



Goal setting and raising the bar: A field experiment[☆]

Max van Lent^{*,a}, Michiel Souverijn^b

^a Department of Economics, Leiden University, the Netherlands

^b Department of Economics, Erasmus University Rotterdam, and Tinbergen Institute, the Netherlands

ARTICLE INFO

Keywords:

Goal setting
Motivation
Education
Field experiments

JEL classification:

C93
I23

ABSTRACT

We study goal setting using a randomized field experiment involving 1092 first-year undergraduate students. Students have private mentor-student meetings during the year. We instructed a random subset of mentors to encourage students to set a course-specific grade goal during one of the mentor-student meetings (goal treatment). A random subset of those mentors was further instructed to challenge students to set more ambitious goals if deemed appropriate (raise treatment). We find that students in the goal treatment perform significantly better as compared to students in the control group, and more so when they performed poorly prior to the experiment. Next, we find that students in the raise treatment do not perform significantly different from the control group, and explore reasons for why this may be the case.

1. Introduction

People often set goals. For example dieters commonly set a target weight, runners aim for a certain time, and managers set goals for employees in the form of targets. Using a series of field experiments, the psychologists (Latham & Locke, 1979) were the first to provide evidence that goals help to increase performance.¹ More recently, goal setting has also been studied by management scientists and economists. Economic theory papers have shown how (non-binding, non-monetary relevant) goals can be used as reference points in order to increase performance for loss averse agents or hyperbolic discounters (see e.g. Suvorov & Van de Ven, 2008; Hsiaw, 2013; Koch & Nafziger, 2011; Koch & Nafziger, 2016, and Koch, Nafziger, & Nielsen, 2014, and that meeting goals can lead to a sense of self-achievement that makes pursuing goals worthwhile (Gómez-Miñambres, 2012). A growing empirical literature tests the effects of goal setting on performance in the laboratory and in the field.

This paper examines whether goal setting can help to increase student performance in an academic course. Furthermore, we are interested to learn whether challenging students to be more ambitious by increasing the goal's difficulty can increase performance further. This is relevant given the widely held belief that many students should be

more ambitious, and the increased use of study counselors and mentors whose job it is to motivate and advise their students.

We test the effects of motivating students to set goals and attempts to raise the goal's difficulty by means of a field experiment among 1092 first-year economics students. Each of these first-year students regularly has meetings with a mentor (who is a senior student). Mentors help students to get used to studying at a university, teach them study skills, help them with their (study) motivation, monitor their performance, and give suggestions in order to increase their study performance. We ran the experiment during the second of three individual meetings between students and their mentor. In one treatment (*goal treatment*) we instructed mentors to ask their students whether they had a specific grade goal in mind for the main course they participated in at that moment, and if not, whether they wanted to set a grade goal. In another treatment (*raise treatment*) mentors received identical instructions as in the goal treatment, and were in addition instructed to encourage students to raise their goal if deemed appropriate. We subsequently measured performance using the grades the student obtained for the course.

We find that students whose mentor was instructed to motivate students to set a goal perform 0.18 better on a 10-point scale (which is 9.5% of a standard deviation) than students in the control group. This effect is driven by students in the goal treatment dropping out less often

[☆] We would like to express our gratitude to Robert Dur and Josse Delfgaauw for their invaluable support and advice throughout the entire research project. We would like to thank Sacha Kapoor, Jan Stoop, as well as seminar participants at EUR, LSE, and Tinbergen Institute, and conference participants at 10th NCBEE, LPEX 2015, Field Days 2016 at WZB, and 2nd SBRE Workshop on EPEO for helpful comments.

^{*} Corresponding author.

E-mail addresses: m.van.lent@law.leidenuniv.nl (M. van Lent), souverijn@ese.eur.nl (M. Souverijn).

¹ Locke and Latham found that goals set by an outsider (a peer or a manager), goals set in cooperation, and self-set goals can all lead to a better performance as compared to not setting goals.

than students in the control group. Students whose mentor was instructed to also ask students to raise their goal do not perform significantly different from the control group.

Setting goals can have adverse effects such as a narrow focus and even unethical behavior (see Ordóñez, Schweitzer, Galinsky, & Bazerman, 2009). One novelty of this paper is that we can accurately measure some of these adverse effects. A concern in our setting is that students may increase effort and performance on the course for which they set a goal by simultaneously substituting away effort from the other course they take at the same time. We estimate the effect of the treatment on performance in the other course, and do not find such a negative effect. This implies that motivating students to set a goal is actually performance increasing overall.

There is a rich literature in psychology studying goal setting and its effects on performance (see Locke, 1996; Locke & Latham, 2002, and Locke & Latham, 2006 for literature reviews). Research in psychology groups goals in roughly three categories: goals set by an outsider, cooperatively set goals, and self-set goals. Our *goal treatment* and *raise treatment* come closest to self-set goals and cooperatively set goals, respectively. Further, the literature shows that other factors such as goal commitment, goal specificity, and how challenging the goal is are important predictors for the success of goals (see for example Hollenbeck, Williams, & Klein, 1989; Locke, 1996, and Seijts, Latham, Tasa, & Latham, 2004). Goals that are specific, measurable and difficult but attainable tend to improve performance most. Our finding that the attempt to raise goals decreases students' performance as compared to goal setting by the student may be explained by a change in commitment to the goal, leading to a decrease in (study) motivation and hence performance.

Our paper is related to a rapidly increasing number of experiments in economics that study the effects of different types of goal setting on performance in various contexts. Experiments range from self-set goals to goals set by others. In some papers goals are combined with monetary incentives (see e.g. Goerg & Kube, 2012; Dalton, Gonzalez Jimenez, & Noussair, 2015; Cornet, Gómez-Miñambres, & Hernán-Gonzalez, 2015, 2018, and Brookins, Goerg, & Kube, 2017) and in other papers goals are set without monetary incentives (see e.g. Goerg & Kube, 2012; Sackett, Wu, White, & Markle, 2014; Brookins et al., 2017, and Clark, Gill, Prowse, & Rush, 2020). These studies typically find that when ambitious but attainable goals are set, goals increase performance, and more so when they are combined with monetary incentives. Our main contributions to this literature are that we incorporate goal setting in a one-on-one mentoring program, and that we investigate the effects of raising goals by increasing its difficulty in a cooperative manner.

Also closely related to this paper is the literature on (non-monetary) incentives for students in education. This literature considers a number of ways besides setting goals, in which students' performance can be increased.² Lavecchia, Liu, and Oreopoulos (2016) review studies of interventions in education designed to improve students' performance. The interventions target a wide range of behaviors, varying from too little focus on the future, overreliance on routines, student self-confidence and the information on and number of choices in education. Further, Sanders and Chonai (2015) show that in education usually (very) small effect sizes are found. The effect we find from goal setting is around the median effect size found in the sample of Sanders and Chonai.³

Goal setting by students has received a lot of attention from psychologists, see e.g. Ames and Archer (1988) and Schunk (1990). Many

of these papers in the psychology literature have tested whether goal setting can increase students' performance (see also Linnenbrink, 2005; Latham & Brown, 2006; Morisano, Hirsh, Peterson, Pihl, & Shore, 2010; Bettinger & Baker, 2014; Schippers, Scheepers, & Peterson, 2015 and Travers, Morisano, & Locke, 2015). These studies are concerned with goal related activities such as coaching, self-reflection and essaying. In our experiment we explicitly ask the student to set a grade goal.

This paper is organized as follows. In the next section we explain the experimental context and describe the data. In Section 4 we present a simple theoretical framework and derive our hypotheses. In Section 5 we explain the empirical strategy. Section 6 presents the descriptive statistics, section 6 the results followed by the conclusion in the final section.

2. Experimental Set-up and data description

2.1. Experimental context

The experiment involved 1092 first-year students enrolled in several undergraduate programmes at Erasmus School of Economics in Rotterdam, The Netherlands during the 2014-2015 academic year. The year is divided into five blocks of eight weeks. In each block students take 12 study credits (ECTS) worth of courses. All courses that students take at this point are obligatory, hence all students within a study programme take the same courses. Our experimental treatments take place during the second block when students have their second individual meeting with their mentor.

