Dec 2018 No.396

Using Goals to Motivate College Students: Theory and Evidence from Field Experiments

Damon Clark, David Gill,
Victoria Prowse and Mark Rush

WORKING PAPER SERIES

Centre for Competitive Advantage in the Global Economy

Department of Economics

Using Goals to Motivate College Students: Theory and Evidence from Field Experiments *

Damon Clark †
David Gill ‡
Victoria Prowse §
Mark Rush ¶

This version: December 19, 2018
First version: October 20, 2016

Abstract

Will college students who set goals for themselves work harder and achieve better outcomes? In theory, setting goals can help present-biased students to mitigate their self-control problem. In practice, there is little credible evidence on the causal effects of goal setting for college students. We report the results of two field experiments that involved almost four thousand college students in total. One experiment asked treated students to set goals for performance in the course; the other asked treated students to set goals for a particular task (completing online practice exams). Task-based goals had robust positive effects on the level of task completion, and task-based goals also increased course performance. We also find that performance-based goals had positive but small effects on course performance. We use a theoretical framework that builds on present bias and loss aversion to interpret our results. Since task-based goal setting is low-cost, scalable and logistically simple, we conclude that our findings have important implications for educational practice and future research.

Keywords: Goal; Goal setting; Higher education; Field experiment; Self-control; Present bias; Time inconsistency; Commitment device; Loss aversion; Reference point; Task-based goal; Performance-based goal; Self-set goal; Performance uncertainty; Overconfidence; Student effort; Student performance; Educational attainment.

JEL Classification: I23, C93.

*Primary IRB approval was granted by Cornell University. We thank Cornell University and UC Irvine for funding this project. We thank Svetlana Beilfuss, Daniel Bonin, Debasmita Das, Linda Hou, Jessica Monnet, Ben Raymond, Mason Reasner, Peter Wagner, Laurel Wheeler and Janos Zsiros for excellent research assistance. Finally, we are grateful for the many helpful and insightful comments that we have received from seminar participants and in private conversations.

†Department of Economics, UC Irvine and NBER; clarkd1@uci.edu.
‡Department of Economics, Purdue University; gill53@purdue.edu.
§Department of Economics, Purdue University; vprowse@purdue.edu.
¶Department of Economics, University of Florida; Mark.Rush@warrington.ufl.edu.

1 Introduction

Researchers and policy-makers worry that college students exert too little effort, with consequences for their learning, their graduation prospects, and ultimately their labor market outcomes. With this in mind, attention has focused on policies and interventions that could increase student effort by introducing financial incentives, such as making student aid conditional on meeting GPA cutoffs and paying students for improved performance; however, these programs are typically expensive and often yield disappointing results (e.g., Henry et al., 2004, Cornwell et al., 2005, Angrist et al., 2009, Cha and Patel, 2010, Leuven et al., 2010, Scott-Clayton, 2011, De Paola et al., 2012, Patel and Rudd, 2012, Castleman, 2014, Cohodes and Goodman, 2014).[1]

In this paper we aim to discover whether goal setting can motivate college students to work harder and achieve better outcomes. We focus on goal setting for three main reasons. First, in contrast to financial incentives, goal setting is low-cost, scalable and logistically simple. Second, students might lack self-control. In other words, although students might set out to exert their preferred level of effort, when the time comes to attend class or study, they might lack the self-control necessary to implement these plans. The educational psychology literature finds that self-control correlates positively with effort, which supports the idea that some students under-invest in effort because of low self-control (e.g., Duckworth and Seligman, 2005, Duckworth et al., 2012). Third, the behavioral economics literature suggests that agents who lack self-control can use commitment devices such as restricted-access savings accounts to self-regulate their behavior (e.g., Wertenbroch, 1998, Ariely and Wertenbroch, 2002, Thaler and Benartzi, 2004, Ashraf et al., 2006, DellaVigna and Malmendier, 2006, Augenblick et al., 2015, Kaur et al., 2015, Patterson, 2016).[2] Goal setting might act as an effective internal commitment device that allows students who lack self-control to increase their effort.[3]

We gather large-scale experimental evidence from the field to investigate the causal effects of goal setting among college students. We study goals that are set by students themselves, as opposed to goals set by another party (such as a counselor or professor), because self-set goals can be personalized to each student's degree of self-control. We study two types of goals: self-set goals that relate to performance in a course (performance-based goals) and self-set goals that relate to a particular study task (task-based goals). The design of our goal interventions builds on prior work. Our performance-based goals can be viewed as a variant of the performance-based incentives discussed above, with the financial incentives removed and with self-set goals added in their place. Our task-based goals build on recent research by Allan and Fryer (2011) and Fryer (2011) that suggests that financial incentives at the K-12 level work well when they are tied to task completion (e.g., reading a book).

[1]See Web Appendix V.1 and the survey by Lavecchia et al. (2016) for more details. A recent study by Lusher (2016) evaluates a program called "CollegeBetter.com" in which students make parimutuel bets that they will raise their GPA by the end of the term. The financial rewards and penalties that the program creates act as an external commitment device. Participating students were more likely to increase their GPA compared to students who wanted to participate but were randomly excluded; however, CollegeBetter.com did not affect average GPA.

[2]See Web Appendix V.2 and the survey by Bryan et al. (2010) for more details.

[3]A small and recent literature in economics suggests that goal setting can influence behavior in other settings (Goerg and Kube, 2012; Harding and Hsiaw, 2014; Corgnet et al., 2015, 2016; Choi et al., 2016); see Web Appendix V.3 and the survey by Goerg (2015) for more details. Although not focused on education, several psychologists argue for the motivational benefits of goals more generally (see, e.g., Locke, 1968, Locke et al., 1981, Mento et al., 1987, Locke and Latham, 2002, and Latham and Pinder, 2005).

In considering both task-based goals and performance-based goals, our aim is not to test which is more effective. Instead, we aim to understand separately the impacts of two goal-setting technologies that could easily be incorporated into the college setting. To do this, we ran two separate experiments, each with its own within-cohort treatment-control comparison. By learning whether each intervention is effective in its own right, we can provide policy makers and educators who are considering introducing a particular form of goal setting with valuable information about the likely impact of the intervention.[4]

We administered two field experiments with almost four thousand college students in total. The subjects were undergraduate students enrolled in an on-campus semester-long introductory course at a public university in the United States. The course was well established prior to our study and has been taught by the same professor for many years. The course is worth four credit hours, and a letter grade of a C or better in the course is required to graduate with a bachelor's degree in the associated subject.

In the performance-based goals experiment, students were randomly assigned to a Treatment group that was asked to set goals for their performance in the course or to a Control group that was not. The performance measures for which goals were set included the overall course letter grade and scores on the midterm exams and final exam. Consistent with the prior work on performance-based incentives discussed above, we find that performance-based goals do not have a significant impact on course performance. Instead, our estimates were positive but small and statistically insignificant.

In the task-based goals experiment, students were randomly assigned to a Treatment group that was asked to set goals for the number of online practice exams that they would complete in advance of each midterm exam and the final exam or to a Control group that was not. We find that task-based goals are effective. Asking students to set task-based goals for the number of practice exams to complete increased the average number of practice exams that students completed by 0.102 of a standard deviation. This positive effect of task-based goals on the level of task completion is statistically significant ($p = 0.017$) and robust. As well as increasing task completion, task-based goals also increased course performance (although the effects are on the margins of statistical significance): asking students to set task-based goals increased average total points scored in the course by 0.068 of a standard deviation ($p = 0.086$) and increased median total points scored by 0.096 of a standard deviation ($p = 0.019$). The obvious explanation for this increase in performance is that it stems from the greater task completion induced by setting task-based goals. If correct, this implies that the task-based goal-setting intervention directed student effort toward a productive activity (completing practice exams). More generally, our results suggest that if tasks are chosen appropriately then task-based goals can improve educational performance as well as induce greater task-specific investments.

Interestingly, we also find that task-based goals were more effective for male students than for female students, both in terms of the impact on the number of practice exams completed and on performance in the course. Specifically, for male students task-based goals increased the

[4] Our experiments are powered to detect plausible treatment-control differences. We did not power our experiments to test directly for differences in the effectiveness of goal setting across experiments for two reasons: first, calculating power ex ante was not realistic because we had little evidence ex ante to guide us regarding the size of such differences; and, second, sample size constraints (that arise from the number of students enrolled in the course) limit power to detect across-experiment differences unless those differences are very large.

average number of practice exams completed by 0.190 of a standard deviation ($p = 0.006$) and increased average total points scored by 0.159 of a standard deviation ($p = 0.013$). In contrast, for female students task-based goals increased the average number of practice exams completed by only 0.033 of a standard deviation and decreased average total points scored by 0.012 of a standard deviation (the treatment effects for women are far from being statistically significant). These gender differences in effect size are in line with prior work showing that males are more responsive to incentives for shorter-term performance (e.g., Gneezy and Rustichini, 2004, Levitt et al., 2011), and contrast with prior work showing that females are more responsive to longer-term performance incentives (e.g., Angrist et al., 2009, Angrist and Lavy, 2009.)

We focus on gender because four strands of literature come together to suggest that the effect of goal setting in education might vary by gender. First, evidence from other educational environments suggests that males have less self-control than females (e.g., Duckworth and Seligman, 2005, Buechel et al., 2014, and Duckworth et al., 2015); summarizing this literature, Duckworth et al. (2015) conjecture that educational interventions aimed at improving self-control may be especially beneficial for males. Second, our theoretical framework implies that goal setting is more effective for present-biased students, while the evidence from incentivized experiments suggests that men are more present biased than women (we survey this literature in Web Appendix V.6). Third, evidence from the laboratory suggests that goal setting is more effective for men: in an experiment in which goals were set by the experimenter rather than by the subjects themselves, Smithers (2015) finds that goals increased the work performance of men but not that of women. Fourth, to the extent that education is a competitive environment, the large literature on gender and competition (that started with Gneezy et al., 2003) suggests that there might be interesting and robust gender differences in the effectiveness of interventions designed to motivate students.

We argue that our findings are consistent with a theoretical framework in which students are present biased and loss averse. This framework builds on Koch and Nafziger (2011) and implies that present-biased students will, in the absence of goals, under-invest in effort. By acting as salient reference points, self-set goals can serve as internal commitment devices that enable students to increase effort. This mechanism can rationalize the positive effects of task-based goal setting (although we do not rule out all other possible mechanisms).[5] We use the framework to suggest three key reasons why performance-based goals might not be very effective in the setting that we studied: performance is realized in the future; performance is uncertain; and students might be overconfident about how effort translates into performance. Consistent with Allan and Fryer (2011)'s explanation for why performance-based financial incentives appear ineffective, our overconfidence explanation implies that students have incorrect beliefs about the best way to increase their academic achievement.[6]

The primary contribution of this paper is to show that a low-cost, scalable and logistically

[5]In related theoretical work, Hsiaw (2013) studies goal setting with present bias and expectations-based reference points. In an educational context, Levitt et al. (2016) find evidence that school children exhibit both loss aversion (incentives framed as losses are more powerful) and present bias (immediate rewards are more effective).

[6]In the case of task-based goals, the first two considerations no longer apply. Furthermore, to the extent that task-based goals direct students toward productive tasks, task-based goal setting mitigates the effect of overconfidence. Plausibly, teachers have better information about which tasks are likely to be productive, and asking students to set goals for productive tasks is one way to improve the power of goal setting for overconfident students.

simple intervention using self-set goals can have a significant effect on performance. As discussed above, prior programs have offered financial incentives for meeting externally set (and usually longer-term) performance targets, but the results of these studies have been modest, especially given their costs and other concerns about using incentives (e.g., crowding out of intrinsic motivation; see Cameron and Pierce, 1994, and Gneezy et al., 2011). We provide experimental evidence that task-based goal setting can increase the effort and performance of college students. We also show that performance-based goals have small and statistically insignificant effects on performance, although any direct comparison of our two interventions should be interpreted with some caution.[7]

Our study represents a substantial innovation on existing experimental evaluations of the effects of goal setting on the effort and performance of college students. In particular, while a handful of papers in psychology use experiments to study the effects of self-set goals among college students (Morgan, 1987; Latham and Brown, 2006; Morisano et al., 2010; Chase et al., 2013), these differ from our analysis in three important respects. First, they rely on much smaller samples. Second, they have not explored the impact of performance-based goals on performance or the impact of task-based goals on performance.[8] Third, they have not studied the effect of task-based goals on task completion and, therefore, have not investigated the mechanism behind any performance effects of task-based goal setting.[9]

Numerous studies in educational psychology report non-causal correlational evidence which suggests that performance-based goal setting has strong positive effects on performance (e.g., Zimmerman and Bandura, 1994, Schutz and Lanehart, 1994, Harackiewicz et al., 1997, Elliot and McGregor, 2001, Barron and Harackiewicz, 2003, Linnenbrink-Garcia et al., 2008 and Darnon et al., 2009). Another contribution of our paper is to cast doubt on this correlational evidence using our experimental finding that performance-based goals have small and statistically insignificant effects on performance. The obvious explanation for the discrepancy between previous correlational estimates and our experimental estimate is that the correlational estimates do not identify the relevant causal effect. We use our sample to explore this possibility. In line with previous correlational studies, in our experiment students who set ambitious performance-based goals performed better: conditional on student characteristics, the correlation in our sample between course performance (measured by total number of points scored out of one hundred) and the level of the goal is 0.203 ($p = 0.000$) for students who set performance-based goals. The difference between the strong positive correlation based on non-experimental variation in our sample and the small and statistically insignificant causal effects that we estimate suggests that correlational analysis gives a misleading impression of the effectiveness of performance-based

[7] In particular, the structure of the practice exams was not exactly the same across the two experiments: practice exams had to be downloaded in the performance-based goals experiment, but could be completed online in the task-based goals experiment. However, we provide evidence that a difference in the saliency of practice exams was not important (see Section 4.3.4).

[8] Morgan (1987) is the exception, but this small-scale study of task-based goal setting does not report a statistical test of the relevant treatment-control comparison. Web Appendix V.4 provides more detail about this paper.

[9] Using a sample of seventy-seven college students, Schunk and Ertmer (1999) studied teacher-set instead of self-set goals: they directed students who were acquiring computer skills to think about outcomes (that the students had already been asked to achieve) as goals. Web Appendix V.5 discusses the literature in psychology on goals and the learning of grade-school-aged children, which focuses on teacher-set goals.

goals.[10]

Our analysis breaks new ground in understanding the impacts of goal setting among college students. In particular, our experimental findings suggest that for these students, task-based goals could be an effective method of mitigating self-control problems. We emphasize that our task-based goal intervention was successful because it directed students toward a productive task. When applying our insights, teachers should attempt to pair goal setting with tasks that they think are productive, while policymakers should publicize new knowledge about which tasks work well with goals.

As we explain in the Conclusion of this paper, our findings have important implications for educational practice and future research. Many colleges already offer a range of academic advising programs, including mentors, study centers and workshops. These programs often recommend goal setting, but only as one of several strategies that students might adopt to foster academic success. Our findings suggest that academic advising programs could give greater prominence to goal setting, and that students could be encouraged to set task-based goals for activities that are important for educational success. Our findings also suggest that individual courses could be designed to give students opportunities to set task-based goals. In courses with some online components (including fully online courses), it would be especially easy to incorporate task-based goal setting into the technology used to deliver course content; in traditional classroom settings, students might be encouraged to set task-based goals in consultation with instructors, who are well placed to select productive tasks. In conjunction with our experimental findings, these possibilities demonstrate that task-based goal setting is a scalable and logistically simple intervention that could help to improve college outcomes at low cost. This is a promising insight, and we argue in the Conclusion that it ought to spur further research into the effects of task-based goal setting in other college contexts (e.g., two-year colleges) and for other tasks (e.g., attending lectures or contributing to online discussions).

The paper proceeds as follows. In Section 2 we describe our field experiments; in Section 3 we present our experimental results; in Section 4 we interpret our results using a theoretical framework that is inspired by present bias and loss aversion; and in Section 5 we conclude by discussing the implications of our findings.

[10] For students who set task-based goals, the correlation between course performance (measured by total number of points scored out of one hundred) and the level of the goal is 0.391 ($p = 0.000$), which is in line with correlational findings from educational psychology (see, e.g., Elliot and McGregor, 2001, Church et al., 2001, and Hsieh et al., 2007).

2 Experimental design and descriptive statistics

2.1 Description of the sample

We ran our field experiments at a large public land-grant university in the United States.[11] Our subjects were undergraduate students enrolled in a large on-campus semester-long introductory course. The course is a mainstream Principles of Microeconomics course that follows a conventional curriculum and assesses student performance in a standard way using quizzes, midterms and a final (see Section 2.2 below). The course was well established prior to our study and has been taught by the same experienced professor for many years. The course is worth four credit hours, and a letter grade of a C or better in the course is required to graduate with a bachelor's degree in the associated subject. Since this is a large course, the live lectures are recorded and placed on the Internet; all students have the choice of watching the lectures as they are delivered live, but many choose to watch online. There are no sections for this course.

At least two features of this course reduce the likelihood of spillovers from the Treatment group to the Control group. First, this is an introductory course in which most of the students are freshmen, and therefore social networks are not yet well established. Second, the absence of sections or organized study groups, and the fact that many students choose to watch the lectures online, reduce the likelihood of in-class spillovers.

As described in Section 2.2, we sought consent from all our subjects (the consent rate was ninety-eight percent). Approximately four thousand students participated in total. We employed a between-subjects design: each student was randomized into the Treatment group or the Control group immediately on giving consent.[12] Students in the Treatment group were asked to set goals while students in the Control group were not asked to set any goals. As described in Section 2.3, in the Fall 2013 and Spring 2014 semesters we studied the effects of performance-based goals on student performance in the course (the 'performance-based goals' experiment). As described in Section 2.4, in the Fall 2014 and Spring 2015 semesters we studied the effects of task-based goals on task completion and course performance (the 'task-based goals' experiment).[13]

Table 1 provides statistics about participant numbers and treatment rates. We have information about participant demographics from the university's Registrar data: Tables A.1, A.2 and A.3 in Web Appendix I summarize the characteristics of our participants and provide evidence that our sample is balanced.[14]

[11] The university is the top-ranked public university in a major state, and is categorized as an R1 (highest research activity) institution by the Carnegie Classification of Institutions of Higher Education. The median SAT score of incoming freshmen is slightly more than 1,300. Around 6,400 full-time first-time undergraduate freshmen students enroll on the main campus each year, of whom around sixty percent are female, around fifty-percent are non-Hispanic white, around twenty percent are Hispanic, around ten percent are Asian, and around five percent are black. Around a third receive Pell grants, and around forty percent receive either a Pell grant or a subsidized Stafford Loan.

[12] When the subject pressed the online consent button, a computerized random draw allocated that subject to the Treatment or Control group with equal probability. The draws were independent across subjects.

[13] We also ran a small-scale pilot in the summer of 2013 to test our software.

[14] For each characteristic we test the null that the difference between the mean of the characteristic in the Treatment Group and the Control group is zero, and we then test the joint null that all of the differences equal zero. The joint test gives p-values of 0.636, 0.153 and 0.471 for, respectively, all semesters, Fall 2013 and Spring 2014 (the performance-based goals experiment), and Fall 2014 and Spring 2015 (the task-based goals experiment). See Tables A.1, A.2 and A.3 for further details.

