial	Tutor	Tutorial	Tutor	Tutorial	Tutor	Tutorial

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. All columns include treatment indicator, gender, and their interaction, controlling for student characteristics (age, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode). Odd numbered columns control for tutorial fixed effects with standard errors clustered by tutor. Even numbered columns control for tutor fixed effects with standard errors clustered by tutorial. Standard errors are reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Table C.21: Effects on Exam Grades: Control for Ability or Procrastination Level

	Week 6 Exam		Week 10 Exam		Final Exam		Average Exam		Overall	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Social Goals vs. Goals										
Treatment (β_1)	-0.194**	-0.205**	-0.116	-0.137	-0.074	-0.078	-0.150	-0.171#	-0.131	-0.136
	(0.086)	(0.078)	(0.122)	(0.140)	(0.109)	(0.124)	(0.089)	(0.101)	(0.090)	(0.106)
Imputed Ability	0.002	-	0.001	-	0.004	-	0.003	-	0.003	-
S.E.	(0.002)	-	(0.002)	-	(0.003)	-	(0.002)	-	(0.002)	-
Procrastination Level	-	-0.063***	-	-0.074***	-	-0.062***	-	-0.079***	-	-0.071***
S.E.	-	(0.010)	-	(0.012)	-	(0.013)	-	(0.012)	-	(0.012)
Obs	613	609	607	603	618	614	602	598	618	614
Treatment Effect on Male ($\beta_1 + \beta_3$)	-0.190	-0.153	-0.140	-0.098	0.080	0.137	-0.106	-0.064	-0.075	0.005
S.E.	0.180	0.167	0.193	0.133	0.113	0.109	0.147	0.125	0.110	0.087
Panel B: Social Goals vs. Control										
Treatment (β_1)	-0.107	-0.106	-0.063	-0.066	-0.055	-0.038	-0.088	-0.091	-0.091	-0.074
	(0.092)	(0.086)	(0.110)	(0.128)	(0.120)	(0.139)	(0.093)	(0.108)	(0.097)	(0.115)
Imputed Ability	0.002	-	0.002	-	0.005*	-	0.004	-	0.005	-
S.E.	(0.003)	-	(0.002)	-	(0.003)	-	(0.003)	-	(0.003)	-
Procrastination Level	-	-0.079***	-	-0.075***	-	-0.062***	-	-0.086***	-	-0.076***
S.E.	-	(0.013)	-	(0.014)	-	(0.015)	-	(0.015)	-	(0.014)
Obs	556	553	545	542	560	557	541	538	560	557
Treatment Effect on Male ($\beta_1 + \beta_3$)	-0.023	0.045	0.004	0.084	0.140	0.232	0.042	0.119	0.029	0.152
S.E.	0.169	0.172	0.115	0.160	0.197	0.218	0.138	0.160	0.163	0.176
Panel C: Goals vs. Control										
Treatment (β_1)	0.099	0.120#	0.066	0.096#	0.053	0.082	0.084	0.113**	0.063	0.095#
	(0.095)	(0.078)	(0.072)	(0.062)	(0.079)	(0.072)	(0.073)	(0.052)	(0.070)	(0.055)
Imputed Ability	0.001	-	0.002	-	0.005**	-	0.003	-	0.004	-
S.E.	(0.003)	-	(0.003)	-	(0.002)	-	(0.003)	-	(0.003)	-
Procrastination Level	-	-0.088***	-	-0.082***	-	-0.063***	-	-0.093***	-	-0.080***
S.E.	-	(0.009)	-	(0.010)	-	(0.012)	-	(0.010)	-	(0.011)
Obs	599	598	592	591	602	601	589	588	602	601
Treatment Effect on Male ($\beta_1 + \beta_3$)	0.225#	0.330**	0.071	0.194	0.257*	0.380**	0.200#	0.328**	0.193#	0.339***
S.E.	0.133	0.140	0.156	0.161	0.170	0.135	0.126	0.123	0.119	0.143
Student Characteristics	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Tutor FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Each panel presents estimates from separate OLS regressions. The dependent variable in column (1)-(2) is first midterm grade (standardized). The dependent variables in columns (3)-(4), (5)-(6), (7)-(8), and (9)-(10) refer to the same outcome formulations for the second midterm, final exam, average exam, and overall course grade. All columns include treatment indicator, gender, and their interaction, controlling for student characteristics (age, dummies for countries of birth groups, a dummy denoting whether a student is enrolled in an Economics degree and a dummy denoting whether a student is in full-time mode). Odd numbered columns additionally control for imputed ability. Even numbered columns additionally control for procrastination level. Standard errors are reported in parentheses. The linear combination of $\beta_1 + \beta_3$ is reported and the corresponding standard error is in brackets. #, *, **, and *** denote significance at the 15%, 10%, 5%, and 1% level, respectively.