Each first-year student has a mentor. Mentors are senior students and are randomly assigned to students enrolled in the same programme at the start of the academic year. All mentors are employed by the university and are paid a flat wage. Our study involves all 84 mentors, and each mentor has 10 to 15 students. Mentors regularly meet with their students, both in groups and individually. Both group and individual meetings are obligatory. The mentor-student meetings are intended to teach students study skills, monitor their motivation, and more generally to provide a point of contact within the university. Motivation and individual prospects are the primary subjects of the three individual mentor-student meetings held over the course of the academic year. All individual mentor-student meetings take approximately 30 min. The first individual mentor-student meeting takes place around the start of the academic year in September, while the second and third take place in November and January, after the results of respectively the first and the second block of courses have been released. Our treatments are administered during the second individual mentor-student meeting.

While the first meeting at the start of the academic year primarily serves to discuss the student's motivation and to detect possible issues, the second and third meetings serve to evaluate results and prospects of the students. Due to university rules and national legislation at the time of the experiment, students with a weak performance record may be better off dropping out before February, which is in the third block of courses. Dropping out on time results in minimal grant loss and additionally allows students to re-enroll in the same programme the following academic year, which students that otherwise fail to meet first year requirements are not allowed to do. Thus, the second meeting is a natural moment to look forward towards the rest of the academic year and to discuss what results are necessary in order to make it sensible for the student to continue their current study programme.

Students participate in two courses in the second block, an introductory course in microeconomics worth 8 ECTS and a programme specific 4 ECTS course.⁴ Our treatment is focused on the microeconomics course. The course is taught in Dutch (824 enrolled students) and English (268 enrolled students). The Dutch and English versions are

² For example changes in the class size (see Angrist & Lavy, 1999 and Bandiera, Larcinese, & Rasul, 2010), providing feedback to students (see Bandiera, Larcinese, & Rasul, 2015), and several financial and non-financial incentives (see Levitt, List, Neckermann, & Sadoff, 2016).

³ While the mean effect size in Sanders and Chonai (2015) is 17% of a standard deviation (and the median effect size is 10%), our almost costless intervention has an effect of 9.5% of a standard deviation.

⁴ Students enrolled in the Economics and Business Economics, Fiscal Economics, and Law and Economics programmes take an ICT course, while Econometrics students take a Calculus course.

identical in all respects except for the lecturers and language spoken. The course follows a standard setup of three non-compulsory plenary lectures each week complemented by two compulsory tutorials taught by teaching assistants. The tutorials serve to review the course material, practise and discuss exercises, and in general to provide students an accessible way to obtain further explanation and clarification of the material. Tutorials are taught in 42 tutorial groups. One tutorial group consists of the students of two mentor groups. Examination of the course follows a standard format with two midterms counting 15% each and a written exam for the remaining 70%. For both midterms and the final exam students receive a grade on a 10 point scale, ranging from 1 to 10 with 10 being the best grade. In addition students could obtain a bonus, which was equal to at most half a point of the final grade, by participating in weekly online tests.

2.2. Experimental Design

We instructed a random subset of 54 of the 84 mentors to motivate their students to set a course specific grade goal during the second individual mentor-student meeting. As discussed before, this second meeting is an excellent opportunity for such a discussion as its purpose is to reflect on past performance and consider what results for the current courses are necessary. This means that discussion of the progress of the current courses is natural, and a focus on microeconomics is expected since it is the most important course in the second block due to its weight in ECTS. Our treatment builds on this discussion.

During meetings with all mentors in the period between 22 and 31 October 2014, we informed the mentors that some of them would be expected to take a somewhat different approach to the second individual meeting. Selected mentors were sent instructions by e-mail (see Appendices 1 and 2) about how to complement the discussion regarding the current courses one and a half week before the meetings. The instructions were accompanied by a simple flow diagram (see Appendix 4).⁵ All 54 selected mentors confirmed that they understood the instructions.

Randomly selected mentors were instructed to ask students whether they have a specific grade goal in mind for microeconomics, and if so to elicit that grade goal. If the student did not have a grade goal in mind, the student was asked whether she wanted to set a goal on the spot, again eliciting the goal. Students were free not to set a goal. Mentors were instructed to write down their evaluation of the student's goal in private after the meeting, evaluating the student's goal as either "too easy", "doable" or "too hard".⁶ The description of the treatment so far describes the *goal treatment*.

A second group of mentors were randomly selected to perform the *raise treatment*. In the raise treatment mentors implement the goal treatment but are in addition requested to attempt to raise the goal (if any) set by the student when deemed appropriate. If the mentor described the goal as "doable" or "too easy" the mentor was instructed to challenge the student by asking whether the student shouldn't be more ambitious and aim for a higher grade, specifically the student's self-set goal + 1 (e.g. if the student's goal was to get a 6 the mentor suggested aiming for a 7). The raise treatment serves to determine whether raising self-set goals can (further) improve study performance. Fig. 1 illustrates the similarities and differences between the goal and raise treatment using a flowchart.

⁵ Note that all mentors (including those assigned to the control group) received an extensive training and instruction from the mentor coordinator regarding how to conduct the mentor-student meetings. This reduces the chance that our instruction to the treatment mentors either signalled the importance of the meetings, or offers a guideline about how to structure the meeting.

⁶ The information that the mentor can use in order to evaluate the difficulty of the goal are the students' grades prior to the experiment, and the experience that the mentor has with the student based on one earlier individual meeting and several group meetings.

We chose to elicit a grade goal instead of other course related goals for multiple reasons. First, the final grade is (one of) the most important motivations to study for many students, hence students might find it more useful to set grade goals as compared to other goals. Second, choosing an output goal (the final grade) instead of an input goal (e.g. study hours) leads to lower measurement error because we cannot perfectly measure study hours. Finally, a grade goal is specific and measurable, which are important factors that influence the success of a goal (Locke & Latham, 2002).

Mentors were asked to record the outcome of the meetings on a form. Mentors record whether students set a goal, what the goal is, and their estimate of the difficulty of the goal. In the raise treatment, mentors further record whether they asked students to raise their goal, the size of the raise, and whether or not the student accepted this higher goal, see Appendix 4 for the form that treated mentors filled in for each student. The mentor's estimate of the difficulty of the initial goal allows us to compare students in the raise treatment whose goal was challenged with similar students in the goal treatment whose goal was not challenged but would have been challenged if they were in the raise treatment.

Besides the forms filled in by the mentors selected to implement the treatments we obtain information on all the students from administrative data from the microeconomics course and the central administrative office. This gives us information on the student's performance in other courses, attendance of microeconomics tutorial sessions, gender, age, study programme, and mentor.⁷ From the administration office we further obtained the mentor's gender and whether the mentor had experience in mentoring in previous years.

Only the mentors and lecturers were aware an experiment was being implemented, although mentors were not explicitly told so. Our introduction to all mentors in a general mentor instruction meeting necessitated that we informed all mentors that some of them would be asked to implement a small change in the upcoming individual mentor-student meetings. However, those not sent specific instructions were not aware of the exact change implemented. We specifically instructed the mentors who were selected for a treatment not to talk to anyone regarding our request. Selected mentors may deduce the purpose of the research but were not informed beyond their own instructions provided in Appendices 1 and 2.

We conducted a power calculation in order to learn about the minimum detectable effect size of our treatments. We use the mean and standard deviation of our mean outcome variable (the rescaled grade), and the intracluster correlation (i.e. the correlation within a mentor group) which equals 0.06 (with an average cluster size of 11 students). Since we have students' GPA prior to the experiment we can control for these in the power calculation. Finally, for the sample size we compare each treatment (separately) with the control group. Using these variables and the conventional 0.05 significance level and 0.2 probability of a type II error, we calculate that the minimum detectable effect size is roughly 0.30 for each treatment. This corresponds to 15.6% of a standard deviation. In a review study Sander and Chonare (2015) report that 17% is the mean effect found in randomized controlled trials in education. Hence, we ex ante expect our study to have sufficient power. In addition, we know in which study program the students participate. Study programs are quite indicative of study performance in general (with the better students choosing the more mathematical programs). However, since study program is measured on a nominal scale, we are not able to include this in a power calculation. Including study program as a control variable in all our regressions likely leads to a lower minimum detectable effect size than 0.30.

⁷ For students in Dutch study programmes who attended a Dutch high school, we also have highschool grades.