	All semesters	Fall 2013 & Spring 2014 (Performance-based goals)	Fall 2014 & Spring 2015 (Task-based goals)
Number of participating students	3,971	1,967	2,004
Number of students in Treatment group	1,979	995	984
Number of students in Control group	1,992	972	1,020
Fraction of students in Treatment group	0.50	0.51	0.49

Notes: The number of participating students excludes: students who did not give consent to participate; students who formally withdrew from the course; students who were under eighteen at the beginning of the semester; students for whom the university's Registrar data does not include SAT or equivalent aptitude test scores; and one student for whom the Registrar data does not include information on gender.

Table 1: Participant numbers and treatment rates.

2.2 Course structure

In all semesters, a student's letter grade for the course was based on the student's total points score out of one hundred. The relationship between total points scored and letter grades was fixed throughout our experiments and is shown in the grade key at the bottom of Figure A.1 in Web Appendix II. The grade key was provided to all students at the start of the course (via the course syllabus) and students were also reminded of the grade key each time they checked their personalized online gradecard (described below).

Points were available for performance in two midterm exams, a final exam and a number of online quizzes. Points were also available for taking an online syllabus quiz and a number of online surveys. For the Fall 2013 semester Figure 1 gives a timeline of the exams, quizzes and surveys, and the number of points available for each. As described in Sections 2.3 and 2.4, the course structure in other semesters was similar.

Each student had access to a private personalized online gradecard that tracked the student's performance through the course and that was available to view at all times. After every exam, quiz or survey, the students received an email telling them that their gradecard had been updated to include the credit that they had earned from that exam, quiz or survey. The gradecards also included links to answer keys for the online quizzes. Figure A.1 in Web Appendix II shows an example gradecard for a student in the Control group in the Fall 2013 semester.

In all semesters, students had the opportunity to complete practice exams that included question-by-question feedback. The opportunity to take practice exams was highlighted on the first page of the course syllabus. In the Fall 2013 and Spring 2014 semesters the students downloaded the practice exams from the course website, and the downloads included answer keys.[15] In the Fall 2014 and Spring 2015 semesters the students completed the practice exams online, and the correct answer was shown to the student after attempting each question. As described below in Section 2.4, the students received emails reminding them about the practice exams in the Fall 2014 and Spring 2015 semesters.

We sought consent from all of our subjects using an online consent form. The consent form appeared immediately after students completed the online syllabus quiz and immediately before the online start-of-course survey. Figure A.2 in Web Appendix II provides the text of the consent

[15] As a result, we have no measure of practice exam completion for the Fall 2013 and Spring 2014 semesters.

form.

Syllabus quiz and start-of-course survey	
Syllabus quiz	2 points for completion
Consent form	For treated and control students
Start-of-course survey	Treated students set goal for letter grade in course 2 points for completion
Online quizzes	
10 online quizzes throughout the semester Each scored from 0 to 3 points	
Midterm exam 1	
Scored from 0 to 30 points Only maximum of midterm 1 & 2 scores counts for letter grade	
Midterm exam 2	
Scored from 0 to 30 points Only maximum of midterm 1 & 2 scores counts for letter grade	
Final exam	
Scored from 0 to 34 points	
End-of-course survey	
2 points for completion	

Figure 1: Fall 2013 semester timeline

2.3 Performance-based goals experiment

In the Fall 2013 and Spring 2014 semesters we studied the effects of performance-based goals on student performance in the course. In the Fall 2013 semester treated students were asked to set a goal for their letter grade in the course. As outlined in Figure 1, treated students were asked to set their goal during the start-of-course survey that all students were invited to take.[16] In the Spring 2014 semester treated students were asked to set goals for their scores in the two midterm exams and the final exam. As outlined in Figure 2, the treated students were asked to set a goal for their score in a particular exam as part of a mid-course survey that all students were invited to take.[17]

Figures A.3 and A.4 in Web Appendix II provide the text of the goal-setting questions. In each case, the treated students were told that their goal would be private and that: "each time you get your quiz, midterm and final scores back, your gradecard will remind you of your goal." Figures A.5 and A.6 illustrate how the goal reminders were communicated to the treated students on the online gradecards. The gradecards, described in Section 2.2, were a popular part of the course: the median number of times students viewed their gradecard during the Fall 2013 and Spring 2014 semesters was twenty-three. In Spring 2014, when the mid-course survey before a particular exam closed, the students received an email telling them that their online gradecard had been updated to include the credit that they had earned from completing that

[16] Treated students set their goal after the quiz on the syllabus. In every semester the syllabus gave the students information about the median student's letter grade in the previous semester.

[17] The students were invited to take the mid-course survey three days before the exam.

mid-course survey; opening the gradecard provided a pre-exam reminder of the treated student's goal for their score in the forthcoming exam.

Syllabus quiz and start-of-course survey	
Syllabus quiz	1 point for completion
Consent form	FOR TREATED AND CONTROL STUDENTS
Start-of-course survey	1 point for completion
Online quizzes	
9 online quizzes throughout the semester Each scored from 0 to 3 points	
Mid-course survey 1	
TREATED STUDENTS SET GOAL FOR SCORE IN MIDTERM EXAM 1 2 points for completion	
Midterm exam 1	
Scored from 0 to 30 points Only maximum of midterm 1 & 2 scores counts for letter grade	
Mid-course survey 2	
TREATED STUDENTS SET GOAL FOR SCORE IN MIDTERM EXAM 2 2 points for completion	
Midterm exam 2	
Scored from 0 to 30 points Only maximum of midterm 1 & 2 scores counts for letter grade	
Mid-course survey 3	
TREATED STUDENTS SET GOAL FOR SCORE IN FINAL EXAM 2 points for completion	
Final exam	
Scored from 0 to 34 points	
End-of-course survey	
1 point for completion	

Figure 2: Spring 2014 semester timeline

2.4 Task-based goals experiment

In the Fall 2014 and Spring 2015 semesters we studied the effects of task-based goals on task completion and course performance. Specifically, we studied the effects of goals about the number of practice exams to complete on: (i) the number of practice exams that students completed (which we call the 'level of task completion'); and (ii) the students' performance in the course. The experimental design was identical across the Fall 2014 and Spring 2015 semesters.

The course structure in the Fall 2014 and Spring 2015 semesters was the same as that outlined in Figure 2 for the Spring 2014 semester, except that before each of the two midterm exams and the final exam, instead of setting performance-based goals, the treated students were asked to set a goal for the number of practice exams to complete out of a maximum of five before that particular exam (recall from Section 2.2 that students had the opportunity to complete practice exams in all four semesters). The treated students were asked to set the goal as part of a mid-course survey that all students were invited to take. Both the treated and control students had the opportunity to complete up to five practice exams online before each exam. The opportunity to take the online practice exams was communicated to the treated and control students in the course syllabus, in the mid-course surveys (see Figure A.7 in Web Appendix II) and in reminder emails before each exam (see Figure A.8). Figures A.9 and A.10 show the practice exam instructions and feedback screens.[18]

Figure A.7 in Web Appendix II provides the text of the goal-setting question. The treated students were told that their goal would be private and that: "when you take the practice exams you will be reminded of your goal." Figures A.9 and A.10 illustrate how the goal reminders were communicated to the treated students when attempting the practice exams. The treated students also received a reminder of their goal in the reminder email about the practice exams that all students received (see Figure A.8). Reminders were not provided on gradecards.

2.5 Descriptive statistics on goals

Table 2 presents some descriptive statistics on the goals that the treated students set and the extent to which they achieved these. Looking at the first row of Panel I, we see that the vast majority of treated students chose to set at least one goal, irrespective of whether the goal was performance based or task based. Looking at the second row of Panel I, we see that on average students in the performance-based goals experiment set performance goals of ninety percent (as explained in the notes to Table 2, all performance goals have been converted to percentages of the maximal performance), while on average students in the task-based goals experiment set task goals of four out of five practice exams. The third row of Panel I tells us that these goals were generally a little ambitious: achievement lagged somewhat behind the goals that the students chose to set. Given that the goals were a little ambitious, many students failed to achieve their goals: the fourth row of Panel I shows that each performance-based goal was reached by about one-quarter of students while each task-based goal was reached by about one-half of students.[19] Panels II and III show that the same patterns hold for both male and female students. We further note that, for students who set a goal related to the first midterm exam and a goal related to the final exam, performance-based goals decreased over the semester by an average of 1.56 percentage points, while task-based goals increased over the semester by an average of 0.60 practice exams; these trends did not vary substantially by gender.

[18]The students were invited to take the mid-course survey five days before the relevant exam. Practice exam reminder emails were sent three days before the exam, at which time the practice exams became active. The practice exams closed when the exam started.

[19]Within the performance-based goals experiment, goals and goal achievement varied little according to whether the students set a goal for their letter grade in the course or set goals for their scores in the two midterm exams and the final exam.

Panel I: All students in the Treatment group		
	Performance-based goals	Task-based goals
Fraction who set at least one goal	0.99	0.98
Mean goal	89.50	4.05
Mean achievement	78.40	3.14
Fraction of goals achieved	0.24	0.53
Panel II: Male students in the Treatment group		
	Performance-based goals	Task-based goals
Fraction who set at least one goal	0.99	0.97
Mean goal	90.35	4.03
Mean achievement	79.50	3.03
Fraction of goals achieved	0.25	0.50
Panel III: Female students in the Treatment group		
	Performance-based goals	Task-based goals
Fraction who set at least one goal	0.99	0.99
Mean goal	88.68	4.07
Mean achievement	77.34	3.23
Fraction of goals achieved	0.24	0.55

Notes: The fraction who set at least one goal is defined as the fraction of students in the Treatment group who set at least one goal during the semester. A student is considered to have set a goal for her letter grade in the course if she chose a goal better than an E (an E can be obtained with a total points score of zero). Other types of goal are numerical, and a student is considered to have set such a goal if she chose a goal strictly above zero. The mean goal, mean achievement and fraction of goals achieved are computed only for the students who set at least one goal. The mean goal is calculated by averaging over the goals set by each student (that is, one, two or three goals) and then averaging over students (goals for the letter grade in the course are converted to scores out of one hundred using the lower grade thresholds on the grade key, and goals for scores in the midterms and final exam are rescaled to scores out of one hundred). Mean achievement is calculated by averaging within students over the outcome that is the object of each set goal and then averaging over students (outcomes that correspond to performance-based goals are converted to scores out of one hundred as described previously for the performance-based goals themselves). The fraction of goals achieved is calculated by averaging within students over indicators for the student achieving each set goal and then averaging over students.

Table 2: Descriptive statistics on goals for students in the Treatment group

3 Experimental results

We now describe the results of our experiments. In Section 3.1 we present the effects on task completion. In Section 3.2 we turn to the effects on course performance.

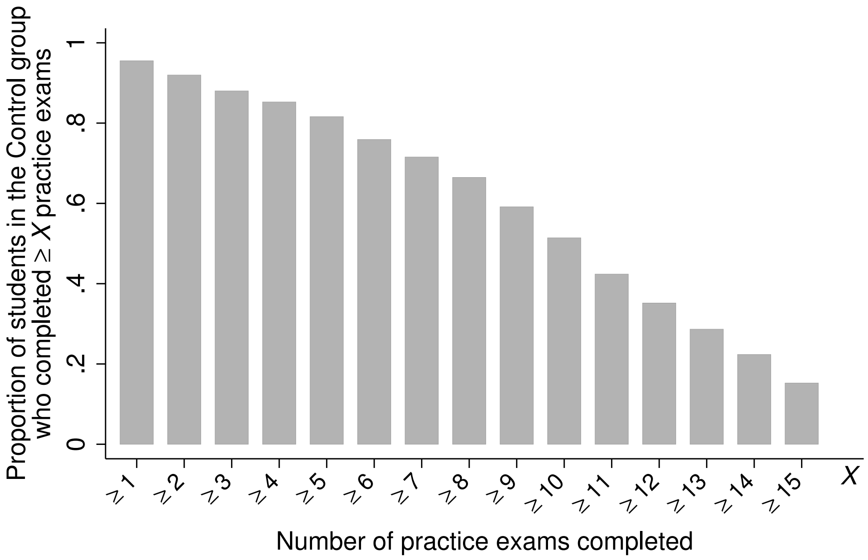
3.1 Impact of task-based goals on task completion

In this section we study the impact of task-based goals on the level of task completion, defined as the number of practice exams that the student completed during the course. Recall that all students in the task-based goals experiment had an opportunity to complete up to five practice exams online before each of two midterms and the final exam, giving a maximum of fifteen practice exams. As explained in Section 2, all students received question-by-question feedback while they completed a practice exam. To preview our results, we find that asking students to set task-based goals for the number of practice exams to complete successfully increased task completion. The positive effect of task-based goals on task completion is large, statistically significant and robust.

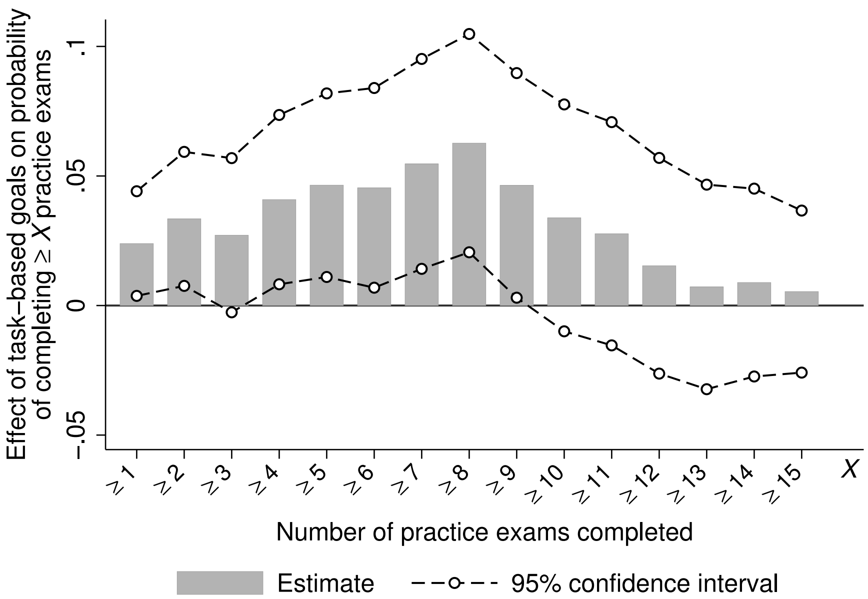
We start by looking at the effects of task-based goals on the pattern of task completion. Figure 3(a) shows the pattern of task completion for the students in the Control group, who were not asked to set goals. For example, Figure 3(a) shows that almost all students in the Control group completed at least one practice exam during the course while around fifteen percent of the students in the Control group completed all fifteen of the available practice exams. Figure 3(b) shows how task-based goal setting changed the pattern of task completion. In particular, Figure 3(b) shows that the task-based goals intervention had significant effects on the bottom and the middle of the distribution of the number of practice exams completed: for example, task-based goals increased the probability that a student completed at least one practice exam by more than two percentage points (p-value $= 0.020$) and increased the probability that a student completed eight or more practice exams by more than six percentage points (p-value $= 0.004$).

Next, we look at how task-based goals changed the average level of task completion. Table 3 reports ordinary least squares (OLS) regressions of the number of practice exams completed during the course on an indicator for the student having been randomly allocated to the Treatment group in the task-based goals experiment. To give a feel for the magnitude of the effects, the second row reports the effect size as a proportion of the standard deviation of the number of practice exams completed in the Control group in the task-based goals experiment, while the third row reports the average number of practice exams completed in the same Control group. The regression in the second column controls for age, gender, race, SAT score, high school GPA, advanced placement credit, Fall semester, and first login time, including linear terms, squares, and interactions of these variables (see the notes to Table 3 for further details on the controls).

From the results in the second column of Table 3, we see that task-based goals increased the mean number of practice exams that students completed by about 0.5 of an exam (the effect has a p-value of 0.017). This corresponds to an increase in practice exam completion of about 0.1 of a standard deviation, or almost six percent relative to the average number of practice exams completed by students in the Control group. From the first column we see that these results are quantitatively similar when we omit the controls for student characteristics.



(a) Number of practice exams completed for students in the Control group of the task-based goals experiment



(b) Effects of task-based goals on the number of practice exams completed

Notes: The effects shown in Panel (b) were estimated using OLS regressions of indicators of the student having completed at least X practice exams for $X \in \{1, .., 15\}$ on an indicator for the student having been randomly allocated to the Treatment group in the task-based goals experiment. The 95% confidence intervals are based on heteroskedasticity-consistent standard errors.

Figure 3: Effects of task-based goals on the pattern of task completion

All students in the task-based goals experiment		
	Number of practice exams completed	
	OLS	OLS
Effect of asking students to set task-based goals	0.479^{**}	0.491^{**}
	(0.208)	(0.205)
	[0.022]	[0.017]
Effect / (SD in Control group)	0.100	0.102
Mean of dependent variable in Control group	8.627	8.627
Controls for student characteristics	No	Yes
Observations	2,004	2,004

Notes: Both columns report OLS regressions of the number of practice exams completed during the course (out of a maximum of fifteen) on an indicator for the student having been randomly allocated to the Treatment group in the task-based goals experiment. 'SD in Control group' refers to the standard deviation of the dependent variable in the Control group. In the first column we do not control for student characteristics. In the second column we control for the student characteristics defined in Table A.1 in Web Appendix I: (i) letting $\mathcal{Q}$ denote the set containing indicators for the binary characteristics other than gender (race-based categories, advanced placement credit, Fall semester) and $\mathcal{Z}$ denote the set containing the non-binary characteristics (age, SAT score, high school GPA, first login time), we include $j \in \mathcal{Q}$, $k \in \mathcal{Z}$, $k \times l$ for $k \in \mathcal{Z}$ and $l \in \mathcal{Z}$, and $j \times k$ for $j \in \mathcal{Q}$ and $k \in \mathcal{Z}$; and (ii) we include gender together with gender interacted with every control variable defined in (i). Heteroskedasticity-consistent standard errors are shown in round brackets and two-sided p-values are shown in square brackets. *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests).

Table 3: Effects of task-based goals on the average level of task completion

As we discussed in the Introduction, evidence from other educational environments suggests that males have less self-control than females. This motivates splitting our analysis by gender to examine whether self-set task-based goals act as a more effective commitment device for male students than for females. In line with this existing evidence on gender differences in self-control, Table 4 shows that the effect of task-based goals is mainly confined to male students. We focus our discussion on the second column of results, which were obtained from OLS regressions that include controls for student characteristics (the first column of results shows that our findings are robust to omitting these controls). Panel I shows that task-based goals increased the number of practice exams that male students completed by about one exam. This corresponds to an increase in practice exam completion of about 0.2 of a standard deviation, or almost eleven percent relative to the average number of practice exams completed by male students in the Control group. This positive effect of task-based goals on the level of task completion for male students is statistically significant at the one-percent level. Panel II shows that for female students task-based goals increased the number of practice exams completed by less than 0.2 of an exam, and this effect is far from being statistically significant.