Conclusion

In this thesis, I have investigated the effect of several institutional changes on pro-social behaviour and education outcomes based on experimental evidence.

Chapter 1 presents results from a laboratory experiment using public goods games with sanctioning opportunities. The public goods game is mainly played among privileged groups. A privileged group is one that has a privileged member who faces a higher per-unit marginal return than do one's group mates. I find that reward opportunities strongly increase both individual contributions and group contributions in those groups. Reward opportunities also significantly mitigate contribution decay over successive periods. The reward institution outperforms the punishment institution in all measured respects. Moreover, reward opportunities give rise to net social welfare improvement whereas punishment opportunities do not. The results suggest that, when group members face different marginal returns to their contributions, a reward institution can give rise to higher contributions than a punishment institution and induce higher level of coordination.

In Chapters 2 and 3, I study the effect of information provision on students' performance, using evidence from one field experiment among nearly a thousand undergraduate students. The field experiment implements the interventions via an online assignment platform. The online assignment always displays absolute performance information, in the form of points, by default. One intervention adds information of milestone-referenced league-referenced goals. Students in this intervention see the league they are in, besides their points. The other intervention, grants students access to relative performance information (rank incentives) along with the milestone-referenced league information. In Chapter 2, I mainly focus on the effect of relative performance information. I find that having access to relative performance information increases the likelihood of a student putting in more effort and achieving beyond the call of duty in the online assignment, compared to their peers exposed only to the milestone-referenced league information. I also find that having access to relative performance information increases low-performing students' exam marks and decreases high-performing students' marks, compared to their peers who only have access to the milestone-based league information. A further investigation in heterogeneous effects shows that treatment effects differ across different subgroups. For example, although it is men that are affected by relative performance to overachieve points in the online assignment, it is high-performing women who have access to the relative performance information that suffer a decrease in their exam marks.

In Chapter 3, I investigate treatment effects by women and men respectively. In the presence of milestone-referenced league information, women with access to the relative performance information perform 0.19 SDs worse in the first midterm, compared to women without this access. In the absence of relative performance information, men with access to the milestone-referenced league information perform 0.26 SDs better in the final exam, compared to men without the access. High-performing females seem to be the most demoralized among the whole sample, which seem to be driven by their increased self-perceived stress. Men's improved exam performance seems driven by their increased effort.

Interestingly, analyses of heterogeneous effects within each gender reveal more subtle pictures. For example, having access to relative performance information positively affects Australian men's exam performance and negatively affects Chinese men's exam performance, compared to their peers in their respective ethnicity. This observation explains the lack of overall significant effects of relative performance information on men's exam performance. In other words, men seem to be significantly affected by the relative performance information as well, although the direction of the effect pertains to ethnicity. This implies that cultural difference might affects people's interpretation of the same information (Alesina and Giuliano, 2015). Investigation on this direction is left to future research.

Overall, this thesis uses experimental approaches to investigate the effect of several institutional changes on subjects' behaviour. On the one hand, experimental approaches have the advantage of maximising researcher control of the research environment and thus more cleanly capture potential treatment effects. On the other hand, experiments, especially lab experiments, suffer from the drawback of limited realism and hence low external validity. Although field experiments are less challenged in this respect, they still face obstacles such as *scalability* of the effects, *representativeness of the population* and *representativeness of the situation*[14]. Another major concern relates to replicability of results. I acknowledge that these general limitations apply to the experiments in this thesis. These limitations can-

[14]see Al-Ubaydli et al. (2019) for detailed explanations of these terms.

not be eliminated, but they can be mitigated with cautious effort. First, while this is not offered in the present work, building a theoretical structure sufficiently detailed and robust to generate theoretical predictions that can be directly compared with experimental findings would deepen the interpretations of results from experimental studies such as this one. Second, this chapter would have benefited from a pre-analysis plan (PAP). I am considering replicating the experiment to address this issue, as part of my future research on this topic. Replication is a possible remedy to experimental findings (Coffman and Niederle, 2015). Third, one may replicate the same design in different situations and for different populations and explore whether similar results hold. Fully addressing these limitations is beyond the scope of this thesis and is left to future research.