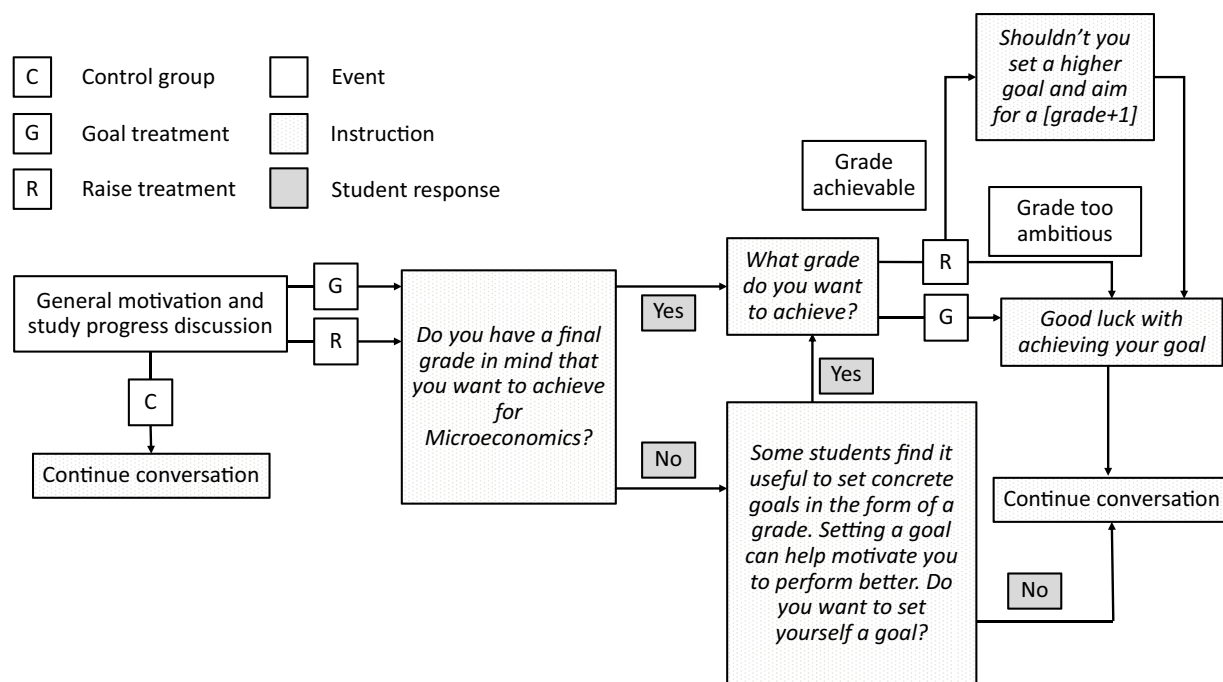


Fig. 1. Flowchart of treatments.

2.3. Assignment Procedure

The assignment of students to both treatments and the control group is randomized at the mentor level. Assignment at the mentor level was chosen in order to increase compliance and prevent contamination. The assignment of mentors to treatment was randomized in a stratified manner as follows. First, given that the tutorial group has a large impact on student performance as it is the main instruction method for many students, we ensure that a tutorial group is always of mixed composition in terms of treatments and control. This serves to create similar conditions for students in all treatments, but comes with the risk of contamination because students from treatment and control are in the same tutorial group. Randomization takes place within the various study programmes offered by the school as the effect of treatment can differ by programme due to the selection of students in a programme and the difficulty of the other course offered. Finally, several teaching assistants teach multiple tutorial groups. We therefore enforce that classes taught by the same teaching assistant have a mix of control and treatment groups.⁸ While our aim was to assign one third of the mentors to each group (treatment 1, 2, and control) we ended up with a slight imbalance as a consequence of our method of stratification. Our randomization procedure is illustrated in Fig. 2.

3. Predictions

Following the early psychology literature (see for instance Locke & Latham, 1990 and Locke2002 for a review) we hypothesize that motivating students to set goals increases their study performance. Goal setting can increase performance for many reasons, including a focus on goal-relevant activities and increased motivation for the task. Further, goals can function as a reference point, increasing performance for individuals who are time inconsistent and loss averse. In addition and specific to this experimental setting, reputational concerns of a student towards their mentor can increase performance. We will not be able to

⁸ For example, if a teaching assistant teaches two groups he teaches four mentor groups of which at least one group is assigned to each treatment and at least one group is a control group.

isolate these mechanisms in this study.

The effect of motivating students to set a goal, and subsequently proposing to raise the goal's difficulty have, to the best of our knowledge, not been studied yet. We will now elaborate on how motivating students to raise their goal can affect study performance.

Raising the goal's difficulty can increase performance if the initial goal was too low, and raising the goal will stimulate the student to increase effort. If the proposed raise is considered too high, the student can reject the raise proposal, and will perform similar to the case where the student would not be asked to raise the goal. However if there is a psychological cost of rejecting the raise, the student may still accept a goal that is too high, leading to an unrealistically high goal which is likely to decrease effort. There is mixed evidence in the psychological literature regarding the origin of the goal on its effects, see e.g. Latham and Marshall (1982) and Hollenbeck et al. (1989). There is a potential pathway for self-set goals to have more of an effect than goals that are assigned or that are set in cooperation due to a higher locus of control. Motivating students to set a goal can be interpreted as a self-set goal. The proposal to raise the goal can be interpreted as a goal that is set in cooperation (or even as a goal that is assigned). Hence the proposal to raise the goal may lead the student to value the goal differently.

Summarizing, while motivating students to set goals is expected to increase study performance, the effects of proposing students to raise their goals is ambiguous, and ultimately an empirical question.

4. Empirical Strategy

We estimate the effects of motivating students to set goals and attempts to raise students' self-set goals by estimating an intention-to-treat effect. The intention-to-treat measures the effect of a mentor being instructed to motivate students to set a grade goal, and in the raise treatment to attempt to raise students' goal. As becomes evident in the next section not all students are asked to set goals, for instance due to more pressing concerns in the meeting such as personal circumstances of the student. Also, not all students who are motivated to set a goal actually set a goal. We estimate:

$$P_i = \beta_0 + \beta_1 G_i + \beta_2 R_i + \beta_3 X_i + \varepsilon_i$$

where P_i is student i 's study performance, X_i a vector of control

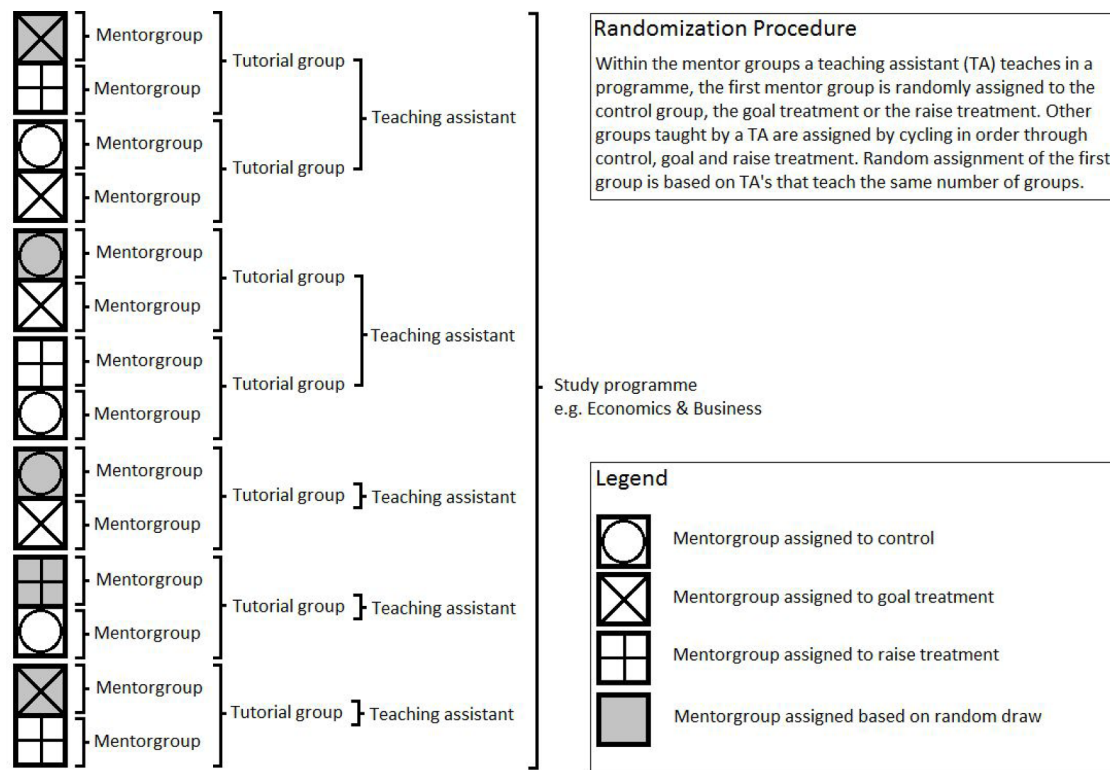


Fig. 2. Randomization procedure.

variables and ε_i the error term. G_i and R_i are treatment dummies indicating whether a student's mentor was assigned to the goal or raise treatment respectively. To be more precise on student performance, P_i is not the final grade of a student. The final grade for the course is composed of two midterm exams (both with weight 15%) and a final exam (with weight 70%). Since the mentor-student meeting is in the same week as the first midterm, we expect that students can hardly change their study behavior for the first midterm, and so we expect the treatment to only affect the later exams of the course. Hence we take as student performance a normalized combination of the second midterm and the final exam. Our performance measure is hence calculated as $(0.15 \cdot \text{midterm2} + 0.7 \cdot \text{final}) / 0.85$.⁹ As a robustness check we also estimate the effect of our treatments on the final grade (i.e. including the first midterm).

