

College Better: Parimutuel Betting Markets as a Commitment Device and Monetary Incentive*

Lester Lusher[†]

University of California, Davis

Job Market Paper

This draft: September 10, 2016

Abstract

Procrastination, an issue linked to poor performance and decreased well-being, is a pervasive problem in education. This paper examines the role of commitment and motivation by evaluating a program called *CollegeBetter.com* which acts as a commitment device and monetary incentive to help college students battle problems of present bias. The zero-sum mechanism is based off a parimutuel betting market, where students join a pool by placing a monetary wager on themselves to achieve the pool's "commitment challenge." Students who successfully commit to the challenge 1) recover their wagers and 2) split losing wagers proportionally. Through a series of lab and field experiments, I find that students interested in the mechanism were low-achieving, overconfident, self-identified procrastinators, while traditional measures of time-preferences (β - δ) were weak predictors of selection. Across all pools, students randomly selected to participate were more likely to achieve the commitment challenge than students who applied for a spot but were randomly excluded. Consistent with loss aversion, having the student risk their own money is a principal contributor to the effectiveness of the mechanism.

Keywords: Higher Education; Commitment Device; Monetary Incentives; Parimutuel Betting

JEL codes: C93, D03, D90, I21

*I am incredibly indebted to the entire faculty at UC Davis for their indispensable help, advice, and mentorship. For feedback and comments, I thank Scott Carrell, Paco Martorell, Dave Rapson, Michal Kurlaender, Sally Sadoff, Timo Rico, Marianne Page, Burkhard Schipper, Barbara Biasi, Felix Friedt, Sharat Ganapati, Brian Duncan, Nancy Jianakoplos, Vasco Yassenov, Michel Grosz, and participants at the University of Chicago Advances in Field Experiments 2016 Conference, the 2016 Spring and Fall UC Davis Applied Microeconomics Brownbag series, the 2016 WEAI Dissertation Workshop, and the 2016 Bay Area Behavioral and Experimental Economics Workshop (BABEEW). For their invaluable research assistance, I thank Rahman Raza, Nicholas Halliwell, Chenyi Luo, Arianne Smith, and Logan Schultheis. This work is supported by the generous funding of the [National Science Foundation \(#1559150\)](#), the [Russell Sage Foundation Behavioral Economics Roundtable](#), and the UC Davis Institute for Social Sciences.

[†]Department of Economics, University of California, Davis, 1 Shields Ave., Davis, CA, 95616. Phone: (510) 938-5223. E-mail: lrusher@ucdavis.edu. Website: lrusher.weebly.com

1 Introduction

An abundance of studies have documented large, increasing returns to attaining a college degree in the United States (Oreopoulos and Petronijevic, 2013). Despite this, young-adult college completion rates in the U.S. have been steadily falling relative to other countries. Since 1995, the U.S. has gone from having the highest young-adult college completion rate among OECD countries to nineteenth.¹ This trend is especially alarming given that college attendance rates in the U.S. have never been higher (Pew Research Center 2014). Thus, more people are entering college than ever before, but degree completion is lagging, where nearly one in two students who enter a four-year college ultimately fail to obtain a degree (Page and Scott-Clayton, 2015). Moreover, the amount of time students take to eventually graduate is also on the rise; since 1972, the average time-to-degree in the U.S. has increased by nearly a year and a half (National Center for Education Statistics 2015).

To understand these trends, it is important to identify the factors that could help explain why students succeed and ultimately graduate from college. An increasing number of studies from economics and education have focused on presenting causal evidence on various university inputs that influence undergraduate success, including capacity constraints and resources (Bound et al., 2010, 2012), professor quality, gender, and race (Hoffmann and Oreopoulos, 2009b; Carrell and West, 2010; Hoffmann and Oreopoulos, 2009a; Carrell et al., 2010; Fairlie et al., 2014), coaching and advising (Bettinger and Baker, 2014; Angrist et al., 2009), academic probation (Lindo et al., 2010), remedial coursework (Martorell and McFarlin Jr, 2011), and teaching assistants (Lusher et al., 2015).

While economists have primarily focused on university inputs, numerous studies from other social sciences have investigated the role of student procrastination in explaining undergraduate success. A meta-analysis from psychologist Steel (2007) suggests that at least 80% of college students suffer from some form of procrastination, where students continually delay studying or completing homework until moments before a deadline. Procrastinating students tend to be more vulnerable to negative consequences such as poor performance and decreased subjective well-being (Schouwenburg, 2004). Furthermore, potentially contributing to the trends in college graduation, studies have documented how students are increasingly facing temptations to procrastinate with the growing popularity of social media sites such as Facebook and Twitter (Knezevic, 2012).