Interestingly, in the Control group female students completed more practice exams than males ($p = 0.000$), and the stronger effect for males of the task-based goals intervention ($p = 0.073$) eliminated most of the gender gap in practice exam completion. Specifically, in the Control group females completed seventeen percent more practice exams than males, while in the Treatment group females completed only seven percent more practice exams than males. Even though females completed more practice exams than males in the Control group, the average marginal effects reported in Table A.7 in Web Appendix I suggest that the marginal productivity of one extra practice exam was similar for males and females, and so it appears that females were not closer to the effort frontier.[20]

Panel I: Male students in the task-based goals experiment		
	Number of practice exams completed	
	OLS	OLS
Effect of asking students to set task-based goals	0.809**	0.893***
	(0.306)	(0.300)
	[0.016]	[0.006]
Effect / (SD in Control group)	0.172	0.190
Mean of dependent variable in Control group	7.892	7.892
Controls for student characteristics	No	Yes
Observations	918	918
Panel II: Female students in the task-based goals experiment		
	Number of practice exams completed	
	OLS	OLS
Effect of asking students to set task-based goals	0.217	0.156
	(0.281)	(0.281)
	[0.882]	[1.000]
Effect / (SD in Control group)	0.045	0.033
Mean of dependent variable in Control group	9.239	9.239
Controls for student characteristics	No	Yes
Observations	1,086	1,086

Notes: The regressions are the same as those reported in Table 3, except that we now split the sample by gender. Heteroskedasticity-consistent standard errors are shown in round brackets and two-sided Bonferonni-adjusted p-values are shown in square brackets. The Bonferonni adjustment accounts for the multiple null hypotheses being considered, i.e., zero treatment effect for men and zero treatment effect for women. *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests based on the Bonferonni-adjusted p-values).

Table 4: Gender differences in the effects of task-based goals on task completion

[20] The estimates of the effect on performance of completing one more practice exam presented in Table A.7 leverage within-student variation in the number of practice exams completed across the two midterms and the final. Since this variation was not experimentally induced, the estimates could be influenced by omitted variable bias; however, we have no evidence that any such bias varies by gender.

3.2 Impact of goals on student performance

We saw in Section 3.1 that task-based goal setting successfully increased the students' level of task completion. Table 5 provides evidence that asking students to set task-based goals also improved student performance in the course, while performance-based goals had only a small and statistically insignificant effect on performance.

Our measure of performance is a student's total points score in the course (out of one hundred) that determines her letter grade. The first and second columns of Table 5 report OLS and unconditional quantile (median) regressions of total points score on an indicator for the student having been randomly allocated to the Treatment group in the task-based goals experiment.[21] The third and fourth columns report OLS and unconditional quantile (median) regressions of total points score on an indicator for the student having been randomly allocated to the Treatment group in the performance-based goals experiment. To give a feel for the magnitude of the effects, the third row reports the effect size as a proportion of the standard deviation of the dependent variable in the relevant Control group, while the fourth row reports the average of the dependent variable in the same Control group. The regressions in Table 5 control for age, gender, race, SAT score, high school GPA, advanced placement credit, Fall semester, and first login time, including linear terms, squares, and interactions of these variables (see the notes to Table 5 for further details on the controls); the results are quantitatively similar but precision falls when we do not condition on student characteristics (see Table A.4 in Web Appendix I).[22]

The first and second columns of Table 5 report results from the task-based goals experiment: asking students to set goals for the number of practice exams to complete improved performance by a little under 0.1 of a standard deviation on average across the two specifications. The median regression gives significance at the five-percent level ($p = 0.019$), while the OLS regression gives significance at the ten-percent level. The tests are two-sided: using one-sided tests would give significance at the one-percent level for the median regression and at the five-percent level for the OLS regression.

The third and fourth columns of Table 5 report results from the performance-based goals experiment: the performance goal experiment shows a non-significant increase in performance. In more detail, asking students to set performance-based goals had positive but small and statistically insignificant effects on student performance in the course. The p-values are not close to the thresholds for statistical significance at conventional levels. Within the performance-based goals experiment, neither goals for letter grades in the course nor goals for scores in the two midterms and the final exam had a statistically significant effect on student performance.[23] For both experiments, we also find that treatment effects did not vary statistically significantly

[21] The median results were obtained using the estimator of Firpo et al. (2009), which delivers the effect of the treatment on the unconditional median of total points score.

[22] Table A.5 in Web Appendix I further shows that average treatment effects do not change when we interact treatment with indicators for SAT score bins (and include SAT score bin controls).

[23] For both specifications reported in the third and fourth columns of Table 5, and using the ten-percent-level criterion, we find no statistically significant effect of either type of performance-based goal, and we find no statistically significant difference between the effects of the two types of goal. For the case of OLS regressions of total points score on the treatment, the p-values for the two effects and the difference are, respectively, $p = 0.234$, $p = 0.856$, and $p = 0.386$.

across exams.[24]

	All students in the task-based goals experiment		All students in the performance-based goals experiment	
	Total points score		Total points score	
	OLS	Median	OLS	Median
Effect of asking students to set task-based goals	0.742^{*} (0.431) [0.086]	1.044^{**} (0.446) [0.019]		
Effect of asking students to set performance-based goals			0.300 (0.398) [0.452]	0.118 (0.459) [0.797]
Effect / (SD in Control group)	0.068	0.096	0.028	0.011
Mean of dependent variable in Control group	83.111	83.111	83.220	83.220
Observations	2,004	2,004	1,967	1,967

Notes: The first and second columns report OLS and unconditional quantile (median) regressions of total points score on an indicator for the student having been randomly allocated to the Treatment group in the task-based goals experiment. The third and fourth columns report OLS and unconditional quantile (median) regressions of total points score on an indicator for the student having been randomly allocated to the Treatment group in the performance-based goals experiment. Total points score (out of one hundred) determines a student's letter grade and is our measure of performance in the course; as explained in Section 2.2, only the maximum of the two midterm exam scores counts toward the total points score. 'SD in Control group' refers to the standard deviation of the dependent variable in the Control group. We control for the student characteristics defined in Table A.1 in Web Appendix I: (i) letting $\mathcal{Q}$ denote the set containing indicators for the binary characteristics other than gender (race-based categories, advanced placement credit, Fall semester) and $\mathcal{Z}$ denote the set containing the non-binary characteristics (age, SAT score, high school GPA, first login time), we include $j \in \mathcal{Q}$, $k \in \mathcal{Z}$, $k \times l$ for $k \in \mathcal{Z}$ and $l \in \mathcal{Z}$, and $j \times k$ for $j \in \mathcal{Q}$ and $k \in \mathcal{Z}$; and (ii) we include gender together with gender interacted with every control variable defined in (i). Heteroskedasticity-consistent standard errors are shown in round brackets and two-sided p-values are shown in square brackets. *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests).

Table 5: Effects of task-based goals and performance-based goals on student performance

In line with previous correlational studies (see the Introduction), we find that students who set ambitious performance-based goals performed better. Conditional on student characteristics, the correlation in our sample between course performance (measured by total number of points scored out of one hundred) and the level of the goal is 0.203 ($p = 0.000$) for students who set performance-based goals. The difference between the strong positive correlation based on non-experimental variation in our sample and the small and statistically insignificant causal effects that we estimate suggests that correlational analysis gives a misleading impression of the effectiveness of performance-based goals.

[24] Using the ten-percent-level criterion, the null hypothesis that there is no difference in the treatment effect on the first midterm exam, the second midterm exam and the final exam cannot be rejected for either experiment. For the effect of task-based goals on the number of practice exams completed, the joint test gives $p = 0.697$; for the effect of task-based goals on total points score, the joint test gives $p = 0.156$; and for the effect of performance-based goals on total points score, the joint test gives $p = 0.628$.

Table 6 repeats the analysis from Table 5 with the sample split by gender.[25] Consistent with our finding in Section 3.1 that task-based goal setting increased task completion only for males, the first and second columns of Table 6 show that task-based goals increased course performance for males but not for females. For male students task-based goals improved performance by over 0.15 of a standard deviation on average across the two specifications, which corresponds to an increase in performance of almost two points. The effects of task-based goal setting on the performance of male students are strongly statistically significant (p-values of 0.013 and 0.015). On the other hand, task-based goals were ineffective in raising performance for female students. On average across the two specifications, task-based goals improved the performance of female students by only 0.02 of a standard deviation, and the effect of task-based goals on the performance of female students is statistically insignificant. In the Control group in the task-based goals experiment, males performed slightly better ($p = 0.642$), and the stronger effect for males of the task-based goal intervention ($p = 0.028$) exacerbated this performance difference (these two p-values are from OLS regressions); thus task-based goal setting closed the gender gap in task completion (see Section 3.1), but increased the gender gap in performance. The third and fourth columns of Table 6 show that we continue to find statistically insignificant effects of performance-based goals on performance when we break the sample down by gender, and there is also no gender difference in the treatment effect ($p = 0.755$).

So far we have shown that task-based goals increased the level of task completion and improved student performance. The obvious explanation for our results is that the increase in task completion induced by task-based goal setting caused the improvement in student performance. A potential concern is that, instead, task-based goals increased students' general engagement in the course. However, we think this is unlikely for two reasons. First, it is hard to understand why only men would become more engaged. Second, we find that task-based goal setting did not affect course participation.[26]

[25] The regressions in Table 6 control for student characteristics. The results are quantitatively similar but precision falls when we do not condition on student characteristics (see Table A.6 in Web Appendix I).

[26] In more detail, we construct an index of course participation, which measures the proportion of course components that a student completed weighted by the importance of each component in determining total points score in the course. We regress our index of course participation on an indicator of the student having been randomly allocated to the Treatment group in the task-based goals experiment. We find that the effects of the treatment on course participation are small and far from being statistically significant. The p-values for OLS regressions of this index on the treatment are 0.668, 0.367 and 0.730 for, respectively, all students, male students, and female students.

	Male students in the task-based goals experiment		Male students in the performance-based goals experiment	
	Total points score		Total points score	
	OLS	Median	OLS	Median
Effect of asking students to set task-based goals	1.787** (0.657) [0.013]	1.714** (0.642) [0.015]		
Effect of asking students to set performance-based goals			0.430 (0.594) [0.937]	0.576 (0.618) [0.703]
Effect / (SD in Control group)	0.159	0.153	0.041	0.055
Mean of dependent variable in Control group	83.285	83.285	83.644	83.644
Observations	918	918	933	933
	Female students in the task-based goals experiment		**Female students in the performance-based goals experiment**	
	Total points score		Total points score	
	OLS	Median	OLS	Median
Effect of asking students to set task-based goals	-0.128 (0.571) [1.000]	0.449 (0.613) [0.929]		
Effect of asking students to set performance-based goals			0.181 (0.536) [1.000]	-0.330 (0.642) [1.000]
Effect / (SD in Control group)	-0.012	0.043	0.017	-0.031
Mean of dependent variable in Control group	82.966	82.966	82.864	82.864
Observations	1,086	1,086	1,034	1,034

Notes: The regressions are the same as those reported in Table 5, except that we now split the sample by gender. Heteroskedasticity-consistent standard errors are shown in round brackets and two-sided Bonferonni-adjusted p-values are shown in square brackets. The Bonferonni adjustment accounts for the multiple null hypotheses being considered, i.e., zero treatment effect for men and zero treatment effect for women. *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests based on the Bonferonni-adjusted p-values).

Table 6: Gender differences in the effects of task-based goals and performance-based goals on student performance

3.3 Benchmarking

In this section, we benchmark the results of our task-bask goals experiment against other experiments in the economics literature. To preview the results of this benchmarking exercise, our estimates are well within the range of those produced by these other experiments. This means that while our estimates are large enough to justify low-cost and scalable interventions, they are not especially large in relation to those found in the prior literature.

First, we benchmark the effects of our task-based goals intervention on the performance of college students by comparing them to prior estimates of the effects of instructor quality, class size, and financial incentives on college grades. As described above, we find that asking students to set task-based goals increased average total points scored in the course by 0.068 of a standard deviation ($p = 0.086$) and increased median total points scored by 0.096 of a standard deviation ($p = 0.019$). Carrell and West (2010) find that a one-standard-deviation increase in instructor quality increased GPA by 0.052 of a standard deviation ($p < 0.05$). Bandiera et al. (2010) find that a one-standard-deviation increase in class size decreased test scores by 0.108 of a standard deviation ($p < 0.01$). When benchmarking against the effects of financial incentives, we restrict attention to the studies listed in Table 1, Panel B (post-secondary education), of the survey by Lavecchia et al. (2016) for which effect sizes are reported in standard deviations. Angrist et al. (2009) find that GPA-based scholarships increased first-year GPA by 0.01 of a standard deviation ($p > 0.10$) and decreased second-year GPA by 0.02 of a standard deviation ($p > 0.10$). Angrist et al. (2009) also find that mentoring combined with a GPA-based scholarship increased first-year GPA by 0.23 of a standard deviation ($p < 0.05$) and increased second-year GPA by 0.08 of a standard deviation ($p > 0.10$). Angrist et al. (2014) find that financial incentives worth up to $1,000 per semester decreased first-year GPA by 0.021 of a standard deviation ($p > 0.10$) and increased second-year GPA by 0.107 of a standard deviation ($p > 0.10$). De Paola et al. (2012) find that performance-based prizes of $1,000 increased exam scores by 0.19 of a standard deviation ($p < 0.05$), while prizes of $350 increased scores by 0.16 of a standard deviation ($p < 0.10$).

Second, we benchmark the effects of our task-based goals intervention on task completion by comparing them to prior estimates of the effects of grading policies, financial incentives, and course format on class attendance. As described above, we find that asking students to set goals for the number of practice exams to complete increased the average number of practice exams completed by 0.102 of a standard deviation ($p = 0.017$). This effect is equivalent to an increase in practice exam completion of 5.691%. Marburger (2006) finds that providing students with credit for class attendance increased attendance by 11.475% ($p < 0.05$). De Paola et al. (2012) find that performance-based prizes of $1,000 increased attendance by 6.145% ($p > 0.10$), while prizes of $350 decreased attendance by 2.509% ($p > 0.10$). Joyce et al. (2015) find the moving from a traditional lecture-based course format to a hybrid course format that combined lectures with online material increased attendance by 1.150% ($p > 0.10$).

4 Using a theoretical framework to interpret our findings

4.1 Motivation

In this section we suggest some hypotheses for our findings in the context of a theoretical framework. Web Appendix III formalizes the discussion and provides further references. Our aim is not to test theory; rather, we use the theoretical framework to guide the analysis and interpretation of our findings.

Our theoretical framework builds on Koch and Nafziger (2011) and is inspired by two key concepts in behavioral economics: present bias and loss aversion. The concept of present bias captures the idea that people lack self-control because they place a high weight on current utility (Strotz, 1956). More specifically, a present-biased discounter places more weight on current utility relative to utility n periods in the future than she does on utility at future time t relative to utility at time $t+n$. This implies that present-biased discounters exhibit time inconsistency, since their time preferences at different dates are not consistent with one another. Present bias has been proposed as an explanation for aspects of many behaviors such as addiction and credit card borrowing (e.g., Gruber and Kőszegi, 2001, Khwaja et al., 2007, Fang and Silverman, 2009, Meier and Sprenger, 2010). In the context of education, a present-biased student might set out to exert her preferred level of effort, but when the time comes to attend class or review for a test she might lack the self-control necessary to implement these plans.[27]

The concept of loss aversion captures the idea that people dislike falling behind a salient reference point (Kahneman and Tversky, 1979). Loss aversion has been proposed as a foundation of a number of phenomena such as the disposition effect and the role of expectations in decision-making (e.g., Genesove and Mayer, 2001, Kőszegi and Rabin, 2006, Gill and Stone, 2010, Gill and Prowse, 2012). In the context of education, a loss-averse student might work particularly hard in an attempt to achieve a salient reference point (e.g., a particular grade in her course).

Together, the literatures on present bias and loss aversion suggest that self-set goals might serve as an effective commitment device. Specifically, self-set goals might act as salient reference points, helping present-biased agents to mitigate their self-control problem and so steer their effort toward its optimal level. Indeed, Koch and Nafziger (2011) developed a model of goal setting based on this idea that we build on here, but unlike us they did not explore the effectiveness of different types of goals (Heath et al., 1999, proposed that goals could act as reference points, but they did not make the connection to present bias).[28]

4.2 Performance-based goal setting

4.2.1 Theoretical framework

We start by describing a theoretical framework that captures performance-based goal setting. In Section 4.2.2 we use the framework to suggest three hypotheses for why performance-based goals might not be very effective in the context that we studied.

At period one the student chooses a goal for performance; we call the student at period

[27] Under standard (i.e., exponential) discounting this self-control problem disappears.

[28] Related theoretical work on goal setting includes Suvorov and Van de Ven (2008), Wu et al. (2008), Jain (2009), Hsiaw (2013), Hsiaw (2016) and Koch and Nafziger (2016).

one the *student-planner*. At period two the student chooses how much effort to exert; we call the student at period two the *student-actor*. At period three performance is realized and the student incurs any disutility from failing to achieve her goal; we call the student at period three the *student-beneficiary*. Performance increases linearly in effort exerted by the student-actor at period two, and the disutility from effort is quadratic in effort. The student-beneficiary is loss averse around her goal: she suffers goal disutility that depends linearly on how far performance falls short of the goal set by the student-planner at period one.

The student is present biased. In particular, the student exhibits quasi-hyperbolic discounting: the student discounts utility n periods in the future by a factor $\beta\delta^n$.[29] Under quasi-hyperbolic discounting the student-planner discounts period-two utility by a factor $\beta\delta$ and period-three utility by a factor $\beta\delta^2$, and so discounts period-three utility by δ relative to period-two utility. The student-actor, on the other hand, discounts period-three utility by $\beta\delta$ relative to immediate period-two utility. Since $\beta\delta < \delta$, the student-planner places more weight on utility from performance at period three relative to the cost of effort at period two than does the student-actor.

As a result of this present bias, and in the absence of a goal, the student-planner's desired effort is higher than the effort chosen by the student-actor: that is, the student exhibits a self-control problem due to time inconsistency. To alleviate her self-control problem, the student-planner chooses to set a goal. Goals work by increasing the student-actor's marginal incentive to work in order to avoid the goal disutility that results from failing to achieve the goal. The optimal goal induces the student to work harder than she would in the absence of a goal

4.2.2 Why might performance-based goals not be very effective?

This theoretical framework suggests that performance-based goals can improve course performance. However, our experimental data show that performance-based goals had a positive but small and statistically insignificant effect on student performance (Table 5). In our view, the theoretical framework sketched above suggests three hypotheses for why performance-based goals might not be very effective in the context that we studied (we view these hypotheses as complementary).

Timing of goal disutility

In the theoretical framework, the student works in period two and experiences any goal disutility from failing to achieve her performance-based goal in period three (i.e., when performance is realized). This temporal distance will dampen the motivating effect of the goal. Even when the temporal distance between effort and goal disutility is modest, the timing of goal disutility dampens the effectiveness of performance-based goals because quasi-hyperbolic discounters discount the near future relative to the present by a factor β even if $\delta \approx 1$ over the modest temporal distance.

[29] Laibson (1997) was the first to apply the analytically tractable quasi-hyperbolic (or 'beta-delta') model of discounting to analyze the choices of present-biased time-inconsistent agents.

Overconfidence

In the theoretical framework, students understand perfectly the relationship between effort and performance. In contrast, the education literature suggests that students face considerable uncertainty about the educational production function, and that this uncertainty could lead to students holding incorrect beliefs about the relationship between effort and performance (e.g., Romer, 1993, and Fryer, 2013). Furthermore, the broader behavioral literature shows that people tend to be overconfident when they face uncertainty (e.g., Weinstein, 1980, Camerer and Lovallo, 1999, Park and Santos-Pinto, 2010). In light of these two strands of literature, suppose that some students are overconfident in the sense that they overestimate how effort translates into performance (and hence think that they need to do less preparation than they actually have to). For an overconfident student, actual performance with goal setting and in the absence of a goal will be a fraction of that expected by the student. As a result, this type of overconfidence reduces the impact of performance-based goal setting on performance.[30]

Performance uncertainty

In the theoretical framework described above, the student knows for sure how her effort translates into performance (i.e., the relationship between effort and performance involves no uncertainty). In practice, the relationship between effort and performance is likely to be noisy. The student could face uncertainty about her own ability or about the productivity of work effort. The student might also get unlucky: for instance, the draw of questions on the exam might be unfavorable or the student might get sick near the exam.