There are two main reasons to include covariates in our regressions. First, since we assign treatment randomly conditional on the student's programme and teaching assistant we include dummies for the tutorial groups which subsume both these categories. Second, we include statistics on past study performance. Past study performance is highly predictive of present study performance and hence including measures of past performance reduces noise in the data, allowing for more precise estimates. We additionally include the student's gender, the mentor's gender and a dummy for the mentor's experience. In order to deal with inter cluster correlation, we cluster the errors of all the regressions at the mentor level.

We assign students who do not complete the course a failing grade for those grade components that they do not complete. The highest failing grade a student could get is a 4.4 and the lowest a 1.0 at a 1 to 10 scale. In our estimations we focus on the lowest grade as the lowest failing grade is the grade that is actually given to students who do not pass the course. Further, for context consider that those who do not take

the final exam but do take the second midterm score a 1.5 on average compared to the overall average of 5.7.

The effect of our treatments is composed of an effect on the intensive margin (performance of students who take the final exam) and an effect on the extensive margin (whether students participate in the final exam). We provide separate estimates for the effect on study performance for those students who complete the course, and for the effects on exam participation.¹⁰

Ideally we would like to estimate the effect of actually setting a goal on student performance using a local average treatment effect (LATE). However there are several reasons why a LATE estimation is infeasible in our setting. First, in order to estimate the effect on students who would be induced to set a goal by our treatment LATE estimation requires knowledge of the actual treatment a student receives, e.g. whether a student set a goal or not. We however don't know whether students in the control group actually set themselves a goal without their mentor explicitly motivating them.¹¹ Since our knowledge of actual treatment is perfectly correlated with treatment assignment we cannot distinguish so-called always-takers (the students who set a goal without the mentor motivating them to do so) from those influenced by our treatment. Thus a LATE estimation would amount to comparing the treated in the treatment group with the full control group. This by itself is expected to result in an underestimation of the effect, since if there is an effect it is also present amongst the always-takers in the control

¹⁰ Note that all students (also those who do not take part in the final exam in December) still have the opportunity to participate in a resit exam which is administered in the summer. As a robustness check we later also estimate the effect of our treatments on the highest grade of the student including the possible resit exam grade.

¹¹ We did not ask mentors in the control group whether their students set themselves goals. If we would have asked control group mentors whether their students set goals during the meeting, mentors could infer the purpose of our study. This could potentially lead mentors in the control group to alter their behavior.

⁹ The weights assigned to the second midterm and the final exam in our performance measure, 0.15 and 0.70 respectively, are the same as the weights used in the composition of the final grade for students.

Table 1
Descriptives by assigned treatment.

	Control		Goal		Raise		C - G	C - R	G - R
	mean	sd	mean	sd	mean	sd	p	p	p
Age	18.72	1.71	18.63	1.39	18.75	1.27	0.57	0.85	0.39
Female	0.29	0.46	0.30	0.46	0.30	0.46	0.87	0.95	0.93
Female mentor	0.52	0.50	0.31	0.46	0.34	0.47	0.00***	0.00***	0.46
Experience mentor	0.31	0.46	0.38	0.49	0.50	0.50	0.03**	0.00***	0.00***
EC Accounting	6.45	1.61	6.13	1.51	5.96	1.73	0.06*	0.01**	0.31
ECX Accounting	5.82	1.72	5.95	1.91	5.74	1.92	0.70	0.79	0.56
EC Math	5.98	1.56	5.87	1.43	5.91	1.63	0.49	0.70	0.79
ECX Math	6.48	1.51	6.64	1.63	6.66	1.38	0.58	0.51	0.96
ET Matrix Alg.	6.48	1.94	6.13	1.88	6.02	2.38	0.32	0.29	0.79
ETX Matrix Alg.	7.54	2.11	6.92	1.81	7.58	2.18	0.28	0.95	0.31
ET Precalculus	6.46	1.50	6.04	1.78	5.72	2.20	0.17	0.05**	0.41
ETX Precalculus	7.42	1.79	6.54	2.29	7.25	1.98	0.13	0.74	0.31
ET Statistics	5.09	2.07	4.45	1.68	4.43	2.23	0.07*	0.13	0.96
ETX Statistics	6.35	2.51	5.50	2.31	5.29	2.17	0.23	0.13	0.78
Microeconomics	6.48	1.76	6.48	1.62	6.29	1.89	0.95	0.19	0.20
Midterm I	4.89	2.25	4.80	2.10	4.78	2.23	0.58	0.55	0.94
Attendance	10.66	2.98	10.66	3.15	10.71	2.85	0.98	0.81	0.84
Dropout	0.14	0.35	0.11	0.31	0.14	0.34	0.21	0.94	0.26
Max. Observations	389		367		336				

Students enroll in an economics (EC) track or an econometrics (ET) track in a Dutch or international (X) programme. Tracks in the Dutch and international programme are identical. Different tracks feature different courses, although some courses (e.g. Microeconomics) are common to all tracks.

group. However, since mentors selectively motivate students to set goals (eg. the mentor may not discuss goals if the student has other issues to discuss) the selection into or out of treatment in the treatment group is biased towards including relatively good students into the treatment. This could lead to an overestimation of the treatment effect since a similar selection does not occur in the control group. As such we do not estimate local average treatment effects.

5. Descriptive Statistics

Our dataset contains information on 1092 students, 824 of whom are enrolled in a Dutch language programme with the remaining 268 students enrolled in an English language programme. Table 1 gives the descriptive statistics for the control (C), goal (G) and raise (R) group, as well as giving the *p*-value for two-sided comparisons of the means of these groups. Although the control and treatment groups appear to be comparable, there are some differences between the groups. Specifically, the characteristics of mentors of students in the treatment and control groups differ.¹² Students in the control group are significantly more likely to have a female mentor whereas students in the raise treatment are more likely to have an experienced mentor.¹³ Furthermore treatment students in a Dutch language economics track (as opposed to students in an English language economics track) scored lower for the 8 credits accounting course in the first block than students in the control group, but there is no such difference regarding the mathematics course, which is more important for microeconomics.¹⁴ In the analysis we control for differences in observables.

Selection into or out of treatment is an issue affecting the generalizability of the results to the whole population. In our experiment there are three sources of selection out of the treatment. First, despite our best efforts to get all mentors to cooperate and ensure their understanding of the instructions, not all mentors assigned to treatment

applied the treatment or took notes when administering the treatment. There are seven mentors for whom we do not have data about what happened during the individual student-mentor meetings. Anecdotal evidence suggests that some mentors have administered the treatment but did not record the results, while others did not administer the treatment. Thus this missing data forms a combination of measurement error and selection out of treatment.

Second, there is some treatment dilution as mentors do not administer the treatment to all students. Mentors assigned to treatment ask students for their grade goal in 93% of the cases although they were instructed to administer the treatment to all. Moreover, mentors are selective in which students they target for treatment. Specifically, students who performed poorly in previous courses are less likely to be asked about their goals as is shown in Table A.1. In cases in which mentors did not ask students about their goals they often noted a lack of time due to the necessity to discuss other issues. Also, conditional on receiving the data from the mentor, we find that more experienced mentors are less likely to administer treatment.

