Information, Goal Setting and Performance: A Field Experiment

Haoran He[*], Marcus Roel[**], Ronghai Su[***], Qian Weng[****]

Abstract

We investigate how information impacts goal setting and performance by conducting a field experiment that varies whether students know or do not know their true relative ability when they set goals in a university physical education course based on relative performance rating. We document that most students set challenging goals, based on either their actual or estimated relative ability, proxied by their relative performance in a baseline test. However, only a small proportion of students estimate their performance accurately; under- and over-estimation are equally common. Consequently, consistent with our theoretical prediction, receiving information about one's relative ability significantly improves goals – increasing (decreasing) goals set by students who underestimate (overestimate) their baseline relative test performance. We find that providing information significantly raises both relative and absolute final test performance of those who underestimate their relative ability. However, we do not find any effect on performance for students who overestimate themselves.

Keywords: information; goal setting; biased beliefs; performance; field experiment

JEL classification: D90, Z29, C93

[*] Business School, Beijing Normal University. No. 19 Xinjiekouwai Street, Haidian District, 100875, Beijing, China. Tel: +86 10 5880 7847, Email: haoran.he@bnu.edu.cn

[**] Business School, Beijing Normal University. Email: mcs.roel@gmail.com

[***] College of P.E. and Sports, Beijing Normal University. Email: sronghai@163.com

[****] Corresponding author: School of Labor and Human Resources, Renmin University of China. No. 59 Zhongguancun Street, Haidian District, Beijing 100872, Beijing, China. Tel: +86 10 82502205, Email: qian.weng@ruc.edu.cn

Acknowledgments: We thank Sherry Xin Li, and conference and seminar participants at the 2019 Asian-Pacific Economic Science Association Conference in Abu Dhabi, Renmin University, Beijing Normal University, and Zhejiang University of Finance and Economics for helpful comments and discussions. We thank every member of our team of research assistants for their hard work and care, especially head research assistant Yu Cao, Shuai Chen and Lei Yue. Financial support from the Ministry of Education in China (No. 21YJA790059), the National Natural Science Foundation of China (No.71973016, No. 72131003, and No. 72273148), and the National Program for Special Support of Top-notch Young Professionals is gratefully acknowledged. All remaining errors are those of the authors.

1. Introduction

People often set goals with the intention to help themselves reach better outcomes. For example, students commonly set grade targets, runners aim for certain finishing times, and firms set annual revenue benchmarks. Since goal setting is ubiquitous in life, a natural question to ask next is how goals can be set most effective in terms of motivating people. Early theoretical studies (Locke, 1968; Locke and Latham, 1985) highlight that goals need to be specific, challenging, and also attainable so as to increase intrinsic motivation and as a result to improve performance. Several empirical studies have examined the impact of these characteristics on performance, and found supportive evidence (e.g., Wu et al., 2008; Goerg and Kube, 2012; Dalton et al., 2015; Brookins et al., 2017; Burdina, 2017; Uetake and Yang, 2018; van Lent and Souverijn, 2020).

Although there seems to exist a clear golden rule for goals, it is not at all easy for individuals to set optimal goals. The reasons are threefold. First, what is a challenging but achievable goal for one person may be out of reach for another person, yet be completely trivial for a third person. In other words, optimal goals are specific to a particular person's ability and circumstances, which may not be constant across time, and often unknown to an outside observer. Such aspects make goal setting and its effects difficult to study empirically. Second, in many contexts, goals are not set for simple and countable tasks or regarding a person's absolute performance but in terms of relative performance, which introduces the performance of others as another, often highly uncertain, dimension. Third, goals are often set well in advance before an activity takes place. For example, grade targets, finishing times, or revenue benchmarks and their respective relative placements are usually set months in advance. As a result, goal setters usually have imperfect knowledge about themselves' and others' future studying capacity, fitness levels, or the competitiveness of the future environment, which all make setting optimal (relative) goals difficult. In view of these difficulties, it is important to further understand how relative-performance goals are set and how these difficulties can, to some extent, be mitigated.

In this paper, we conduct a field experiment to investigate how information impacts relative-performance goal setting and performance and develop a theoretical framework to illustrate the possible consequences. In particular, we try to address the following research questions: What is the effect of setting goals without and with knowing one's ability relative to others' on both absolute and relative performance? How are goals set? Do individuals with different beliefs about their relative ability set different goals, and how does information about such relative ability affect their goal setting and performance?

Our field experiment is implemented in a university physical education course to address these questions. This course uses a relative grading scheme to determine students' test scores for holding a plank – the rank of their plank time for each gender in each class. Students are randomly assigned to one of two treatments which ask them to set goals for the final test relative performance, or to a control group which do not set goals. Goals are self-set and non-binding,[1] and performance-based.[2] The two goal-setting treatments differ solely in terms of whether students are informed of their relative performance in the baseline test, which provides them a signal about their relative ability, *before* or *after* they are asked to set goals for their relative performance in the final test. In both treatments and the control group, students are also asked to predict their relative performance in the baseline test before goal setting.

Our theoretical framework illustrates how and when information about one's relative ability influences goal setting and consequently improves performance, both relative and absolute. In our model, a present-biased decision maker (she) aims to set a relative performance goal in order to align the lazier future self's preferences with her own preferences over relative (and absolute) performance. Goals act as an internal commitment device to improve performance. We show that performance follows an inverted V-shape with goal and that optimal goals are challenging but achievable. As goals are set in terms of relative performance, the optimal goal depends on the decision maker's belief about her relative ability (compared to others), or, equivalently, her beliefs about other's absolute performance. When the goal setter underestimates her relative ability, she sets a goal that is too easy to achieve. Information about her true relative ability allows her to adjust the goal upwards, pushing her future-self towards higher performance. When the goal setter overestimates her relative ability, she sets her goal higher than her ideal target. Now, information results in a downward adjustment of the goal, and the implication on performance depends on whether the goal setter can or cannot motivate her future-self to perform at her most preferred performance level, assuming her beliefs about relative ability were correct. If goals cannot fully undo the negative effects from present bias on performance, optimal performance is characterized by an interior. In this case, information always improves performance as a too challenging goal is adjusted downward, becomes achievable and more motivating. In the other case, which is a corner-solution, information may actually result in a (welfare improving) lower

[1] Goals can broadly be categorized into two often related dimensions – binding or non-binding, and self-set or exogenously given goals. Binding vs. non-binding refer to whether monetary rewards (punishment) related to the goal achievement (failure) are involved or not. Self-set vs. exogenously given refer to by whom the goals are set. Exogenously given binding goals are mostly used in the context of principal-agent problems where monetary rewards are used to elicit effort from workers (e.g., Goerg and Kube, 2012; Gosnell et al., 2020; Kuhn and Yu, 2021). Recent studies have also seen cases in which principals set non-binding goals for agents (e.g., Corgnet et al., 2015, 2018; Smithers, 2015; Fan and Gómez-Miñambres, 2020; Sauer et al., 2018). Self-set non-binding goals are often studied in situations involving self-control problems (e.g., van Lent and Souverijn, 2020, Clark et al., 2020, on course grade; Sackett et al., 2014 on running time; Uetake and Yang, 2018 on weight loss; Brookins et al., 2017 on principal-agent problem; Harding and Hsiaw, 2014 on energy conservation).

[2] In terms of the target, goals can also be broadly categorized into performance-based or task-based goals. Performance-based goals are set for particular performance outcomes, whereas task-based goals are set for the completion of particular tasks. We chose performance-based goals for the following reasons. Firstly, performance-based goals are more suitable to address our research question on how information impacts goal-setting level, and consequently are more directly linked with our outcome of interest. Secondly, they are also easier to be executed in our experimental setting (introduced below). Thirdly, existing empirical findings regarding the effect of performance-based goals on performance is mixed (discussed below). We thus try to provide some additional evidence.

performance as, in the absence of information, the future-self will perform above the goal setter's ideal performance level in order to reduce her psychological losses from missing the goal set when her belief was not too far from the truth. Finally, information does neither affect the goals nor the performance of those decision makers who hold accurate beliefs about their relative performance.

We find that the vast majority of students set goals above either their actual or estimated baseline performance, respectively, when being informed or not being informed about their relative performance in the baseline test. Only a small proportion of students predict their performance accurately; under- and over-estimation are equally common. Consequently, consistent with our theoretical prediction, receiving information about one's relative ability significantly improves goals – increasing goals set by students who underestimate their relative ability while decreasing goals set by the ones who overestimate their relative ability, whereas information does not impact goals set by students who accurately predict their performance. Compared to setting goals without information, providing information significantly raises final test performance for students who underestimate own relative abilities. However, we do not find any effect on performance for students who overestimate, which dilutes the overall effect of information in the entire sample. There is also evidence that students who underestimate their relative ability perform worse when they set goals without information compared to the respective students in the control group, which is likely driven by a demotivating effect from too low goals set without information compared to unelicited personal goals set by at least some students in the control group. The aforementioned impact on performance holds regardless of whether the absolute or relative performance measure is used. We further find that reaching goals raises students' satisfaction with test performance.

The present paper contributes to four strands of literature. First, it contributes to understanding theoretically how goals can be better set to promote performance. In the theoretical literature, goals are conventionally modelled as reference points. Heath et al. (1999) are the first to model goals in this respect using the value function of Prospect Theory (Kahneman and Tversky, 1979) and provide survey evidence to justify this approach. Their model is extended by Wu et al. (2008), who show that output is first increasing in goals but then decreasing when goals become too challenging. Koch and Nafzier (2011) model the process of goal setting in a discrete setup and suggest that goals are useful to overcome self-control problems (Strotz, 1955; Laibson, 1997; O'Donoghue and Rabin, 1999). Suvorov and van de Ven (2008) and Hsiaw (2013) also model goals as a similar commitment device, but adopt the equilibrium framework of Köszegi and Rabin (2006), which requires goals to be consistent with equilibrium behavior. Our model broadly follows Koch and Nafziger (2011) and is similar to the recent model of Clark et al. (2020). Like Clark et al., we allow for continuous choices and focus only on losses that arise from failing goals. Our model differs from Clark et al. in three aspects. In contrast to their linear specification, our model features diminishing sensitivity to losses. This has the crucial implication that goals that are too high can be demotivating, which not only makes goal setting more difficult for the goal setter but also more realistic.[3] In addition, our main focus is on relating goal setting to relative ability, especially in situations where information on ability is incomplete, which is not a topic of Clark et al.[4] Finally, our model generalizes goal setting to a relative performance setting, which, to our knowledge, has not been modelled before. It can, however, be easily adapted to a pure decision problem with absolute instead of relative ability, translating our insights regarding the impact of information for underestimating one's ability (information is performance improving) and overestimating one's ability (information may improve or deteriorate performance).

The closest paper to ours is a purely empirical paper by van Lent (2019), who conducts a field experiment to investigate the effect of goal setting and goal revision on exam grades in a university course. The similarity arises from the fact that goal revision allows goal setters to adjust their goals based on information about the mid-term exam grade. As goals can be revised after some time, this setting introduces the potential confounding effect of time as goals are (re-) set closer to the final exam, however.[5] More generally, knowing the opportunity to revise goals or the fact that goals are actually revised may also affect one's commitment to goals themselves - from the onset or after the revision itself. Finally, in their study, students were only given an option to set any arbitrary goals, which could be either performance-based or task-based goals.[6] The above factors could make goal setting no longer an effective motivation device and thus contribute to their findings of no or negative impact of goal setting on performance.

Second, this paper contributes to the literature on the interaction among self-confidence in terms of over- or underestimating one's own ability relative to others', goal choice and behavior. In the psychological literature where the concept of effective goal setting originated, to our best of knowledge, there is no study focusing on this interaction.[7] Only a recent study in economics considers this

---

[3] With a linear loss function, goals that are too high do not demotivate and hence result in the same additional effort as goals that are optimal.

[4] Clark et al. mostly focus on complete information to explain differences between a performance-based goal and a task-based goal. In a simple extension, they allow for output uncertainty – which is only resolved at the end of the game, where output is either unaffected, or equal to zero. As uncertainty is not resolved at the time of effort provision, this reduces the marginal incentive of effort as expected and marginal output are lower than when there is no uncertainty. Similar to Koch and Nafziger (2011), it also makes goals riskier as the decision maker may fall short of her goal and thus may incur a psychological loss. This results in an incentive to reduce goals. To explain why task-based goals are more effective than performance-based goals, Clark et al. also suggest that overconfidence can make performance-based goals less effective: when students overestimate how effort translates into exam performance, they work too little and unexpectedly fail to achieve their goals.

[5] The typical model of goal setting relies on time (inconsistent) preferences. Resetting a goal closer to the final exam may thus push a goal setter's preference over effort closer towards those of her future self's. Hence, it is not theoretically obvious whether re-setting goals can improve performance. Holding more accurate beliefs about her own ability may allow the goal setter to set more realistic goals (which may improve performance) but which may not be challenging enough as they are set closer to the time of the exam (deteriorating performance).

[6] For task-based goals, it is not theoretically obvious whether a goal setter would actually want to update her goal in response to new information, if, for example, she simply wants to motivate her future-self, who is lazy, to study a certain amount of time each day.

[7] There are two papers loosely related to this interaction, but whose focus as well as concept of confidence are fundamentally different from ours. Beckmann et al. (2009) study how interactions between feedback, self-confidence and goal orientation affect performance. Hadwin and Webster (2013) study whether students'

interaction. Avdeenko et al. (2019) conduct a field experiment with smallholder farmers in rural Ethiopia to study the effect of saving goals on savings behavior, with personalized feedback consisting of recommendations to self-set goals provided to a subsample of the participants. They find that feedback strongly increases savings of underconfident farmers in terms of financial literacy, and the increase offsets savings deficiency relative to overconfident farmers. Our contribution to this literature lies in analyzing the possible heterogeneities in self-confidence on goal choice and performance. Our finding highlights that decision makers who underestimate their own relative ability can be helped by being provided with factual and verifiable information about it.

Third, this paper contributes to the research on whether and how goal setting boosts educational outcomes. Clark et al. (2020) conduct field experiments in an introductory economics course in a US university and find that performance goals for exams and the overall course grade have little effect while task-based goals lead students to engage more with the tasks and reach better performance on exams. Dobronyi et al. (2019) conduct a similar experiment in a Canadian university and find no evidence of an effect of goal setting on GPA, course credits, or second year persistence. In contrast, van Lent and Souverijn (2020) find that setting a target grade for a course in a university in the Netherlands has positive effects on grade, but the effect becomes negative if goals are raised by taking others' suggestions. Given the mixed findings, it is important to understand how to make goals play positive roles. By explicitly connecting the effectiveness of goals to a student's relative ability, which is often unknown to researchers, our paper helps uncover why the literature finds mixed effects of goals. Our field experiment conducted in a physical education course also extends this line of research to a different educational scenario.

Finally, this paper contributes to the discussion of the performance impact of goal setting in the relative-performance evaluation scheme. Several recent studies examining the motivational strengths of rank-order relative-performance evaluation scheme on performance have obtained heterogeneous findings depending on gender (e.g., Jalava et al., 2015; Czibor et al., 2020), position in the ability distribution (e.g., Jalava et al., 2015), or position in the relative performance distribution (e.g., Gill et al., 2019). There are also recent field experiments that investigate the impact of providing individualized information about student's (relative) ability to the students or their parents on educational investments (Dizon-Ross, 2019; Gan, 2022), school choice (Bobba and Frisancho, 2019), and various other academic decisions (Franco, 2019). These papers find that individuals misperceive their ability and that providing such information generally leads to improvements in the outcomes of interest, with improvements being quite heterogeneous across subjects. However, to our best knowledge, no study

judgments of confidence is better calibrated with self-evaluations of past or current goal attainment, and whether calibration between judgments of confidence and self-evaluations of goal attainment are related to overall academic success.

has investigated the impact of goal setting, under a relative-performance evaluation scheme, not to mention the potential heterogeneous effect.

2. Experimental Design

2.1 Experimental setting - the physical education course and the subjects

We conducted the field experiment in a physical education course (PEC) called the "strength and conditioning training course (SCTC)" during the 2018-2019 academic year at Beijing Normal University, an elite Chinese University with more 20,000 students. During four years of undergraduate studies, students are required to take at least one PEC in four out of eight semesters, each of which is worth one credit and included for GPA calculation. The university offers more than 20 types of PECs.[8] The SCTC is an optional course that aims to improve general fitness. It runs for an entire semester of 16 weeks and has one training session per week, which lasts 1.5 hours. The same instructor (a professional PE teacher) offered three and four identical classes, respectively, in the fall and spring semesters. Twenty-five up to a maximum of 30 students attended one of these seven classes, in total 196 students. The same course content has been used for several years. The course was structured as follows. Training sessions in the first two weeks consisted of an introduction to the course arrangements and its grading criteria as well as general training. The following 12 weeks were divided into three modules of equal number of weeks for improving physical capacities related to aerobic endurance, speed, and power, respectively. The course concluded with two general training sessions. For the three modules, there was a separate test session each week in addition to the training session. We conducted our experiment in module 1, which aimed to improve muscle endurance of the core. The test task was the plank. Scores for each module were determined independently. At the end of the course, one of the three modules was randomly selected (the same module for all students) to determine the overall course grade.

2.2 Treatments

We implemented a 1 × 3 between-subjects design, in which we varied whether or not a goal was set and the information under which the goal was set.