Bibliography

Acemoglu, D., S. Johnson, and J. A. Robinson (2002). Reversal of fortune: Geography and institutions in the making of the modern world income distribution. *The Quarterly Journal of Economics 117*(4), 1231–1294.

Adams, R. B. (2016). Women on boards: The superheroes of tomorrow? *The Leadership Quarterly 27*(3), 371–386.

Adams, R. B., S. Gray, and J. Nowland (2011). Does gender matter in the boardroom? evidence from the market reaction to mandatory new director announcements. *Evidence from the Market Reaction to Mandatory New Director Announcements (November 2, 2011)*.

AEI (2019). International student data 2019. Technical report, Department of Education, Skills and Employment.

Akın, Z. and E. Karagözoğlu (2017). The role of goals and feedback in incentivizing performance. *Managerial and Decision Economics 38*(2), 193–211.

Al-Ubaydli, O., M. S. Lee, J. A. List, C. Mackevicius, and D. Suskind (2019). How can experiments play a greater role in public policy? 12 proposals from an economic model of scaling. *University of Chicago, Becker Friedman Institute for Economics Working Paper* (2019-131).

Alesina, A. and P. Giuliano (2015). Culture and institutions. *Journal of Economic Literature 53*(4), 898–944.

Andreoni, J., W. Harbaugh, and L. Vesterlund (2003). The carrot or the stick: Rewards, punishments, and cooperation. *The American Economic Review 93*(3), 893–902.

Asiedu, E. and E. Gross (2017). Can differences in benefits affect group investment into irrigation projects? experimental evidence from northern ghana. Technical report, Courant Research Centre: Poverty, Equity and Growth-Discussion Papers.

Azmat, G. and N. Iriberri (2010). The importance of relative performance feedback information: Evidence from a natural experiment using high school students. *Journal of Public Economics 94*(7-8), 435–452.

Balafoutas, L. and M. Sutter (2010). Gender, competition and the efficiency of policy interventions. IZA Discussion Paper No. 4955.

Balliet, D., L. B. Mulder, and P. A. Van Lange (2011). Reward, punishment, and cooperation: a meta-analysis. *Psychological Bulletin 137*(4), 594.

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

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

Barankay, I. (2012). Rank incentives: Evidence from a randomized workplace experiment. Retrieved from `https://repository.upenn.edu/bepp_papers/75`.

Beghini, V., U. Cattaneo, and E. Pozzan (2019). A quantum leap for gender equality: For a better future of work for all. Technical report, Gallup Inc. and International Labor Organisation.

Benabou, R. and J. Tirole (2006). Incentives and prosocial behaviour. *The American Economic Review 96*(5), 1652–1678.

Bertocchi, G. and M. Bozzano (2019). Gender gaps in education. CEPR Discussion Paper No. DP14082.

Bhanot, S. P. (2017). Rank and response: A field experiment on peer information and water use behavior. *Journal of Economic Psychology 62*, 155–172.

Bisin, A. and K. Hyndman (2020). Present-bias, procrastination and deadlines in a field experiment. *Games and Economic Behavior 119*, 339–357.

Blader, S., C. M. Gartenberg, and A. Prat (2016). The contingent effect of management practices. CEPR Discussion Paper No. DP11057.

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

Borah, B. J. and A. Basu (2013). Highlighting differences between conditional and unconditional quantile regression approaches through an application to assess medication adherence. *Health Economics 22*(9), 1052–1070.

Bouton, L. and G. Kirchsteiger (2015). Good rankings are bad: Why reliable rankings can hurt consumers. Working Paper 21083, National Bureau of Economic Research.

Bowles, S. (2009). *Microeconomics: behavior, institutions, and evolution*. Princeton University Press.

Burdina, M., R. S. Hiller, and N. E. Metz (2017). Goal attainability and performance: Evidence from boston marathon qualifying standards. *Journal of Economic Psychology 58*, 77–88.