Finally, in the raise treatment mentors were instructed to attempt to raise the student's goal when they deemed the goal to be either too easy or doable. Of the 193 students setting a goal in the raise treatment 163 set a goal that met this requirement. However, mentors attempt to raise the goal in only 95 of these cases (58%), including all 47 cases where the goal is deemed too easy.¹⁵ Overall students who are asked to raise their goal have slightly higher grades than those not asked, although differences are largely insignificant.¹⁶ See Table A.2 for more descriptives of the comparison between students asked and not asked to raise their goal.

It is of interest to note that 270 of the 492 students (55%) asked to set a goal already had a grade they wanted to achieve in mind. We see that those students who have a goal in mind are on average the better students (in terms of GPA), see also Table A.3. The average initial goal set by the student is 6.9, a histogram of the goals set is shown in Fig. 3. As expected higher (lower) goals are more likely to be deemed too hard

¹² Note that there are only 84 mentors in the sample. In addition, at the time of the randomization the information on mentor characteristics was not available to us. Hence we could not stratify our randomization on mentor characteristics.

¹³ We define an experienced mentor as a mentor who mentored students in earlier years.

¹⁴ In Table 1 we tested for differences between control and treatment groups using *t*-tests. We obtain similar results if we use nonparametric tests.

¹⁵ In addition, there are 9 instances where the mentor asks a student to raise the goal even though she estimated the goal to be too difficult.

¹⁶ The low number of observations for the Dutch econometrics courses is due to the fact that three of the four Dutch econometrics mentors assigned to the raise treatment failed to provide data.

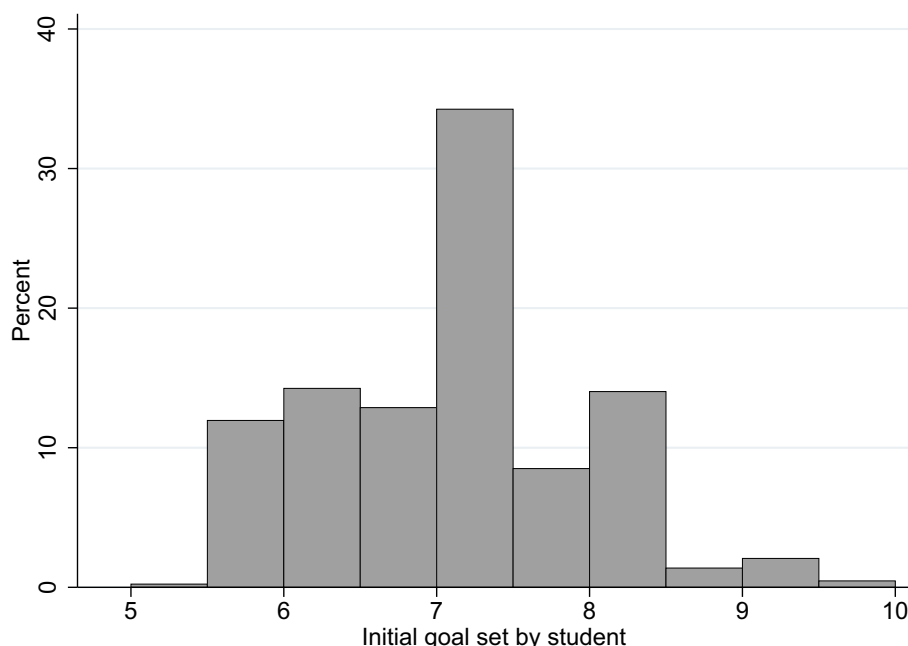


Fig. 3. Histogram of goals initially set by students (prior to any attempts to raise goal).

(too easy) to achieve for the student by their mentor.

Mentors appear to be able to gauge goal difficulty, as a regression of the difference between the final grade achieved and the initial goal set (both treatments) on the estimate of the mentor shows in Table A.4. Goals that were expected to be too difficult were not achieved on average whereas goals that were too easy are indeed beaten by a significant margin. Furthermore, all point estimates of the judgement categories differ significantly from each other, indicating that mentors differentiate well between the three categories. On average students failed to meet their goal by 0.4 of a point, which suggests that mentors are slightly overoptimistic about their students.¹⁷

Also 12% (59) of the 492 students asked for a goal do not set a goal. While students are more likely to set a goal if they have a female mentor there are no significant differences between those setting and not setting a goal in terms of past performance as shown in Table A.5. Of all students asked to raise their goal half accept a higher goal. Again there is no significant difference in terms of past study results, but students are less likely to accept a raise from more experienced mentors (p -value 0.03), see Table A.6. Furthermore, the level of the initial goal set has no influence on the acceptance of a suggested raise of the goal. Of the students who were asked to raise their goal we see that 50 percent (i.e. 52 students) reject the goal the mentor proposed.¹⁸

6. Results

We first provide the total effect of the treatments, imputing a failing grade for students who did not complete the course. We then provide the results for students who complete the course before turning our attention to the results for students who did not complete the course.

¹⁷ There may be a concern that mentors assigned to the raise treatment are more likely to report that they expect the student's goal to be too difficult, in order to avoid challenging the students to raise their goal. We test whether the distribution of the mentor's estimates of the students' goal differs across treatments. We do not find evidence of such an effect.

¹⁸ The average goal proposed does not differ (two-sided p -value of 0.62) between students who accept and reject the goal.

6.1. Total effect

We estimate the total effect of the treatments by imputing the highest and lowest possible grade that would result in failing to pass the course for those graded aspects of the course that the student did not complete. As discussed in Section 5, we focus on the case in which we impute a missing grade as 1.0, as this appears to be the most relevant case.

Table 2 gives the result of the intention to treat estimations. We find a positive effect of 0.18 of a grade point (i.e. 9.5% of a standard deviation) for students in the goal treatment and an insignificant negative effect of the raise treatment.¹⁹ The positive effect of assignment to the goal treatment is in line with our hypothesis that setting goals improves student performance. The insignificant negative effect in the raise treatment shows that attempts to raise a goal backfire, resulting in performance similar to students in the control group. Columns 2 and 3 show that the results are mainly driven by female students. Further, columns 4 and 5 of Table 2 provide separate estimates for students in Dutch and international programmes. These estimates show a widely divergent response to the treatments in these two groups, with Dutch programme students responding positively to the goal treatment while students in international programmes respond to the raise treatment in a strongly negative manner. This difference in response may be explained by differences in the initial study motivation of students (students in the Dutch program are admitted based on only a high school diploma while International students are besides selected on grades also selected on the basis of a motivation letter). The overall picture that emerges from the intention to treat estimates is that motivating students to set goals increases performance, but attempting to raise goals undoes any positive effect of setting goals and may even result in worse performance as compared to students that were assigned to the control group.

Because mentor characteristics (i.e. gender and prior experience as a mentor) are not balanced across treatments, we include some additional

¹⁹ As a robustness check we use randomization inference, implemented using the "ritest" package in Stata, see Hess (2017), and find that the significant positive effect of the goal treatment is no longer statistically significant ($p=0.220$).

Table 2
Intention to treat effect: total effect.

Grade	Overall	Male	Female	Dutch	International
Missinggrade=1.0					
T: Goal	0.181** (0.091)	0.152 (0.111)	0.394** (0.187)	0.339*** (0.115)	-0.011 (0.125)
T: Raise	-0.165 (0.105)	-0.031 (0.127)	-0.265* (0.155)	0.019 (0.118)	-0.428** (0.177)
Observations	967	682	285	724	243
Mentorgroups	84	84	82	60	24
R ²	0.619	0.627	0.679	0.618	0.627

Standard errors (clustered at the mentor level) in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$ Control variables: study program, tutorial group, mentor gender and experience, student gender, 1st midterm grade, prior GPA

analysis in order to explore the extend to which mentor characteristics can explain students' performance. First, we add the interaction of mentor gender and experience. This allows the effect of experience of a mentor to be different for male and female mentors. Table A.7 shows that the positive effect of the goal treatment is not affected, but that the raise treatment now has a significant negative effect on student performance. As a second check we explore how much mentors affect performance after controlling for mentors' observable characteristics. For the control group, we regress students' performance on mentor fixed effects while controlling for student characteristics. Then we regress the estimated fixed effects from that first regression on mentor characteristics. This regression shows that only 3.3% of the variation in the mentor fixed effect is explained by mentors' observable characteristics. This gives the impression that the imbalance in mentor characteristics turns out to be not very important empirically.