- *NoGoal*: the standard SCTC without goal setting; it represents the control group.
- *NoInfoGoal*: identical to the *NoGoal* treatment, except that students were requested to set goals for their final test score (and associated ranking range, defined below) *before* the score (and associated ranking range) from the baseline test was announced.

[8] The PECs include fitness courses (quality development/expansion, strength and conditioning training, bodybuilding), sports courses (volleyball, basketball, tennis, badminton, football, soccer, swimming, taekwondo), and leisure courses (basic ballet, ballroom dance, aerobics, yoga, aikido, women's self-defense).

- *InfoGoal*: identical to the *NoGoal* treatment, except that students were requested to set goals *after* the score (and associated ranking range) from the baseline test was announced.

The intervention was implemented through personalized survey questions (discussed below). We randomly assigned each student of a given gender and class to one of the treatments or the control group. Since the experiment lasted for a total of eight weeks across two semesters, a stratified random assignment ensured that any temporal differences across classes, such as temperature or air quality, would be balanced across treatments. Since students were from 19 out of in total 28 different schools, they were typically unfamiliar with each other and seldomly communicated with one another so that information spillover/ treatment contamination should not be a concern.[9]

Our experiment is based on a *relative* grading scheme, which is a common way in which students at the university are graded. Specifically, we ranked the plank time (in seconds in descending order) for each gender in each class, computed in which range each ranking was located by dividing the gender-specific number of students in that class, and then associated each range with a typical score between 60 and 100. Specifically, the top 10% students got a full score of 100, those ranked in (10%, 20%] a 94, (20%, 30%] an 88, (30%, 40%] an 84, (40%, 50%] an 80, (50%, 60%] a 76, (60%, 70%] a 72, (70%, 80%] a 68, (80%, 90%] a 64, (90%, 100%] a 60 (see Appendix Figure A1).[10] Students in the two goal-setting treatments also set goals for their final rank.

2.3 Experimental procedure

Before the start of our experiment in week 3,[11] the instructor informed students of the following. In week 1, apart from learning about the basic course arrangements, students were informed that the course was part of an education reform project approved by the university's academic affairs office to improve teaching quality.[12] As a result, they would be required to answer a number of online surveys, which would be administered by wjx.cn (China's Qualtrics.com) and sent via individual WeChat[13] messages by a project assistant, who was unrelated to the course. In view of the university course selection rules, this gave them the opportunity to opt out the course and the surveys. Students were ensured that survey

---

[9] Based on their responses from the 2nd survey, students only knew the goals of about 2% of other students in their class.

[10] The structure of scores, which are not equally spaced, is common in university physical education courses in China.

[11] There are two reasons to start the experiment in week 3: first, since students could freely exit the course in the first two weeks, we started our intervention in the third week to avoid the attrition/self-selection problem; second, in order to minimize the effect of any possible confounding factors, we started the intervention at the earliest feasible time given the course arrangement. Only one registered student dropped out after week 1.

[12] The SCTC's instructor received a grant to conduct an education reform project aiming to improve the course quality before we implemented the experiment.

[13] WeChat is China's ubiquitous instant-messaging program, which is used for text-messaging, photo/file sharing, etc., as well as for digital payments.

responses would be kept confidential from the instructor, the teaching assistants and other students.[14] Since education reform projects are frequently implemented in various courses at the university, students have often participated in them before. Hence, they would be unaware of being involved in a scientific study. This ensures that our experiment keeps the features of a natural field experiment (Harrison and List, 2004).

In week 2, the instructor introduced the aim and arrangements of the first module. Specifically, students were informed that each week would consist of a test session and a training session. The test sessions were conducted individually and independently in a separate time slot of a student's choice before the training sessions. During the tests, the instructor or teaching assistant ensured that each student satisfied the required body position of the plank and recorded the duration held. No performance-related information was revealed to the student at the time of the test.[15] The instructor then announced and posted the module's grade composition (see Appendix Figure A2) and the grading criteria for the module with an example of score and ranking range association (see Appendix Figure A1). The first three tests each accounted for 8%, while the final test accounted for 60%; class attendance accounted for 10%, and two surveys for 3% each. All these components scored on a scale of 0-100. The weighted average constituted the module grade. The instructor then demonstrated the essentials of the test task plank.[16] Afterwards, the project assistant added each student as a contact in WeChat for further communication. Students also signed up for their test slots.

The purpose of the training sessions was to improve core-muscle strength and endurance. The training phase included various training tasks, such as crunches, leg lifts, etc. The instructor and teaching assistants measured several fitness indicators at the beginning of the first training session as the basic measure for students' physical fitness levels, including height, weight, body mass index (BMI), body fat rate, metabolic rate; oxygen saturation, heart rate, and respiratory rate.[17]

On the day after the first training session, each student was sent a personalized link via WeChat to the first survey (see Appendix Figure A3 for the goal-setting related questions). In the survey, among other questions, the student was first asked to predict his/her baseline test score (and associated ranking range). The next questions varied with the treatment. In the *NoGoal* treatment, the student was presented with a grade card showing his/her actual test score (and associated ranking range). In the *InfoGoa*l treatment, the student was also presented with his/her grade card and then asked to set a goal for the

[14] Chinese universities, typically do not have Institutional Review Boards (IRBs) to approve the ethics of economic experiments using human subjects. However, to the best of our understanding, our design falls under the "minimal risk" exemption from IRB approval.

[15] While they could probably estimate their absolute performance fairly accurately – or simply re-run the test again themselves privately – they crucially did not know anything about other students' absolute performance and hence hardly knew their relative performance.

[16] The plank was chosen as the test task because it is one of the common tasks used to test endurance and it was convenient for implementing *relative* grading.

[17] Height and weight were measured by a height and weight meter; body fat rate and metabolic rate were measured by a body fat meter; oxygen saturation and heart rate were measured by a pulse oximeter; respiratory rate was measured for 30 seconds and self-reported by students.

final test. In the *NoInfoGoal* treatment, the order of the grade card presentation and goal setting was reversed, namely, the goal was elicited before they were informed about the baseline test score, while the text remained the same.[18]

On the day after the training sessions 2, 3 and 4, each student was sent a personalized grade card with updated test scores via WeChat message. In the two goal-setting treatments, the grade card also reminded students of their goals. Appendix Figure A4 provides an example of the WeChat message for the two goal-setting treatments. After the last training session, each student was also sent a personalized link via WeChat to a second survey. Among other questions, risk preference, time preference and other-regarding preferences such as trust, reciprocity, fairness, altruism were elicited (Falk et al., 2018). We also asked questions to measure students' willingness to compete (Bönte et al., 2017), self-regulation (self-control) (Gaumer Erickson et al., 2018), as well as socioeconomic background (see Appendix Figure A5). In the spring semester, we also elicited students' satisfaction with their final test performance (see Appendix C). The timeline of our four-week long experiment is summarized in Figure 1, where week 1 of the experiment corresponds to week 3 of the course.

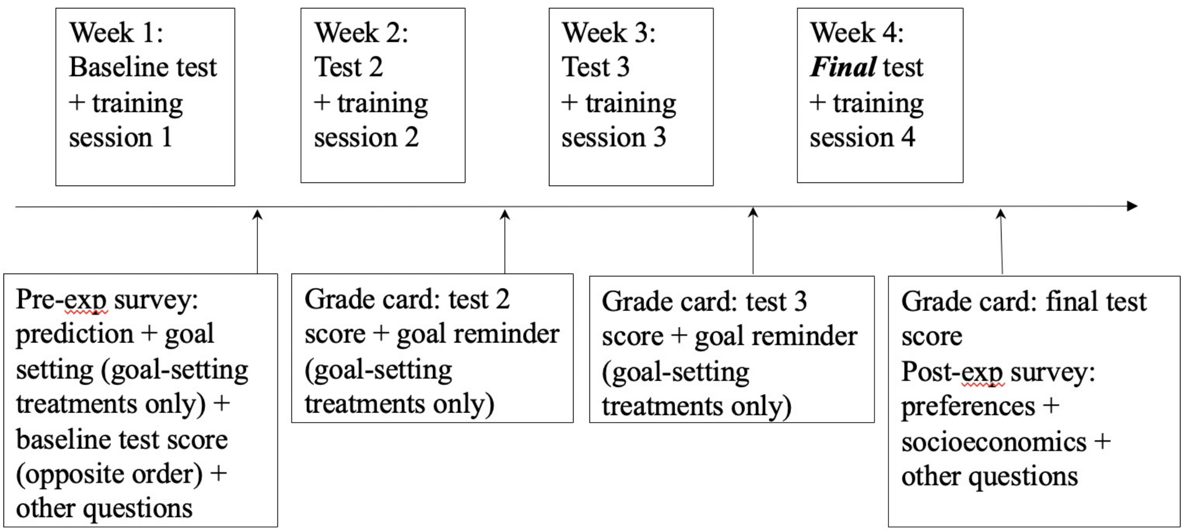


Figure 1. Timeline of our experiment

In order to avoid possible confounds from experimenter demand effects, we implemented some blindness into the design. All intervention and feedback were conducted via WeChat by an independent education reform project assistant. The instructor and the teaching assistants were not informed of the treatment assignment or any responses to the survey questions throughout the course. Throughout the experiment, there were no communications between the students and the experimenters, except for the carefully scripted written messages via WeChat.

[18] The goal-setting question read as "Please set a goal for your score in the final test. Note that you will be reminded of your goal when you are informed of your score for each test through WeChat messages."

3. The Model

To highlight how information enables better goal setting and performance, we will now sketch out a three-period model, in which a present-biased decision maker (Laibson, 1997; O'Donoghue and Rabin, 1999) first sets a goal and then chooses a level of performance (output) that results in delayed benefits. All proofs can be found in Appendix B.

3.1 Setup

At $t = 1$, the decision maker (she) sets a goal $g \in R^+$ regarding her *relative* performance y^r in a test. At $t = 2$, she chooses her *absolute* performance $y \in R^+$, which is realized at $t = 3$, and pays an immediate cost $c(y)$ that satisfies $c'(y) > 0$, $c''(y) > 0$, $c(0) = 0$, $c'(0) = 0$, and $lim_{y \to \infty} c'(y) = \infty$. Her relative performance, over which she has linear preferences, relates her absolute performance to the (average) performance of others, denoted by $\bar{y}$. Consequently, relative performance can be expressed as a function of the decision maker's absolute performance y and $\bar{y}$: $r(y, \bar{y})$, which is increasing (decreasing) in her own (others') performance, i.e., $r_y > 0$ and $r_{\bar{y}} < 0$. We make the additional assumption that relative performance is linear in both terms, i.e., $r(y, \bar{y}) = y - \bar{y}$, which imposes the simple feature that the marginal benefit of the decision maker's own performance is constant for any level of relative performance (and the performance of others). This is a natural starting point for any model of relative performance *with* goals, representing a base-line against which models with more intricate relative performance measures can be compared.[19]

The decision maker has reference-dependent preferences. In addition to caring directly about her relative performance and costs, she also evaluates her relative performance against her goal and incurs a loss at $t = 3$ whenever her performance falls short. The loss is described by $v(g - y^r) \in R^+$ and satisfies $v(z) = 0$ for $z \leq 0$, and $v(z) > 0, v'(z) > 0$, and $v''(z) < 0$ for all $z > 0$. The right derivative of $v(z)$ at zero is bounded.[20] In order to distinguish between the losses imposed from failing one's goal and the direct benefits and cost of performance, we will, depending on the context, often refer to the former by *psychological* and the latter by *material* utility, benefits, or costs. Finally,

[19] There are further generalizations that could prove interesting in future analyses. One may generalize $r(\cdot)$ and allow it to be either convex or concave in y depending on the performance of others. For instance, the marginal effect of increasing absolute performance may be increasing when the decision maker's performance is far below those of others but decreasing if she is relatively far ahead. One may also allow preferences over relative performance to be more general, capturing more complex motivations. While these are certainly interesting generalizations, we focus on the simplest possible model to understand the role and implications of goal-setting itself in such a context instead of exploring secondary effects of relative performance (measures)

[20] Focusing only on losses simplifies the analysis of goal-setting compared to using a more general s-shaped gain-loss function as in Kahneman and Tversky (1979), Tversky and Kahneman (1991), Köszegi and Rabin (2006), or Wu et al. (2008). When the decision maker also benefits from surpassing her own goal, the goal-setter may prefer the 'free utility' that a goal of $g = 0$ could achieve over its motivational benefit. It seems empirically implausible, however, that such low goals could provide real utility to the agent. An alternative solution to such a problem would be to require goals to be consistent with optimal behavior as done in Köszegi and Rabin (2006) equilibrium framework for reference dependent preferences: for a goal to be accepted, it must also be reached in equilibrium.

let $\delta \in (0,1]$ denote the standard time discount factor, $\beta \in (0,1)$ the present-bias discount factor, and $\alpha \in [0,1]$ the degree that goals matter to the decision maker.

3.2 Optimal Performance

The decision maker's problem can be solved by backward induction. At time $t = 2$, the decision maker (doer) solves for the optimal absolute performance:

$$\max_{y \in R^+} \quad \beta\delta \cdot r(y, \bar{y}) - c(y) - \alpha \cdot \beta\delta \cdot v(g - r(y, \bar{y})) \tag{1}$$

If the solution to this problem is non-unique, we assume that the decision maker chooses the larger performance.[21] We start by discussing the standard case when the decision maker is not motivated by goals, i.e., $\alpha = 0$. Since the cost function is convex and relative performance is linear in y, optimal performance is captured by the usual first order condition:

$$\beta\delta \cdot r_y(y_l, \bar{y}) = c'(y_l) \tag{2}$$

where we denote the optimal absolute performance by y_l. Since the impact of goals on performance is always weakly positive in our setting, y_l represents a lower bound on absolute performance for any α and any g. Setting $\beta = 1$ in equation (2) yields another special case, namely the performance level that the doer, who does not care about goals and who is not present-biased, wants to exert. We will refer to this performance level as y_h. Clearly $y_h > y_l$. From the perspective of the $t = 1$ self, y_h is the ideal performance level – any performance above it is inefficient as it exceeds the cost of providing it.[22]

Due to her inherent laziness at $t = 2$, the doer needs additional motivation in order to increase her performance from y_l towards y_h, which is achieved through goals. Whether the doer actually wants to reach y_h depends on the goal and how much she is motivated by it relatively to the additional cost she incurs from performing at above y_l. In particular, for $\alpha > 0$, there is an additional psychological benefit of increasing performance whenever $r(y, \bar{y}) < g$ as it reduces the psychological costs that she incurs when she falls short of her goal. We now characterize the $\alpha > 0$ case, describing the optimal absolute performance as a function of the goal: $y^*(g)$.

Proposition 1: For $g \leq r(y_l, \bar{y})$, $y^*(g) = y_l$, for $g \in (r(y_l, \bar{y}), \hat{g}]$, $r(y^*(g), \bar{y}) = g$, and for $g > \hat{g}$, $y_l < y^*(g) < y^*(\hat{g})$ and is decreasing in g.

Since the doer always wants to perform at a level of at least y_l, the goal must exceed $r(y_l, \bar{y})$ in order to affect her behavior; goals below $r(y_l, \bar{y})$ do not fundamentally change her motivation. Raising

---

21 This technical tie-breaker assumption ensures that a goal that maximizes absolute performance exists in the case when optimal performance $y^*(g)$ is not continuous in g. For more detail, see footnote 24 and the respective proof of proposition 1.

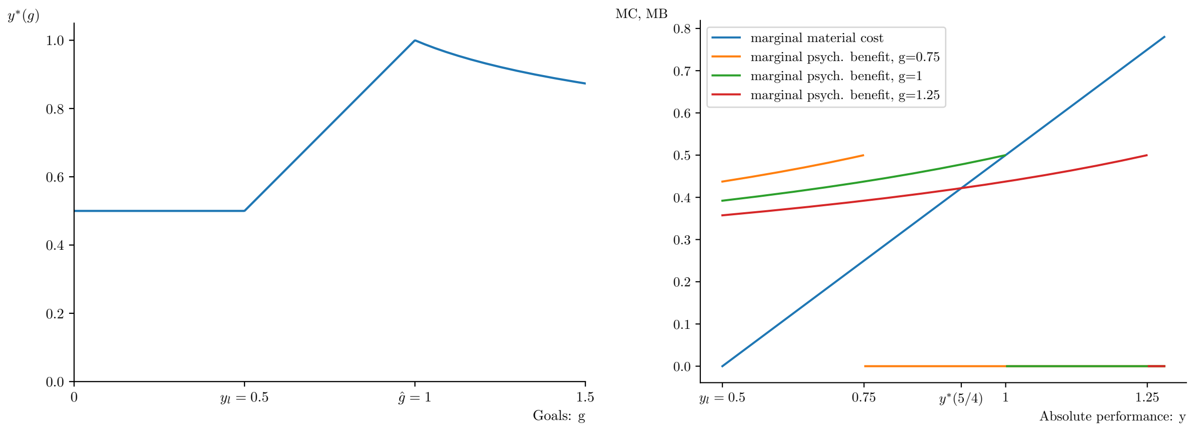
22 To see this, note that preferences over performance, equation 1, evaluated from the perspective of $t = 1$, setting $\alpha = 0$, becomes $\beta\delta^2 \cdot r(y, \bar{y}) - \beta\delta \cdot c(y)$. Since β multiplies all terms, it no longer affects preferences.

the goal slightly above $r(y_l, \bar{y})$ increases the additional benefits of absolute performance, which in turn motivates her to reach her goal. At some point, however, goals lose their motivational force as they are too costly to achieve, which happens when $g > \hat{g}$. After all, when choosing y, the doer balances the psychological benefits against the material cost, and the latter exceed the former already before the goal is reached. Moreover, performance no longer increases but decreases in goals. This decrease is driven by the decision maker's diminishing sensitivity to psychological losses, which causes the marginal psychological benefit of performance to fall as the goal increases.