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

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

Caudill, S. B. and D. M. Gropper (1991). Test structure, human capital, and student performance on economics exams. *The Journal of Economic Education 22*(4), 303–306.

Charness, G., D. Masclet, and M. C. Villeval (2011). Competitive preferences and status as an incentive: Experimental evidence. CIRANO-Scientific Publications 2011s-07. Available at SSRN: `https://ssrn.com/abstract=1752213orhttp://dx.doi.org/10.2139/ssrn.1752213`.

Choi, J.-K. and T. Ahn (2013). Strategic reward and altruistic punishment support cooperation in a public goods game experiment. *Journal of Economic Psychology 35*, 17–30.

Clark, D., D. Gill, V. Prowse, and M. Rush (2016). Using goals to motivate college students: Theory and evidence from field experiments. *Review of Economics and Statistics*, 1–45.

Coffman, L. C. and M. Niederle (2015). Pre-analysis plans have limited upside, especially where replications are feasible. *Journal of Economic Perspectives 29*(3), 81–98.

Cohen, S. and G. Williamson (1988). Perceived stress in a probability sample of the united states. *In S. Spacapam & S. Oskamp (Eds.), The Social Psychology of Health: Claremont Symposium on Applied Social Psychology*.

Collins, C. A. and L. Gan (2013). Does sorting students improve scores? an analysis of class composition. Working Paper 18848, National Bureau of Economic Research.

Dawes, R. M., J. M. Orbell, R. T. Simmons, and A. J. Van De Kragt (1986). Organizing groups for collective action. *American Political Science Review 80*(4), 1171–1185.

Dobrescu, L., M. Faravelli, A. Motta, and R. Megalokonomou (2019). Rank incentives and social learning: Evidence from a randomized controlled experiment. IZA Discussion Paper No. 12437.

Elsner, B. and I. E. Isphording (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics 35*(3), 787–828.

Eriksson, T., A. Poulsen, and M. C. Villeval (2009). Feedback and incentives: Experimental evidence. *Labour Economics 16*(6), 679–688.

Fan, J. and J. Gómez-Miñambres (2019). Nonbinding goals in teams: A real effort coordination experiment. *Manufacturing & Service Operations Management*.

Fehr, E. and S. Gachter (2000). Cooperation and punishment in public goods experiments. *The American Economic Review 90*(4), 980–994.

Fehr, E. and K. M. Schmidt (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

Ferguson, H. and H. Sherrell (2019). Overseas students in australian higher education: A quick guide. Parliamentary library research paper, Canberra, ACT, Australia: Parliament of Australia.

Firpo, S., N. M. Fortin, and T. Lemieux (2009). Unconditional quantile regressions. *Econometrica 77*(3), 953–973.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics 10*(2), 171–178.

Fischer, C. (2001). Read this paper later: Procrastination with time-consistent preferences. *Journal of Economic Behavior & Organization 46*(3), 249–269.

Fischer, M. and V. Wagner (2018). Effects of timing and reference frame of feedback: Evidence from a field experiment. IZA Discussion Paper No. 11970. Available at SSRN: `https://ssrn.com/abstract=3301725`.

Fisher, J., R. M. Isaac, J. W. Schatzberg, and J. M. Walker (1995). Heterogenous demand for public goods: Behavior in the voluntary contributions mechanism. *Public Choice 85*(3-4), 249–266.

Foster, G. (2012). The impact of international students on measured learning and standards in australian higher education. *Economics of Education Review 31*(5), 587–600.

Foster, G., P. Frijters, and B. Greiner (2017). *Investors and Exploiters in Ecology and Economics: Principles and Applications*, Chapter 10, pp. 171–187. Cambridge: MIT Press.

French, M. T., J. F. Homer, I. Popovici, and P. K. Robins (2015). What you do in high school matters: High school gpa, educational attainment, and labor market earnings as a young adult. *Eastern Economic Journal 41*(3), 370–386.

Gibbons, R. (2005). Incentives between firms (and within). *Management Science 51*(1), 2–17.

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

Global Education Monitoring Report Team [541] (2020). #hereducationourfuture. Technical report, UNESCO. Available at `https://unesdoc.unesco.org/ark:/48223/pf0000372963?locale=en`.