The total effect of our treatments discussed above consists of two effects. First the treatments may have an impact on the students who complete the course inducing them to alter their efforts. This is the effect on the intensive margin. Second, our treatments may affect the decision to participate in the course. These two effects cannot be interpreted separately as this risks the confusion of selection effects on the extensive margin for the effects of treatment on the intensive margin. We turn to these two effects now.

6.2. Intensive margin

Table 3 gives the intention-to-treat estimates for those students that complete the course (i.e. the students who participate in the final exam). The results are largely in line with the overall estimates provided in Table 2. We find that the overall positive effect of being assigned to the goal treatment is no longer significant for the students who complete the course. The results confirm the overall impression that motivating students to set goals can (somewhat) improve student performance, and that proposing to raise the goal has an insignificant negative effect on performance.

Table 3
Intention to treat effect: intensive margin

Grade	Overall	Male	Female	Dutch	International
Missinggrade=1.0					
T: Goal	0.112 (0.077)	0.069 (0.092)	0.361** (0.178)	0.265*** (0.090)	0.011 (0.125)
T: Raise	-0.147 (0.092)	-0.007 (0.103)	-0.262* (0.151)	0.054 (0.091)	-0.428** (0.177)
Observations	940	661	279	697	243
Mentorgroups	84	84	82	60	24
R ²	0.609	0.626	0.674	0.605	0.627

Standard errors (clustered at the mentor level) in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$ Control variables: study program, tutorial group, mentor gender and experience, student gender, 1st midterm grade, prior GPA

Table 4
Intention to treat effect: dropout

Dropout	Overall	Male	Female
T: Goal	-0.021*** (0.008)	-0.216** (0.110)	-0.016 (0.013)
T: Raise	0.004 (0.008)	-0.000 (0.012)	0.009 (0.011)
Observations	967	682	285
Mentorgroups	84	84	82
R ²	0.165	0.172	0.267

Standard errors (clustered at the mentor level) in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Control variables: study program, tutorial group, mentor gender and experience, student gender, 1st midterm grade, prior GPA.

6.3. Extensive margin

The results above indicate a positive effect of the goal treatment on course performance for those students who complete the course. Differences between the total effect estimates and the estimates on the intensive margin may be due to selection effects induced by the treatments. For instance, our treatments may affect course completion by creating greater commitment. To study the effects of our treatments on course completion we estimate a linear probability model in much the same way as above.

In the control group 6.2 percent (24) of the students having attended at least 3 sessions dropped out of the course. This dropout rate is lowered by 2 percentage points on average in the goal treatment as can be seen in Table 4 providing the intention to treat estimates on the dropout rate.²⁰ The results on the dropout rate are similar to those on course performance given course completion in that the goal treatment again has a positive effect on performance (i.e. dropout decreased) whereas the raise treatment does not have an effect but has an oppositely signed coefficient. In contrast to the effect on the course grade however, the reduction in dropouts is concentrated among men rather than women. This is most likely due to the fact that women have a substantially lower baseline dropout rate than men (6.8% for men compared to 2.8% for women in the control group, two-sided p -value 0.12).

In our setting students cannot retake the entire course. However, at the end of the academic year there is a resit for the final exam. In order to see whether dropping out of the course more often is undesirable (and dropping out less often is desirable), we estimate the final grade taking the maximum of the final exam grade from the regular exam and the final exam grade from the resit. Using this approach we find that the positive effect of the goal treatment becomes smaller, and the negative effect of the raise treatment becomes stronger and significant, see Table A.8. This implies that it is unlikely that students in the raise treatment strategically dropout in order to receive a higher grade using the resit exam.

6.4. Further results

In this section we present a number of additional analyses that shed light on further questions. To start, we consider whether there is a heterogeneous effect of treatment based on GPA prior to the experiment. We measure GPA by taking the average of the grades achieved in the first block, and centering this grade average by subtracting the overall mean average score of 6.2 (std. dev. 1.65). We then interact the ability measure with students' treatment assignment. The intention-to-treat estimates in Table 5 show that students who performed better in previous courses respond less to the goal treatment. Thus our

²⁰ No estimates on subsamples for the Dutch and international programmes are provided as there are too few dropouts from the international programmes, resulting in collinearity.

Table 5
ITT interaction treatment with GPA in preceding block

	Overall	Male	Female	Dutch	International
T: Goal	0.172** (0.075)	0.123 (0.090)	0.389* (0.198)	0.323*** (0.094)	0.048 (0.109)
T: Goal × GPA	− 0.103** (0.050)	− 0.169*** (0.061)	− 0.103 (0.128)	− 0.142** (0.061)	− 0.418 (0.087)
T: Raise	− 0.085 (0.079)	− 0.063 (0.096)	− 0.109 (0.142)	0.093 (0.081)	− 0.361** (0.142)
T: Raise × GPA	− 0.053 (0.058)	0.020 (0.076)	0.080 (0.099)	− 0.016 (0.063)	0.208* (0.112)
GPA	1.064*** (0.263)	1.248*** (0.242)	0.607 (0.065)	0.556** (0.243)	1.035*** (0.308)
Observations	914	644	270	675	239
Mentorgroups	84	84	81	60	24
R ²	0.627	0.682	0.568	0.633	0.624

GPA is the mean centered GPA of students calculated over all courses they participated in prior to microeconomics. Standard errors (clustered at the mentor level) in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Control variables: study program, tutorial group, mentor gender and experience, student gender, 1st midterm grade.

intervention had a stronger positive effect on weaker students than it did on top students. There is no such heterogeneous effect regarding the raise treatment.²¹

Second, as we have seen earlier, students set higher goals in front of a female mentor (two-sided p -value of 0.012). This suggests that the motivation to set a goal may differ depending on the gender of the mentor. It is thus natural to analyze whether our treatments have heterogeneous effects depending on the mentor's gender. Overall there is no sign of a heterogeneous treatment effect depending on the gender of the administering mentor as shown in Table 6.

The desire to impress a member of the opposite gender may result in a heterogeneous treatment effect based on the gender combination of mentor and student. Estimates on male and female students reveal no such significant interaction effect, although the signs of the point estimates do conflict in the expected manner which is consistent with such an effect.

One potential explanation for the fact that the raise treatment has a negative effect on performance is that the proposed raises are too high. In order to explore this we can make use of the fact that some student-mentor meetings take place after the first midterm. After the midterm, students as well as mentors have more information about what goals would be realistic for the student. Hence we expect goals as well as raises to be more realistic after the midterm than before. We test for the effects of differences in timing of the meeting, and find that timing of the student-mentor meeting does not affect students' willingness to set a goal, and neither does timing affect the probability that the student is asked to raise the goal. Finally, the timing of the meeting does not seem to have an effect on student performance.²²

In our main specification the outcome variable is a weighted average of the second midterm and the final exam. However, some students in both the control group and treatment groups had their mentor-student meeting shortly before the first midterm. Therefore the intervention could, for some students, have influenced even their first midterm grade. This is especially the case if students cram their study effort very close to the exam. Hence, as a robustness check we estimate our main specifications using the final grade that the students achieved. We find similar results using this final grade as the outcome variable.

We find a consistent pattern in our data that female students respond stronger to our treatments than male students. At first sight this

²¹ For students who attended a Dutch high school we also have high school grades. If we use high school grades (as a measure of ability) for this subsample we find a qualitatively similar result but a decrease in power because of the smaller sample size.

²² However, we have to note that students are allowed to pick a timeslot for the student-mentor meeting. Hence we cannot make any causal claims on the effects of timing of the meeting.