The relationship between performance and goals is described by an inverted v-shape, see, for illustration, Figure 2(a). The particular graph, as well as Figure 2(b), is based on the following basic parameterization: $\alpha = \delta = 1$ and $\beta = 0.5$; $c(y) = y^2/2$, $v(z) = [(1+z)^\gamma - 1]/\gamma$, with $\gamma = 0.4$, for $z \geq 0$.[23] $\bar{y}$ is normalized to be zero, which has the consequence that the decision maker reaches her goal whenever $y = g$.[24] Figure 2(b) graphs the marginal benefits and costs for performance levels above the minimum level y_l.[25] For $y > y_l$, the marginal material utility is negative as the decision maker finds it costly to exert a performance above y_l. We depict this cost as the marginal material costs in the graph.[26] The marginal psychological benefit, on the other hand, is the derivative of the loss function.[27] Due to diminishing sensitivities to losses, the marginal reduction of psychological losses increases as the decision maker gets closer to her goal (from below). In other words, the marginal psychological benefits are increasing until the decision maker reaches her goal (and become zero thereafter). In the graph, for example, the psychological marginal benefits increase until $y = 0.75$ when $g = 0.75$, which is exactly the point at which the goal is reached. Next, note that since the marginal psychological benefits exceed the marginal material cost for any $y \leq 0.75$, the optimal performance is at the corner: $y^*(g = 0.75) = 0.75$. We also observe that an increase in the goal shifts the marginal psychological benefit curve to the right: for a higher goal, e.g., $g = 1$, the performance needed to reach such goal must also higher. Moreover, for a given absolute performance level, the marginal psychological benefit is smaller due to diminishing sensitivities of the loss function. In Figure 2(b), we also see that when the goal is equal to 1, the doer's marginal psychological benefits exceed the marginal material costs whenever the goal is not yet reached, and that they are exactly equal at $r(y, \bar{y}) = g$. It follows that for any goal in $[y_l, 1]$, the decision maker must prefer the corner solution, exerting an absolute performance that reaches her goal. For more ambitious goals above $g = 1$, e.g.,

[23] The goal loss function is a normalized power function. The normalization ensures that the loss is zero at $z = 0$ and the (right) derivate at zero is 1.
[24] The exact shape of $y^*(g)$ for $g > \hat{g}$ depends on the particular functions for $c(\cdot)$ and $v(\cdot)$. It should be noted that $y^*(g)$ may not be continuous at $\hat{g}$: optimal performance may drop significantly as g increases to $\hat{g} + \epsilon$. For more details, consult the respective proof of proposition 1.
[25] While obviously specific to the example, the graphs do capture the doer's general behavior and incentives. For more technical details, as well as some special cases, we refer the interested reader to the proofs.
[26] Formally, the material marginal cost is $c'(y) - \beta\delta \cdot r_y(y, \bar{y})$.
[27] Specifically, $\partial[-\alpha \cdot \beta\delta \cdot v(g - r(y, \bar{y}))]/\partial y$.

$g = 1.25$, performance does not increase further, however. The marginal psychological benefit curve crosses the marginal material cost curve at a point below 1, resulting in a worse performance than with a goal of 1.



(a) Optimal absolute performance for a given goal **(b) Material marginal costs and psychological marginal benefits**

Figure 2. Optimal absolute performance for a given goal and its marginal costs and benefits

3.3 Optimal Goals

At time $t = 1$, the decision maker (goal-setter) takes her future-self's behavior as given and solves

$$\max_{g \in R^+} \; \beta\delta^2 \cdot r(y^*(g), \bar{y}) - \beta\delta \cdot c\big(y^*(g)\big) - \alpha \cdot \beta\delta^2 \cdot v(g - r(y^*(g), \bar{y})) \tag{3}$$

Since all payoffs accrue in the future, the goal-setter's preference over performance is unaffected by her present-bias: her ideal performance level is y_h whereas, without any additional incentives, her future-self only performs at y_l. To push her future-self towards y_h, the goal-setter can set a goal above $r(y_l, \bar{y})$ to shape her future self's preferences. From our characterization of $y^*(g)$, we know that any goal g' above $\hat{g}$, the goal that maximizes performance, is inefficient. This is because the very same performance can be induced by some lower goal $g'' \in (r(y_l, \bar{y}), \hat{g})$ but which, unlike g', is reached and thereby does not impose additional psychological costs. It follows that the goal-setter will select the goal from the interval $(r(y_l, \bar{y}), \hat{g}]$ that induces a performance level that is as close as possible to her ideal performance level y_h. If the highest performance that she can induce is less than her ideal, $y^*(g) < y_h$, she will set the goal precisely at $\hat{g}$. If, however, the maximum implementable performance is above y_h, she will simply select the goal that induces y_h. After all, her intention is not to maximize performance but to motivate her future-self to overcome her laziness, to perform at the level that she, the goal-setter, views optimal. Note that our model reiterates the typical prescription for goals: a good goal is challenging but achievable (Locke,1968; Locke and Latham, 1985). It effectively incentivizes effort without imposing additional (psychological) losses. We summarize these observations in the following result:

Proposition 2: Let $\hat{g} = arg\,max_{g \in R^+}\, r(y^*(g), \bar{y})$. If $y^*(\hat{g}) \leq y_h$ then the optimal goal is $g^* = \hat{g}$. If $y^*(\hat{g}) > y_h$ then g^* is the lowest goal that implements y_h.

So far, we have solved the decision problem (of setting the optimal goal and the respective optimal performance level) given (the decision maker's belief about) the performance of others. Hence, we have neglected the fact that the performance of others (i.e., that of the group itself) is determined in equilibrium. For our particular linear specification of relative performance, this turns out to be without loss as y_l and y_h are independent of $\bar{y}$, meaning that we can effectively treat $\bar{y}$ as a parameter. The optimal performance induced by the optimal goal is the same regardless of (explicitly) incorporating such equilibrium effect in the model or not (albeit the relative performance goal g^* would need to be adjusted accordingly).

Of course, we understand that, more generally, in a setting where relative performance matters, the absolute performance of everyone is determined in equilibrium. This may lead to an increase for each person's performance. Moreover, if goal setting is effective and used by at least some people, it will impact the group's overall performance and thereby each decision maker's relative performance and incentives. We leave it for future models to generalize $r(y, \bar{y})$ and explore the respective equilibrium effect in more detail. Instead, we focus on the primary effect of goals on enabling higher-level performance and, importantly, how information interacts with goals in shaping performance - which follows in section 3.4 below. The theoretical results from the latter analysis should extend to a more general formulation of the problem.[28]

3.4 Information and Goal Setting

We now turn to the importance of information for setting goals. In our *InfoGoal* treatment, participants are informed about their relative performance in a baseline test prior to setting a goal while in the *NoInfoGoal* treatment they are only informed about it after. Since participants can observe their own performance or at least be decently well aware of it, the information about their relative performance mainly teaches participants about the performance of others.[29] As all participants hold the correct information about $\bar{y}$ after this initial information provision state, their $t = 2$ problem is the same. However, they may solve it with different goals at hand. It follows

[28] Note that given the limited number of classes within the course, it is not sensible to randomize by class in a way to test for such an equilibrium effect on performance.

[29] Of course, the decision maker may not be perfectly aware of her own performance, or more generally of her own ability, even after the baseline test. If such a lack of knowledge about one's own ability was important, the decision maker could and would easily 're-test' her performance in a private setting without much effort. It is the performance of others that cannot be learned without outside information, however. More specifically, by observing the past performance level of their classmates, a student leans more about his/her classmate's overall ability to perform the plank, which in turn will affect their future performance in the final test.

that if participants are not motivated by goals, $\alpha = 0$, their absolute and consequently relative performance must be identical across treatments. This represents our Null Hypothesis.

To model the $t = 1$ problem under different information conditions, we take the following approach: the decision maker initially believes that the performance level of others is $\bar{y}_{belief}$, which may or may not be equal to their true performance $\bar{y}_{true}$. Hence, she may overestimate others' performance, $\bar{y}_{belief} > \bar{y}_{true}$, underestimate it, $\bar{y}_{belief} < \bar{y}_{true}$, or has accurate beliefs, $\bar{y}_{belief} = \bar{y}_{true}$. In the *NoInfoGoal* treatment, she sets her goal based on this initial belief. In the *InfoGoal* treatment, she updates her belief after learning the truth and sets her goal given the updated, correct beliefs.[30]

Underestimating one's relative ability: $\bar{y}_{belief} > \bar{y}_{true}$

If the decision maker overestimates the performance of others, which is equivalent to underestimating her own relative ability,[31] it becomes easier for her to reach the goal that she had set based on her initial belief when she learns the truth. After all, she only needs to exert a lower absolute performance in order to attain the same level of relative performance. Moreover, from the doer's point of view, reaching this goal is optimal. Figure 3 depicts the marginal costs and benefits for this case. In the graph, the performance level y_0 captures the optimal performance in response to the goal that was set based on the initial belief, i.e., when the decision maker still believes that others perform at $\bar{y}_{belief}$. Learning that the true $\bar{y}$ is lower shifts the psychological marginal benefits curve to the left. The optimal performance level under $\bar{y}_{true}$ given her initial goal is at y_1.[32] It follows that a decision maker in the *NoInfoGoal* treatment performs at y_1.

[30] Modelling incorrect beliefs using degenerate beliefs, i.e., $\bar{y}_{belief}$, instead of a more general subjective probability distribution over $\bar{y}$ is (1) simple and convenient, (2) a common approach in the literature, and (3) is not rejected by our data. The assumption that students rely on their best guess to set goals can be tested by comparing the goals set by accurate types in the *NoInfoGoal* treatment against (a) accurate types in the *InfoGoal* treatment, or more generally against (b) everyone in the *InfoGoal* treatment. We find no differences in either analysis, with or without controlling for different ability levels across subjects, see section 4.4.

[31] In view of our results section later, we refer to students who overestimate (underestimate) other's performance as those who underestimate (overestimate) their own relative ability / performance in the baseline test. The two types of notions are equivalent. This could be modelled more formally by both introducing first absolute abilities for the decision maker as well as for others and then introducing a notion of relative ability.

[32] Note, the difference between the goal and the relative performance is smaller under the new compared to the initial belief: $g - r(y, \bar{y}_{true}) = g - (y - \bar{y}_{true}) < g - (y - \bar{y}_{belief})$, which shifts the psychological marginal benefit curve to the left.

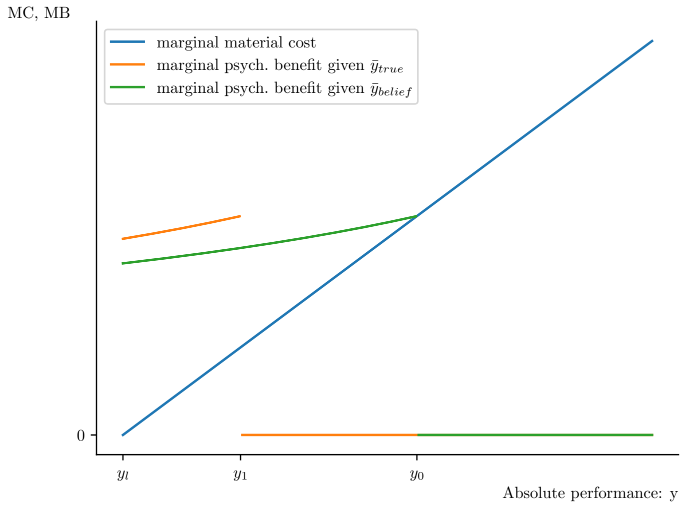


Figure 3. Underestimating one's relative ability: material marg. costs and psych. marg. benefits

In Figure 3, we can think of a decision maker in the *InfoGoal* treatment in terms of the following thought experiment: first, she sets a goal based on her initial belief. Her plan is thus to perform at y_0. Once she learns $\bar{y}_{true}$, the initial goal only induces the doer to perform at y_1, however. As the goal-setter only commits to a goal after she learns the true value of $\bar{y}$, she can adjust her goal in response to this new information. In particular, she can shift the marginal psychological benefit curve back to the right by increasing the goal. As she wants to undo her future-self's present bias and implement a performance level as close as possible to y_h, which was originally at y_0, she will increase her goal in such a way that the doer once again performs at the initial plan of y_0.[33] Since relative performance is increasing in absolute performance, it follows that

Hypothesis 1.1: information improves the relative and absolute performance of students who underestimate their relative ability, i.e., $y^r_{InfoGoal} > y^r_{NoInfoGoal}$ and $y_{InfoGoal} > y_{NoInfoGoal}$ if $\bar{y}_{belief} > \bar{y}_{true}$.

Hypothesis 1.2: information results in an increase in goals of students who underestimate their relative ability, i.e., $g_{InfoGoal} > g_{NoInfoGoal}$ if $\bar{y}_{belief} > \bar{y}_{true}$.

Overestimating one's relative ability: $\bar{y}_{belief} < \bar{y}_{true}$

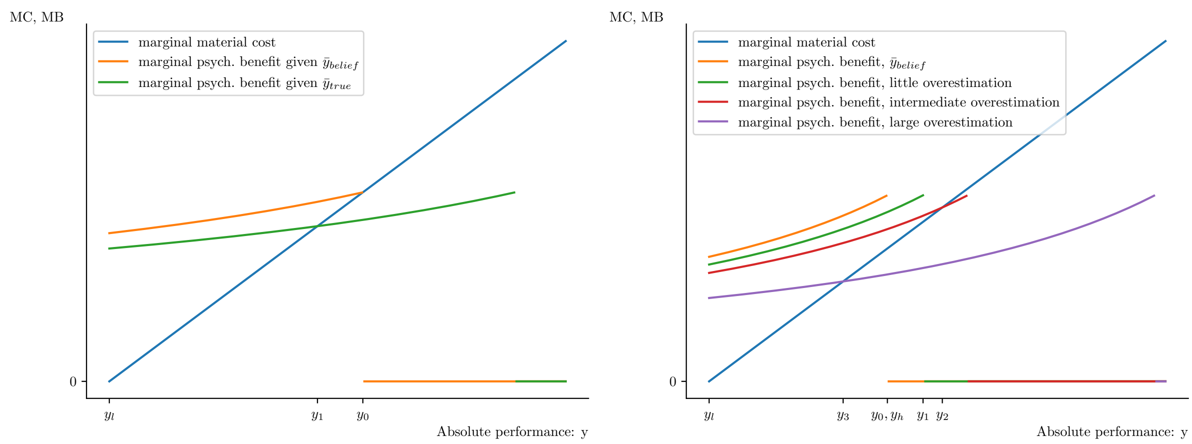
If the decision maker underestimates the performance of others, it becomes more challenging to reach her initial goal when she learns the truth. She needs to perform at a higher absolute level in order to attain the same level of relative performance. This does not mean that the doer actually

[33] Note that if the goal-setter found it optimal to implement y_0 for $\bar{y}_{belief}$, then she will also find it optimal to implement this very performance under the updated belief $\bar{y}_{true}$ To see this, recall that the ideal performance level y_h, and, more generally, the marginal material benefits and costs are independent of $\bar{y}$ in our setting.

increases her performance, however, as we can see from Figure 4(a). Learning that the true $\bar{y}$ is higher shifts the psychological marginal benefits curve to the right.[34] The goal becomes too challenging to reach, resulting in the lower performance of y_1 for a decision maker in the *NoInfoGoal* treatment. In contrast, a decision maker in the *InfoGoal* treatment can re-adjust her goal after learning $\bar{y}_{true}$. She modifies her goal downward, making it less challenging, shifting the marginal psychological benefit curve back to the left, and inducing her future-self to perform at the higher level of y_0. As in the underestimation case, incorrect beliefs result in lower absolute and relative performance:

Hypothesis 2.1a: information improves the relative and absolute performance of students who overestimate their relative ability, i.e., $y^r_{InfoGoal} > y^r_{NoInfoGoal}$ and $y_{InfoGoal} > y_{NoInfoGoal}$ if $\bar{y}_{belief} < \bar{y}_{true}$.

Hypothesis 2.2: information results in a decrease in goals of students who overestimate their relative ability, i.e., $g_{InfoGoal} < g_{NoInfoGoal}$ if $\bar{y}_{belief} < \bar{y}_{true}$.



(a) $y_0 \leq y_h$ **(b)** $y_0 = y_h$

Figure 4. Overestimating one's relative ability: material marg. costs and psych. marg. benefits

It turns out that our discussion of the overestimation case is incomplete, however, see Figure 4(b). When the optimal goal given the initial belief induces the future-self to perform at the *ideal performance level* of y_h, the psychological marginal benefit exceeds the material marginal cost at $y_0 = y_h$.[35] In this corner-solution case, the implications of information actually depend on how incorrect the decision maker's initial belief was. Upon learning $\bar{y}_{\mathrm{true}}$, the decision maker realizes that she needs to perform

[34] In this case, the difference between the goal and the relative performance increases: $g - r(y, \bar{y}_{true}) = g - (y - \bar{y}_{true}) > g - (y - \bar{y}_{belief})$.