To introduce uncertainty about performance in a straightforward way, suppose that with known probability performance falls to some baseline level (since we assume that this probability is known, the student is neither overconfident nor underconfident).[31] The uncertainty directly reduces the student-actor's marginal incentive to exert effort, which reduces both the student's goal and her choice of effort with and without goal setting. However, this reduction in the expected value of effort is not the only effect of uncertainty: performance-based goals also become risky because when performance turns out to be low the student fails to achieve her performance-based goal and so suffers goal disutility that increases in the goal.[32] Anticipating the goal disutility suffered when performance turns out to be low, the student-planner further scales back the performance-based goal that she sets for the student-actor, which reduces the effectiveness of performance-based goal setting.[33]

[30] A naive student who does not understand her present bias would be overconfident about her level of effort. However, such a student would not understand how to use goals to overcome her lack of self-control, and so our discussion focuses on sophisticated students who understand their present bias.

[31] We can think of this baseline level as the performance that the student achieves with little effort even in the absence of goal setting.

[32] It is this second effect that drives the prediction that uncertainty reduces the effectiveness of performance-based goal setting. If we assumed that only the variance of performance changed, this second effect would still operate, but the formal analysis in Web Appendix III would become substantially more involved.

[33] This scaling back of goals is not necessarily at odds with the fact that the performance-based goals that we see in the data appear ambitious. First, the goal will appear ambitious relative to average achievement because, as noted above, when performance turns out to be low the student fails to achieve her goal. Second, without any scaling back the goals might have been even higher. Third, the overconfidence that we discuss above could keep the scaled-back goal high. Fourth, we explain in Web Appendix III.3.4 that students likely report as their goal an 'aspiration' that is only relevant if, when the time comes to study, the cost of effort turns out to be particularly low: the actual cost-specific goal that the student aims to hit could be much lower than this aspiration.

4.3 Task-based goal setting

4.3.1 Theoretical framework

We now extend our theoretical framework to task-based goal setting. At period one the student-planner chooses a goal for the number of units of the task to complete. At period two the student-actor chooses the level of task completion, and the loss-averse student-actor suffers goal disutility that depends linearly on how far the level of task completion falls short of the goal set by the student-planner at period one. At period three performance is realized. Performance increases linearly in the level of task completion, and the disutility from task completion is quadratic in the level of task completion.

The present-biased student exhibits quasi-hyperbolic discounting as described in Section 4.2.1. In the absence of a goal the present-biased student exhibits a self-control problem due to time inconsistency: the student-actor chooses a level of task completion that is smaller than the student-planner's desired level of task completion. As a result, the student-planner chooses to set a goal to alleviate her self-control problem. The optimal goal increases the level of task completion above the level without a goal, which in turn improves course performance.

4.3.2 Why were task-based goals effective?

Our experimental data show that task-based goals improved task completion and course performance (see Table 3 for the effect on task completion and Table 5 for the effect on course performance).[34] How might we account for these findings, given our discussion of why performance-based goals might not be very effective? In our view, an obvious answer is that with task-based goal setting, the three factors that reduce the effectiveness of performance-based goals (Section 4.2.2) are of lesser importance or do not apply at all.

Timing of goal disutility

In the case of task-based goal setting, any goal disutility from failing to achieve the task-based goal is suffered immediately when the student stops working on the task in period two. Thus, unlike the case of performance-based goal setting discussed in Section 4.2.2, there is no temporal distance that dampens the motivating effect of the goal.

Overconfidence

As discussed in Section 4.2.2, overconfident students overestimate how effort translates into performance, which reduces the effectiveness of goal setting. In the case of task-based goal setting, this effect is mitigated if practice exams direct students toward productive tasks. Plausibly, teachers have better information about which tasks are likely to be productive, and asking

[34] It is possible that some students in the Control group (who were not invited to set goals) might already use goals as a commitment device. However, since we find that task-based goals are successful at increasing performance, we conclude that many students in the Control group did not use goals or set goals that were not fully effective. We note that asking students to set goals might make the usefulness of goal setting as a commitment device more salient and so effective. Reminding students of their goal, as we did, might also help to make them more effective.

students to set goals for productive tasks is one way to improve the power of goal setting for overconfident students.[35]

Performance uncertainty

Even with uncertainty about performance, the student faces no uncertainty about the level of task completion because the student-actor controls the number of units of the task that she completes. Thus, unlike the case of performance-based goals with uncertainty, the student has no reason to scale back her task-based goal to reduce goal disutility in the event that the goal is not reached.

4.3.3 Why were task-based goals more effective for men than for women?

Our data show that task-based goals are more effective for men than for women. More specifically: in the Control group without goal setting men completed fewer practice exams than women (Table 4); and task-based goals increased performance and the number of practice exams completed more for men than for women (Tables 6 and 4 respectively). In the context of our theoretical framework, a higher degree of present bias among men can explain both of these findings, and existing empirical evidence supports the idea that men have less self-control and are more present biased than women (see Web Appendix V.6 for a survey of this evidence).[36]

4.3.4 Saliency of the task

If practice exams were less salient in the performance-based goals experiment, and if goals work better when students have access to salient practice exams, then the lower saliency could help to explain why task-based goals were effective, while performance-based goals were not. There were some differences in the practice exams across experiments (most notably, practice exams had to be downloaded in the performance-based goals experiment, while they could be completed online in the task-based goals experiment; see the penultimate paragraph of Section 2.2). However, we do not think that a difference in saliency was important, for three reasons. First, in both experiments the first page of the course syllabus highlighted the practice exams, and the syllabus quiz at the start of each semester made the syllabus itself salient. Second, analysis of the course evaluations shows that students mentioned that the practice exams were helpful at a similar rate

[35] Instead of improving the power of goal setting by directing overconfident students toward productive tasks, it is conceivable that task-based goals improved performance via another channel: signaling to students in the Treatment group that practice exams were an effective task. But we think this is highly unlikely. First, we were careful to make the practice exams as salient as possible to the Control group. Second, students in the Control group in fact completed many practice exams. Third, it is hard to understand why only men would respond to the signal.

[36] Two alternative explanations for the gender differences that we find seem inconsistent with our data. The first alternative explanation is based on the idea that women are closer to the effort frontier. However, we report that the marginal productivity of practice exams was similar by gender (see Section 3.1). The second alternative explanation posits that because women perform worse in higher stakes environments (Ors et al., 2013, Azmat et al., 2016), the high stakes might make women care less about completing more practice exams in response to goal setting. However, if high stakes make women care less about completing more practice exams in response to goal setting, we should also expect the stakes to make women care less about completing practice exams in the Control group (where practice exams are also salient), but in fact our data show that women complete more practice exams in the control.

in the two experiments.[37] Third, course performance in the Control groups was almost identical across the two experiments (Table 5), which suggests that any difference in the saliency of the practice exams was not an important determinant of performance.

[37]In the performance-based goals experiment, 3.2 percent of the 557 students who made comments mentioned that the practice exams were helpful; in the task-based goals experiment 2.8 per cent of 532 did so. We do not have data on practice exam downloads in the performance-based goals experiment.

5 Conclusion

Our experimental findings suggest that task-based goal setting is an intervention that can improve college outcomes: asking students to set goals for the number of practice exams to complete increased the number of practice exams that students completed and increased course performance. We emphasize that our task-based goal worked because goal setting directed students toward a productive task. In an educational context, teachers should pair goal setting with tasks that they think are productive, while policymakers should disseminate new knowledge to teachers about which tasks work well with goals.

One of the challenges in applying our findings to other settings is that policymakers often do not know the production function for course performance. Future research that examines the effects of self-set goals for other tasks such as attending class, contributing to online discussions, or working through textbook chapters, can advance our knowledge of the production function. Specifically, if task-based goal setting increases course performance only through the effects of goal setting on task-specific investments, then assignment to the goals treatment is an instrument that can be used to identify the performance effects of these investments.[38]

As well as measuring the effectiveness of different tasks, it would also be interesting to conduct similar goal-setting experiments in other types of colleges and educational environments. For example, our subjects (who attend a four-year college) are likely more able than two-year college students. If they also possess more self-control than these two-year college students, then goal setting might be more effective at two-year colleges.

The most direct way to incorporate task-based goals into the college environment would be for instructors to design courses that promote task-based goal setting. For example, in a course that required students to complete certain tasks online (e.g., homework assignments or class discussion), the opportunity to set goals could be built into the technology used to deliver these course components. Academic advising services could also give greater prominence to task-based goal setting and encourage students to set task-based goals in consultation with course instructors who can give advice about the tasks most likely to be productive.[39] Ideally, this advice would rest on a solid base of evidence.

To summarize, we believe that our study marks an important step toward a better understanding of the role that self-set goals could play in motivating college students to work harder and perform better. Research in psychology and economics provides reason to expect that college students, like other agents in the economy, will lack self-control. Our results suggest that self-set goals can act as an effective commitment device that helps college students to self-regulate behavior and mitigate these self-control problems. Provided that students set goals for productive tasks, task-based goal setting can also improve student performance. Since task-based goal setting could easily be incorporated into the college environment, our findings have important implications for educational practice. As noted above, future research should probe the effects of task-based goal setting in other contexts and for other tasks.

[38] There is already a small literature on the performance effects of attending class. For example, Dobkin et al. (2010) and Arulampalam et al. (2012) exploit quasi-experiments to estimate the effects of attendance on college course performance.

[39] CUNY's Accelerated Study in Associate Programs (ASAP) encourages first-year students to think about goal setting as one of many strategies that students might try (Scrivener et al., 2015).

References

References for the main text and the Web Appendix.

Akın, Z. and **Karagözoğlu, E.** (forthcoming). The role of goals and feedback in incentivizing performance. *Managerial and Decision Economics*

Allan, B.M. and **Fryer, R.G.** (2011). The power and pitfalls of education incentives. *Policy Paper, Hamilton Project*

Angrist, J., **Lang, D.**, and **Oreopoulos, P.** (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 1(1): 136–163

Angrist, J. and **Lavy, V.** (2009). The effects of high stakes high school achievement awards: Evidence from a randomized trial. *American Economic Review*, 99(4): 1384–1414

Angrist, J., **Oreopoulos, P.**, and **Williams, T.** (2014). When opportunity knocks, who answers? New evidence on college achievement awards. *Journal of Human Resources*, 49(3): 572–610

Ariely, D. and **Wertenbroch, K.** (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, 13(3): 219–224

Arulampalam, W., **Naylor, R.A.**, and **Smith, J.** (2012). Am I missing something? The effects of absence from class on student performance. *Economics of Education Review*, 31(4): 363–375

Ashraf, N., **Karlan, D.**, and **Yin, W.** (2006). Tying Odysseus to the mast: Evidence from a commitment savings product in the Philippines. *Quarterly Journal of Economics*, 121(2): 635–672

Augenblick, N., **Niederle, M.**, and **Sprenger, C.** (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics*, 130(3): 1067–1115

Azmat, G., **Calsamiglia, C.**, and **Iriberri, N.** (2016). Gender differences in response to big stakes. *Journal of the European Economic Association*, 14(6): 1372–1400

Bandiera, O., **Larcinese, V.**, and **Rasul, I.** (2010). Heterogeneous class size effects: New evidence from a panel of university students. *Economic Journal*, 120(549): 1365–1398

Barron, K.E. and **Harackiewicz, J.M.** (2003). Revisiting the benefits of performance-approach goals in the college classroom: Exploring the role of goals in advanced college courses. *International Journal of Educational Research*, 39(4): 357–374

Barrow, L., **Richburg-Hayes, L.**, **Rouse, C.E.**, and **Brock, T.** (2014). Paying for performance: The education impacts of a community college scholarship program for low-income adults. *Journal of Labor Economics*, 32(3): 563–599

Bauer, M. and **Chytilová, J.** (2013). Women, children and patience: Experimental evidence from Indian villages. *Review of Development Economics*, 17(4): 662–675

Bell, D.E. (1985). Disappointment in decision making under uncertainty. *Operations Research*, 33(1): 1–27

Brunnermeier, M.K. and **Parker, J.A.** (2005). Optimal expectations. *American Economic Review*, 95(4): 1092–1118

Bryan, G., **Karlan, D.**, and **Nelson, S.** (2010). Commitment devices. *Annual Review of Economics*, 2(1): 671–698

Buechel, B., **Mechtenberg, L.**, and **Petersen, J.** (2014). Peer effects and students' self-control. *SFB 649 Discussion Paper 2014-024, Humboldt University*

Camerer, C. and **Lovallo, D.** (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review*, 89(1): 306–318

Cameron, J. and **Pierce, W.D.** (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational research*, 64(3): 363–423

Carrell, S.E. and **West, J.E.** (2010). Does professor quality matter? Evidence from random assignment of students to professors. *Journal of Political Economy*, 118(3): 409–432

Castleman, B.L. (2014). The impact of partial and full merit scholarships on college entry and success: Evidence from the Florida Bright Futures Scholarship program. *EdPolicyWorks Working Paper, University of Virginia*

Cha, P. and **Patel, R.** (2010). Rewarding progress, reducing debt: Early results from Ohio's performance-based scholarship demonstration for low-income parents. *Technical Report, MDRC*

Chase, J.A., **Houmanfar, R.**, **Hayes, S.C.**, **Ward, T.A.**, **Vilardaga, J.P.**, and **Follette, V.** (2013). Values are not just goals: Online ACT-based values training adds to goal setting in improving undergraduate college student performance. *Journal of Contextual Behavioral Science*, 2(3): 79–84

Choi, J.J., **Haisley, E.**, **Kurkoski, J.**, and **Massey, C.** (2016). Small cues change savings choices. *Mimeo, Yale University*

Church, M.A., **Elliot, A.J.**, and **Gable, S.L.** (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology*, 93(1): 43–54

Cohodes, S.R. and **Goodman, J.S.** (2014). Merit aid, college quality, and college completion: Massachusetts' Adams scholarship as an in-kind subsidy. *American Economic Journal: Applied Economics*, 6(4): 251–285

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 **Hernán-Gonzalez, R.** (2016). Goal setting in the principal-agent model: Weak incentives for strong performance. *CEDEX Discussion Paper 2016-09, University of Nottingham*

Cornwell, C.M., **Lee, K.H.**, and **Mustard, D.B.** (2005). Student responses to merit scholarship retention rules. *Journal of Human Resources*, 40(4): 895–917

Darnon, C., **Butera, F.**, **Mugny, G.**, **Quiamzade, A.**, and **Hulleman, C.S.** (2009). Too complex for me! Why do performance-approach and performance-avoidance goals predict exam performance? *European Journal of Psychology of Education*, 24(4): 423–434

De Paola, M., **Scoppa, V.**, and **Nisticò, R.** (2012). Monetary incentives and student achievement in a depressed labor market: Results from a randomized experiment. *Journal of Human Capital*, 6(1): 56–85

DellaVigna, S.D. and **Malmendier, U.** (2006). Paying not to go to the gym. *American Economic Review*, 96(3): 694–719

Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford: Oxford University Press

Diamond, P. and **Kőszegi, B.** (2003). Quasi-hyperbolic discounting and retirement. *Journal of Public Economics*, 87(9): 1839–1872

Dobkin, C., **Gil, R.**, and **Marion, J.** (2010). Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment. *Economics of Education Review*, 29(4): 566–575

Duckworth, A.L., **Quinn, P.D.**, and **Tsukayama, E.** (2012). What No Child Left Behind leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, 104(2): 439–451

Duckworth, A.L. and **Seligman, M.E.** (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16(12): 939–944

Duckworth, A.L., **Shulman, E.P.**, **Mastronarde, A.J.**, **Patrick, S.D.**, **Zhang, J.**, and **Druckman, J.** (2015). Will not want: Self-control rather than motivation explains the female advantage in report card grades. *Learning and Individual Differences*, 39: 13–23

Dynarski, S. (2008). Building the stock of college-educated labor. *Journal of Human Resources*, 43(3): 576–610

Elliot, A.J. and **McGregor, H.A.** (2001). A 2$\times$ 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80(3): 501–519

Fang, H. and **Silverman, D.** (2009). Time-inconsistency and welfare program participation: Evidence from the NLSY. *International Economic Review*, 50(4): 1043–1077

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

Fryer, R.G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *Quarterly Journal of Economics*, 126(4): 1755–1798

Fryer, R.G. (2013). Information and student achievement: Evidence from a cellular phone experiment. *Working Paper 19113, NBER*

Genesove, D. and **Mayer, C.** (2001). Loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics*, 116(4): 1233–1260

Gill, D. and **Prowse, V.** (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review*, 102(1): 469–503

Gill, D. and **Stone, R.** (2010). Fairness and desert in tournaments. *Games and Economic Behavior*, 69(2): 346–364

Giné, X., **Goldberg, J.**, **Silverman, D.**, and **Yang, D.** (forthcoming). Revising commitments: Field evidence on the adjustment of prior choices. *Economic Journal*

Gneezy, U., **Meier, S.**, and **Rey-Biel, P.** (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, 25(4): 191–209

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

Gneezy, U. and **Rustichini, A.** (2004). Gender and competition at a young age. *American Economic Review: Papers and Proceedings*, 94(2): 377–381

Goerg, S.J. (2015). Goal setting and worker motivation. *IZA World of Labor*, 178: 1–10

Goerg, S.J. and **Kube, S.** (2012). Goals (th) at work–goals, monetary incentives, and workers' performance. *Preprint 2012/19, Max Planck Institute for Research on Collective Goods*

Gossner, O. and **Steiner, J.** (2016). Optimal illusion of control and related perception biases. *Discussion Paper 11478, CEPR*

Griffee, D.T. and **Templin, S.A.** (1997). Goal setting affects task performance. In *Proceedings of the JALT 1997 International Conference on Language Teaching and Learning*, 21–26

Gruber, J. and **Kőszegi, B.** (2001). Is addiction rational? Theory and evidence. *Quarterly Journal of Economics*, 116(4): 1261–1303

Harackiewicz, J.M., **Barron, K.E.**, **Carter, S.M.**, **Lehto, A.T.**, and **Elliot, A.J.** (1997). Predictors and consequences of achievement goals in the college classroom: Maintaining interest and making the grade. *Journal of Personality and Social Psychology*, 73(6): 1284–1295

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

Harrison, G.W., **Lau, M.I.**, and **Rutström, E.E.** (2005). Dynamic consistency in Denmark: A longitudinal field experiment. *Economics Working Paper 05-02, University of Central Florida*

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

Henry, G.T., **Rubenstein, R.**, and **Bugler, D.T.** (2004). Is HOPE enough? Impacts of receiving and losing merit-based financial aid. *Educational Policy*, 18(5): 686–709

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

Hsiaw, A. (2016). Goal bracketing and self-control. *Mimeo, Brandeis University*

Hsieh, P., **Sullivan, J.R.**, and **Guerra, N.S.** (2007). A closer look at college students: Self-efficacy and goal orientation. *Journal of Advanced Academics*, 18(3): 454–476

Jain, S. (2009). Self-control and optimal goals: A theoretical analysis. *Marketing Science*, 28(6): 1027–1045

Johnson, D.D. and **Fowler, J.H.** (2011). The evolution of overconfidence. *Nature*, 477(7364): 317–320

Joyce, T., **Crockett, S.**, **Jaeger, D.A.**, **Altindag, O.**, and **O'Connell, S.D.** (2015). Does classroom time matter? *Economics of Education Review*, 46: 64–77