Glöckner, A., B. Irlenbusch, S. Kube, A. Nicklisch, and H.-T. Normann (2011). Leading with (out) sacrifice? a public-goods experiment with a privileged player. *Economic Inquiry 49*(2), 591–597.

Gneezy, U., M. Niederle, and A. Rustichini (2003). Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics 118*(3), 1049–1074.

Goerg, S. J. and S. Kube (2012). Goals (th) at work–goals, monetary incentives, and workers' performance. *MPI Collective Goods Preprint* (2012/19).

Gómez-Miñambres, J. (2012). Motivation through goal setting. *Journal of Economic Psychology 33*(6), 1223–1239.

Goulas, S. and R. Megalokonomou (2015). Knowing who you are: The effect of feedback information on short and long term outcomes. Warwick Economics Research Paper Series No. 1075.

Goy, S. C., Y. L. Wong, W. Y. Low, S. N. M. Noor, Z. Fazli-Khalaf, N. Onyeneho, E. Daniel, S. Azizan, M. Hasbullah, and A. GinikaUzoigwe (2018). Swimming against the tide in stem education and gender equality: a problem of recruitment or retention in malaysia. *Studies in Higher Education 43*(11), 1793–1809.

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association 1*(1), 114–125.

Harding, M. and A. Hsiaw (2014). Goal setting and energy conservation. *Journal of Economic Behavior & Organization 107*, 209–227.

Heath, C., R. P. Larrick, and G. Wu (1999). Goals as reference points. *Cognitive Psychology 38*(1), 79–109.

Herrmann, B., C. Thöni, and S. Gächter (2008). Antisocial punishment across societies. *Science 319*(5868), 1362–1367.

Hirt, E. and S. McCrea (2009). Man smart, woman smarter? Getting to the root of gender differences in self-handicapping. *Social and Personality Psychology Compass 3(3)*, 260–274.

Hsiaw, A. (2013). Goal-setting and self-control. *Journal of Economic Theory 148*(2), 601–626.

Hurley, P. and N. Van Dyke (2020). Australian investment in education: higher education. Technical report, Mitchell Institute, Melbourne.

ILO (2019). Global wage report 2018/19. Technical report, International Labor Organisation.

ILO-Gallup (2017). Towards a better future for women and work: Voices of women and men. Technical report, ILO-Gallup Report.

Jurajda, Š. and D. Münich (2011). Gender gap in performance under competitive pressure: Admissions to czech universities. *The American Economic Review 101*(3), 514–18.

Kagel, J. H. and A. E. Roth (2016). *The handbook of experimental economics*, Volume 2. Princeton university press.

Kahneman, D. (2011). *Thinking, Fast and Slow* (2012 Reprint ed.). Penguin Books.

Katreniakova, D. (2014). *Social Comparison, Rewards and Incentives to Learn: A randomized control trial in Uganda*. Center for Economic Research and Graduate Education.

Koestner, R., M. Zuckerman, and J. Koestner (1989). Attributional focus of praise and children's intrinsic motivation: The moderating role of gender. *Personality and Social Psychology Bulletin 15(1)*, 61–72.

Latham, G. P. and E. A. Locke (1979). Goal setting—a motivational technique that works. *Organizational Dynamics 8*(2), 68–80.

Lyubomirsky, S. and H. Lepper (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research 46*, 137–155.

Mas, A. and E. Moretti (2009). Peers at work. *The American Economic Review 99*(1), 112–145.

McCalley, L. and C. J. Midden (2002). Energy conservation through product-integrated feedback: The roles of goal-setting and social orientation. *Journal of Economic Psychology 23*(5), 589–603.

Moldovanu, B., A. Sela, and X. Shi (2007). Contests for status. *Journal of Political Economy 115*(2), 338–363.

Murphy, R. and F. Weinhardt (2014). Top of the class: The importance of ordinal rank. CEP Discussion Paper No. 1241.

Niederle, M. and L. Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics 122*(3), 1067–1101.

Niederle, M. and L. Vesterlund (2010). Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives 24*(2), 129–44.

Niederle, M. and L. Vesterlund (2011). Gender and competition. *Annual Review of Economics 3*(1), 601–630.

North, D. C. et al. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.