[35] Unless, of course, we are in the non-generic case where the two are equal at y_h.

at a higher level. We see that unlike before, however, she is actually willing to increase her performance when her initial belief is not too incorrect. For small deviations in beliefs, the rightward shift in the marginal psychological benefit curve is small so that the marginal psychological benefits remain above the marginal material costs. Consequently, the doer still finds it optimal to reach her initial goal, which now requires a performance above y_h, e.g., y_1. While this level of performance is inefficient from the perspective of the goal-setter, it is in the interest of the doer to fully eliminate the psychological costs of falling short of her goals. For intermediate levels of overestimation, the marginal benefit curve intersects the marginal cost curve above y_0, e.g., at y_2, still resulting in a higher performance. The doer falls short of her goal, however, as she only finds it ideal to eliminate some but not all psychological losses. When the degree of overestimation becomes relatively large, performance drops below y_h as the motivation from goals become relatively small. It follows that the performance of a decision maker in the *NoInfoGoal* treatment may increase, e.g., y_1 or y_2, remain unchanged, or decrease, e.g., y_3. In contrast, the goal-setter in the *InfoGoal* treatment is able to induce the ideal performance y_h by reducing her goal, shifting the marginal psychological benefit curve back to the left. It follows that Hypothesis 2.1 needs to be updated to:

Hypothesis 2.1b: information may improve, not affect, or decrease the relative and absolute performance of students who overestimate their relative ability, i.e., $y^r_{InfoGoal} > or = or < y^r_{NoInfoGoal}$ and $y_{InfoGoal} > or = or < y_{NoInfoGoal}$ if $\bar{y}_{belief} < \bar{y}_{true}$.

Whether hypothesis 2.1a or 2.1b is applicable depends on how powerful goals are and/or how much laziness must be overcome. Naturally this will depend on the setting of analysis and may particularly depend on the timeframe over which these goals operate.[36]

Accurate beliefs: $\bar{y}_{belief} = \bar{y}_{true}$

The final case of accurate beliefs does not require much explanation. If initial beliefs are correct, our model predicts no differences between the treatment groups.

Hypothesis 3.1: information does not affect the relative and absolute performance of students who have accurate beliefs, i.e., $y_{InfoGoal} = y_{NoInfoGoal}$ and $y^r_{InfoGoal} = y^r_{NoInfoGoal}$ if $\bar{y}_{belief} = \bar{y}_{true}$.

Hypothesis 3.2: information does not affect the goals of students who have accurate beliefs, i.e., $g_{InfoGoal} = g_{NoInfoGoal}$ if $\bar{y}_{belief} = \bar{y}_{true}$.

[36] Indeed, these two cases also exist in the underestimation case. There, they do not result in different implications, however. For further details, consult the proofs in Appendix A.

Finally, note that our experiment only elicits goals from students in the two goal-setting treatments. This does not mean that students in the control group do not set their own personal goals. If they all do, then they could do so after learning the performance of others (or update their goals in response). Consequently, their performance (as well as the unobservable goals) should be similar to the *InfoGoal* treatment group. If instead, students in the control group do not set their own personal goals, they would perform worse than those in both goal-setting treatments. For some intermediate case (i.e., only some set goals or those that set goals themselves are less committed than if they were asked to set goals in the survey, which is the typical assumption behind many goal experiments), the control group should feature lower performance than the *InfoGoal* group but possibly better performance than the *NoInfoGoal* group.

4. Results

The key outcome variables collected from the experiment include the prediction of the baseline test score, the actual plank time (i.e., the absolute performance measure) and the corresponding actual rank/score (i.e., the relative performance measure) in the baseline test, the goal for the score in the final test in the *NoInfoGoal* and *InfoGoal* treatments, and the actual time and the corresponding actual score in the final test. We also collect data on basic individual characteristics, socioeconomic characteristics, physical fitness status, and preferences, as well as data about the environment when the tests were conducted. Appendix Table A2 reports the definition and summary statistics for all these variables and the students' self-confidence types we construct by comparing the prediction and the actual score of the baseline test.[37,38] Below we present results to address our research questions.

4.1 What is the effect of setting goals without and with knowing relative ability on performance?

Before we discuss the regression results, it is important to discuss how our intervention affects the relative performance measure. Since for each gender, a student's score is computed based on his/her relative performance in his/her class, it is not only determined by his/her own performance but also indirectly by everyone else's performance, both of which may be affected by our treatments. As the treatment assignment was also randomized for each gender in each class, the indirect effect is the same for all students in the evaluation group and thus is independent of the treatments. Hence, our analysis that uses the relative performance measure is unbiased. We will also report the results based on absolute performance measure, which is not affected by the indirect effect, and, as we will see shortly, paints a

[37] One student who did not participate in the baseline test is excluded from the analysis.
[38] We run pairwise (across treatments) Kolmogorov-Smirnov tests of equality of distributions of the control variables to assess whether our randomization worked well in most dimensions. The results in Appendix Table A3 show no significant differences for most variables at conventional significance levels, only except for baseline test rank, BMI index, patience, and positive reciprocity. In the regressions we control for these variables anyway, and find that our estimated treatment effects do not vary with whether or not we include these controls.

very similar picture. Finally, if there were to exist a general equilibrium effect on performance from our intervention, it would also be the same across treatments due to the randomization within class.

Table 1 reports the differential effects between *InfoGoal* and *NoInfoGoal* treatments on relative and absolute performance in the top panel, based on the regression results from ordinary least squares (OLS) estimations in the bottom panel. Relative performance is measured by the rank associated with each decile of the ranking range in the final test for each gender in each class, equal to 10 if in (0, 10%], 9 if in (10%, 20%], etc.,[39] in columns (1) - (4). Since students in the *NoGoal* treatment might also have set their own personal goals (i.e., compliance was imperfect), our estimator takes the assigned goal-setting status as treatment status, yielding an intention-to-treat result and a lower bound of the effect compared to the case that no one had set a goal.[40]

Column (1) includes only the treatment dummy variables, with the *NoGoal* treatment as the reference group. Columns (2), (3) and (4) further add control for ability captured by 9 dummies for the rank in the baseline test, basic individual characteristics, physical fitness and test environment characteristics, and socioeconomic characteristics and preferences, respectively. The results in the top panel show that there are no significant differences in final test rank between the two goal-setting treatments in any specification. The results in the bottom panel highlight that neither setting a goal without knowing one's ability in the *NoInfoGoal* treatment nor with knowing ability in the *InfoGoal* treatment significantly raises final test rank relative to the *NoGoal* treatment on average. Columns (1') - (4') present the analysis for absolute performance, i.e., the dependent variable is the absolute time (in seconds) that students hold the plank in the final test. All the controls in (2') - (4') are identical to those in (2) - (4) except that the ability control is replaced by the absolute time the student held the plank in the initial test. All results are qualitatively similar as those of relative performance.

In short, we find no overall effect for goal setting in terms of relative and absolute test performance. This can be driven by [1] knowing relative ability does not affect the goals set and hence does not affect performance; [2] knowing relative ability does affect the goals set but does not affect performance; or [3] the effects are present but heterogeneous and cancel each other out across students who over- and underestimate their relative ability. We next provide evidence along the consideration of these reasons.

Table 1. Treatment effect on relative and absolute performance in the final test

Dependent variable	Final test rank	Final test time

[39] We describe a student's rank with an ascending instead of a descending variable to facilitate easier interpretation of the estimates – an improvement in performance is captured by a positive estimate. As students simultaneously set a goal for their final test score and the associated ranking range (rank is in the descending order of ranking range), either could be used as the dependent variable. The results are qualitatively similar. We choose to report results based on rank for ease of interpretation.

[40] Since not everyone in the *NoGoal* treatment probably set own personal goals, and even if they do we cannot observe such goals, we cannot use the instrumental variable approach to estimate the local average treatment effect of goal on performance.

	(1)	(2)	(3)	(4)	(1’)	(2’)	(3’)	(4’)
Differential treatment effect								
(i) *InfoGoal - NoInfoGoal*	-0.143	0.125	0.182	0.119	2.691	5.894	7.694	6.818
	[0.779]	[0.754]	[0.662]	[0.807]	[0.854]	[0.541]	[0.439]	[0.557]
Regression results								
NoInfoGoal	0.462	0.214	0.237	0.211	8.435	0.750	1.342	-1.097
	(0.587)	(0.447)	(0.430)	(0.533)	(13.132)	(9.808)	(9.415)	(11.710)
InfoGoal	0.318	0.339	0.420	0.330	11.127	6.644	9.036	5.721
	(0.641)	(0.499)	(0.487)	(0.477)	(16.051)	(11.368)	(11.270)	(11.978)
Baseline test performance	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Basic individual characteristics	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Physical fitness and test environment	No	No	Yes	Yes	No	No	Yes	Yes
Socioeconomic characteristics and preferences	No	No	No	Yes	No	No	No	Yes
Observations	195	195	184	182	195	195	184	182

Notes: This table reports the estimates from OLS models in the bottom panel, and the implied differential treatment effects in the top panel. One student who did not participate in the baseline test is excluded from the analysis. Estimates for control variables are not reported; the full regression results are available upon request. Indicators for “baseline test performance” refers to the 9 dummies for the baseline test rank in columns (1) – (4) and refers to time (in seconds) of holding the plank in the baseline test in columns (1’) – (4’). Individual basic characteristics include gender, and age; physical fitness characteristics include height, weight, BMI, fat rate, and metabolic rate, pre-course exercise length, pre-course exercise times; test environment characteristics include test temperature, and test humidity; socioeconomic characteristics include Han, affiliated, urban *Hukou*, boarding school, monthly income, no. of siblings, birth order, healthy, and GPA; preferences include survey measures for self-regulation, willingness to compete, willingness to take risk, patience, positive reciprocity, negative reciprocity, altruism, and trust. Robust standard errors allowing for heteroskedasticity and clustered at the semester-class-gender level are reported in parentheses. Wald test p-values for the differential treatment effect are reported in brackets.

4.2 How are goals set?

Figure 5 depicts the distribution of goals set by students over the 10 possible ranks measured in the same manner as relative test performance by treatment. The red line represents the uniform distribution of test ranks, which is also the implied distribution from our relative grading rule, that is, exactly 10% of students will be assigned to each rank for their final performance. Comparing the uniform and actual distribution indicates that in general, goals are set too high, implying that many students will fail to achieve their goals. Moreover, this is particularly evident for students in the *NoInfoGoal* treatment, who set goals in the highest two ranks much more often than students in the *InfoGoal* treatment, resulting in the distribution of goals in the *InfoGoal* and *NoInfoGoal* treatments being significantly different (p-value=0.082, two-sample Kolmogorov-Smirnov test).

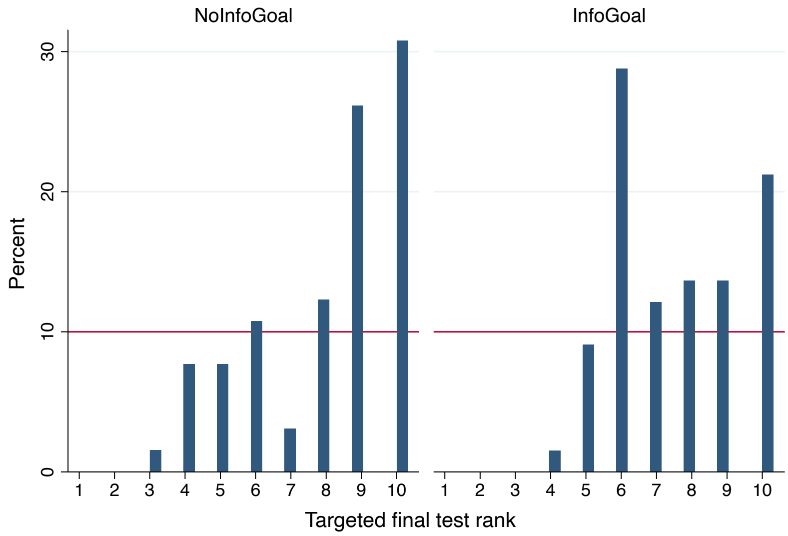


Figure 5. Distribution of goals by treatment

To provide more rigorous evidence on the effect of knowing ability on goal setting and how this effect can vary with the information at hand when setting goals, we conduct a regression analysis. Table 2 reports regression results from OLS estimations in the top panel, and the estimated differences in the bottom panel. Column (1) includes only the treatment dummy variable, with the *NoInfoGoal* treatment as the reference group, and relative ability, captured by 9 rank dummies for the baseline test; column (2) adds all remaining controls as in column (4) of Table 1. The results in both columns show that, on average, goals are not affected by knowing one's relative ability. Columns (3) and (4) further add the predicted and actual baseline test ranks (as linear measures), respectively, and the interactions between these two variables and the treatment dummy. Rank dummies for the baseline test are consequently omitted. From the top panel we find that, for the *NoInfoGoal* treatment, the actual baseline test rank does not correlate with goals, while the predicted rank significantly and positively correlates with goals. When the actual rank is known in the *InfoGoal* treatment, the effect of the actual baseline test rank significantly increases relative to its effect in the *NoInfoGoal* treatment, while the effect of the predicted rank significantly decreases relative to the *NoInfoGoal* treatment. The bottom panel of Table 2 shows that in the *InfoGoal* treatment both the actual and the predicted baseline test rank significantly and positively correlate with goals, and the former has a significantly larger magnitude than the latter (Wald tests, p-values < 0.01 in columns (3) and (4), respectively. The above differences across treatments suggest that students set goals based on the information at hand, which is the predicted rank when the actual rank is not available. When the actual rank becomes available, its impact on goals is much stronger than that of the predicted rank.

Table 2. Determinants of goals

Dependent variable: goal for final test rank	(1)	(2)	(3)	(4)
InfoGoal	-0.243	-0.518	0.049	-0.520
	(0.360)	(0.300)	(0.488)	(0.504)
Baseline test rank			0.101	0.099
			(0.072)	(0.073)
Baseline test rank × *InfoGoal*			0.398***	0.433***
			(0.074)	(0.075)
Predicted baseline test rank			0.523***	0.520***
			(0.074)	(0.076)
Predicted baseline test rank × *InfoGoal*			-0.374***	-0.327***
			(0.074)	(0.108)
Dummies for baseline test rank	Yes	Yes	No	No
Basic individual characteristics	No	Yes	No	Yes
Physical fitness and test environment	No	Yes	No	Yes
Socioeconomic characteristics and preferences	No	Yes	No	Yes
Observations	131	122	131	122
Estimated differences				
(i) *InfoGoal*: baseline test rank			0.499***	0.532***
			[<0.001]	[<0.001]
(ii) *InfoGoal*: predicted baseline test rank			0.148***	0.192**
			[0.005]	[0.022]

Notes: This table reports the estimates from OLS models in the top panel, and the estimated differences in the bottom panel. One student who did not participate in the baseline test is excluded from the analysis. Control variables are identical to those in Table 1. Estimates for control variables are not reported; the full regression results are available upon request. Robust standard errors allowing for heteroskedasticity and clustered at the semester-class-gender level are reported in parentheses. Wald test *p* values are reported in brackets. *** indicate statistical significance at the 1% level.

We then look at how accurate students' beliefs are with regards to their baseline test performance. Figure 6 depicts the distribution of predicted rank against actual rank for each treatment, with the 45-degree line indicating accurate prediction. We can see that in all treatments only a small proportion of students accurately predict their actual ranks and that both under- and over-prediction are prevalent. The pairwise correlation coefficients between predicted and actual ranks in the *NoGoal, NoInfoGoal* and *InfoGoal* treatments are 0.255, 0.160, 0.402, respectively. The low correlation implies that the predictions are inaccurate, and hence informing relative ability potentially have a big effect on correcting biased belief

4.3 How do students who accurately, under- or over-estimate their relative ability set goals and how does knowing relative ability affect their goal setting?

Motivated by our theory, we now classify students into three belief types, those who underestimate, accurately estimate and overestimate their relative ability based on whether their actual baseline test rank is higher than, equal to, or lower than their predicted rank. In our sample 48% overestimate and 39% underestimate their initial relative performance, while only 13% make accurate predictions.[41] The distributions look rather similar across treatments (p-value>0.1, Kolmogorov-Smirnov tests of equality of distributions of predicted over actual baseline test rank), indicating successful randomization of our treatment assignments.