Kahneman, D., **Knetsch, J.L.**, and **Thaler, R.H.** (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of Political Economy*, 98(6): 1325–1348

Kahneman, D. and **Tversky, A.** (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2): 263–291

Kaur, S., **Kremer, M.**, and **Mullainathan, S.** (2015). Self-control at work. *Journal of Political Economy*, 123(6): 1227–1277

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

Khwaja, A., **Silverman, D.**, and **Sloan, F.** (2007). Time preference, time discounting, and smoking decisions. *Journal of Health Economics*, 26(5): 927–949

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

Koch, A.K. and **Nafziger, J.** (2016). Goals and bracketing under mental accounting. *Journal of Economic Theory*, 162: 305–351

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2): 443–477

LaPorte, R.E. and **Nath, R.** (1976). Role of performance goals in prose learning. *Journal of Educational Psychology*, 68(3): 260–264

Latham, G.P. and **Brown, T.C.** (2006). The effect of learning vs. outcome goals on self-efficacy, satisfaction and performance in an MBA program. *Applied Psychology*, 55(4): 606–623

Latham, G.P. and **Pinder, C.C.** (2005). Work motivation theory and research at the dawn of the twenty-first century. *Annual Review of Psychology*, 56(1): 485–516

Lavecchia, A.M., **Liu, H.**, and **Oreopoulos, P.** (2016). Behavioral economics of education: Progress and possibilities. In E.A. Hanushek, S. Machin, and L. Woessman, editors, *Handbook of the Economics of Education, Vol. 5*, 1–74. North Holland

Leuven, E., **Oosterbeek, H.**, and **van der Klaauw, B.** (2010). The effect of financial rewards on students' achievement: Evidence from a randomized experiment. *Journal of the European Economic Association*, 8(6): 1243–1265

Levitt, S.D., **List, J.A.**, **Neckermann, S.**, and **Sadoff, S.** (2011). The impact of short-term incentives on student performance. *Mimeo, University of Chicago*

Levitt, S.D., **List, J.A.**, **Neckermann, S.**, and **Sadoff, S.** (2016). The behavioralist goes to school: Leveraging behavioral economics to improve educational performance. *American Economic Journal: Economic Policy*, 8(4): 183–219

Linnenbrink-Garcia, L., **Tyson, D.F.**, and **Patall, E.A.** (2008). When are achievement goal orientations beneficial for academic achievement? A closer look at main effects and moderating factors. *Revue Internationale de Psychologie Sociale*, 21(1): 19–70

Locke, E.A. (1968). Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance*, 3(2): 157–189

Locke, E.A. and **Latham, G.P.** (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9): 705–717

Locke, E.A., **Shaw, K.N.**, **Saari, L.M.**, and **Latham, G.P.** (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin*, 90(1): 125–152

Lusher, L. (2016). College Better: Parimutuel betting markets as a commitment device and monetary incentive. *Natural Field Experiments 561, www.fieldexperiments.com*

Marburger, D.R. (2006). Does mandatory attendance improve student performance? *Journal of Economic Education*, 37(2): 148–155

McLeish, K.N. and **Oxoby, R.J.** (2007). Gender, affect and intertemporal consistency: An experimental approach. *Discussion Paper 2663, IZA*

Meier, S. and **Sprenger, C.** (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, 2(1): 193–210

Mento, A.J., **Steel, R.P.**, and **Karren, R.J.** (1987). A meta-analytic study of the effects of goal setting on task performance: 1966–1984. *Organizational Behavior and Human Decision Processes*, 39(1): 52–83

Miller, C., **Binder, M.**, **Harris, V.**, and **Krause, K.** (2011). Staying on track: Early findings from a performance-based scholarship program at the university of New Mexico. *Technical Report, MDRC*

Morgan, M. (1987). Self-monitoring and goal setting in private study. *Contemporary Educational Psychology*, 12(1): 1–6

Morisano, D., **Hirsh, J.B.**, **Peterson, J.B.**, **Pihl, R.O.**, and **Shore, B.M.** (2010). Setting, elaborating, and reflecting on personal goals improves academic performance. *Journal of Applied Psychology*, 95(2): 255–264

Ors, E., **Palomino, F.**, and **Peyrache, E.** (2013). Performance gender gap: Does competition matter? *Journal of Labor Economics*, 31(3): 443–499

Park, Y.J. and **Santos-Pinto, L.** (2010). Overconfidence in tournaments: Evidence from the field. *Theory and Decision*, 69(1): 143–166

Patel, R. and **Rudd, T.** (2012). Can scholarships alone help students succeed? Lessons from two New York City community colleges. *Technical Report, MDRC*

Patterson, R.W. (2016). Can behavioral tools improve online student outcomes? Experimental evidence from a massive open online course. *Mimeo, United States Military Academy at West Point*

Pollak, R.A. (1968). Consistent planning. *Review of Economic Studies*, 35(2): 201–208

Prince, J.T. and **Shawhan, D.L.** (2011). Is time inconsistency primarily a male problem? *Applied Economics Letters*, 18(6): 501–504

Rabin, M. (2000). Risk aversion and expected-utility theory: A calibration theorem. *Econometrica*, 68(5): 1281–1292

Romer, D. (1993). Do students go to class? Should they? *Journal of Economic Perspectives*, 7(3): 167–174

Schunk, D.H. (1983). Developing children's self-efficacy and skills: The roles of social comparative information and goal setting. *Contemporary Educational Psychology*, 8(1): 76–86

Schunk, D.H. (1984). Enhancing self-efficacy and achievement through rewards and goals: Motivational and informational effects. *Journal of Educational Research*, 78(1): 29–34

Schunk, D.H. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Research Journal*, 33(2): 359–382

Schunk, D.H. and **Ertmer, P.A.** (1999). Self-regulatory processes during computer skill acquisition: Goal and self-evaluative influences. *Journal of Educational Psychology*, 91(2): 251–260

Schunk, D.H. and **Rice, J.M.** (1991). Learning goals and progress feedback during reading comprehension instruction. *Journal of Literacy Research*, 23(3): 351–364

Schunk, D.H. and **Swartz, C.W.** (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology*, 18(3): 337–354

Schutz, P.A. and **Lanehart, S.L.** (1994). Long-term educational goals, subgoals, learning strategies use and the academic performance of college students. *Learning and Individual Differences*, 6(4): 399–412

Scott-Clayton, J. (2011). On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *Journal of Human Resources*, 46(3): 614–646

Scrivener, S., **Weiss, M.J.**, **Ratledge, A.**, **Rudd, T.**, **Sommo, C.**, and **Fresques, H.** (2015). Doubling graduation rates: Three-year effects of CUNY's Accelerated Study in Associate Programs (ASAP) for developmental education students. *Technical Report, MDRC*

Shih, S.S. and **Alexander, J.M.** (2000). Interacting effects of goal setting and self- or other-referenced feedback on children's development of self-efficacy and cognitive skill within the Taiwanese classroom. *Journal of Educational Psychology*, 92(3): 536–543

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

Strotz, R.H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23(3): 165–180

Suvorov, A. and **Van de Ven, J.** (2008). Goal setting as a self-regulation mechanism. *Working Paper 122, CEFIR NES*

Tanaka, T., **Camerer, C.F.**, and **Nguyen, Q.** (2010). Risk and time preferences: Linking experimental and household survey data from Vietnam. *American Economic Review*, 100(1): 557–571

Thaler, R.H. and **Benartzi, S.** (2004). Save more tomorrow^{TM}: Using behavioral economics to increase employee saving. *Journal of Political Economy*, 112(S1): S164–S187

Weinstein, N.D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5): 806–820

Wertenbroch, K. (1998). Consumption self-control by rationing purchase quantities of virtue and vice. *Marketing Science*, 17(4): 317–337

Wu, G., **Heath, C.**, and **Larrick, R.** (2008). A prospect theory model of goal behavior. *Mimeo, University of Chicago*

Zimmerman, B.J. and **Bandura, A.** (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, 31(4): 845–862

Web Appendix

(Intended for Online Publication)

Web Appendix I Tables

	Mean value		Treatment-Control difference		
	Treatment group	Control group	Difference	S.E.	p-value
Age	0.005	-0.005	0.010	0.032	0.764
Male	0.477	0.455	0.022	0.016	0.171
Black	0.064	0.051	0.012	0.007	0.091
Non-Hispanic white	0.604	0.619	-0.015	0.015	0.328
Hispanic	0.193	0.192	0.000	0.013	0.984
Asian	0.102	0.092	0.009	0.009	0.328
SAT score	0.001	-0.001	0.002	0.032	0.945
High school GPA	-0.016	0.016	-0.032	0.032	0.320
Advanced placement credit	0.759	0.756	0.003	0.014	0.800
Fall semester	0.620	0.607	0.012	0.015	0.435
First login time	-0.004	0.004	-0.007	0.032	0.820

Notes: The Treatment and Control groups contain 1,979 and 1,992 students respectively. Information about age, gender, race, SAT scores, high school GPA and advanced placement credit was obtained from the university's Registrar data. Age is measured on the first of the month in which the semester began and is rounded down to the nearest whole month. The variable SAT score is the sum of the student's scores on the verbal, analytic and numerical components of the primary aptitude test in the Registrar data (these are SAT scores for the majority of students). Advanced placement credit is an indicator for the student having entered the university with advanced placement credit. Fall semester is an indicator for the student having participated in the course in the Fall semester. First login time is the elapsed time between when the first email invitation to take the syllabus quiz was sent and when the student first logged into the course webpage. Each of the non-binary characteristics (age, SAT score, high School GPA and first login time) has been standardized to have a mean of zero and a variance of one within the Fall 2013 and Spring 2014 semesters combined (the performance-based goals experiment) and within the Fall 2014 and Spring 2015 semesters combined (the task-based goals experiment). The standardization of SAT score is stratified to ensure that this variable has the same mean and the same variance among students taking each type of aptitude test. S.E. is the standard error of the difference between the characteristic mean in the Treatment group and the characteristic mean in the Control group and is obtained assuming independent samples with equal variances. p-value is the two-sided p-value for the null hypothesis that the magnitude of the difference between the characteristic mean in the Treatment group and the characteristic mean in the Control group is zero. The joint significance of the characteristics is tested using a χ-squared test based on the results of a probit regression of an indicator for treatment on an intercept and the eleven characteristics listed in this table: the p-value for the joint null hypothesis that none of the eleven characteristics predicts treatment is 0.636.

Table A.1: Characteristics of students across all semesters

	Mean value		Treatment-Control difference		
	Treatment group	Control group	Difference	S.E.	p-value
Age	0.007	-0.007	0.014	0.045	0.761
Male	0.491	0.457	0.035	0.023	0.124
Black	0.069	0.047	0.022	0.011	0.037
Non-Hispanic white	0.628	0.644	-0.016	0.022	0.464
Hispanic	0.175	0.175	0.000	0.017	0.999
Asian	0.094	0.085	0.009	0.013	0.482
SAT score	-0.048	0.050	-0.098	0.045	0.030
High school GPA	-0.044	0.045	-0.089	0.045	0.049
Advanced placement credit	0.762	0.770	-0.008	0.019	0.686
Fall semester	0.575	0.591	-0.016	0.022	0.482
First login time	-0.003	0.004	-0.007	0.045	0.876

Notes: The Treatment and Control groups contain 995 and 972 students respectively. The p-value for the joint null hypothesis that none of the eleven characteristics predicts treatment is 0.153. Also see the notes to Table A.1.

Table A.2: Characteristics of students in Fall 2013 & Spring 2014 semesters (performance-based goals experiment)

	Mean value		Treatment-Control difference		
	Treatment group	Control group	Difference	S.E.	p-value
Age	0.003	-0.003	0.005	0.045	0.903
Male	0.462	0.454	0.008	0.022	0.704
Black	0.058	0.055	0.003	0.010	0.769
Non-Hispanic white	0.579	0.595	-0.016	0.022	0.472
Hispanic	0.210	0.209	0.002	0.018	0.932
Asian	0.109	0.099	0.010	0.014	0.476
SAT score	0.051	-0.049	0.101	0.045	0.024
High school GPA	0.013	-0.012	0.025	0.045	0.579
Advanced placement credit	0.756	0.742	0.014	0.019	0.472
Fall semester	0.665	0.624	0.041	0.021	0.055
First login time	-0.004	0.004	-0.007	0.045	0.868

Notes: The Treatment and Control groups contain 984 and 1,020 students respectively. The p-value for the joint null hypothesis that none of the eleven characteristics predicts treatment is 0.471. Also see the notes to Table A.1.

Table A.3: Characteristics of students in Fall 2014 & Spring 2015 semesters (task-based goals experiment)

	All students in the task-based goals experiment		All students in the performance-based goals experiment	
	Total points score		Total points score	
	OLS	Median	OLS	Median
Effect of asking students to set task-based goals	0.743	0.924*		
	(0.474)	(0.475)		
	[0.117]	[0.052]		
Effect of asking students to set performance-based goals			-0.237	-0.360
			(0.458)	(0.494)
			[0.605]	[0.466]
Effect / (SD in Control group)	0.068	0.085	-0.022	-0.034
Mean of dependent variable in Control group	83.111	83.111	83.220	83.220
Observations	2,004	2,004	1,967	1,967

Notes: The regressions are the same as those reported in Table 5, except that we no longer include controls for student characteristics. Heteroskedasticity-consistent standard errors are shown in round brackets and two-sided *p*-values are shown in square brackets. *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests).

Table A.4: Effects of task-based goals and performance-based goals on student performance without controls for student characteristics

	Effect of task-based goals on the average level of task completion	Effect of task-based goals on student performance	Effect of performance-based goals on student performance
	Panel I: Original specification (Tables 3 and 5)		
Treatment effect	0.491**	0.742*	0.300
	(0.205)	(0.431)	(0.398)
	[0.017]	[0.086]	[0.452]
Effect/ (SD in Control Group)	0.102	0.068	0.028
	Panel II: Specification with treatment interacted with SAT score bins		
Average treatment effect	0.516**	0.752*	0.278
	(0.207)	(0.431)	(0.395)
	[0.012]	[0.081]	[0.481]
Average effect/ (SD in Control Group)	0.108	0.069	0.026
Treatment × SAT score bin 1	-0.283	1.420	0.402
	(0.430)	(1.096)	(0.985)
	[0.510]	[0.195]	[0.683]
Treatment × SAT score bin 2	0.933**	0.726	0.868
	(0.471)	(0.962)	(0.999)
	[0.048]	[0.451]	[0.385]
Treatment × SAT score bin 3	1.046**	0.715	0.497
	(0.469)	(0.997)	(0.880)
	[0.026]	[0.473]	[0.572]
Treatment × SAT score bin 4	-0.098	0.832	0.180
	(0.462)	(0.936)	(0.809)
	[0.832]	[0.374]	[0.824]
Treatment × SAT score bin 5	0.984**	0.067	-0.554
	(0.462)	(0.769)	(0.844)
	[0.033]	[0.931]	[0.512]
Mean of dependent variable in Control group	8.627	83.111	83.220
Observations	2,004	2,004	1,967

Notes: Panel I repeats the OLS regression results with controls for student characteristics from Tables 3 and 5. The first column of Panel II shows results from an OLS regression of the number of practice exams completed on an indicator for the student having been randomly allocated to the Treatment group in the task-based goals experiment interacted with indicators for SAT score bin (bin 1 contains the lowest twenty per cent of SAT scores, bin 2 contains SAT scores greater than the 20th percentile and less than or equal to the 40th percentile, and so forth). The second column of Panel II is the same as the first column except that the dependent variable is total points score. The third column of Panel II shows results from an OLS regression of total points score on an indicator for the student having been randomly allocated to the Treatment group in the performance-based goals experiment interacted with indicators for SAT score bin. All three regressions in Panel II include controls for SAT score bin, along with all of the controls described in the notes to Table 3 except for those controls that are based on SAT score. For each regression in Panel II, the average treatment effect is the average of the five coefficients on the treatment-SAT-score-bin interactions. Heteroskedasticity-consistent standard errors are shown in round brackets and two-sided p-values are shown in square brackets. Standard errors for the average treatment effect are calculated using the delta method. *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests).

Table A.5: Effects of task-based goals and performance-based goals by SAT score bin

	Male students in the task-based goals experiment		Male students in the performance-based goals experiment	
	Total points score		Total points score	
	OLS	Median	OLS	Median
Effect of asking students to set task-based goals	1.581*	1.700**		
	(0.706)	(0.674)		
	[0.050]	[0.024]		
Effect of asking students to set performance-based goals			-0.223	0.041
			(0.672)	(0.665)
			[1.000]	[1.000]
Effect / (SD in Control group)	0.141	0.151	-0.021	0.004
Mean of dependent variable in Control group	83.285	83.285	83.644	83.644
Observations	918	918	933	933
	Female students in the task-based goals experiment		**Female students in the performance-based goals experiment**	
	Total points score		Total points score	
	OLS	Median	OLS	Median
Effect of asking students to set task-based goals	0.017	0.471		
	(0.637)	(0.652)		
	[1.000]	[0.941]		
Effect of asking students to set performance-based goals			-0.304	-0.810
			(0.625)	(0.689)
			[1.000]	[0.480]
Effect / (SD in Control group)	0.002	0.045	-0.029	-0.076
Mean of dependent variable in Control group	82.966	82.966	82.864	82.864
Observations	1,086	1,086	1,034	1,034

Notes: The regressions are the same as those reported in Table 6, except that we no longer include controls for student characteristics. Heteroskedasticity-consistent standard errors are shown in round brackets and two-sided Bonferonni-adjusted p-values are shown in square brackets. The Bonferonni adjustment accounts for the multiple null hypotheses being considered, i.e., zero treatment effect for men and zero treatment effect for women. *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests based on the Bonferonni-adjusted p-values).

Table A.6: Gender differences in the effects of task-based goals and performance-based goals on student performance without controls for student characteristics

Students in the Control group of the task-based goals experiment			
	Points scored in one of the midterms or the final exam		
	(1)	(2)	(3)
	Male students	Female students	Male-Female difference
Completed practice exams	3.342	3.571	-0.229
	(0.475)	(0.443)	(0.650)
	[0.000]	[0.000]	[0.725]
Completed practice exams squared	-0.387	-0.386	-0.001
	(0.082)	(0.074)	(0.110)
	[0.000]	[0.000]	[0.995]
Average marginal effect	1.220	1.080	0.140
	(0.127)	(0.088)	(0.154)
	[0.000]	[0.000]	[0.363]
Mean of dependent variable in Control group	23.491	23.349	
Student fixed effects	Yes	Yes	
Observations (student-exam pairs)	1,389	1,671	

Notes: Columns (1) and (2) report fixed effects panel regressions of points scored in one of the midterms or the final exam on the number of practice exams completed in preparation for that midterm or final exam and the square of this variable. Each student-exam pair is an observation (giving three observations per student). We include a fixed effect for each student, and the fixed effects absorb any effects of student characteristics on student performance. The sample includes only students from the Control group of the task-based goals experiment, who were not asked to set goals. Average marginal effects are obtained as follows: for each student exam-pair in the Control group of the task-based goals experiment, we calculate the marginal effect of completing one extra practice exam on points scored in that midterm or final exam (the marginal effect for a student exam-pair is zero if the student completed the maximum number of practice exams for that exam); and we then average across all student exam-pairs in the Control group of the task-based goals experiment. Column (3) reports the male-female difference in the regression coefficients and the average marginal effect. Heteroskedasticity-consistent standard errors (with clustering at the student level) are shown in round brackets and two-sided p-values are shown in square brackets. Standard errors for the average marginal effect are calculated using the delta method. *, ** and *** denote, respectively, significance at the 10%, 5% and 1% levels (two-sided tests).