Table 6
ITT interaction treatment with gender of mentor

Grade	Overall	Male	Female	Dutch	International
Female mentor	− 0.0002 (0.134)	− 0.053 (0.148)	− 0.146 (0.300)	0.159 (0.113)	− 0.105 (0.282)
T: Goal	0.171 (0.113)	0.002 (0.128)	0.554** (0.262)	0.387*** (0.082)	0.007 (0.156)
T: Goal × Fem. mentor	− 0.198 (0.228)	0.085 (0.241)	− 0.711 (0.535)	− 0.520** (0.258)	0.002 (0.427)
T: Raise	− 0.178 (0.111)	− 0.121 (0.133)	− 0.430** (0.214)	− 0.091 (0.092)	− 0.186 (0.271)
T: Raise × Fem. mentor	0.070 (0.178)	0.262 (0.212)	0.336 (0.395)	0.449** (0.205)	− 0.487 (0.292)
Observations	940	661	279	697	243
Mentorgroups	84	84	82	60	24
R ²	0.609	0.626	0.678	0.609	0.629

Standard errors (clustered at the mentor level) in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Control variables: study program, tutorial group, mentor experience, student gender, 1st midterm grade, prior GPA

may seem surprising. There is a very rich literature on heterogeneous gender effects to monetary incentives and non-monetary incentives when there is a competition element. Many papers find that males respond stronger to competitive incentives or information about their (relative) ranking, see for example Gneezy, Niederle, and Rustichini (2003), Barankay (2011) and Niederle and Vesterlund (2011), while others find that there is no gender difference, see Dreber, Von Essen, and Ranehill (2011) and Delfgaauw, Dur, Sol, and Verbeke (2013). An important difference between these incentives and the incentives in our treatments is that goal setting in this experiment does not have any competitive element, which might drive the gender effect in the current literature. Apicella, Demiral, and Mollerstrom (2017) for instance find that males are more willing to compete against others than females, but find no gender difference when people compete against themselves. However, a recent paper by Clark et al. (2020) who study goal setting among university students using surveys, find that goal setting is more effective for male students. Similarly in a laboratory experiment Smithers (2015) finds that goals improve male but not female performance. Hence, the way in which goals are elicited (during a personal meeting or using surveys) may also have an impact on how effective goal setting is.

A potential concern of goal setting in multitasking environments is that goals lead subjects away from other (non-incentivized) task. In our setting, students take one other course at the same time as Microeconomics. In order to test whether our intervention crowds out performance in the other course, we run our main regression with the

Table 7
Grade on the other course

Grade	Overall	Male	Female	Dutch	International
T: Goal	−0.020 (0.076)	−0.067 (0.104)	0.107 (0.116)	−0.111 (0.111)	0.162 (0.099)
T: Raise	0.038 (0.086)	0.078 (0.105)	0.046 (0.150)	−0.040 (0.123)	0.232 (0.139)
Observations	894	626	268	661	233
Mentorgroups	84	84	80	60	24

Standard errors (clustered at the mentor level) in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$ Control variables: study program, tutorial group, mentor gender and experience, student gender, 1st midterm grade, prior GPA.

other course's grade as the outcome variable. The results, presented in Table 7, show that the performance in the other course is not affected. Hence, the positive effect of goal setting in this study is a net increase in performance.

7. Conclusion

We conducted a field experiment in order to test the effects of encouraging students to set goals and encouraging students to increase the ambitiousness of their goals during mentor-student meetings in a university study programme. We designed two treatments. In the *goal treatment* we instructed mentors to encourage students to set a grade specific goal. In the *raise treatment* we gave mentors the same instruction and in addition instructed them to raise this goal if deemed appropriate.

We find that students in the goal treatment perform better than students in the control group. Students whose mentor was assigned to the goal treatment score 0.18 grade points (i.e. 9.5% of a standard deviation) higher than students in the control group. Students in the raise treatment perform similarly to students in the control group, although there are some indications that their performance is even lower. This is true in terms of both the dropout rate and the grades achieved that are conditional on participating in the final exam of the course. The null effect of the raise treatment is in line with the goal becoming unacceptable due to the raise, indicating that the size of the raise was too high.

An alternative explanation for the result that students in the raise treatment perform worse than students in the goal treatment is the nature of the goal. While in the goal treatment students set themselves a goal, a proposal to raise this self-set goal can be seen as a goal of a different kind, namely a cooperatively set goal (or even an assigned goal). Changing the nature of the goal can change the commitment of

the student to the goal (see Hollenbeck et al., 1989), which implies that the intrinsic motivation (i.e. the utility gain when reaching the goal) changes between the two treatments. As a consequence students perform worse in the raise treatment than in the goal treatment. Further, if some students in the control group set themselves a goal then this could even lead to a lower performance of students in the raise treatment as compared to the control group.²³ Finally, we find that students who performed poorly prior to the experiment benefit most from being motivated to set goals in a one-on-one meeting with their mentor.

It is interesting to learn to what extent our results are generalizable to other university teaching programmes. First, the treatments are conducted within an existing mentor program, not all universities have such programmes. The effects of goal setting using for instance surveys is mixed. This implies that a goal setting treatment in an existing program can be successful, but that this may not work in the exact same way in an online environment. Second, students in our setting have an option of dropping out of the course, and taking a resit at the end of the academic year. It remains to be seen what the effect of our treatments would be if the students have only one opportunity to pass their courses.

Our paper is (relatively) silent on the mechanisms that drive goal setting. It is interesting to learn to what extent present bias preferences and loss aversion, as is posited in economic theory papers as important drivers that make goals work, are predictors of the success of goal setting. For example our result that goal setting works mostly for initially poor performing students may be explained by poor performing students having stronger present bias preferences, are more loss averse, or by the fact that extra effort more easily increases performance when initial performance is low. van Lent (2019) takes a first step in disentangling these effects by measuring students' motivation, effort, and time preferences and relating this to the performance of students who are motivated to set goals in a randomized field experiment.

> Appendix A

Appendix 1

Dear X,

Following our introduction during the tutor instruction session, we request you to adjust the progress meetings with your students. Your participation contributes to research regarding the possibilities to increase students' study success by improving the tutor meetings.

The instructions regarding the progress meetings that you conduct in the week of 17th to 21st November follow.

After you have discussed the general motivation and study progress of the student, you are expected to ask some additional questions while discussing the current courses the student follows. These questions relate to Microeconomics. The intention is to motivate the students to set a goal. Ask the questions in italics.

Do you have a final grade in mind that you want to achieve for Microeconomics?

If YES: *What grade do you want to achieve?*

(if the student answers: I want to pass the course, then try to specify this, for example: Are you aiming for 5,5 or a 6?)

If NO: *Some students find it useful to set concrete goals in the form of a grade. Setting a goal can help motivate you to perform better. Do you want to set yourself a goal?*

If YES: *What final grade do you want to achieve for Microeconomics?*

²³ Our finding that 55% of students that are asked about goals already have a goal in mind supports this idea.

If NO: continue the conversation as usual.

If the student set a goal:

Good luck with achieving your goal.

It is important that you follow the instructions as much as possible. However, do try to incorporate the questions into the conversation naturally. Attached, you find a flowchart summarizing the script. You can use this flowchart to refresh your memory prior to the meeting.

We request you to complete the attached form after the meeting with each student. You can print this form yourself, or pick up a copy at H8-23 or H8-24. It can be useful to make notes. Please, read the form carefully before the meetings.

After the meetings we would like to receive the completed forms. The completed forms can be handed in at H8-23 or H8-24 or can be emailed to vanlent@ese.eur.nl or souverijn@ese.eur.nl. We request that you hand in the form at Friday 28th November at the latest.

For research purposes we request that you do not discuss these instructions with others.

If you have any questions, do not hesitate to contact us. You can find us in H8-23 or H8-24, and you can reach us at vanlent@ese.eur.nl, phone: 010 408 1793 or souverijn@ese.eur.nl, phone: 010 408 9038.

Max van Lent

Michiel Souverijn

P.S. Could you please confirm to us by email that you have received this email and that you have read the instructions.

Appendix 2

Dear X,

Following our introduction during the tutor instruction session, we request you to adjust the progress meetings with your students. Your participation contributes to research regarding the possibilities to increase students' study success by improving the tutor meetings.

The instructions regarding the progress meetings that you conduct in the week of 17th to 21st November follow.

After you have discussed the general motivation and study progress of the student, you are expected to ask some additional questions while discussing the current courses the student follows. These questions relate to Microeconomics. The intention is to motivate the students to set a goal. Ask the questions in italics.

Do you have a final grade in mind that you want to achieve for Microeconomics?

If YES: *What grade do you want to achieve?*

(if the student answers: I want to pass the course, then try to specify this, for example: Are you aiming for 5,5 or a 6?)

If NO: *Some students find it useful to set concrete goals in the form of a grade. Setting a goal can help motivate you to perform better. Do you want to set yourself a goal?*

If YES: *What final grade do you want to achieve for Microeconomics?*

If NO: continue the conversation as usual.

If you (as a tutor) think the goal (grade) set is achievable:

Shouldn't you set a higher goal and aim for a [grade + 1]? [So if the student chooses a 6 as a goal and you think this is achievable, propose to aim for a 7.]

If the student set a goal:

Good luck with achieving your goal.

It is important that you follow the instructions as much as possible. However, do try to incorporate the questions into the conversation naturally. Attached, you find a flowchart summarizing the script. You can use this flowchart to refresh your memory prior to the meeting.