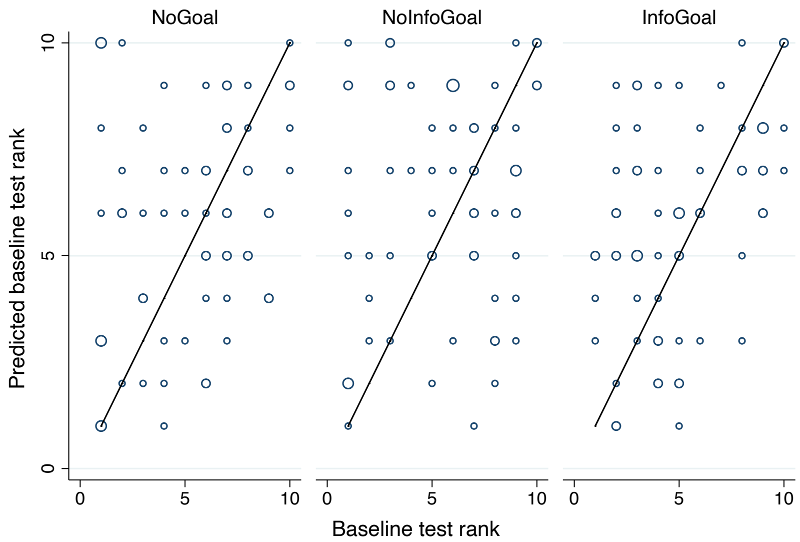


Figure 6. Distribution of predicted against actual baseline test ranks by treatment

We test the goal-setting hypotheses 1.2, 2.2 and 3.2 graphically and by conducting regression analysis. Figure 7(a) and 7(b) plot goals against the predicted and actual baseline test ranks by treatment and student type, respectively, with the 45-degree line implying equality between goal and relative performance. We find that almost all types of students set goals higher than their predicted ranks in the *NoInfoGoal* treatment (first row of Figure 7(a)), and higher than their actual ranks in the *InfoGoal* treatment (second row of Figure 7(b)). Recall that, when setting goals, students did not know their relative ability in the *NoInfoGoal* treatment but did know it in the *InfoGoal* treatment. This result thus suggests that students intended to set challenging goals based on the information at hand as the reference point, and provides evidence for why goals are set too high on average. The first row of Figure 7(b) and the second row of Figure 7(a) in turn highlight that information about relative ability helps students set better goals. In the first row of Figure 7(b), a significant proportion of students who underestimate their relative performance set goals lower than their actual baseline test ranks, which makes such goals too easy to achieve and ineffective in raising performance. In contrast, the second row of Figure 7(a) shows that, after learning their relative performance, a considerable proportion of students who had overestimated their rank now set goals below their predicted baseline test ranks (although all goals are

[41] Figure 6 provides a more detailed overview of their predictions relatively to their actual baseline test performance. Appendix Table A4 shows the distribution of students by treatment and type.

still higher than their actual baseline test ranks), which otherwise would cause the goals to be too difficult to achieve and also ineffective in raising performance.

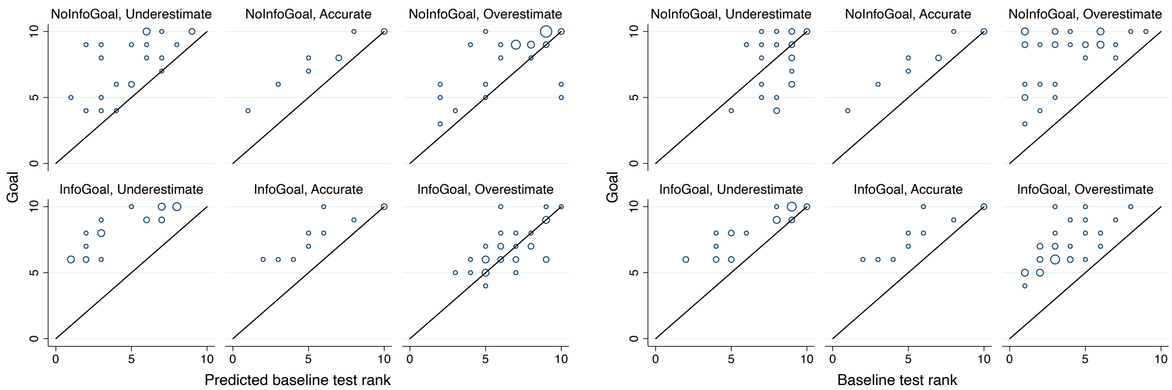


Figure 7(a) and 7(b). Distribution of goals against predicted and actual baseline test ranks by treatment and student type

To investigate the effect of knowing one's relative ability on goals for different types of students, we extend our specifications in columns (1) and (2) of Table 2 by adding indicators for whether students under- or overestimate their relative ability (as well as respective interaction terms with the treatments), with accurate estimating students in the *NoInfoGoal* treatment as the reference group. The top panel of Table 3 provides the Wald test results for the estimated effects of providing information on goal setting for the three types based on regression results from OLS estimations in the bottom panel. The top panel shows that knowing one's relative ability significantly increases goals set by students who underestimate their own relative ability and significantly reduces goals set by those who overestimate their relative ability. Information does not alter goals for those who hold accurate beliefs. This result corroborates our insights from Figures 7(a) and 7(b) and confirms the goal-setting hypotheses 1.2, 2.2 and 3.2. It highlights how information improves goal setting.

The bottom panel further highlights that in the *NoInfoGoal* treatment, under- and overestimating types set significantly lower and higher goals, respectively, compared to those who have accurate estimates. Moreover, notice that the opposing effects of information on goals for under- vs. overestimating students balance out, which explains why the overall treatment effect of ability information on goals is not significantly different from zero in Table 2.

When deriving hypotheses 1.2, 2.2 and 3.2, we assumed that students only rely on their best estimate over relative ability when setting goals (which is likely poor in the absence of information but rather precise upon receiving information) and disregarded any potential second order impact of uncertainty on goal setting. One possible influence of uncertainty on goals is, for example, for goals to be set more cautious.[42] We can test for the impact of uncertainty on goals by comparing the goals set

[42] For example, students may want to set slightly less ambitious goals (relatively to their expected relative ability) as more ambitious goals are relatively costlier if their beliefs turn to be (ex-post) too optimistic than if they turn out to be too pessimistic. In the former case, the cost of a too high goal arises from (1) lower motivational benefit

by accurate types in the *NoInfoGoal* treatment either against accurate types in the *InfoGoal* treatment, or more generally against everyone in the *InfoGoal* treatment, whose beliefs are accurate and without uncertainty after the information provision. Appendix Table A5 reports the regression results. We find no differences in either analysis, with or without controlling for initial ability. In other words, we find no evidence that students respond to uncertainty over relative ability by, for example, setting more cautious goals.[43]

Table 3. Treatment effect on goals by student type

Dependent variable: goal for final test rank		(1)	(2)
Differential treatment effects			
InfoGoal-NoInfoGoal	(i) underestimate	1.432***	1.220***
		[0.005]	[0.005]
	(ii) *accurate*	0.253	0.275
		[0.508]	[0.741]
	(iii) *overestimate*	-1.268***	-1.812***
		[0.005]	[0.004]
Regression results			
InfoGoal		0.253	0.275
		(0.372)	(0.812)
Underestimation		-1.135**	-1.553**
		(0.496)	(0.594)
Overestimation		1.682***	1.344**
		(0.456)	(0.617)
InfoGoal × underestimation		1.179*	0.945
		(0.557)	(0.899)
InfoGoal × overestimation		-1.521**	-2.086
		(0.531)	(1.199)
Dummies for baseline test rank		Yes	Yes
Basic individual characteristics		No	Yes
Physical fitness and test environment		No	Yes
Socioeconomic characteristics and preferences		No	Yes
Observations		131	122

Notes: This table reports the estimates from OLS models in the bottom panel, and the implied differential treatment effects in the top panel. One student who did not participate in the baseline test is excluded from the analysis. Control variables are identical to those in Table 1. Estimates for control variables are not reported; the full regression results are available upon request. Robust standard errors allowing for heteroskedasticity and clustered at the semester-class-gender level are reported in parentheses. Wald test p values are reported in brackets. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

and (2) a psychological loss from falling short of such goal. In the latter, the cost of too low goal is also a lower motivational benefit. However, there is no psychological upside from meeting (or surpassing) this goal. Ex-Ante, the one-sided psychological downsides from falling short of fairly ambitious goal may therefore induce the goal setter to set a relatively more cautious goal.

43 One interesting direction for future work could be to explore how subjects with accurate beliefs respond to uncertainty about their (relative) ability when setting goals. To estimate such effects precisely, a much larger sample will be required.

4.4 How does knowing ability affect the performance of different types of students?

Finally, we explore the effect of information on goal setting for the three belief-types of students. To tackle this question, we introduce the type indicators and the interactions between the treatment and the type indicators into the specifications of Table 1, with the accurate type in the *NoGoal* treatment as the reference group. The top panel in Table 4 presents the differential effects of providing information on test performance for each of the three types, based on the regression results from OLS estimations in the bottom panel. Our discussion will mainly focus on regression specification with full control shown in columns (4) and (4').

The estimated effects in row (vii) of both columns (4) and (4') show that providing information in the *InfoGoal* treatment significantly raises final test rank and absolute time for students who underestimate their own relative ability, compared to those in the *NoInfoGoal* treatment.[44] This result lends support to Hypothesis 1.1. Information allows those who underestimate themselves to set a more sensible goal. In addition, there seems to be some evidence, as shown in row (i), that setting goals without knowing ability hurts these students who underestimate their own relative ability – reducing their final test rank and absolute time compared to those in the *NoGoal* treatment. This could be driven by some form of demotivating effect from too low goals or by students in the control group who set their own personal goals based on the full information. Notice that the latter idea may also explain why we do not see any difference between students who underestimate in the *InfoGoal* and those in the *NoGoal* treatments in row (iv). After all, if sufficiently many students in the *NoGoal* treatment also set goals, they can set after learning their relative performance.

Next, while goals set by students who overestimate their relative ability are improved by information as found in Table 3, their performance is not, as shown in row (ix). Consequently, Hypothesis 2.1a is not supported. However, the behavior can still be consistent with Hypothesis 2.1b, which says that for those who overestimate their relative ability, performance may increase, remain the same, or decrease in response to information.[45] For the accurate type, our results in row (viii) support Hypothesis 3.1.[46]

[44] We verify whether the results are robust to alternative ability measures. In columns (1) and (2) in Appendix Table A6 we repeat the regressions from columns (1) and (4) of Table 4 using a simple quadratic specification, i.e., combining the linear measure of baseline test rank, instead of the 9 ability rank dummies, and its quadratic term. In columns (3) and (4), we repeat the regressions from columns (1') and (4') of Table 4 with a quadratic specification for baseline test time. In Appendix Table A7 we control for initial performance using both the baseline test ranks and the baseline test time for the two outcome measures, respectively. The results from both tables are qualitatively similar compared to columns (1) and (4) and columns (1') and (4') in Table 4.

[45] With our limited data, we unfortunately cannot test the theory behind hypothesis 2.1b further, which would predict that performance first increases and eventually decreases in the degree of overestimation.

[46] This result should be treated with some caution given the small number of observations in this group. Consequently, we also redefine students as accurate type both when they correctly predicted their rank and when they mispredicted their rank within only one rank (from above or below), which by construction increases the sample size of the accurate type. The results shown in Table A8 are qualitatively similar to the results shown in Table 4 for the underestimating type. It also provides further evidence for Hypothesis 3.1, i.e., there is no effect for accurate type.

Overall, it appears that information has a heterogenous effect on performance across types and that it particularly benefits those who underestimate their relative ability. Variation in the effects across types, which introduces noise and partially counteracts each other, explains why we do not find any overall effects of goal setting with compared to without knowing one's relative ability on performance in Table 1.

We also investigate how satisfaction with final test performance relates to personal goals. The results (reported and discussed in more detail in Appendix C) show that the satisfaction brought from reaching one's goal is equivalent to that from increasing the test performance by more than 3 ranks. This evidence suggests the importance of goals in shaping preference over outcomes and links goals to motivation for training harder and performing better in our setting of a PE course.

Table 4. Treatment effect on relative and absolute performance in the final test for various types

Dependent variable		Final test rank				Final test time			
		(1)	(2)	(3)	(4)	(1')	(2')	(3')	(4')
Differential treatment effect									
NoInfoGoal-NoGoal	(i) underestimate	-0.667	-0.613	-0.477	-0.866**	-25.288*	-23.938*	-18.528	-27.146***
		[0.303]	[0.333]	[0.427]	[0.041]	[0.076]	[0.077]	[0.124]	[0.005]
	(ii) accurate	-0.224	-0.032	-0.504	-0.209	-22.722	-20.836	-31.528	-23.655
		[0.676]	[0.956]	[0.439]	[0.831]	[0.426]	[0.484]	[0.341]	[0.472]
	(iii) overestimate	0.868	0.815	0.872	0.993	24.334	23.286	23.142	23.157
		[0.245]	[0.279]	[0.257]	[0.315]	[0.159]	[0.177]	[0.190]	[0.280]
InfoGoal-NoGoal	(iv) underestimate	0.307	0.386	0.362	0.019	4.508	6.081	8.189	-0.016
		[0.660]	[0.554]	[0.541]	[0.972]	[0.773]	[0.687]	[0.599]	[0.999]
	(v) accurate	-1.144	-1.305	-1.744	-1.325	-16.349	-20.579	-28.154	-26.138
		[0.148]	[0.136]	[0.102]	[0.260]	[0.559]	[0.480]	[0.381]	[0.375]
	(vi) overestimate	0.870	0.823	1.040	1.066	16.281	15.153	19.456	18.764
		[0.116]	[0.119]	[0.103]	[0.121]	[0.257]	[0.274]	[0.206]	[0.240]
InfoGoal-NoInfoGoal	(vii) underestimate	0.974**	0.999**	0.839*	0.885*	29.796**	30.019**	26.717*	27.130*
		[0.037]	[0.030]	[0.068]	[0.058]	[0.039]	[0.036]	[0.091]	[0.083]
	(viii) accurate	-0.920	-1.273*	-1.239	-1.115	6.373	0.256	3.374	-2.483
		[0.156]	[0.095]	[0.211]	[0.263]	[0.675]	[0.988]	[0.878]	[0.918]
	(ix) overestimate	0.002	0.007	0.167	0.073	-8.053	-8.134	-3.686	-4.393
		[0.998]	[0.992]	[0.824]	[0.937]	[0.708]	[0.703]	[0.850]	[0.861]
Regression results									
NoInfoGoal		-0.224	-0.032	-0.504	-0.209	-22.722	-20.836	-31.528	-23.655
		(0.523)	(0.579)	(0.632)	(0.960)	(27.657)	(28.919)	(31.894)	(31.969)
InfoGoal		-1.144	-1.305	-1.744	-1.325	-16.349	-20.579	-28.154	-26.138
		(0.743)	(0.820)	(0.991)	(1.124)	(27.234)	(28.289)	(31.048)	(28.422)
Underestimation		-0.398	-0.321	-0.930	-0.555	-7.649	-7.383	-21.099	-13.890
		(0.693)	(0.699)	(0.787)	(0.929)	(25.417)	(25.562)	(29.291)	(23.649)

Overestimation	-0.633	-0.567	-1.022	-0.748	-15.327	-14.377	-27.395	-22.593
	(0.611)	(0.564)	(0.727)	(1.149)	(23.438)	(23.182)	(29.356)	(31.495)
NoInfoGoal × underestimation	-0.443	-0.580	0.0278	-0.656	-2.565	-3.103	13.000	-3.491
	(0.722)	(0.775)	(0.869)	(0.986)	(30.682)	(31.732)	(35.237)	(33.414)
NoInfoGoal × overestimation	1.091	0.848	1.377	1.202	47.057	44.122	54.670*	46.812
	(0.713)	(0.744)	(0.808)	(1.276)	(28.822)	(29.201)	(29.772)	(33.661)
InfoGoal × underestimation	1.451*	1.691**	2.106*	1.344	20.857	26.660	36.342	26.122
	(0.754)	(0.757)	(1.010)	(1.138)	(28.859)	(29.320)	(34.782)	(33.503)
InfoGoal × overestimation	2.014***	2.128***	2.783***	2.391**	32.630	35.732	47.610	44.902
	(0.648)	(0.671)	(0.871)	(0.933)	(28.234)	(28.455)	(34.227)	(27.263)
Baseline test performance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Basic individual characteristics	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Physical fitness and test environment	No	No	Yes	Yes	No	No	Yes	Yes
Socioeconomic char. and preferences	No	No	No	Yes	No	No	No	Yes
Observations	195	195	184	182	195	195	184	182

Notes: This table reports the estimates from OLS models in the bottom panel, and the implied differential treatment effects in the top panel. One student who did not participate in the baseline test is excluded from the analysis. Control variables are identical to those in Table 1. Estimates for control variables are not reported; the full regression results are available upon request. Robust standard errors allowing for heteroskedasticity and clustered at the semester-class-gender level are reported in parentheses. Wald test *p* values are reported in brackets. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

5. Conclusions

Self-set goals are simple motivational tools that can be used to push one's future self to work harder. By the very nature of this setup – goals are often set well in advance – the goal setter tends to have less information regarding the future self's and, in the case of relative goals, others' future capacity in doing the work. This, in turn, usually leads to ineffective goals. In this paper, we explore the role of information in relative-performance goal setting and performance – using data from a field experiment conducted in a university physical education course that uses relative grading. We generate random variation in whether goals were elicited and whether students were informed about their relative performance in the baseline test prior to setting goals. By eliciting their beliefs about their baseline relative performance, we further test for heterogeneous effect of information on goal setters who predicted their performance accurately, under- or overestimated it.

Our theoretical model predicts that performance follows an inverted V-shape with goal and that optimal goals are challenging but achievable. Information about relative ability enables the goal setter to increase or decrease goals depending on whether she under- or overestimates her ability relative to others'. This is shown to always improve performance if she underestimates her relative ability. If she overestimates her relative ability, information always improves performance in the (interior solution) case where the goal setter is only able to implement a performance below her most preferred performance level but may decrease in the alternative (corner solution) case. In our model, information neither affects the goals nor the performance of those decision makers who hold accurate beliefs about their relative performance. Our field experiment show that the majority of students set goals above their initial performance levels, either known or predicted. Our findings largely confirm our theoretical predictions by highlight that providing information about one's relative ability significantly improves goals – increasing goals set by underestimating students while decreasing goals set by overestimating ones, and that this results in improved performance for underestimating students.