Table A.7: Estimates of the gender difference in marginal productivity

Web Appendix II Figures

Component	Points available	Points scored	Answer key
Syllabus Quiz	2	2	N/A
Start-Of-Course Survey	2	2	N/A
Quiz 1	3	2	Answer Key
Quiz 2	3	3	Answer Key
Quiz 3	3	2	Answer Key
Quiz 4	3	2	Answer Key
Quiz 5	3		
Quiz 6	3		
Quiz 7	3		
Quiz 8	3		
Quiz 9	3		
Quiz 10	3		
Best Midterm	30		
Midterm 1			
Midterm 2			
Final Exam	34		
End-Of-Course-Survey	2		
Total Points	**100**	**13**	

Grade Key

Total Points Scored (out of 100)	Letter Grade
91 and above	A
90 to 88	A-
87 to 86	B+
85 to 81	B
80 to 78	B-
77 to 76	C+
75 to 70	C
69 to 66	D
65 and below	E

Figure A.1: Example gradecard for a student in the Control group (Fall 2013 semester)

Consent Form for Cornell University Research Team Study on Course Performance

Before you start the survey, I want to tell you about a "Cornell University Research Team" that is conducting research to evaluate which factors contribute to good performance on this course.

Research Method
The team will use:
- Survey responses.
- Grades from this course.
- Information held by the [University name] registrar (e.g., admissions data, demographic information).

Confidentiality
- All the information will be made anonymous.
- This means that your name will never be seen by the Cornell University Research Team and will not be associated with the findings.

What you will be asked to do in this study
Nothing.

Risks
There are no risks to you.

Right to withdraw from the study
You have the right to withdraw from the study at any time during the semester. If you withdraw there will be no consequences for you; your academic standing, record, or relationship with the university will not be affected. Details of how to withdraw are available from the course webpage.

Who to contact if you have questions about the study:
Cornell Research Team: [curt@cornell.edu]
Full contact details are available from the course webpage.

Who to contact about your rights as a participant in this study:
Cornell Institutional Review Board, Ithaca NY. Email: irbhp@cornell.edu, phone: 607-255-5138; website: www.irb.cornell.edu. Concerns/complaints can also be anonymously reported through Ethicspoint (web: www.hotline.cornell.edu, phone (toll-free): 1-866-293-3077). Full contact details are available from the course webpage.

The Cornell University Research Team would be very grateful if you'd be willing to consent to your data being used in this study. Remember that your name will never be seen by the Research Team and there is nothing you need to do. (If you choose not to consent, you will still receive [1%][2%] towards your score for this course from completing the survey).

Yes, I consent	No, I don't consent

Figure A.2: Consent form

Please set a goal for your grade in this course.

Think carefully before setting your goal.

The professor and the TA will not see your goal. However, each time you get your quiz, midterm and final scores back, your gradecard will remind you of your goal.

My goal for this course is:

- ⊙ A
- ⊙ A−
- ⊙ B+
- ⊙ B
- ⊙ B−
- ⊙ C+
- ⊙ C
- ⊙ D
- ⊙ E
- ⊙ Prefer not to say

Figure A.3: Fall 2013 semester goal-setting question in start-of-course survey

Please set a goal for your score in the [Midterm 1][Midterm 2][Final] Exam.

Think carefully before setting your goal.

The professor and the TA will not see your goal. However, each time you get your quiz, midterm and final exam scores back, your gradecard will remind you of your goal.

My goal for my score in the [Midterm 1][Midterm 2][Final] Exam is:

☐ out of [30][30][34]

⊙ Prefer not to say

Figure A.4: Spring 2014 semester goal-setting question in mid-course surveys

The goal that you set for this course is: $\boxed{\text{X}}$

(You set your goal in the start-of-course survey)

Component	Points available	Points scored	Answer key
Syllabus Quiz	2	2	N/A
Start-Of-Course Survey	2	2	N/A
Quiz 1	3	2	Answer Key
Quiz 2	3	3	Answer Key
Quiz 3	3	2	Answer Key
...	...	...	...

Figure A.5: Fall 2013 semester goal reminder on gradecard

Component	Points available	Points scored	Answer key
... ...	... ...	... ...	... ...
Best Midterm	30	24	
Midterm 1		22	**Your goal(*): X**
Midterm 2		24	**Your goal(*): Y**
Final Exam	34		**Your goal(*): Z**
End-Of-Course-Survey	1		
Total Points	**100**	**41**	

* You set this goal as part of a Mid-Course Survey

Grade Key

Total Points Scored (out of 100)	Letter Grade
91 and above	A
...	...

Figure A.6: Spring 2014 semester goal reminder on gradecard

{TREATED AND CONTROL STUDENTS}

Practice Exams

There will be 5 practice exams for the [Midterm 1][Midterm 2][Final] Exam. Each practice exam will contain the same number of questions as the [Midterm 1][Midterm 2][Final] Exam.

The practice exams for the [Midterm 1][Midterm 2][Final] Exam will become active when Mid-Course Survey [1][2][3] closes. You will receive a reminder email at that time.

{TREATED STUDENTS ONLY}

Question 6

Please set a goal for the number of practice exams that you will complete out of the 5 practice exams for the [Midterm 1][Midterm 2][Final] Exam.

Think carefully before setting your goal.

The professor and the TA will not see your goal. However, when you take the practice exams you will be reminded of your goal.

My goal is to complete

[ ] out of the 5 practice exams for the [Midterm 1][Midterm 2][Final] Exam

⊙ Prefer not to say

Figure A.7: Fall 2014 & Spring 2015 semesters practice exams information and goal-setting question in mid-course surveys

Dear Econ2023 Students,

The 5 practice exams for the [Midterm 1][Midterm 2][Final] Exam are now active.

{TREATED STUDENTS ONLY} Your goal is to complete X out of the 5 practice exams (you set this goal as part of Mid-Course Survey [1][2][3]).

To go to the practice exams, please go to the course webpage and follow the link.

Note that you will not receive any further reminders about the practice exams for the [Midterm 1][Midterm 2][Final] Exam and the practice exams will close when the [Midterm 1][Midterm 2][Final] Exam begins.

This is an automated email from the ECO 2023 system.

Figure A.8: Fall 2014 & Spring 2015 semesters practice exams reminder email

Practice Exams for the [Midterm 1][Midterm 2][Final] Exam

You have completed X out of the 5 practice exams for the [Midterm 1][Midterm 2][Final] Exam

{TREATED STUDENTS ONLY} **Your goal is to complete Z out of the 5 practice exams** (you set this goal as part of Mid-course Survey [1][2][3]).

Instructions:

- To take one of the practice exams, click on the link below.
- You can only take each practice exam once.
- There is no time limit.
- After answering each question, you will be given the correct answer.
- **This will be your only opportunity to see the correct answer.**
- You will not be able to go back to previous questions.

Practice Exam X+1

Figure A.9: Fall 2014 & Spring 2015 semesters practice exams introductory screen

Practice Exam X for the [Midterm 1][Midterm 2][Final] Exam

Your score was **Y** out of [30][30][34]

You have now completed X out of the 5 practice exams for the
[Midterm 1][Midterm 2][Final] Exam

{Treated students only} **Your goal is to complete Z out of the 5 practice exams**
(you set this goal as part of Mid-course Survey [1][2][3]).

Return To Practice Exams Screen

Figure A.10: Fall 2014 & Spring 2015 semesters practice exams feedback screen

Web Appendix III Using theory to interpret our findings

Web Appendix III.1 Motivation

In this section we suggest some hypotheses for our findings. For each type of goal (performance- and task-based), our approach is to write down a simple model of goal setting and then use this to generate possible hypotheses. We acknowledge that these models are not the only ones that we could have used, but we are not aiming to test theory. Rather, we are using theory to guide the analysis and interpretation of our findings.

Our models build on Koch and Nafziger (2011) and are inspired by two key concepts in behavioral economics: present bias and loss aversion. The concept of present bias captures the idea that people lack control because they place a high weight on current utility (Strotz, 1956).[40] More specifically, a present-biased discounter places more weight on current utility relative to utility n periods in the future than she does on utility at future time t relative to utility at time $t + n$. This implies that present-biased discounters exhibit time inconsistency, since their time preferences at different dates are not consistent with one another. In the context of education, a present-biased student might set out to exert her preferred level of effort, but when the time comes to attend class or review for a test she might lack the self-control necessary to implement these plans.[41] Strotz (1956) and Pollak (1968) were the first to analyze how time-inconsistent agents make choices anticipating the different time preferences of their future selves. Building on this insight, Strotz (1956) noted that present-biased agents can mitigate their self-control problem by using commitment devices to bind their future self.[42]

The concept of loss aversion captures the idea that people dislike falling behind a salient reference point (Kahneman and Tversky, 1979).[43] In the context of education, a loss-averse student might work particularly hard in an attempt to achieve a salient reference point (e.g., a particular grade in her course).

Together, the literatures on present bias and loss aversion suggest that self-set goals might serve as an effective commitment device. Specifically, self-set goals might act as salient reference points, helping present-biased agents to mitigate their self-control problem and so steer their effort toward its optimal level. Indeed, Koch and Nafziger (2011) developed a model of goal setting based on this idea that we build on here. Unlike us, however, Koch and Nafziger (2011)

[40]Present bias has been proposed as an explanation for aspects of many behaviors such as addiction (Gruber and Kőszegi, 2001), early retirement (Diamond and Kőszegi, 2003), smoking (Khwaja et al., 2007), welfare program participation (Fang and Silverman, 2009) and credit card borrowing (Meier and Sprenger, 2010). See Dhami (2016) for a recent comprehensive survey of the literature on present bias.

[41]Under standard (i.e., exponential) discounting this self-control problem disappears.

[42]We provide examples of such commitment devices in Web Appendix V.2.

[43]Loss aversion has been proposed as a foundation of a number of phenomena such as the endowment effect (Kahneman et al., 1990), small-scale risk aversion (Rabin, 2000), the disposition effect (Genesove and Mayer, 2001), and the role of expectations in single-agent decision-making (Bell, 1985; Kőszegi and Rabin, 2006) and in strategic interactions (Gill and Stone, 2010; Gill and Prowse, 2012).

did not explore the effectiveness of different types of goals.[44]

Web Appendix III.2 Performance-based goal setting

Web Appendix III.2.1 Simple model

We start by building a simple model of performance-based goal setting. In Web Appendix III.2.2 we use the model to suggest three hypotheses for why performance-based goals might not be very effective in the context that we studied.

At period one the student sets a goal $g \geq 0$ for performance $f \geq 0$; we call the student at period one the *student-planner*. At period two the student chooses effort $e \geq 0$; we call the student at period two the *student-actor*. The student-actor incurs a cost of effort $C(e) = ce^2/2$, with $c > 0$. At period three performance is realized and the student incurs any disutility from failing to achieve her goal; we call the student at period three the *student-beneficiary*. Performance increases one-to-one in effort exerted by the student-actor at period two, i.e., $f(e) = e$, and the student-beneficiary's utility increases one-to-one in performance.[45] The student-beneficiary is loss averse around her goal: she suffers goal disutility that depends linearly on how far performance falls short of the goal set by the student-planner at period one. The student-beneficiary's goal disutility is given by $-l \max\{g - f(e), 0\}$. The parameter $l > 0$ captures the strength of loss aversion, which in our context we call the 'strength of goal disutility'.[46,47]

The student is present biased. In particular, the student exhibits quasi-hyperbolic discounting, with $\beta \in (0, 1)$ and $\delta \in (0, 1]$: the student discounts utility n periods in the future by a factor $\beta\delta^n$.[48] Under quasi-hyperbolic discounting the student-planner discounts period-two utility by a factor $\beta\delta$ and period-three utility by a factor $\beta\delta^2$, and so discounts period-three utility by δ relative to period-two utility. The student-actor, on the other hand, discounts period-three utility by $\beta\delta$ relative to immediate period-two utility. Since $\beta\delta < \delta$, the student-planner places more weight on utility from performance at period three relative to the cost of effort at period two than does the student-actor.

As a result of this present bias, and in the absence of a goal, the student-planner's desired effort is higher than the effort chosen by the student-actor: that is, the student exhibits a

[44] Koch and Nafziger (2011)'s model also differs from our models in that agents in their model choose from only two possible effort levels, while our models allow students to choose both effort and goals from a continuum. Again without exploring the effectiveness of different types of goals, Jain (2009) also studies theoretically how present-biased agents can use goals as reference points; in Jain (2009)'s model utility is discontinuous at the reference point, rather than kinked as in Kahneman and Tversky (1979)'s model of loss aversion that we and Koch and Nafziger (2011) use. Heath et al. (1999) and Wu et al. (2008) linked goals to loss aversion, but did not make the connection to present bias. Finally, a related theoretical literature studies expectations as goals (Suvorov and Van de Ven, 2008; Hsiaw, 2013; Hsiaw, 2016; Koch and Nafziger, 2016).

[45] Using one-to-one relationships instead of more general linear relationships is without loss of generality.

[46] Specifically, l measures the psychological loss from failing to achieve the goal relative to 'material' utility.

[47] This formulation implies that the student suffers disutility from failing to achieve her goal. However, it also implies that she enjoys no elation from exceeding the goal. This latter assumption can be justified in two ways. First, the more parsimonious one-parameter model of loss aversion allows us to gain useful insights into the effectiveness of goal setting. Second, if students enjoyed elation from exceeding their goals, they would have a strategic incentive to set low goals in order to enjoy the utility boost from exceeding them, but we do not see evidence that this motivation is an important driver of behavior in our data.

[48] Laibson (1997) was the first to apply the analytically tractable quasi-hyperbolic (or 'beta-delta') model of discounting to analyze the choices of present-biased time-inconsistent agents. Like us, Laibson (1997) finds the equilibria of a dynamic game among a sequence of temporal selves.

self-control problem due to time inconsistency.[49] We formalize this as follows:

Remark A.1
In the absence of a goal the student exhibits time inconsistency:
(i) The student-actor chooses effort $\underline{e} = \beta\delta/c$.
(ii) The student-planner would like the student-actor to exert effort $\hat{e} = \delta/c > \underline{e}$.

Proof. See Web Appendix IV.1. ■

To alleviate her self-control problem due to time-inconsistency, the student-planner might choose to set a goal. Goals can be effective by increasing the student-actor's marginal incentive to work in order to avoid the goal disutility that results from failing to achieve the goal.

To demonstrate this point, we solve for the subgame-perfect Nash equilibria of the game outlined above, in which the players are the student-planner and student-actor. We do so by backward induction. First, we analyze the effort choice of the student-actor at period two for any goal set by the student-planner at period one. The student-actor's utility is given by:

$$u_{act}(e|g) = \beta\delta[f(e) - l\max\{g - f(e), 0\}] - C(e) \quad (1)$$

$$= \beta\delta[e - l\max\{g - e, 0\}] - \frac{ce^2}{2}. \quad (2)$$

Proposition A.1 shows how the student-actor's effort responds to the goal.

Proposition A.1
Let $\overline{e} = \beta\delta(1 + l)/c$ *and recall from Remark A.1 that* $\underline{e} = \beta\delta/c < \overline{e}$ *denotes the student-actor's effort in the absence of a goal.*
(i) When $g \leq \underline{e}$*, the student-actor exerts effort* $e^* = \underline{e}$.
(ii) When $g \in [\underline{e}, \overline{e}]$*, the student-actor exerts effort* $e^* = g$.
(iii) When $g \geq \overline{e}$*, the student-actor exerts effort* $e^* = \overline{e}$.

Proof. See Web Appendix IV.1. ■

Proposition A.1 tells us that, perhaps unsurprisingly, the goal does not raise effort when it is set lower than the student-actor's optimal level of effort in the absence of a goal $\underline{e}$. Intermediate goals are effective: intermediate goals induce the student-actor to work hard enough to achieve the goal in order to avoid disutility from falling short of the goal. Beyond a certain point the marginal cost of effort outweighs the marginal reduction in goal disutility, and so the goal induces an increase in effort only to an upper bound $\overline{e}$. Goals above the upper bound leave the student-actor to suffer some goal disutility. This upper bound increases as the time-inconsistency problem becomes less severe (higher β) and as the strength of goal disutility l goes up.

Having established how the student-actor's effort responds to any goal set by the student-planner, we now consider the student-planner's optimal choice of goal. Letting $e^*(g)$ represent

[49] When $g = 0$, goal disutility is zero since $\max\{g - f(e), 0\} = 0$, and so $g = 0$ is equivalent to the absence of a goal.

the student-actor's optimal effort given a goal g, the student-planner's utility is given by:

$$\begin{aligned} u_{plan}(g|e^*(g)) &= \beta\delta^2[f(e^*(g)) - l\max\{g - f(e^*(g)), 0\}] - \beta\delta C(e^*(g)) \quad (3) \\ &= \beta\delta^2[e^*(g) - l\max\{g - e^*(g), 0\}] - \beta\delta\frac{c[e^*(g)]^2}{2}. \quad (4) \end{aligned}$$

Proposition A.2
Recall from Remark A.1 that $\underline{e} = \beta\delta/c$ *and* $\hat{e} = \delta/c$ *denote, respectively, student-actor effort and student-planner desired effort in the absence of a goal.*
Recall from Proposition A.1 that $\overline{e} = \beta\delta(1+l)/c$ *denotes maximal student-actor effort in the presence of a goal.*
(i) The optimal choice of goal for the student-planner is given by $g^* = \min\{\hat{e}, \overline{e}\}$.
(ii) When $\beta(1+l) < 1$, $g^* = \overline{e}$.
(iii) When $\beta(1+l) \geq 1$, $g^* = \hat{e}$.
(iv) Effort of the student-actor $e^* = g^* > \underline{e}$*, and so the student-actor works harder than in the absence of goal.*

Proof. See Web Appendix IV.1. ■

We know from Proposition A.1 that goals in the range $[\underline{e}, \overline{e}]$ induce the student-actor to work to achieve the goal, but that higher goals are ineffective in raising effort above $\overline{e}$. Thus, the student-planner will never set a goal higher than $\overline{e}$, since higher goals are not effective but leave the student-actor to suffer some goal disutility from failing to achieve the goal. If the student-planner could simply impose a level of effort on the student-actor, then we know from Remark 1 that she would choose $\hat{e}$. When $\beta(1+l)$ is big enough, $\hat{e} \leq \overline{e}$, and so the student-planner achieves her desired level of effort by setting $g^* = \hat{e}$. This case holds when the time-inconsistency problem is not too severe (high β) or the strength of goal disutility l is sufficiently high. When her desired level of effort is not achievable, the student-planner sets $g^* = \overline{e}$, and so the student-planner uses the goal to induce as much effort from the student-actor as she is able to. In either case, the optimal goal induces the student to work harder than she would in the absence of a goal and the student always achieves her goal in equilibrium.

Web Appendix III.2.2 Why might performance-based goals not be very effective?

This model of performance-based goal setting suggests that performance-based goals can improve course performance. However, our experimental data show that performance-based goals had a positive but small and statistically insignificant effect on student performance (Table 5). In our view, the simple model sketched above suggests three hypotheses for why performance-based goals might not be very effective in the context that we studied (we view these hypotheses as complementary).

Timing of goal disutility

In the simple model, the student works in period two and experiences any goal disutility from failing to achieve her performance-based goal in period three (i.e., when performance is realized).