We request you to complete the attached form after the meeting with each student. You can print this form yourself, or pick up a copy at H8-23 or H8-24. It can be useful to make notes. Please, read the form carefully before the meetings.

After the meetings we would like to receive the completed forms. The completed forms can be handed in at H8-23 or H8-24 or can be emailed to vanlent@ese.eur.nl or souverijn@ese.eur.nl. We request that you hand in the form at Friday 28th November at the latest.

For research purposes we request that you do not discuss these instructions with others.

If you have any questions, do not hesitate to contact us. You can find us in H8-23 or H8-24, and you can reach us at vanlent@ese.eur.nl, phone: 010 408 1793 or souverijn@ese.eur.nl, phone: 010 408 9038.

Max van Lent

Michiel Souverijn

P.S. Could you please confirm to us by email that you have received this email and that you have read the instructions.

Appendix 3

Table A.1

Students asked and not asked to set a goal within the treatment groups.

	Not asked to set goal		Asked to set goal		p-value
	Mean	Observations	Mean	Observations	
Age	18.50	26	18.71	308	0.43
Female	0.30	37	0.27	483	0.69
Female mentor	0.22	37	0.34	492	0.14
Exp. mentor	0.62	37	0.46	492	0.06
EC Math	5.02	19	5.87	294	0.02
ECX Math	5.60	6	6.71	78	0.08
ET Precalculus	4.60	3	6.20	64	0.14

(continued on next page)

Table A.1 (continued)

	Not asked to set goal		Asked to set goal		<i>p</i> -value
	Mean	Observations	Mean	Observations	
ETX Precalculus	5.50	2	7.43	22	0.24
EC Accounting	4.55	19	6.12	293	0.00
ECX Accounting	4.85	6	5.88	77	0.20
ET Matrix Alg.	4.47	3	6.35	63	0.10
ETX Matrix Alg.	6.70	2	7.85	22	0.46
ET Statistics	2.77	3	4.83	61	0.06
ETX Statistics	3.50	2	5.95	22	0.14

See Table 1 for an explanation of terms.

Table A.2

Students asked and not asked to raise their goal in the raise treatment.

	No raise proposed		Raise proposed		<i>p</i> -value
	Mean	Observations	Mean	Observations	
Age	18.79	84	18.59	75	0.29
Female	0.26	127	0.25	100	0.87
Female mentor	0.29	128	0.38	104	0.17
Exp. mentor	0.58	128	0.45	104	0.06
EC Math	5.63	70	6.04	74	0.13
ECX Math	6.32	21	7.24	17	0.03
ET Precalculus	6.26	7	6.97	3	0.73
ETX Precalculus	7.31	13	7.88	4	0.63
EC Accounting	5.69	69	6.24	75	0.06
ECX Accounting	5.25	21	6.31	17	0.10
ET Matrix Alg.	6.31	7	6.40	3	0.97
ETX Matrix Alg.	7.58	13	8.80	4	0.30
ET Statistics	5.36	7	7.40	3	0.22
ETX Statistics	5.44	13	6.47	4	0.37

See Table 1 for an explanation of terms.

Table A.3

Students who have a goal in mind when asked.

	Goal in mind		No goal in mind		<i>p</i> -value
	Mean	Observations	Mean	Observations	
Age	18.78	177	18.61	157	0.23
Female	0.27	265	0.27	255	0.90
Female mentor	0.30	271	0.36	258	0.16
Exp. mentor	0.53	271	0.41	258	0.01
GPA	6.06	256	5.71	238	0.02

See Table 1 for an explanation of terms.

Table A.4

Mentor estimates are accurate.

	grade - goal
Goal too hard	- 1. 507*** (0.261)
Goal too easy	0.337*** (0.412)
Baseline: doable	- 0. 446*** (0.261)
Observations	406
R^2	0.111

Standard errors (clustered at the mentor level) in parentheses.

* $p < .1$, ** $p < .05$, *** $p < .01$

Table A.5

Students who set and do not set goals when asked to set a goal.

	Set Goal		No goal set		p
	Mean	Observations	Mean	Observations	
Age	18.74	265	18.58	43	0.48
Female	0.27	424	0.27	59	0.94
Female Mentor	0.35	433	0.22	59	0.05
Exp. Mentor	0.47	433	0.41	59	0.39
EC Math	5.85	254	5.99	40	0.60
ECX Math	6.67	73	7.16	5	0.49
ET Precalculus	6.28	57	5.57	7	0.32
ETX Precalculus	7.62	20	5.55	2	0.19
EC Accounting	6.12	252	6.17	41	0.85
ECX Accounting	5.96	72	4.78	5	0.18
ET Matrix Alg.	6.28	57	7.00	6	0.39
ETX Matrix Alg.	7.89	20	7.40	2	0.74
ET Statistics	4.90	55	4.18	6	0.38
ETX Statistics	6.03	20	5.20	2	0.62
Midterm I	4.77	426	4.47	55	0.32

See Table 1 for an explanation of terms.

Table A.6

Students who accept and reject raise when asked to raise their goal.

	Reject raise		Accept raise		p-value
	Mean	Observations	Mean	Observations	
Age	18.53	38	18.65	37	0.60
Female	0.22	50	0.28	50	0.49
Female mentor	0.46	52	0.29	52	0.07
Exp. mentor	0.56	52	0.35	52	0.03
EC Math	5.93	37	6.15	37	0.57
ECX Math	6.78	9	7.75	8	0.10
ET Precalculus	4.70	1	8.10	2	.
ETX Precalculus	8.85	2	6.90	2	0.34
EC Accounting	6.52	38	5.95	37	0.13
ECX Accounting	6.12	9	6.51	8	0.70
ET Matrix Alg.	4.60	1	7.30	2	.
ETX Matrix Alg.	9.90	2	7.70	2	0.33
ET Statistics	6.10	1	8.05	2	.
ETX Statistics	7.25	2	5.70	2	0.54
Initial goal	6.60	52	6.53	52	0.62

See Table 1 for an explanation of terms.

Table A.7

Total effect tutor characteristics interacted.

Grade	Overall
T: Goal	0.163* (0.088)
T: Raise	− 0.205*** (0.101)
Observations	967
Mentorgroups	84

Standard errors (clustered at the mentor level) in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$ Control variables: study program, tutorial group, mentor gender and experience, student gender, 1st midterm grade, prior GPA.

Table A.8
Intention to treat effect: maximum grade including
resit

Standard errors (clustered at the mentor level) in parentheses. * $p < .1$, ** $p < .05$, *** $p < .01$. Control variables: study program, tutorial group, mentor gender and experience, student gender, 1st midterm grade, prior GPA.

Appendix 4

Figs. A.1 and A.2.

[illegible]

Fig. A.1. Form for mentors in goal treatment.

- Applied Psychology 95 (2), 255–264.
- Niederle, M., Vesterlund, L., 2011. Gender and competition. *Annual Review of Economics*, 3 (1), 601–630.
- Ordóñez, L.D., Schweitzer, M.E., Galinsky, A.D., Bazerman, M.H., 2009. Goals gone wild: The systematic side effects of overprescribing goal setting. *The Academy of Management Perspectives* 23 (1), 6–16.
- Sackett, A.M., Wu, G., White, R.J., Markle, A., 2014. Harnessing optimism: How eliciting goals improves performance. Working Paper
- Sanders, M., Chonaire, A.N., 2015. “powered to detect small effect sizes” : You keep saying that. I do not think it means what you think it means (no. 15/337). Working Paper Department of Economics, University of Bristol, UK
- Schippers, M.C., Scheepers, A.W., Peterson, J.B., 2015. A scalable goal-setting intervention closes both the gender and ethnic minority achievement gap. *Palgrave Communications* 1.
- Schunk, D.H., 1990. Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, 25 (1), 71–86.
- Seijts, G.H., Latham, G.P., Tasa, K., Latham, B.W., 2004. Goal setting and goal orientation: An integration of two different yet related literatures. *Academy of Management Journal* 47 (2), 227–239.
- Smithers, S., 2015. Goals, motivation and gender. *Economics Letters* 131, 75–77.
- Suvorov, A., Van de Ven, J., 2008. Goal setting as a self-regulation mechanism. Working Paper
- Travers, C.J., Morisano, D., Locke, E.A., 2015. Self-reflection, growth goals, and academic outcomes: A qualitative study. *British Journal of Educational Psychology* 85 (2), 224–241.