There are three natural directions in which this research could be fruitfully extended. First, while we focus on the case of relative goal setting, much of our (theoretical) work and insights translate to absolute goal setting with uncertainty about the goal setter's own (future) ability or her future environment.[47] This could be a fruitful setting for future empirical work. Second, it would be interesting to analyze our analytical predictions for the overestimation case further and see in which settings performance is going to increase or to decrease, and how this relates

[47] To do so, the key change is substituting relative performance $r(y, \bar{y})$ by an absolute performance measure $f(e, \theta) = e + \theta$. In this model, the decision maker chooses effort e (instead of y) given some ability level θ (instead of performance of others $\bar{y}$). Costs are defined over effort and the goal is set in terms of absolute performance.

to the extent the decision maker's beliefs are incorrect. Third, given that we find possible demotivation effect of goal setting without information, it could also be interesting to consider the possible detrimental effects of information on goal setting and subsequent information avoidance.

In general, goal setting is low-cost, scalable, logistically simple, and shown to be an effective commitment device across a variety of settings. This can be particularly important for policy making in developing countries like China, which are less capable of using financial incentives, for example, to promote healthier lifestyle despite dramatically rising health-related problems such as obesity, hypertension, and diabetes.[48] Exploring the potential impacts of non-monetary incentives in more detail, could provide tremendous long-term benefits.

[48] For instance, Zhang and He (2016) document a continuing trend of increasing obesity rate and myopia for Chinese adolescents and a declining physical fitness such as endurance, speed, power and strength from 2000 to 2014.

References

Avdeenko, A., Bohne, A. and Frölich, M. (2019). Linking savings behavior, confidence and individual feedback: A field experiment in Ethiopia. Journal of Economic Behavior & Organization, 167, 122-151.

Beckmann, N., Beckmann, J. F. and Elliott, J. G. (2009). Self-confidence and performance goal orientation interactively predict performance in a reasoning test with accuracy feedback. Learning and Individual Differences, 19(2), 277-282.

Bobba, M. and V. Frisancho (2019): "Perceived Ability and School Choices," TSE Working Papers 16-660, Toulouse School of Economics (TSE).

Bönte, W., Lombardo, S. and Urbig, D. (2017). Economics meets psychology: experimental and self-reported measures of individual competitiveness. Personality and Individual Differences, 116, 179-185.

Brookins, P., S. J. Goerg, and Kube, S. (2017). Self-chosen goals, incentives, and effort.

Burdina, M., Hiller, R. S. and Metz, N. E. (2017). Goal attainability and performance: evidence from Boston marathon qualifying standards. Journal of Economic Psychology, 58.

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

Czibor, E., Onderstal, S., Sloof, R. and van Praag, C. M. (2020). Does relative grading help male students? Evidence from a field experiment in the classroom. Economics of Education Review, 75, 101953.

Corgnet, B., Gómez-Miñambres, J. and Hernán-Gonzalez, R. (2015). Goal setting and monetary incentives: When large stakes are not enough. Management Science, 61(12), 2926-2944.

Corgnet, B., Gómez-Miñambres, J. and R. Hernán-González. (2018). Goal setting in the principal–agent model: weak incentives for strong performance. Games and Economic Behavior, 109, 311–326.

Dalton, P.S., Gonzalez Jimenez, V.H. and Noussair, C.N. (2015). Paying with self-chosen goals: incentives and gender differences. CentER Discussion Paper, 2015-021. Tilburg: Economics.

Dizon-Ross, R. (2019). Parents' beliefs about their children's academic ability: Implications for educational investments. American Economic Review, 109(8), 2728-65.

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

Eil, D., and Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. American Economic Journal: Microeconomics, 2(3), 114–138.

Gaumer Erickson, A. S., Monroe, K., Soukup, J., and Noonan, P. M. (2018). Self-regulation formative questionnaire. In P. Noonan & A. Gaumer Erickson. The skills that matter: Teaching interpersonal and intrapersonal competencies in any classroom (p. 177–178). Corwin.

Falk, A., Becker, A., Dohmen, T., Enke, B., Huffman, D. and Sunde, U. (2018). Global evidence on economic preferences. Quarterly Journal of Economics, 133(4), 1645-1692.

Fan, J. and Gómez-Miñambres, J. (2020). Non-binding goals in teams: A real effort coordination experiment. Manufacturing & Service Operations Management, 22(5), 1026-1044.

Franco, C. (2019): "How does relative performance feedback affect beliefs and academic decisions?" Working paper, Norwegian School of Economics.

Gan, T. (2022). How Parents' Beliefs About Their Children's Academic Ability Affect Educational Investments. Working paper, University of Maryland.

Gill, D., Kissová, Z., Lee, J. and Prowse, V. (2019). First-place loving and last-place loathing: how rank in the distribution of performance affects effort provision. Management Science 65, 494–507.

Goerg, S. J. and Kube, S. (2012). Goals (th)at work, goals, monetary incentives, and workers' performance. Discussion Paper Series of the Max Planck Institute for Research on Collective Goods 2012_19, Max Planck Institute for Research on Collective Goods.

Gosnell, G. K., List, J. A. and Metcalfe, R. D. (2020). The impact of management practices on employee productivity: A field experiment with airline captains. Journal of Political Economy, 128(4), 1195-1233.

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

Hadwin, A. F. and Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. Learning and Instruction, 24, 37-47.

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

Harrison, G.W., and List, J.A. (2004). Field experiments. Journal of Economic Literature, 42(4), 1009-55.

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

Jalava, N., Joensen, J. S. and Pellas, E. (2015). Grades and rank: Impacts of non-financial incentives on test performance. Journal of Economic Behavior & Organization, 115, 161-196.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 47(2), 263-292.

Koch, A.K. and Nafziger, J. (2011). Self-regulation through goal setting. Scandinavian Journal of Economics, 113(1), 212-227.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. The Quarterly Journal of Economics, 121(4), 1133-1165.

Kuhn, P. J. and Yu, L. (2021). Kinks as goals: Accelerating commissions and the performance of sales teams (No. w28487). National Bureau of Economic Research.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. The Quarterly Journal of Economics, 112(2), 443-478.

Locke, E. A. (1968). Toward a theory of task motivation incentives. Organizational Behavior and Human Performance, 3, 157–189.

Locke, E. A. and Latham, G. P. (1985). The application of goal setting to sports. Journal of Sport Psychology, 7, 205–222.

Mobius, M. M., Niederle, M., Niehaus, P., and Rosenblat, T. S. (2014). Managing self-confidence. working paper, 2014.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. American economic review, 89(1), 103-124.

Sackett, A. M., Wu, G., White, R. J. and Markle, A. (2014). Harnessing optimism: How eliciting goals improves performance. Working Paper. Available at SSRN: https://ssrn.com/abstract=2544020 or http://dx.doi.org/10.2139/ssrn.2544020

Sauer, S. J., Rodgers, M. and Becker, W. J. (2018). The effects of goals and pay structure on managerial reporting dishonesty. Journal of Accounting, Ethics and Public Policy, 19(3).

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

Suvorov, A. and Van de Ven, J. (2008). Goal setting as a self-regulation mechanism. Working Paper. Available at SSRN 1286029.

Strotz, R. H. (1955), Myopia and Inconsistency in Dynamic Utility Maximization, Review of Economic Studies 23, 165–180.

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. The Quarterly Journal of Economics, 106(4), 1039–1061.

Uetake, K. and Yang, N. (2018). Success breeds success: weight loss dynamics in the presence of short-term and long-term goals (No. 170002).

van Lent, M. and Souverijn, M. (2020). Goal setting and raising the bar: A field experiment. Journal of Behavioral and Experimental Economics, 87, 1-16.

van Lent, M. (2019). Goal setting, information, and goal revision: A field experiment. German Economic Review, 20(4), e949-e972.

Wu, G., Heath, C. and Larrick, R. (2008). A prospect theory model of goal behavior. Unpublished manuscript.

Zhang, Y., and He, L. (2016). Dynamic analysis of adolescent physical health in China: based on four national physical health monitoring data from 2000 to 2014. China Youth Research (in Chinese), 24(6), 5-12.

Appendix A. Appendix tables and figures

Figure A1. Association between test score and ranking range

Score	Male/Female ranking range
100	(0, 10%]
94	(10%, 20%]
88	(20%, 30%]
84	(30%, 40%]
80	(40%, 50%]
76	(50%, 60%]
72	(60%, 70%]
68	(70%, 80%]
64	(80%, 90%]
60	(90%, 100%]

Figure A2. Grade composition for the module

Task	Weight (Highest possible score)	Test score (points)	Male/Female ranking range
Test 1	8	88	(20%, 30%]
Test 2	8		
Test 3	8		
Final test	60		
Class attendance	10		
Survey 1	3		
Survey 2	3		
Total	100		

Figure A3. First survey goal-setting related questions in the order for the *InfoGoal* treatment

1. Please predict your score on last week's plank test (test 1) (choose one option in the following table).

Note: Please note that you will earn 2 out of 3 points by completing all other questions in this survey; the other 1 point depends on how accurate your prediction is in this question. The scoring criteria are as follows: if your prediction ranking range is the same as the actual ranking range, you will earn 1 point; if your prediction is 1 range higher or lower than the actual ranking range, you will earn 0.5 points; if your prediction is more than 1 range higher or lower than the actual ranking range, you will get 0 points.

	Score	Male/Female ranking range
○	100	(0, 10%]
○	94	(10%, 20%]
○	88	(20%, 30%]
○	84	(30%, 40%]
○	80	(40%, 50%]
○	76	(50%, 60%]
○	72	(60%, 70%]
○	68	(70%, 80%]
○	64	(80%, 90%]
○	60	(90%, 100%]

2. Your actual score in last week's plank test (test 1) is shown in the following table:

Task	Weight (Highest possible score)	Test score (points)	Male/Female ranking range
Test 1	8	88	(20%, 30%]
Test 2	8		
Test 3	8		
Final test	60		
Class attendance	10		
Survey 1	3		
Survey 2	3		
Total	100		

Please check the box below

□ I already know my "Test 1" score

3. Please set a goal for your score in the final test (choose one option in the following table). Note that you will be reminded of your goal when you are informed of your score for each test through WeChat messages.

	Score	Male/Female ranking range
○	100	(0, 10%]
○	94	(10%, 20%]
○	88	(20%, 30%]
○	84	(30%, 40%]
○	80	(40%, 50%]
○	76	(50%, 60%]
○	72	(60%, 70%]
○	68	(70%, 80%]
○	64	(80%, 90%]
○	60	(90%, 100%]

Figure A4. An example of the WeChat message for the two goal-setting treatments

Hello, XXX

The goal that you set for your score in the final test of the plank is: 94 - (10%,20%].

Your actual score on last week's plank test 2 is shown in the table below.

Task	Weight (Highest possible score)	Test score (points)	Male/Female ranking range
Test 1	8	88	(20%, 30%]
Test 2	8	84	(30%, 40%]
Test 3	8		
Final test	60		
Class attendance	10		
Survey 1	3		
Survey 2	3		
Total	100		

Please reply when you receive this message. Thank you!

Education reform project assistant
Date

Figure A5. Survey questions related to socioeconomic background in the second survey

1. What is your ethnicity?

○Han ○Minority

2. What is your political affiliation?

○Not politically affiliated ○Youth League member ○Communist Party member ○Democratic Party member

3. Which kind of Hukou did you hold before enrolling into university?

○Non-agricultural ○Agricultural ○No Hukou

4. Have you been to a boarding elementary, middle or high school?

○Yes ○No

5. What is your average total after-tax monthly income from all sources (including from your parents, etc.)? _____ Chinese yuan

6. Do you have any siblings (including half-siblings and siblings adopted by parents)?

○No ○Yes

7. [Only shown when choosing "yes" in question 6] How many siblings do you have excluding yourself (including half-siblings and siblings adopted by parents)? ______

8. [Only shown when choosing "Yes" in question 6] please specify your order in the siblings. ______

9. What is your health status?

○Very healthy ○Rather healthy ○Fair ○Rather unhealthy ○Very unhealthy

10. What was your grade point average (on a 100 points scale) for all courses, including physical education and optional courses, but excluding additional credits earned from extracurricular activities, in the previous academic year (2017-2018)? ______

Table A2. Variable definition and summary statistics

Variable	Definition	No. of Obs.	Mean	Std. Dev.	Min.	Max.
Dependent variable						
Rank	rank in the final test, =10 if (0, 10%]/ score 100; =9 if (10%, 20%] / score 94, etc.	195	5.26	2.87	1	10
Time	time (in seconds) of holding the plank in the final test	195	200.51	81.50	49.13	489.80
Goal	goal for rank in the final test, =10 if (0, 10%]/ score 100; =9 if (10%, 20%] / score 94, etc.	131	7.77	1.94	3	10
Treatment variable						
NoGoal	=1 if in *NoGoal* treatment; 0 otherwise	195	0.33	0.47	0	1
NoInfoGoal	=1 if in *NoInfoGoal* treatment; 0 otherwise	195	0.33	0.47	0	1
InfoGoal	=1 if in *InfoGoal* treatment; 0 otherwise	195	0.34	0.47	0	1
Variables related to type						
Underestimating	=1 if score is higher than prediction for the baseline test; 0 otherwise	195	0.38	0.49	0	1
Accurately-predicted	=1 if score is equal to prediction for the baseline test; 0 otherwise	195	0.13	0.34	0	1
Overestimating	=1 if score is lower than prediction for the baseline test; 0 otherwise	195	0.48	0.50	0	1
Control variable						
Baseline test time	time (in seconds) of holding the plank in the baseline test	195	135.83	59.51	20.79	388.61
Baseline test rank	=10 if (0, 10%]/ score 100; =9 if (10%, 20%] / score 94, etc.	195	5.27	2.86	1	10
Prediction for baseline test score	predicted rank for the baseline test, =10 if (0, 10%]/ score 100; =9 if (10%, 20%] / score 94, etc.	195	5.94	2.55	1	10
Basic controls						
Female	=1 if the student is female; 0 otherwise	195	0.69	0.46	0	1

Age	Age of the student	195	19.58	0.99	15.55	20.79
Health and environment controls						
Height	height (in centimeters)	192	166.03	7.87	151.10	192.10
Weight	weight (in kilograms)	192	58.42	10.07	37.40	93.90
BMI index	weight (in kilograms/height in meters^2)	192	21.31	3.55	14.70	52.60
Fat rate	body fat rate	192	23.02	5.55	5.80	49.80
Metabolic rate	basic metabolic rate	192	1486.53	190.83	1198.00	2181.00
Pre-course exercise length	total number of minutes the student trained on the playground or in the gym on campus the semester before enrolling in the course	187	1264.92	498.92	0.00	3563.12
Pre-course exercise times	total number of times the student train on the playground or in the gym on campus the semester before enrolling in the course	187	26.64	9.98	0.00	63.00
Test temperature	temperature in centigrade in the room for the plank test	195	19.08	3.46	13.60	23.70
Test humidity	humidity degree in the room for the plank test	195	27.68	3.15	22.00	34.00
Socioeconomic characteristic controls						
Han	=1 if the student's ethnicity is Han; 0 otherwise	195	0.90	0.30	0	1
Affiliated	=1 if the student is a Communist Party member or a Youth League member; 0 otherwise	195	0.92	0.27	0	1
Urban *Hukou*	=1 if the student has non-agricultural *Hukou* before enrolling in university; 0 otherwise	195	0.80	0.40	0	1
Boarding school	=1 if the student has ever attended boarding school before enrolling in university; 0 otherwise	195	0.59	0.49	0	1
Monthly income	monthly disposable income (in CNY)	195	2204.82	2222.51	0	30000
No. of siblings	number of siblings	193	0.61	1.00	0	7
Birth order	order of birth in all siblings	195	1.22	0.67	1	7
Healthy	=1 if the student is very healthy or healthy in a 5-level scale; =0 otherwise	195	0.75	0.43	0	1
GPA	=GPA score in previous academic year (in 100 points)	195	84.01	10.44	2.05	95
Preference controls						

Self-regulation	the average of 22 questions, each with a 5-level scale from "1: not applicable at all" to "5: very applicable", using 6 subtracting the level if the level ordering is reversed.	195	3.74	0.42	2.47	4.78
Willingness to compete	the average of 4 questions, each with a 5-level scale from "1: not applicable at all" to "5: very applicable", using 6 subtracting the level if the level ordering is reversed.	195	2.79	0.83	1	5
Willingness to take risk	answer to the question "How willing are you to take risks?", an 11-level scale from "0: very willing" to "10: completely unwilling".	195	5.96	2.11	0	10
Patience	answer to the question "How willing are you to give up something that is beneficial for you today in order to benefit more from that in the future?", an 11-level scale from "0: very willing" to "10: completely unwilling".	195	5.19	2.62	0	10
Positive reciprocity	answer to the question "When someone does me a favor I am willing to return it", an 11-level scale from "0: describes me perfectly" to "10: does not describe me at all".	195	5.04	3.85	0	10
Negative reciprocity	answer to the question "If I am treated very unjustly, I will take revenge at the first occasion, even if there is a cost to do so.", an 11-level scale from "0: describes me perfectly" to "10: does not describe me at all".	195	5.68	2.49	0	10
Altruism	answer to the question "How willing are you to give to good causes without expecting anything in return?", a 11-level scale from "0: very willing" to "10: completely unwilling".	195	5.12	2.52	0	10
Trust	answer to the question "I assume that people have only the best intentions", an 11-level scale from "0: describes me perfectly" to "10: does not describe me at all".	195	5.23	2.66	0	10