This temporal distance will dampen the motivating effect of the goal. Formally, the student-actor discounts goal disutility by a factor $\beta\delta < 1$; in the expression for maximal student-actor effort in the presence of a goal, $\overline{e}$, the parameter measuring the strength of goal disutility, l, is multiplied by this discount factor (see Proposition A.2). Even when the temporal distance between effort and goal disutility is modest, the timing of goal disutility dampens the effectiveness of performance-based goals because quasi-hyperbolic discounters discount the near future relative to the present by a factor β even if $\delta \approx 1$ over the modest temporal distance.

Overconfidence

In the simple model, students understand perfectly the relationship between effort and performance. In contrast, the education literature suggests that students face considerable uncertainty about the educational production function, and that this uncertainty could lead to students holding incorrect beliefs about the relationship between effort and performance (e.g., Romer, 1993, and Fryer, 2013). Furthermore, the broader behavioral literature shows that people tend to be overconfident when they face uncertainty (e.g., Weinstein, 1980, Camerer and Lovallo, 1999, and Park and Santos-Pinto, 2010). The behavioral literature also provides a number of theoretical underpinnings for overconfidence (e.g., Brunnermeier and Parker, 2005, Johnson and Fowler, 2011, and Gossner and Steiner, 2016).

Suppose that some students are overconfident. An overconfident student believes that the production function is given by $f(\cdot)$, when in fact performance is given by $hf(\cdot)$ with $h \in (0, 1)$ for any value of effort. An overconfident student will act as if the production function is given by $f(\cdot)$, and so the model in Web Appendix III.2.1 describes her choice of goal and effort. However, the student's actual performance with goal setting and in the absence of a goal will be a proportion h of that expected by the student. As a result, the impact of performance-based goal setting on performance will be reduced by this proportion h. Furthermore, the overconfident student will unexpectedly fail to achieve her performance-based goal.

Performance uncertainty

In the simple model, the student knows for sure how her effort translates into performance (i.e., the relationship between effort and performance involves no uncertainty). As such, her goal is always achieved in equilibrium. In practice, the relationship between effort and performance is likely to be noisy and, as in our experiment, performance-based goals are not always reached. The student could face uncertainty about her own ability or about the productivity of work effort. The student might also get unlucky: for instance, the draw of questions on the exam might be unfavorable or the student might get sick near the exam.

To introduce uncertainty about performance in a straightforward way, suppose that the student is risk neutral and that with known probability $\pi \in (0, 1)$ her effort translates into performance according to $f(\cdot)$ as in Web Appendix III.2.1, while with probability $1 - \pi$ performance $f = 0$ (since we assume that π is known, the student is neither overconfident nor underconfident).[50]

[50] We can think of $f = 0$ as a baseline level of performance that the student achieves with little effort even in the absence of goal setting.

The formal details of the analysis are relegated to Web Appendix IV.3. The uncertainty directly reduces the student-actor's marginal incentive to exert effort, which reduces by a factor π the equilibrium goal and equilibrium effort with and without goal setting. However, this general reduction in incentives is not the only effect of uncertainty: performance-based goals become risky because when performance turns out to be low the student fails to achieve her performance-based goal and so suffers goal disutility that increases in the goal (as in the simple model, goals are never exceeded in equilibrium). In contrast to the case of overconfidence discussed above, goal failure is not unexpected: the student facing uncertainty anticipates that she will not always achieve her performance-based goal.

Anticipating the goal disutility suffered when performance turns out to be low, the student-planner further scales back the performance-based goal that she sets for the student-actor, which reduces the effectiveness of performance-based goal setting. Formally, this extra effect adds the second term to the numerator in the expression for $\tilde{e}$ in Proposition A.6 in Web Appendix IV.3.[51,52]

Web Appendix III.3 Task-based goal setting

Web Appendix III.3.1 Simple model

We now model task-based goal setting. At period one the student-planner sets a goal $g \geq 0$ for the number of units of the task to complete $a \geq 0$. We call a the 'level of task completion'. At period two the student-actor chooses the level of task completion a. The student-actor incurs a cost of task completion $C(a) = \kappa a^2/2$, with $\kappa > 0$. Furthermore, at period two the loss-averse student-actor suffers goal disutility that depends linearly on how far the level of task completion falls short of the goal set by the student-planner at period one: she suffers goal disutility of $-\lambda \max\{g - a, 0\}$, where the loss parameter λ captures the strength of goal disutility.[53] At period three performance is realized. Performance increases linearly in the level of task completion: $f(a) = \theta a$, with $\theta > 0$; while the student-beneficiary's utility increases one-to-one in performance.[54] The present-biased student exhibits quasi-hyperbolic discounting

[51] Proposition A.6 in Web Appendix IV.3 focuses on the case in which uncertainty is not too big. When the student faces a lot of uncertainty, the extra effect could lead the student-planner to prefer not to set a goal at all.

[52] This scaling back of goals is not necessarily at odds with the fact that the performance-based goals that we see in the data appear ambitious. First, the goal will appear ambitious relative to average achievement because, as noted above, when performance turns out to be low the student fails to achieve her goal. Second, without any scaling back the goals might have been even higher. Third, the overconfidence that we discuss above could keep the scaled-back goal high. Fourth, we explain in Web Appendix III.3.4 below that students likely report as their goal an 'aspiration' that is only relevant if, when the time comes to study, the cost of effort turns out to be particularly low: the actual cost-specific goal that the student aims to hit could be much lower than this aspiration.

[53] As explained in footnote 47, we do not include any elation from exceeding the goal. We use new notation for the parameter that measures the strength of goal disutility λ because the units used to measure the level of task completion are different from the performance units in Web Appendix III.2.1. Note also that goal disutility is incurred at period two here because the student-actor observes how far she is from the task-based goal immediately when she stops working on the task. For performance-based goals in Web Appendix III.2.1 goal disutility is incurred at period three when performance is realized.

[54] Units of performance and units of task completion are both defined externally, and so we need to introduce the parameter θ to model a linear relationship between them.

as described in Web Appendix III.2.1. Thus the student-actor's utility is given by:

$$\begin{aligned} u_{act}(a|g) &= \beta\delta f(a) - [\lambda \max\{g-a,0\} + C(a)] && (5) \\ &= \beta\delta\theta a - \left[\lambda \max\{g-a,0\} + \frac{\kappa a^2}{2}\right]; && (6) \end{aligned}$$

and, letting $a^*(g)$ represent the student-actor's optimal level of task completion given a goal g, the student-planner's utility is given by:

$$\begin{aligned} u_{plan}(g|a^*(g)) &= \beta\delta^2 f(a^*(g)) - \beta\delta[\lambda \max\{g-a^*(g),0\} + C(a^*(g))] && (7) \\ &= \beta\delta^2\theta a^*(g) - \beta\delta\left[\lambda \max\{g-a^*(g),0\} + \frac{\kappa[a^*(g)]^2}{2}\right]. && (8) \end{aligned}$$

When we solve the game by backward induction, we get results that are qualitatively similar to those for performance-based goals in Web Appendix III.2.1. The formal results and proofs are relegated to Web Appendix IV.2. The three relevant thresholds now become:

$$\underline{a} = \frac{\beta\delta\theta}{\kappa}; \qquad \hat{a} = \frac{\delta\theta}{\kappa} > \underline{a}; \qquad \overline{a} = \frac{\beta\delta\theta + \lambda}{\kappa} > \underline{a}. \tag{9}$$

Mirroring Remark A.1, in the absence of a goal the present-biased student exhibits a self-control problem due to time inconsistency: the student-actor chooses a level of task completion $\underline{a}$, which is smaller than the student-planner's desired level of task completion $\hat{a}$. The upper bound on student-actor task completion in the presence of a goal is given by $\overline{a}$. Mirroring Proposition A.1, this upper bound increases as the time-inconsistency problem becomes less severe (higher β) and as the strength of goal disutility λ goes up. Mirroring Proposition A.2, the optimal choice of goal for the student-planner is given by $g^* = \min\{\hat{a}, \overline{a}\}$ and the optimal goal induces a level of task completion by the student-actor of $a^* = g^* > \underline{a}$; the optimal goal induces a higher level of task completion than in the absence of a goal, and the student always achieves her goal in equilibrium.[55] The goal increases the level of task completion as well as improving course performance.

Web Appendix III.3.2 Why were task-based goals effective?

Our experimental data show that task-based goals improved task completion and course performance (see Table 3 for the effect on task completion and Table 5 for the effect on course performance).[56] How might we account for these findings, given our analysis of why performance-based goals might not be very effective? In our view, an obvious answer is that with task-based goal setting, the three factors that reduce the effectiveness of performance-based goals (Web Appendix III.2.2) are of lesser importance or do not apply at all.

[55]When $\beta\delta\theta + \lambda \geq \delta\theta$, $\hat{a} \leq \overline{a}$, and so the student-planner achieves her desired level of task completion by setting $g^* = \hat{a}$. Similarly to Proposition A.2, this case holds when the time-inconsistency problem is not too severe (high β) or the strength of goal disutility λ is sufficiently high.

[56]It is possible that some students in the Control group (who were not invited to set goals) might already use goals as a commitment device. However, since we find that task-based goals are successful at increasing performance, we conclude that many students in the Control group did not use goals or set goals that were not fully effective. We note that asking students to set goals might make the usefulness of goal setting as a commitment device more salient and so effective. Reminding students of their goal, as we did, might also help to make them more effective.

Timing of goal disutility

In the case of task-based goal setting, any goal disutility from failing to achieve the task-based goal is suffered immediately when the student stops working on the task in period two. Thus, unlike the case of performance-based goal setting discussed in Web Appendix III.2.2, there is no temporal distance that dampens the motivating effect of the goal. Formally, in the expression for maximal task completion in the presence of a goal, $\overline{a}$, the parameter measuring the strength of goal disutility, λ, is undiscounted (see (9)).

Overconfidence

As discussed in Web Appendix III.2.2, overconfidence reduces the effectiveness of goal setting. Recall that an overconfident student acts as if the production function is given $f(\cdot)$, when in fact performance is given by $hf(\cdot)$, which reduces the impact of goal setting on performance by the proportion h. In the case of task-based goal setting, this effect is mitigated if practice exams direct students toward productive tasks: in that case h goes up. Plausibly, teachers have better information about which tasks are likely to be productive, and asking students to set goals for productive tasks is one way to improve the power of goal setting for overconfident students.

Instead of improving the power of goal setting by directing overconfident students toward productive tasks, it is conceivable that task-based goals improved performance via another channel: signaling to students in the Treatment group that practice exams were an effective task. But we think this is highly unlikely. First, we were careful to make the practice exams as salient as possible to the Control group. Second, students in the Control group in fact completed many practice exams. Third, it is hard to understand why only men would respond to the signal.

Performance uncertainty

In Web Appendix III.2.2, we introduced uncertainty about performance. It is straightforward to add performance uncertainty into the simple model of task-based goal setting outlined in Web Appendix III.3.1. The formal details of the analysis are relegated to Web Appendix IV.3. The important point to note is that even with uncertainty about performance, the student continues to achieve her task-based goal: there is no 'task uncertainty'. The reason is that the student-actor controls the number of units of the task that she completes and so can guarantee to hit her task-based goal. Thus, unlike the case of performance-based goals with uncertainty, the student has no reason to scale back her task-based goal to reduce goal disutility in the event that the goal is not reached.

Web Appendix III.3.3 Why were task-based goals more effective for men than for women?

Our data show that task-based goals are more effective for men than for women. More specifically: in the Control group without goal setting men completed fewer practice exams than women (Table 4); and task-based goals increased performance and the number of practice exams completed more for men than for women (Tables 6 and 4 and respectively). In the context of our simple model of task-based goal setting (Web Appendix III.3.1), one hypothesis that could explain this finding is that the male students in our sample are more present biased than the female students (i.e., the men have a lower β parameter). Existing empirical evidence supports the idea that men may have less self-control and be more present biased than women (see Web Appendix V.6 for a survey of this evidence.) Interestingly, in a laboratory experiment in which goals were set by the experimenter rather than by the subjects themselves, Smithers (2015) finds that goals increased the work performance of men but not that of women.

To understand the role of present bias, first note that the student-actor's level of task completion in the absence of a goal $\underline{a}$ is increasing in β: the more present biased the student, the fewer practice exams he or she completes without a goal. Thus, if men are more present biased than women, then their higher degree of present bias will push down their level of task completion in the Control group relative to that of women. Second, the increase in task completion induced by goal setting also depends on the degree of present bias: in particular, the difference between the student-planner's desired level of task completion $\hat{a}$ and the student-actor's level of task completion in the absence of a goal $\underline{a}$ is decreasing in β. Thus, if men are more present biased than women, then goal setting will tend to be more effective at increasing the number of practice exams that men complete, which in turn feeds into a larger effect on performance.

Web Appendix III.3.4 Why were task-based goals not always achieved?

In the simple models outlined in Web Appendix III.2.1 and Web Appendix III.3.1 goals are always achieved. In Web Appendix III.2.2 we explained how overconfidence and performance uncertainty can result in a student's failure to achieve her performance-based goal. A puzzle remains: even though task-based goals are more frequently achieved than performance-based goals, task-based goals are not always achieved (Table 2).

Failure to achieve task-based goals emerges naturally if we relax the assumption that costs are known with certainty. In particular, suppose that the student-actor's cost parameter (c or κ) can be high or low, and that the student-actor draws her cost parameter in period two before she decides how hard to work (the analysis extends naturally to any number of possible cost draws). For example, the cost uncertainty could be driven by uncertainty about the set of leisure activities available to the student during the time that she planned to study. Anticipating this cost uncertainty, we allow the student-planner to set a goal for both possible cost draws. The optimal goal for a given cost draw is just as in the models in Web Appendix III.2.1 and Web Appendix III.3.1 with no cost uncertainty, and the student-actor always works hard enough to achieve the cost-specific goal. Of course, we ask the student to report only one goal: we assume here that the student-planner reports to us only her goal for the low-cost draw. This goal is like an aspiration: if the cost turns out to be high, the goal is scaled down to reflect the higher cost. Because we as the experimenter observe only the reported aspiration, when the cost is high we observe a failure to achieve the reported aspiration, even though the student achieves her cost-specific goal.

Web Appendix IV Formal results and proofs

Web Appendix IV.1 Proofs for Web Appendix III.2.1

Proof of Remark A.1.

In the absence of a goal, the student-actor's utility and the student-planner's utility are given by, respectively:

$$
\begin{aligned}
u_{act}(e) &= \beta\delta f(e) - C(e) && (10)\\
&= \beta\delta e - \frac{ce^2}{2}; && (11)\\
u_{plan}(e) &= \beta\delta^2 f(e) - \beta\delta C(e) && (12)\\
&= \beta\delta^2 e - \beta\delta\frac{ce^2}{2}. && (13)
\end{aligned}
$$

Both utilities are strictly concave since $c > 0$. Straightforward maximization then gives the result. ■

Proof of Proposition A.1.

Using (2), on the range $e \in [0, g]$:

$$
\begin{aligned}
\frac{\partial u_{act}}{\partial e} &= \beta\delta(1 + l) - ce; && (14)\\
\frac{\partial^2 u_{act}}{\partial e^2} &= -c < 0; \quad \text{and so} && (15)\\
e^* &= \min\{\overline{e}, g\}. && (16)
\end{aligned}
$$

Using (2), on the range $e \in [g, \infty)$:

$$
\begin{aligned}
\frac{\partial u_{act}}{\partial e} &= \beta\delta - ce; && (17)\\
\frac{\partial^2 u_{act}}{\partial e^2} &= -c < 0; \quad \text{and so} && (18)\\
e^* &= \max\{\underline{e}, g\}. && (19)
\end{aligned}
$$

(i) When $g \leq \underline{e}$, on the range $e \in [0, g]$, $e^* = g$, and on the range $e \in [g, \infty)$, $e^* = \underline{e}$. Thus, on the range $e \in [0, \infty)$, $e^* = \underline{e}$.

(ii) When $g \in [\underline{e}, \overline{e}]$, on the range $e \in [0, g]$, $e^* = g$, and on the range $e \in [g, \infty)$, $e^* = g$. Thus, on the range $e \in [0, \infty)$, $e^* = g$.

(iii) When $g \geq \overline{e}$, on the range $e \in [0, g]$, $e^* = \overline{e}$, and on the range $e \in [g, \infty)$, $e^* = g$. Thus, on the range $e \in [0, \infty)$, $e^* = \overline{e}$. ■

Proof of Proposition A.2.

On the range $g \in [0, \underline{e}]$, $e^*(g) = \underline{e}$ from Proposition A.1, and so $\partial e^*/\partial g = 0$ and $\max\{g - e^*(g), 0\} = 0$. Using (4), $du_{plan}/dg = 0$ and so any $g \in [0, \underline{e}]$ is optimal (including $\underline{e}$).

On the range $g \in [\underline{e}, \overline{e}]$, $e^*(g) = g$ from Proposition A.1, and so $\partial e^*/\partial g = 1$ and

$\max\{g - e^*(g), 0\} = 0$. Using (4), and noting that $\hat{e} > \underline{e}$ and $\overline{e} > \underline{e}$:

$$\frac{du_{plan}}{dg} = \beta\delta^2 - \beta\delta c g; \tag{20}$$

$$\frac{d^2 u_{plan}}{dg^2} = -\beta\delta c < 0; \quad \text{and so} \tag{21}$$

$$g^* = \min\{\hat{e}, \overline{e}\} > \underline{e}. \tag{22}$$

On the range $g \in [\overline{e}, \infty)$, $e^*(g) = \overline{e}$ from Proposition A.1, and so $\partial e^*/\partial g = 0$ and $\max\{g - e^*(g), 0\} = g - \overline{e}$. Using (4), $du_{plan}/dg = -\beta\delta^2 l < 0$ and so $g^* = \overline{e}$.

Stitching the ranges together gives $g^* = \min\{\hat{e}, \overline{e}\} > \underline{e}$. Parts (ii) and (iii) follow, given that $\overline{e} < \hat{e} \Leftrightarrow \beta(1 + l) < 1$. Finally, (iv) follows immediately from Proposition A.1. ■

Web Appendix IV.2 Results and proofs for Web Appendix III.3.1

Remark A.2

In the absence of a goal the student exhibits time inconsistency:
(i) The student-actor chooses a level of task completion $\underline{a} = \beta\delta\theta/\kappa$.
(ii) The student-planner would like the student-actor to choose a level of task completion $\hat{a} = \delta\theta/\kappa > \underline{a}$.

Proof of Remark A.2.

In the absence of a goal, the student-actor's utility and the student-planner's utility are given by, respectively:

$$\begin{aligned}
u_{act}(a) &= \beta\delta f(a) - C(a) && (23)\\
&= \beta\delta\theta a - \frac{\kappa a^2}{2}; && (24)\\
u_{plan}(a) &= \beta\delta^2 f(a) - \beta\delta C(a) && (25)\\
&= \beta\delta^2\theta a - \beta\delta\frac{\kappa a^2}{2}. && (26)
\end{aligned}$$

Both utilities are strictly concave since $\kappa > 0$. Straightforward maximization then gives the result. ■

Proposition A.3

Let $\overline{a} = (\beta\delta\theta + \lambda)/\kappa$ *and recall from Remark A.2 that* $\underline{a} = \beta\delta\theta/\kappa < \overline{a}$ *denotes the student-actor's level of task completion in the absence of a goal.*
(i) When $g \leq \underline{a}$*, the student-actor chooses a level of task completion* $a^* = \underline{a}$.
(ii) When $g \in [\underline{a}, \overline{a}]$*, the student-actor chooses a level of task completion* $a^* = g$.
(iii) When $g \geq \overline{a}$*, the student-actor chooses a level of task completion* $a^* = \overline{a}$.