O'Donoghue, T. and M. Rabin (1999). Doing it now or later. *The American Economic Review 89*(1), 103–124.

Olson, M. (1965). *The logic of collective action*, Volume 124. Harvard University Press.

O'Neill, T., L. Hambley, and A. Bercovich (2014). Prediction of cyberslacking when employees are working away from the office. *Computers in Human Behavior 34*, 291–298.

Opstrup, N. and A. R. Villadsen (2015). The right mix? gender diversity in top management teams and financial performance. *Public Administration Review 75*(2), 291–301.

Palfrey, T. R. and J. E. Prisbrey (1997). Anomalous behavior in public goods experiments: How much and why? *The American Economic Review*, 829–846.

Reuben, E. and A. Riedl (2009). Public goods provision and sanctioning in privileged groups. *Journal of Conflict Resolution 53*(1), 72–93.

Reuben, E. and A. Riedl (2013). Enforcement of contribution norms in public good games with heterogeneous populations. *Games and Economic Behavior 77*(1), 122–137.

Reuben, E. and F. van Winden (2008). Social ties and coordination on negative reciprocity: The role of affect. *Journal of Public Economics 92*(1-2), 34–53.

Schraw, G., T. Wadkins, and L. Olafson (2007). Doing the things we do: A grounded theory of academic procrastination. *Journal of Educational Psychology 99*(1), 12.

Schultz, P. W., J. M. Nolan, R. B. Cialdini, N. J. Goldstein, and V. Griskevicius (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science 18*(5), 429–434.

Sefton, M., R. Shupp, and J. M. Walker (2007). The effect of rewards and sanctions in provision of public goods. *Economic Inquiry 45*(4), 671–690.

Smith, A. (1759). *The Theory of Moral Sentiments* (2010 Reprint ed.). Penguin Books.

Smithers, S. (2015). Goals, motivation and gender. *Economics Letters 131*, 75–77.

Smithers, S. F. T. (2016). *Essays in Motivation*. Ph. D. thesis, University of Leicester.

Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller (2007). *Climate Change 2007: Working Group I: The Physical Science Basis*. Cambridge: Cambridge University Press.

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin 133*(1), 65.

Steel, P., F. Svartdal, T. Thundiyil, and T. Brothen (2018). Examining procrastination across multiple goal stages: a longitudinal study of temporal motivation theory. *Frontiers in Psychology 9*, 327.

Sunstein, C. (2011). Empirically informed regulation. *University of Chicago Law Review 78*, 1349–1429.

Sutter, M., S. Haigner, and M. G. Kocher (2010). Choosing the carrot or the stick? endogenous institutional choice in social dilemma situations. *The Review of Economic Studies 77*(4), 1540–1566.

Thaler, R. H. and C. R. Sunstein (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin Books.

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

Van Hooft, E., C. Wanberg, and G. Van Hoye (2013). Moving beyond job search quantity towards a conceptualization and self-regulatory framework of job search quality. *Organizational Psychology Review 3(1)*, 3–40.

Van Lent, M. and M. Souverijn (2017). Goal setting and raising the bar: A field experiment. *Tinbergen Institute Discussion Paper 17-001/VII.*

Villeval, M. C. (2020). Performance feedback and peer effects. Technical report, GLO Discussion Paper.

Vyrastekova, J. and D. Van Soest (2008). On the (in) effectiveness of rewards in sustaining cooperation. *Experimental Economics 11*(1), 53–65.

Walker, J. M. and M. A. Halloran (2004). Rewards and sanctions and the provision of public goods in one-shot settings. *Experimental Economics 7*(3), 235–247.

Wan, H., L. Downey, and C. Stough (2014). Understanding non-work presenteeism: relationships between emotional intelligence, boredom, procrastination and job stress. *Personality and Individual Differences 65*, 86–90.

Wise, D. A. (1975). Academic achievement and job performance. *The American Economic Review 65*(3), 350–366.

World Bank (2019). The little data book on gender. Technical report, World Bank. "https://openknowledge.worldbank.org/bitstream/handle/10986/31689/LDB-Gender-2019.pdf?sequence=2&isAllowed=y".

Wozniak, D., W. T. Harbaugh, and U. Mayr (2014). The menstrual cycle and performance feedback alter gender differences in competitive choices. *Journal of Labor Economics 32*(1), 161–198.