Table A3. Pairwise randomization tests between treatments

	NoGoal			*NoInfoGoal*			*InfoGoal*		
	(1)			(2)			(3)		
Variable	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.	Obs.	Mean	Std. Dev.
Baseline test rank	64	5.20	2.91	65	5.63[(3)*]	2.95	66	4.97	2.71
Baseline test time	64	131.95	62.12	65	139.79	55.56	66	135.69	61.31
Predicted baseline test rank	64	5.76	2.63	65	6.31	2.61	66	5.74	2.42
Female	64	0.70	0.46	65	0.69	0.47	66	0.68	0.47
Age	64	19.61	0.96	65	19.64	0.90	66	19.50	1.09
Height	62	165.90	6.94	65	166.26	8.08	65	165.92	8.58
Weight	62	58.89	9.00	65	59.05	10.33	65	57.35	10.82
BMI index	62	21.41[(2)*]	2.32	65	21.38	2.93	65	21.16	4.89
Fat rate	62	23.41	4.71	65	23.02	5.23	65	22.76	6.58
Metabolic rate	62	1486.24	176.36	65	1497.60	201.46	65	1476.85	195.51
Pre-course exercise length	62	1199.47	477.58	62	1283.31	502.02	63	1305.62	518.45
Pre-course exercise times	62	26.13	9.43	62	26.87	10.72	63	26.90	9.88
Test temperature	64	19.39	1.92	65	19.68	2.60	66	18.97	2.20
Test humidity	64	31.31	5.17	65	30.37	5.04	66	31.53	5.42
Han	64	0.88	0.33	65	0.91	0.29	66	0.92	0.27
Affiliated	64	0.89	0.31	65	0.94	0.24	66	0.94	0.24
Urban Hukou	64	0.77	0.43	65	0.82	0.39	66	0.82	0.39
Boarding school	64	0.58	0.50	65	0.57	0.50	66	0.64	0.48
Monthly income	64	2057.34	782.32	65	2050.31	1116.64	66	2500.00	3576.03
No. of siblings	64	0.72	1.25	63	0.41	0.56	66	0.70	1.05
Birth order	64	1.26	0.88	65	1.11	0.44	66	1.29	0.63
Healthy	64	0.67	0.47	65	0.78	0.41	66	0.80	0.40

GPA	64	83.45	12.55	65	84.76	7.65	66	83.82	10.69
Self-regulation	64	3.77	0.45	65	3.70	0.40	66	3.75	0.39
Willingness to compete	64	2.68	0.85	65	2.83	0.84	66	2.86	0.81
Willingness to take risk	64	5.89	2.12	65	5.75	2.09	66	6.24	2.13
Patience	64	4.98	2.79	65	$5.89^{(3)**}$	2.40	66	4.71	2.55
Positive reciprocity	64	$4.76^{(2)*}$	3.91	65	$6.17^{(3)**}$	3.70	66	4.44	3.73
Negative reciprocity	64	5.83	2.53	65	5.51	2.25	66	5.70	2.70
Altruism	64	4.77	2.73	65	5.71	2.36	66	4.88	2.39
Trust	64	5.41	2.55	65	5.31	2.52	66	5.06	2.92

Notes: This table reports the Pairwise randomization tests between treatments. The superscript next to the mean of each treatment shows the column number to which treatment (column) is compared, and the asterisks mark the significance level of the difference following the conventional manner. If, for a given variable, two treatments are not significantly different at conventional levels, no superscript is added. This comparison is only conducted to the "right" to avoid double counting, i.e., column (1) is compared to columns (2)-(3), column (2) is compared to columns (3). *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Table A4. Distribution of students by treatment and type

Treatment/ type	Underestimating	Accurate	Overestimating	All types
NoGoal	28	7	29	64
NoInfoGoal	23	9	33	65
InfoGoal	24	10	32	66
All treatments	75	26	94	195

Table A5. Impact of uncertainty on goals

Sample	Accurate type in *NoInfoGoal* against accurate type in *InfoGoal*		Accurate type in *NoInfoGoal* against everyone in *InfoGoal*		
Dependent variable: goal for final test rank	(1)	(2)	(3)	(4)	(5)
InfoGoal	0.111	-0.167	-0.359	0.255	-0.190
	(0.716)	(0.293)	(0.624)	(0.199)	(0.673)
Dummies for baseline test rank	No	Yes	No	Yes	Yes
Basic individual characteristics	No	No	No	No	Yes
Physical fitness and test environment	No	No	No	No	Yes
Socioeconomic characteristics and preferences	No	No	No	No	Yes
Observations	19	19	75	75	70

Notes: This table reports estimates from OLS models. The samples covered are indicated in the column heading. Control variables are identical to those in Table 1. Estimates for control variables are not reported (available upon request). Robust standard errors allowing for heteroskedasticity and clustered at the semester-class-gender level are reported in parentheses.

Table A6. Treatment effect on performance in the final test for various types (robustness check #1: quadratic baseline test controls)

Dependent variable		Final test rank		Final test time	
		(1)	(2)	(3)	(4)
Implied treatment effects					
NoInfoGoal-NoGoal	(i) underestimate	-0.683	-0.881**	-25.915*	-27.259***
		[0.288]	[0.045]	[0.066]	[0.006]
	(ii) accurate	-0.726	-0.532	-24.400	-23.903
		[0.316]	[0.562]	[0.426]	[0.485]
	(iii) overestimate	0.952	1.001	24.185	23.108
		[0.245]	[0.321]	[0.161]	[0.281]
InfoGoal-NoGoal	(iv) underestimate	0.527	0.178	4.621	0.034
		[0.439]	[0.764]	[0.771]	[0.998]
	(v) accurate	-1.073	-1.298	-17.356	-26.182
		[0.167]	[0.190]	[0.547]	[0.376]
	(vi) overestimate	0.100*	1.145	16.000	18.713
		[0.078]	[0.122]	[0.260]	[0.248]
InfoGoal-NoInfoGoal	(vii) underestimate	1.210**	1.059**	30.536**	27.292*
		[0.029]	[0.027]	[0.028]	[0.065]
	(viii) accurate	-0.347	-0.766	7.045	-2.279
		[0.568]	[0.392]	[0.647]	[0.927]
	(ix) overestimate	0.048	0.144	-8.185	-4.395
		[0.951]	[0.872]	[0.705]	[0.862]
Regression results					
NoInfoGoal		-0.726	-0.532	-24.400	-23.903
		(0.697)	(0.895)	(29.662)	(33.241)
InfoGoal		-1.073	-1.298	-17.356	-26.182
		(0.733)	(0.939)	(28.064)	(28.577)
Underestimation		-0.603	-0.690	-9.913	-14.357
		(0.705)	(0.742)	(27.390)	(25.193)
Overestimation		-0.837	-0.978	-16.813	-22.785
		(0.663)	(1.011)	(25.188)	(32.217)
NoInfoGoal × underestimation		0.0431	-0.349	-1.515	-3.356
		(0.730)	(0.908)	(32.089)	(34.113)
NoInfoGoal × overestimation		1.678*	1.533	48.585	47.011
		(0.905)	(1.222)	(30.808)	(34.879)
InfoGoal × underestimation		1.600*	1.476	21.977	26.216
		(0.805)	(1.110)	(29.285)	(33.483)
InfoGoal × overestimation		2.073**	2.443***	33.356	44.894
		(0.708)	(0.795)	(29.178)	(27.389)
Baseline test rank/time		0.214	0.0941	1.171***	1.010***
		(0.261)	(0.349)	(0.235)	(0.315)

Baseline test rank/time squared	0.047*	0.055*	-0.0003	-0.00006
	(0.023)	(0.030)	(0.0006)	(0.0008)
Basic individual characteristics	No	Yes	No	Yes
Physical fitness and test environment	No	Yes	No	Yes
Socioeconomic characteristics and preferences	No	Yes	No	Yes
Observations	195	182	195	182

Notes: See notes to Table 4. “Baseline test rank” is a continues variable ranging from 1 to 10.

Table A7. Treatment effect on performance in the final test for various types (robustness check #2: score *and* time baseline controls)

Dependent variable		Final test rank		Final test time	
		(1)	(2)	(3)	(4)
Estimated differences					
NoInfoGoal-NoGoal	(i) underestimate	-3.115	-4.029**	-27.021*	-29.900**
		[0.261]	[0.035]	[0.092]	[0.023]
	(ii) accurate	0.080	-0.090	-14.149	-16.505
		[0.967]	[0.984]	[0.590]	[0.642]
	(iii) overestimate	3.721	4.085	22.994	23.544
		[0.237]	[0.307]	[0.146]	[0.270]
InfoGoal-NoGoal	(iv) underestimate	1.112	0.015	-1.189	-3.085
		[0.718]	[0.995]	[0.945]	[0.852]
	(v) accurate	-4.103	-3.577	-17.488	-29.565
		[0.180]	[0.493]	[0.523]	[0.451]
	(vi) overestimate	3.367	4.083	14.668	18.445
		[0.153]	[0.174]	[0.335]	[0.206]
InfoGoal-NoInfoGoal	(vii) underestimate	4.227**	4.044*	25.832*	26.814*
		[0.029]	[0.069]	[0.077]	[0.080]
	(viii) accurate	-4.182	-3.487	-3.338	-13.060
		[0.121]	[0.413]	[0.831]	[0.617]
	(ix) overestimate	-0.354	-0.002	-8.326	-5.099
		[0.918]	[1.000]	[0.700]	[0.839]
Regression results					
NoInfoGoal		0.080	-0.090	-14.149	-16.505
		(1.884)	(4.358)	(25.566)	(34.674)
InfoGoal		-4.103	-3.577	-17.488	-29.565
		(2.892)	(5.076)	(26.663)	(38.086)
Underestimation		-0.355	-0.018	-4.461	-13.730
		(2.773)	(4.706)	(25.776)	(31.484)
Overestimation		-1.993	-1.435	-11.052	-16.714
		(2.162)	(5.321)	(24.033)	(37.518)
NoInfoGoal × underestimation		-3.194	-3.939	-12.871	-13.395
		(2.737)	(4.534)	(30.795)	(37.651)
NoInfoGoal × overestimation		3.641	4.175	37.144	40.049
		(3.153)	(5.497)	(24.994)	(37.331)
InfoGoal × underestimation		5.214*	3.592	16.299	26.479
		(2.769)	(5.526)	(25.807)	(39.527)
InfoGoal × overestimation		7.469***	7.660	32.156	48.010
		(2.208)	(4.386)	(27.115)	(37.156)
Baseline test performance (rank dummies & time)		Yes	Yes	Yes	Yes
Basic individual characteristics		No	Yes	No	Yes
Physical fitness and test environment		No	Yes	No	Yes

Socioeconomic characteristics and preferences	No	Yes	No	Yes
Observations	195	182	195	182

Notes: See notes to Table 4. "Baseline test performance" refers to 9 dummies for the baseline test rank and a continues variable for the baseline test time.

Table A8. Treatment effect on performance in the final test for various types (robustness check #3: broader accurate type)

Dependent variable:		Final test rank		Final test time	
		(1)	(2)	(3)	(4)
Estimated differences					
NoInfoGoal-NoGoal	(i) underestimate	-0.585	-0.593	-36.274**	-36.231**
		[0.435]	[0.395]	[0.039]	[0.033]
	(ii) accurate	0.028	-0.087	6.319	2.744
		[0.954]	[0.915]	[0.594]	[0.902]
	(iii) overestimate	0.960	1.027	22.373	21.384
		[0.209]	[0.320]	[0.183]	[0.285]
InfoGoal-NoGoal	(iv) underestimate	0.218	-0.133	-2.190	-9.111
		[0.778]	[0.834]	[0.883]	[0.584]
	(v) accurate	0.170	-0.114	12.050	3.423
		[0.803]	[0.856]	[0.394]	[0.811]
	(vi) overestimate	0.721	1.141	10.297	19.685
		[0.203]	[0.147]	[0.542]	[0.302]
InfoGoal-NoInfoGoal	(vii) underestimate	0.802*	0.460	34.084**	27.120
		[0.061]	[0.318]	[0.035]	[0.116]
	(viii) accurate	0.143	-0.027	5.731	0.680
		[0.833]	[0.977]	[0.668]	[0.973]
	(ix) overestimate	-0.239	0.114	-12.076	-1.698
		[0.763]	[0.881]	[0.593]	[0.940]
Regression results					
NoInfoGoal		0.0276	-0.0875	6.319	2.744
		(0.465)	(0.801)	(11.56)	(21.89)
InfoGoal		0.170	-0.114	12.05	3.423
		(0.670)	(0.618)	(13.66)	(14.06)
Underestimation (< -1)		0.414	0.331	26.26*	22.83
		(0.684)	(0.793)	(14.22)	(21.43)
Overestimation (> +1)		-0.0363	-0.140	8.263	3.218
		(0.582)	(0.863)	(10.30)	(17.64)
NoInfoGoal × underestimation -1		-0.612	-0.506	-42.59**	-38.97

	(0.763)	(1.164)	(16.68)	(29.29)
NoInfoGoal × overestimation +1	0.933	1.114	16.05	18.64
	(0.763)	(1.136)	(19.48)	(25.60)
InfoGoal × underestimation - 1	0.0474	-0.0185	-14.24	-12.53
	(0.595)	(0.534)	(11.48)	(13.51)
InfoGoal × overestimation +1	0.551	1.255	-1.753	16.26
	(0.612)	(0.866)	(16.78)	(20.30)
Baseline test performance	Yes	Yes	Yes	Yes
Basic individual characteristics	No	Yes	No	Yes
Physical fitness and test environment	No	Yes	No	Yes
Socioeconomic characteristics and preferences	No	Yes	No	Yes
Observations	195	182	195	182

Notes: See notes to Table 4. "Broader accurate type" refers to the group of students who are classified as accurate type both when they correctly predicted their rank and when they mispredicted their rank within only one rank (from above or below).

Appendix B: Proofs

Proof of Proposition 1. Denote the material utility by $a(y) = \beta\delta \cdot r(y,\bar{y}) - c(y)$. As $r(y,\bar{y})$ is linear in y and $c(y)$ is strictly convex, $a(y)$is strictly concave, and, together with $c'(0) = 0$ and $lim_{y\to\infty} c'(y) = \infty$, the interior solution y_l that satisfies $a'(y_l) = 0$ exists, is unique, and is both necessary and sufficient. Next, denote the psychological utility by $b(y,g) := -\alpha \cdot \beta\delta \cdot v(g - r(y,\bar{y}))$. $b_y > 0$ whenever $r(y,\bar{y}) \leq g$ and $b(\cdot) = 0$ otherwise. Note that $b(y,g)$ is strictly convex in y for $r(y,\bar{y}) \leq g$. Finally, recognize that the doer's utility is $a(y) + b(y,g)$.

Case 1: $g \leq r(y_l,\bar{y})$. As $b_y(y,g) \geq 0$ for any y, g, the optimal absolute performance always satisfies $y \geq y_l$. For any goal g (weakly) below $r(y_l,\bar{y})$, $b_y(y,g) = 0$ for any $y \geq y_l$ and so the optimal effort in this case must be $y^*(g) = y_l$.

Case 2: $g > r(y_l,\bar{y})$. Since $a(y)$ reaches its maximum at y_l, increasing y above y_l imposes additional (marginal) costs of $a_y(\cdot)$ on the doer. Implementing a higher absolute performance may be beneficial if the (marginal) benefit $b_y(\cdot)$ is sufficiently large, i.e., reducing the costs from falling short of one's target.[49] b_y satisfies the following properties: (1) For any $g > r(y_l,\bar{y})$, $b_y > 0$ at $y = y_l$ since $v(z)$ is strictly increasing whenever $z > 0$. (2) Since $b(y,g)$ is strictly convex in y (when $r(y,\bar{y}) \leq g$), $b_y(y,g)$ is increasing in y. (3) As $r(y,\bar{y})$ is linear in y, $b_y(y,g)$ is constant at the performance level that reaches the goal. (4) An increase in g increases the losses imposed by $v(g - r(y,\bar{y}))$ for a given y, $\bar{y}$, which, given the diminishing sensitivity of $v(\cdot)$, leads to a decrease in b_y for all y when $r(y,\bar{y}) < g$.

Implications: since $a'(y_l) = 0$, it must be that when g is sufficiently close to $r(y_l,\bar{y})$, $b_y(y,g) + a'(y) > 0$ for any y that induces a relative performance $r(y,\bar{y})$ in $[r(y_l,\bar{y}), g]$. For such goals, the optimal y must hence be at the corner: $r(y,\bar{y}) = g$. Henceforth, we will use the term corner solution to refer to the optimal absolute performance that ensures that the goal is just met, i.e., $r(y^*,\bar{y}) = g$.

Since goals only affect $b_y(\cdot)$ but not $a'(y)$, and since $b_y(y,g)$ is bounded, increasing in y but decreasing in g, and $c'(y) \to \infty$ as $y \to \infty$, there must exist some $\tilde{g}$ such that $b_y(\cdot)$ either (a) starts to cross $-a'(y)$ for the first time from above, or (b) becomes tangent to $-a'(\cdot)$ for the first time from above.

[49] For ease of exposition, whenever we refer to the derivative of $v(z)$ at $z = 0$, we refer to the 'right' derivative.

Case 2a: Since $b_y(\cdot)$ and $-a'(\cdot)$ do not have kinks, they can only cross for the first time, *without* being tangent beforehand, at the corner solution. Further increasing g by a small amount, lowers $b_y(\cdot)$ without affecting $a'(y)$, with the consequence that the two functions intersect at a lower y. In other words, the optimal performance will ensure that the goal is reached for goals up to and including some $\hat{g}$ (which is equal to $\tilde{g}$ for case 2a) and decreases thereafter as long as there is only a single intersection.