Proof of Proposition A.3.

Using (6), on the range $a \in [0, g]$:

$$\begin{aligned}
\frac{\partial u_{act}}{\partial a} &= \beta\delta\theta + \lambda - \kappa a; && (27)\\
\frac{\partial^2 u_{act}}{\partial a^2} &= -\kappa < 0; \quad \text{and so} && (28)\\
a^* &= \min\{\overline{a}, g\}. && (29)
\end{aligned}$$

Using (6), on the range $a \in [g, \infty)$:

$$\begin{aligned}
\frac{\partial u_{act}}{\partial a} &= \beta\delta\theta - \kappa a; && (30)\\
\frac{\partial^2 u_{act}}{\partial a^2} &= -\kappa < 0; \quad \text{and so} && (31)\\
a^* &= \max\{\underline{a}, g\}. && (32)
\end{aligned}$$

(i) When $g \leq \underline{a}$, on the range $a \in [0, g]$, $a^* = g$, and on the range $a \in [g, \infty)$, $a^* = \underline{a}$. Thus, on the range $a \in [0, \infty)$, $a^* = \underline{a}$.

(ii) When $g \in [\underline{a}, \overline{a}]$, on the range $a \in [0, g]$, $a^* = g$, and on the range $a \in [g, \infty)$, $a^* = g$. Thus, on the range $a \in [0, \infty)$, $a^* = g$.

(iii) When $g \geq \overline{a}$, on the range $a \in [0, g]$, $a^* = \overline{a}$, and on the range $a \in [g, \infty)$, $a^* = g$. Thus, on the range $a \in [0, \infty)$, $a^* = \overline{a}$. ∎

Proposition A.4

Recall from Remark A.2 that $\underline{a} = \beta\delta\theta/\kappa$ *and* $\hat{a} = \delta\theta/\kappa$ *denote, respectively, student-actor task completion and student-planner desired task completion in the absence of a goal.*

Recall from Proposition A.3 that $\overline{a} = (\beta\delta\theta + \lambda)/\kappa$ *denotes maximal student-actor task completion in the presence of a goal.*

(i) The optimal choice of goal for the student-planner is given by $g^* = \min\{\hat{a}, \overline{a}\}$.

(ii) When $\beta\delta\theta + \lambda < \delta\theta$, $g^* = \overline{a}$.

(iii) When $\beta\delta\theta + \lambda \geq \delta\theta$, $g^* = \hat{a}$.

(iv) Student-actor task completion $a^* = g^* > \underline{a}$, *and so the level of task completion is higher than in the absence of goal.*

Proof of Proposition A.4.

On the range $g \in [0, \underline{a}]$, $a^*(g) = \underline{a}$ from Proposition A.3, and so $\partial a^*/\partial g = 0$ and $\max\{g - a^*(g), 0\} = 0$. Using (8), $du_{plan}/dg = 0$ and so any $g \in [0, \underline{a}]$ is optimal (including $\underline{a}$).

On the range $g \in [\underline{a}, \overline{a}]$, $a^*(g) = g$ from Proposition A.3, and so $\partial a^*/\partial g = 1$ and $\max\{g - a^*(g), 0\} = 0$. Using (8), and noting that $\hat{a} > \underline{a}$ and $\overline{a} > \underline{a}$:

$$\frac{du_{plan}}{dg} = \beta\delta^2\theta - \beta\delta\kappa g; \tag{33}$$

$$\frac{d^2u_{plan}}{dg^2} = -\beta\delta\kappa < 0; \quad \text{and so} \tag{34}$$

$$g^* = \min\{\hat{a}, \overline{a}\} > \underline{a}. \tag{35}$$

On the range $g \in [\overline{a}, \infty)$, $a^*(g) = \overline{a}$ from Proposition A.3, and so $\partial a^*/\partial g = 0$ and $\max\{g - a^*(g), 0\} = g - \overline{a}$. Using (8), $du_{plan}/dg = -\beta\delta\lambda < 0$ and so $g^* = \overline{a}$.

Stitching the ranges together gives $g^* = \min\{\hat{a}, \overline{a}\} > \underline{a}$. Parts (ii) and (iii) follow, given that $\overline{a} < \hat{a} \Leftrightarrow \beta\delta\theta + \lambda < \delta\theta$. Finally, (iv) follows immediately from Proposition A.3 ∎

Web Appendix IV.3 Performance uncertainty

Web Appendix IV.3, Part A: Task-based goals (see Web Appendix III.3.2)

For task-based goals, when we add uncertainty as described in the third part of Web Appendix III.3.2 the student-actor's utility (given by (5) and (6) with no uncertainty) and the student-planner's utility (given by (7) and (8) with no uncertainty) become, respectively:

$$Eu_{act}(a|g) = \beta\delta[\pi f(a) + (1-\pi)(0)] - [\lambda \max\{g-a,0\} + C(a)] \quad (36)$$

$$= \beta\delta\pi\theta a - \left[\lambda \max\{g-a,0\} + \frac{\kappa a^2}{2}\right]; \quad (37)$$

$$Eu_{plan}(g|a^*(g)) = \beta\delta^2[\pi f(a^*(g)) + (1-\pi)(0)] - \beta\delta[\lambda \max\{g-a^*(g),0\} + C(a^*(g))] \quad (38)$$

$$= \beta\delta^2\pi\theta a^*(g) - \beta\delta\left[\lambda \max\{g-a^*(g),0\} + \frac{\kappa[a^*(g)]^2}{2}\right]. \quad (39)$$

Comparing (37) and (39) to (6) and (8) without uncertainty, the only difference is that every θ in the case without uncertainty has been replaced by $\pi\theta$. Since $\theta \in (0,1)$ and $\pi\theta \in (0,1)$, the results for task-based goals without uncertainty described in Web Appendix IV.2 continue to hold with uncertainty, replacing every θ with $\pi\theta$.

Web Appendix IV.3, Part B: Performance-based goals (see Web Appendix III.2.2)

For performance-based goals, when we add uncertainty as described in the third part of Web Appendix III.2.2 the student-actor's utility (given by (1) and (2) with no uncertainty) and the student-planner's utility (given by (3) and (4) with no uncertainty) become, respectively:

$$Eu_{act}(e|g) = \beta\delta\{\pi[f(e) - l\max\{g-f(e),0\}] + (1-\pi)[0 - l\max\{g-0,0\}]\} - C(e) \quad (40)$$

$$= \beta\delta\{\pi[e - l\max\{g-e,0\}] - (1-\pi)lg\} - \frac{ce^2}{2}; \quad (41)$$

$$Eu_{plan}(g|e^*(g)) =$$
$$\beta\delta^2\{\pi[f(e^*(g)) - l\max\{g-f(e^*(g)),0\}] + (1-\pi)[0 - l\max\{g-0,0\}]\} - \beta\delta C(e^*(g)) \quad (42)$$

$$= \beta\delta^2\{\pi[e^*(g) - l\max\{g-e^*(g),0\}] - (1-\pi)lg\} - \beta\delta\frac{c[e^*(g)]^2}{2}. \quad (43)$$

Remark A.3
In the absence of a goal the student exhibits time inconsistency:
(i) The student-actor chooses effort $\underline{e} = \beta\delta\pi/c$.
(ii) The student-planner would like the student-actor to exert effort $\hat{e} = \delta\pi/c > \underline{e}$.

Proof of Remark A.3.

In the absence of a goal, the student-actor's utility and the student-planner's utility are given by, respectively:

$$\begin{aligned} Eu_{act}(e) &= \beta\delta[\pi f(e) + (1-\pi)(0)] - C(e) && (44)\\ &= \beta\delta\pi e - \frac{ce^2}{2}; && (45)\\ Eu_{plan}(e) &= \beta\delta^2[\pi f(e) + (1-\pi)(0)] - \beta\delta C(e) && (46)\\ &= \beta\delta^2\pi e - \beta\delta\frac{ce^2}{2}. && (47) \end{aligned}$$

Both utilities are strictly concave since $c > 0$. Straigthforward maximization then gives the result. ■

Proposition A.5

Let $\overline{e} = \beta\delta\pi(1+l)/c$ and recall from Remark A.3 that $\underline{e} = \beta\delta\pi/c < \overline{e}$ denotes the student-actor's effort in the absence of a goal.

(i) When $g \leq \underline{e}$, the student-actor exerts effort $e^* = \underline{e}$.

(ii) When $g \in [\underline{e}, \overline{e}]$, the student-actor exerts effort $e^* = g$.

(iii) When $g \geq \overline{e}$, the student-actor exerts effort $e^* = \overline{e}$.

Proof of Proposition A.5.

Using (41), on the range $e \in [0, g]$:

$$\begin{aligned} \frac{\partial Eu_{act}}{\partial e} &= \beta\delta\pi(1+l) - ce; && (48)\\ \frac{\partial^2 Eu_{act}}{\partial e^2} &= -c < 0; \quad \text{and so} && (49)\\ e^* &= \min\{\overline{e}, g\}. && (50) \end{aligned}$$

Using (41), on the range $e \in [g, \infty)$:

$$\begin{aligned} \frac{\partial Eu_{act}}{\partial e} &= \beta\delta\pi - ce; && (51)\\ \frac{\partial^2 Eu_{act}}{\partial e^2} &= -c < 0; \quad \text{and so} && (52)\\ e^* &= \max\{\underline{e}, g\}. && (53) \end{aligned}$$

(i) When $g \leq \underline{e}$, on the range $e \in [0, g]$, $e^* = g$, and on the range $e \in [g, \infty)$, $e^* = \underline{e}$. Thus, on the range $e \in [0, \infty)$, $e^* = \underline{e}$.

(ii) When $g \in [\underline{e}, \overline{e}]$, on the range $e \in [0, g]$, $e^* = g$, and on the range $e \in [g, \infty)$, $e^* = g$. Thus, on the range $e \in [0, \infty)$, $e^* = g$.

(iii) When $g \geq \overline{e}$, on the range $e \in [0, g]$, $e^* = \overline{e}$, and on the range $e \in [g, \infty)$, $e^* = g$. Thus, on the range $e \in [0, \infty)$, $e^* = \overline{e}$. ■

Proposition A.6

Recall from Remark A.3 that $\underline{e} = \beta\delta\pi/c$ *denotes student-actor effort in the absence of a goal. Recall from Proposition A.5 that* $\overline{e} = \beta\delta\pi(1+l)/c$ *denotes maximal student-actor effort in the presence of a goal.*

Let $\tilde{e} = [\delta\pi - \delta(1-\pi)l]/c$ *and recall from Remark A.3 that* $\hat{e} = \delta\pi/c > \tilde{e}$ *denotes the student-planner's desired effort in the absence of a goal.*

There exists a $\overline{\pi} \in (0,1)$ *such that for all* $\pi \in [\overline{\pi}, 1)$*:*

(i) The optimal choice of goal for the student-planner is given by $g^* = \min\{\tilde{e}, \overline{e}\}$*.*

(ii) When $\beta(1+l) + l(1-\pi)/\pi < 1$, $g^* = \overline{e}$*.*

(iii) When $\beta(1+l) + l(1-\pi)/\pi \geq 1$, $g^* = \tilde{e}$*.*

(iv) Effort of the student-actor $e^* = g^* > \underline{e}$*, and so the student-actor works harder than in the absence of goal.*

Proof of Proposition A.6.

On the range $g \in [0, \underline{e}]$, $e^*(g) = \underline{e}$ from Proposition A.5, and so $\partial e^*/\partial g = 0$ and $\max\{g - e^*(g), 0\} = 0$. Using (43), $dEu_{plan}/dg = -\beta\delta^2(1-\pi)l < 0$ and so $g^* = 0$. Note also that $\lim_{\pi\to 1}(dEu_{plan}/dg) = 0$, and so:

$$\lim_{\pi\to 1}\left[Eu_{plan}(g = \underline{e}) - Eu_{plan}(g = 0)\right] = 0. \tag{54}$$

On the range $g \in [\underline{e}, \overline{e}]$, $e^*(g) = g$ from Proposition A.5, and so $\partial e^*/\partial g = 1$ and $\max\{g - e^*(g), 0\} = 0$. Using (43), and noting that $\overline{e} > \underline{e}$:

$$\frac{dEu_{plan}}{dg} = \beta\delta^2[\pi - (1-\pi)l] - \beta\delta c g; \tag{55}$$

$$\frac{d^2Eu_{plan}}{dg^2} = -\beta\delta c < 0; \quad \text{and so} \tag{56}$$

$$g^* = \min\{\max\{\tilde{e}, \underline{e}\}, \overline{e}\}. \tag{57}$$

Note also that $\lim_{\pi\to 1}(\tilde{e} - \underline{e}) > 0$ and $\lim_{\pi\to 1}(\overline{e} - \underline{e}) > 0$. Thus, on the range $g \in [\underline{e}, \overline{e}]$, $g^* = \min\{\tilde{e}, \overline{e}\} > \underline{e}$ for π sufficiently close to 1.

When $\pi = 1$, from the proof of Proposition A.2, $g^* = \min\{\hat{e}, \overline{e}\}$ gives the student-planner strictly more utility than $g = \underline{e}$. Furthermore, $\lim_{\pi\to 1}(55) = (20)$, $\lim_{\pi\to 1}\tilde{e} = \hat{e}_{|\pi=1}$, $\lim_{\pi\to 1}\overline{e} = \overline{e}_{|\pi=1}$ and $\lim_{\pi\to 1}\underline{e} = \underline{e}_{|\pi=1}$. Thus:

$$\lim_{\pi\to 1}\left[Eu_{plan}(g = \min\{\tilde{e}, \overline{e}\}) - Eu_{plan}(g = \underline{e})\right] > 0. \tag{58}$$

On the range $g \in [\overline{e}, \infty)$, $e^*(g) = \overline{e}$ from Proposition A.5, and so $\partial e^*/\partial g = 0$ and $\max\{g - e^*(g), 0\} = g - \overline{e}$. Using (43), $dEu_{plan}/dg = -\beta\delta^2 l < 0$ and so $g^* = \overline{e}$.

Stitching the ranges together, and using (54) and (58), there exists a $\overline{\pi} \in (0,1)$ such that $g^* = \min\{\tilde{e}, \overline{e}\} > \underline{e}$ for all $\pi \in [\overline{\pi}, 1)$. Parts (ii) and (iii) follow, given that $\overline{e} < \tilde{e} \Leftrightarrow \beta(1+l) + l(1-\pi)/\pi < 1$. Finally, (iv) follows immediately from Proposition A.5. ■

Web Appendix V More details regarding related literature

Web Appendix V.1 Details referred to in footnote 1

Studies using randomized experiments and natural experiments to evaluate the effects of financial incentives on the performance of college students have been inconclusive: Henry et al. (2004), Cha and Patel (2010), Scott-Clayton (2011), De Paola et al. (2012) and Castleman (2014) report positive effects; while Cornwell et al. (2005), Angrist et al. (2009), Leuven et al. (2010), Patel and Rudd (2012) and Cohodes and Goodman (2014) do not find significant effects. Although there is little consensus on the reason behind the failure of many incentive programs, Dynarski (2008) notes that the incentives may be irrelevant for many students, and Angrist et al. (2014) report that one-third of the students in their study failed to fully understand a relatively simple grade-based incentive scheme. In other experiments on college students, academic support services have been combined with financial incentives. Results on the performance effects of these interventions are again mixed: Angrist et al. (2009) and Barrow et al. (2014) report strong effects; Angrist et al. (2014) find weak effects; and Miller et al. (2011) find no significant effects. Financial incentives are also controversial due to concerns that they might crowd out intrinsic incentives to study (see, e.g., Cameron and Pierce, 1994, and Gneezy et al., 2011). See Lavecchia et al. (2016) for a survey of financial incentives in higher education.

Web Appendix V.2 Details referred to in footnote 2

Commitment devices include purchase-quantity rationing of vice goods (Wertenbroch, 1998), deadlines (Ariely and Wertenbroch, 2002), commitments to future savings (Thaler and Benartzi, 2004), long-term gym membership contracts (DellaVigna and Malmendier, 2006), restricted access savings accounts (Ashraf et al., 2006), penalties for failing to hit work targets (Kaur et al., 2015) and Internet blockers (Patterson, 2016), while Augenblick et al. (2015) show that experimental subjects in the laboratory who are more present biased in the domain of work effort are more likely to use a commitment device. See Bryan et al. (2010) for a survey.

Web Appendix V.3 Details referred to in footnote 3

A small and recent literature in economics suggests that goal setting can influence behavior in other settings. Harding and Hsiaw (2014) find that goal setting can influence consumption: energy savings goals reduced energy consumption. Choi et al. (2016) find that goal setting can affect savings: goal-based cues increased savings into 401k accounts. Finally, goals can increase worker performance even in the absence of monetary incentives for achieving the goal: see Corgnet et al. (2015, 2016) for laboratory evidence (although Akın and Karagözoğlu, forthcoming, find no effect of goals), Goerg and Kube (2012) for field evidence and Goerg (2015) for a concise survey.

Web Appendix V.4 Details referred to in footnote 8

Morgan (1987) ran an experiment using one-hundred and eighty college students split into two control groups and three treatment groups. One treatment group set themselves goals for study time, pages to read and topics to cover. The second treatment group self-monitored but did not set goals. The third treatment group did both. The study tracked performance but did not report task completion. Average performance of the treated subjects was higher than that of control subjects. However, the paper does not report a statistical test of the performance difference between subjects who set goals and subjects in the controls.

Web Appendix V.5 Details referred to in footnote 9

A literature in psychology uses small-scale experiments to look at the effects of teacher-set goals on the learning of grade-school-aged children (e.g., LaPorte and Nath, 1976, Schunk, 1983, Schunk, 1984, Schunk and Rice, 1991, Schunk and Swartz, 1993, Schunk, 1996, and Griffee and Templin, 1997). There are important differences between teacher-set goals for grade-school-aged children and self-set goals for college students: first, college students can use self-set goals to regulate optimally their own behavior given their private information about the extent of their self-control problem; and second, in the school environment children are closely monitored by teachers and parents, which gives extrinsic motivation to reach goals (for instance, children might worry about explicit and implicit penalties, monetary or otherwise, for failing to achieve the goal set for them). Using a sample of eighty-four fourth-grade children, Shih and Alexander (2000) explore experimentally the effects of self-set goals (in particular, they study the effects of self-set goals for the number of fractions to solve in class on the ability to solve fractions in a later test).

Web Appendix V.6 Details referred to at the end of Section 4.3.3

In the Introduction we referred to evidence from educational environments that females have more self-control than males (e.g., Duckworth and Seligman, 2005, Buechel et al., 2014, and Duckworth et al., 2015). Consistent with gender differences in self-control, incentivized experiments suggest that men may be more present biased than women. When the earliest payment is immediate (no 'front-end delay'), which provides a test of quasi-hyperbolic discounting, McLeish and Oxoby (2007), Meier and Sprenger (2010) and Prince and Shawhan (2011) find that men are more present biased than women, while Tanaka et al. (2010) find no gender difference in present bias for rural Vietnamese (in Meier and Sprenger, 2010, when controls that include endogenous behavioral measures are included men remain more present biased than women, but the effect is no longer statistically significant). When rural Malawians are given an unexpected opportunity to reverse an earlier commitment, Giné et al. (forthcoming) find that men are more likely to reverse their earlier choice and instead choose an earlier but smaller payment. When the earliest payment is not immediate ('front-end delay'), no gender differences have been found (see Bauer and Chytilová, 2013, for rural Indians and Harrison et al., 2005, where the earliest payment is delayed by a month).