Case 2b: If the two become tangent at some y^t for the first time, then $b_y(\cdot)$ must lie above $-a'(\cdot)$ or all $y \leq y'$, where $r(y', \bar{y}) = \tilde{g}$. Hence y^t is a saddle point and the absolute performance that ensures the goal is reached is still optimal. Increasing g slightly above $\tilde{g}$, turns the saddle point into a local maximum[50], as it causes $b_y(\cdot)$ to cross $a'(\cdot)$ (from above). Hence, we need to check whether the corner solution, y^c, or the local maximum, y^{loc} is globally optimal.[51] The difference in utility between the two is:

$$U\,(y^c) - U\left(y^{loc}\right) = b(y^c, g) + a(y^c) - b(y^{loc}, g) - a(y^{loc})$$

Differentiating with respect to g using $b_y\left(y^{loc}, g\right) + a_y(y^{loc}) = 0$ yields

$$d\frac{U\,(y^c)\; - \;U\,(y^{loc})}{dg} = \left[b_y(y^c, g) + \; a'(y^c)\right] \cdot \frac{\partial y^c}{\partial g} + \; b_g(y^c, g) - b_g(y^{loc}, g)$$

Using the fact that at the corner $\partial y^c / \partial g = 1$ and that $b_y(y^c, g) = -b_g(y^c, g)$ given our linearity assumption for $r(\cdot)$, the derivative of the utility differences simplifies to

$$a'(y^c) + b_y(y^{loc}, g)$$

Adding and subtracting $a'(y^{loc})$ and making once again use of the observation that the derivative is zero at the local maximum, we get

$$a'(y^c) + b_y(y^{loc}, g) \; + a'(y^{loc}) \; - \; a'(y^{loc}) = a'(y^c) \; - \; a'(y^{loc})$$

Since $y^c > y^{loc} > y_l^*$, $a'(y^c) < a'\left(y^{loc}\right) < 0$ and so $d\frac{u(y^c) - u(y^{loc})}{dg} < 0$. The key implication of this is that once y^{loc} becomes optimal, it always remains better than the corner solution. Since $b_y(y, g)$ is bounded and $c'(y) \to \infty$ as $y \to \infty$, an interior solution exists for some g (sufficiently large) and there also exists some g for which the corner solution can no longer be optimal. As $b_y(\cdot)$ crosses $-a'(\cdot)$ from above at the local maximum, any further (small) increase in g lowers $b_y(\cdot)$and so y^{loc} decreases in g. In other words, when the local maximum is also the global maximum, y^* is decreasing in g. As a result, optimal

---

[50] More specifically, the saddle point separates into a local maximum and a local minimum, with the former being the focus of our analysis.

[51] Obviously, y^c and y^{loc} depend on g. Just like for many other functions before, we do not make these relationships explicit to avoid overly cluttered notation.

absolute performance ensures the goal is met up to a $\hat{g}$ and is decreasing thereafter. As the optimal performance jumps from the corner solution to the local optimum, there is a discontinuous decrease in y^* as the goal increases from $\hat{g} - \epsilon' \text{ to } \hat{g} + \epsilon''$. Finally, recognize that at $\hat{g}$, the doer is indifferent between two levels of performances. In the model, we made the technical assumption that she prefers the higher performance level, in this case, the corner solution. This assumption ensures that a goal that maximizes performance exists.[52]

Case 2c: So far, we assumed that only a single local maximum in addition to the corner solution exists. As we made relatively few assumptions with regards to the shape of the first derivatives of $c(\cdot)$ and $v(\cdot)$, it is possible that b_y and $-a'(\cdot)$ intersect at more than one level of y for a given goal.[53] Moreover, the number of local maxima may change as g increases. Using a combination of our previous tools, we will now show that our previous characterization of $y^*(g)$ is not affected by such technicalities.

Take two local maxima, y^i and y^j, with $y^i < y^j$. Following the ideas of case 2b, it is easy to see that the lower maximum becomes relatively better when g increases:

$$U(y^j) - U(y^i) = b(y^j, g) + a(y^j) - b(y^i, g) - a(y^i)$$

$$d\frac{U(y^j) - U(y^i)}{dg} = b_g(y^j, g) - b_g(y^i, g) = b_y(y^i, g) - b_y(y^j, g) < 0 \qquad (4)$$

where the inequality follows from the fact that b_y is increasing in y and $y^i < y^j$. In other words, the doer would never want to choose the higher performance level y^j when she previously opted for y^i and the goal is increased. Since any local maximum is decreasing in g, it follows that as long as the set of local maxima remains unchanged and one of them is chosen (which is true by definition when $g > \hat{g}$), y^* is decreasing in g.

Next, suppose an increase in g to g^n introduces a new local y^n maximum. Take any local maximum $y^i < y^n$ and the next highest local maximum or corner solution y^k relatively to y^n ($y^n < y^k$). Repeating the ideas from Case 2b, the new local maximum must have been a saddle point at $g^n - \epsilon$. By continuity, it follows that the doer also prefers y^k relatively to y^n at g^n. If the doer also prefers y^i over y^k, then, by equation 4, we know that as g increases, she must still prefer y^i over y^n and so the introduction of a new local maximum cannot result in an increase of $y^*(g)$ in g.

[52] As will become evident from our analysis behind Proposition 2, the goal-setter would never set such a goal in the first place if the doer were to choose the lower of the two performance levels at $\hat{g}$. Instead, she would set $\hat{g} - \epsilon$ if $y^*(\hat{g} - \epsilon) \leq y_h$ or the lowest goal that implements y_h otherwise.

[53] In particular, there will have to be at least three intersections. If the two curves only intersect twice, the lower intersection represents a local maximum while the higher intersection point is a local minimum. Drawing b_y as an 'increasing wobbly line' in the material marginal costs and psychological marginal benefits figure, i.e., allowing it to be concave and convex at various points, it is easy to see that multiple local maxima are theoretically possible. Of course, stronger assumptions, on, for example, the third derivative could rule out such cases.

Lastly, the possibility that a local maximum y^j is eliminated when g increases to, say g^e is not a concern for our analysis, as in such a case, the local maximum would not have been preferred by the doer at $g^e - \epsilon$. After all, y^j becomes a tangency point, but now with b_y being below $-a'(\cdot)$ in the vicinity of y^j, and so the doer prefers the previous (highest) local maximum below y^j over y^j at g^e and thus also at $g^e - \epsilon$. Moreover, such previous point exists since $b_y(y_l, g) > 0 = c'(y_l)$.

Having covered all cases, we finish the overall proof, by noting that since $b_y(y, g) > 0$ for any y such that $r(y, \bar{y}) < g$, $y^*(g) > y_l$, and so $y_l < y^*(g) < y^*(\hat{g})$ for any $g > \hat{g}$.

Proof of Proposition 2. The first statement follows immediately from proposition 1 and the description in the main text. For the second statement, notice that due to the inverse v-shape of effort, there may be two goals that implement y_h. As the higher goal imposes a psychological loss, the lower goal must be optimal

Proof of Hypotheses. Denote the optimal goal and the resulting absolute performance for some initial belief $\bar{y}_{\text{belief}}$ by g_0 and y_0. From proposition 2, we know that goal g_0 is reached when $\bar{y} = \bar{y}_{\text{belief}}$. For what follows, we assume that the decision makers in each treatment group are motivated by goals, i.e., $\alpha > 0$. The hypotheses obviously extend if only some fraction of students are motivated by goals.

Case 1. overestimating others' performance: $\bar{y}_{belief} > \bar{y}_{true}$

Since $g_0 = y_0 - \bar{y}_{belief}$, the highest possible y that the decision maker could possibly be motivated to achieve given g_0 knowing $\bar{y}_{true}$ is $max\{g_0 + \bar{y}_{true}, y_l\}$, which is less than y_0. In contrast, the optimal goal for $\bar{y}_{true}$ is $g_1 = y_0 - \bar{y}_{true}$, which implements y_0. To see this, note that it results in the same marginal benefits b_y for every y as the original goal, $b_y(g_1 - (y - \bar{y}_{true})) = b_y(y_0 - y) = b_y(g_0 - (y - \bar{y}_{belief}))$, and since $a'(y)$ is independent of $\bar{y}$, it thus leads to the same absolute level of performance as g_0 given $\bar{y}_{belief}$. Next, g_1 must be optimal given $\bar{y}_{true}$ for otherwise some better g_2 would result in an absolute performance of $y^*(g_2)$ that is closer to y_h (which is independent of $\bar{y}$) than $y^*(g_1) = y_0$ (Proposition 2). But then, y_0 could not have been optimal given $\bar{y}_{belief}$ in the first place as $y^*(g_2)$ can be implemented for the initial belief using the previous argument. It follows that $g_{InfoGoal} > g_{NoInfoGoal}$, $y_{InfoGoal} > y_{NoInfoGoal}$ and consequently also $y^r_{InfoGoal} > y^r_{NoInfoGoal}$

Case 2. underestimating others' performance: $\bar{y}_{belief} < \bar{y}_{true}$

Case 2a: Suppose $y(\hat{g}) \leq y_h$, which implies that $y_0 = y(\hat{g})$ when $\bar{y} = \bar{y}_{belief}$. From proposition 1, we know that any increase in g beyond $\hat{g}$ results in a decrease in performance. As the difference between the goal and the relative performance equals $g - r(y, \bar{y}) = g + \bar{y} - y$, an increase in $\bar{y}$ from $\bar{y}_{belief}$ to $\bar{y}_{true}$ has the equivalent impact on performance as an equal-sized increase in the goal (recall $a'(y)$ is independent of $\bar{y}$). It follows that the decision maker in the *NoInfoGoal* treatment performs at a level below y_0. Repeating the argument from Case 1, it can easily be shown that the decision maker in the *InfoGoal* treatment sets a lower goal and performs at y_0. Hypothesis 2.1a and 2.2 follow.

Case 2b: Suppose next that $y(\hat{g}) > y_h$. In this case, the goal-setter sets the lowest goal that implements y_h. The doer operates at the corner solution, meaning that $b_y > a'(y)$ at $y = y_h$ (for details, consult proposition 1). But then, a small increase in $\bar{y}$ (which is technically equivalent to an equal-sized increase in g) will cause the doer's performance to increase above y_o for the given g_o. In other words, the decision maker in the *NoInfoGoal* may perform above y_o. In contrast, decision maker in the *InfoGoal* treatment will still find it optimal to perform at $y_o = y_h$ (just as in case 2a). Hypothesis 2.1b follows.

Case 3. accurate beliefs: $\bar{y}_{belief} = \bar{y}_{true}$

As people in the *InfoGoal* and *NoInfoGoal* treatment hold the same beliefs, they solve the same problem, both at $t = 1$ and $t = 2$. The hypotheses follow.

Appendix C. Student Satisfaction and Goals

In this section, we explore how satisfaction with test performance is related to test scores and goals. By doing so, we provide evidence that students take their goals into account when they evaluate their final relative performance, which in turn suggests that goals affect their motivation. Table C1 provides an overview of how many students missed (i.e., fell short of), met (i.e., exactly hit), or exceeded their goal in their final test for the *NoInfoGoal* and *InfoGoal* treatments, respectively.

Table C1. Performance relative to Goal

		Missed Goal		Hit Goal		Exceeded Goal	
Treatment	Obs	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
Full sample							
NoInfoGoal	65	0.77	0.42	0.06	0.24	0.17	0.38
InfoGoal	66	0.74	0.44	0.09	0.29	0.17	0.38
Spring-term sample only							
NoInfoGoal	38	0.82	0.39	0.08	0.27	0.11	0.31
InfoGoal	38	0.79	0.41	0.08	0.27	0.13	0.34

Notes: This table reports summary statistics regarding whether a goal was missed, hit, or exceeded in the final test performance. There are no statistical differences of having missed, hit, or exceeded one's goal between the two goal treatments at conventional significant levels.

For the second wave of our experiment, which occurred during the spring semester of 2019, we also elicited students' satisfaction with their test performance in the second survey.[54] Table C2 provides OLS regression results on how satisfaction is influenced by goals and final test scores. Not having elicited the outcome measures for the full sample has the unfortunate consequence that the total number of observations are relatively small. In view of this constraint, we opt to control for baseline test performance using a linear control instead of the usual dummies.

Table C2. Satisfaction, Performance, and Goals

Dependent variable: satisfaction with test performance	(1)	(2)	(3)	(4)
Final test rank ≥ Goal	0.521*	0.529*	0.438	0.383
	(0.254)	(0.247)	(0.420)	(0.401)
Final test rank - Goal	0.139**	0.141**	0.079	0.070
	(0.042)	(0.042)	(0.047)	(0.060)
max {Final test rank - Goal, 0}		-0.011		0.078
		(0.068)		(0.137)

[54] The question reads "How satisfied are you with your performance on the plank tests?" (5-level scale, coded as 1 for being not satisfied at all, up to 5 for being very satisfied). The mean (median) answer to this question is 2.57 (2), with a standard deviation of 1.13.

Baseline test rank (linear control)	Yes	Yes	Yes	Yes
Basic individual characteristics	No	No	Yes	Yes
Physical fitness and test environment	No	No	Yes	Yes
Socioeconomic characteristics and preferences	No	No	Yes	Yes
Observations	76	76	69	69

Notes: This table reports estimates from OLS models. Estimates for control variables are not reported; the full regression results are upon request. Control variables are identical to those in Table 1, except that we using a continuous variable instead of the usual dummies to control for the baseline test rank. Robust standard errors allowing for heteroskedasticity and clustered at the semester-class-gender level are reported in parentheses. *, and **, *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

Column (1) shows that having met or exceeded one's goal is significantly correlated with higher satisfaction and that satisfaction increases in the difference between the final test score and one's goal. [55] Motivated by the idea of loss aversion, we allow for the difference between the test score and one's goal to have a differential impact for positive and negative levels in column (2). Given the small sample, especially for those who meet or exceed their goal, it is not surprising that we find no significant difference in the slopes. In Figure C1, we plot student's satisfaction against the difference between the final test score and the goal and provide the fitted equation based on regression results in column (2). The graph highlights the large impact that meeting (and vice versa failing) one's goal has on raising people's satisfaction with their test performance, which is depicted by the jump at 0 in the graph.

[55] While using the difference between the test score and one's goal as the second explanatory variable is motivated by the literature on loss aversion, or more generally reference dependence, to capture the general effect of performance and goals on satisfaction, there are obvious alternatives – none of which change the conclusion that meeting or exceeding one's goal increases one's satisfaction. For example, one can control for the final test scores and goals separately instead of their difference (the latter imposes the restriction on coefficients of equal magnitude but opposite signs). Alternatively, we could also use dummies for each final test score, which would allow for non-linear effects of scores on satisfaction.

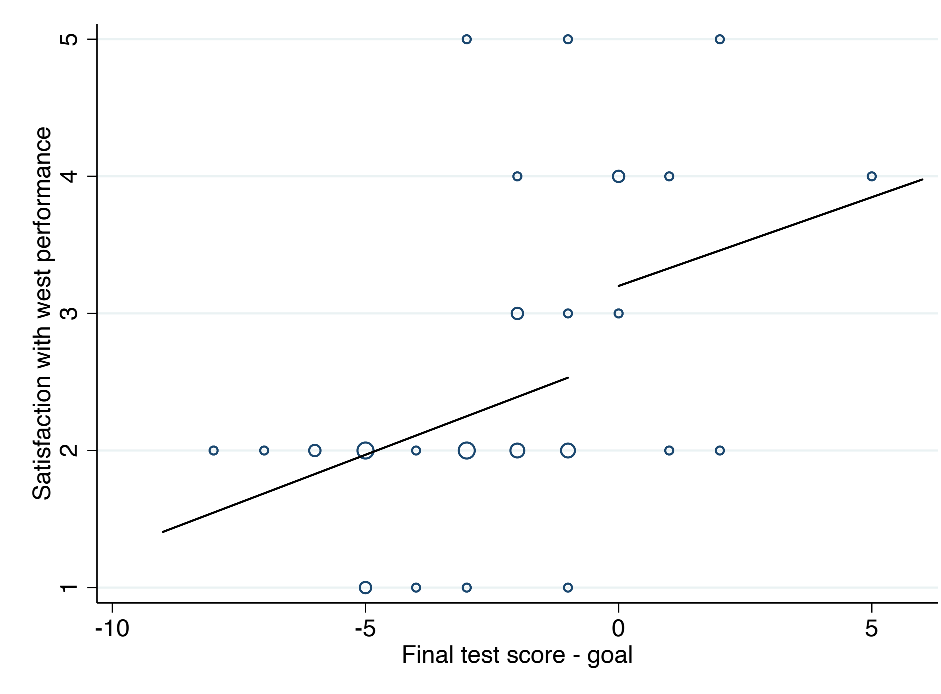


Figure C1. Satisfaction as a function of scores and goals

It is also noteworthy how strong the effect of meeting one's goal is relative to increasing one's score by 1 rank—reaching one's goal is equivalent to increasing the score-goal difference measure by 3.75 ranks.[56] In columns (3) and (4), we also control for the full set of control variables – an endeavor that makes estimating our main variables of interest tricky given the small sample and the relatively large set of controls. However, the estimated coefficient of meeting one's goal does not change materially, giving us some confidence that goals still matter.

[56] Namely, it can be calculated from either column (1) estimates (=0.521/0.139) or column (2) estimates (=0.529/0.141).