Economists typically think of procrastination as a consequence of present-biased preferences (Laibson, 1997). Under models with present-bias, an agent's preferences change over time in such a way that the agent's choices become inconsistent between two different points in time. For

¹OECD (2014), Education at a Glance 2014, Chart A3.2. Twenty-eight member countries in 2012 were considered for the study. Young-adult college completion rate is defined as the fraction of 18-25 year olds in the population who attain a post-secondary degree.

instance, an agent today may decide that it's optimal to exercise tomorrow; then, come tomorrow, the agent decides that it is no longer in his best interest to exercise. Under certain circumstances, the agent could fall into a cycle of continual deferment of costly activities. To help dynamically inconsistent agents, behavioral economists have focused on the commitment device, a mechanism which lets an agent restrict his future choice set by making "sub-optimal" deviations more costly (O'Donoghue and Rabin, 1999). For example, an agent who wants to exercise tomorrow and recognizes he may reverse his decision could sign a contract to burn \$10 if he fails to exercise. Commitment devices have had varying degrees of success across a multitude of settings, including rotating savings and credit associations (Gugerty, 2007; Ambec and Treich, 2007; Basu, 2011), savings accounts (Thaler and Benartzi, 2004; Ashraf et al., 2006; Beshears et al., 2013), exercise and weight loss (Royer et al., 2015), and smoking cessation (Giné et al., 2010).

This paper investigates the role of commitment in education by evaluating a mechanism that acts as both a commitment device and monetary incentive for college students. Many interventions in education have turned toward using monetary incentives as a means to provide an extra motivation for students (see section 2 for background). While extrinsic incentives have occasionally improved student outcomes, the cost-effectiveness of such interventions is often unclear and ambiguous. More recent work from psychology and economics has shown that behaviorally-inspired incentive schemes can have profound effects on educational performance. For example, Levitt et al. (2016) find that consistent with present bias, rewards had positive effects on primary and secondary student performance, but only when the rewards were given without a delay. Fryer Jr et al. (2012) found evidence of loss averse preferences in teachers, where teacher quality significantly improved when teachers were paid in advance and asked to give the money back if their students did not succeed, relative to teachers who received a deferred reward.

The mechanism outlined in this paper, called *CollegeBetter*, is based off of a parimutuel betting market, where bets are placed together in a pool and payoff odds are calculated by sharing the pool among all winning bettors.² With this mechanism, students commit to their own academic performance by buying into a pool, where each pool has a predefined, common "commitment challenge". If a student achieves the pool's commitment challenge, then the student splits the pot of money with other students who also achieved the commitment challenge. For example, a pool could be defined by the commitment challenge "raise your cumulative grade point average (GPA) by the end of the term." Students then place parimutuel wagers on themselves to achieve this goal. Students who fail to achieve the goal do not recover their wager, while those who successfully raise their cumulative GPAs get to split the pot of money. This mechanism acts as both a commitment device and a monetary incentive in that those who fail to reach the academic goal forego their

²I am the founder of the *CollegeBetter* program, which continues to be offered to undergraduates through the website www.CollegeBetter.com.

wager, while those who achieve the goal make an expected profit.

From a series of lab and field experiments conducted at the University of California Davis, this paper 1) measures selection into the mechanism, 2) estimates causal effectiveness, and 3) explores channels through which the mechanism works. Students interested in the mechanism were more likely to be low-achieving, overconfident, self-identified procrastinators. Traditional measures of time-preferences (β - δ , Laibson, 1997) elicited in the lab via convex time budgets (Andreoni and Sprenger, 2012) were weak predictors of selection. Students randomly selected to participate (i.e. treatment students) were more likely to commit to the challenge of improving their cumulative GPA than students who applied for a spot to participate but were randomly excluded from the pool (i.e. control students). This increased probability of goal attainment was primarily driven by the lowest achieving students entering the term, or students who were at the largest risk of being placed on academic probation.

Furthermore, interesting distributional shifts underline the treatment effects. While students randomly selected to participate were more likely to commit to the pool's challenge, the average difference in GPA between the treatment and control groups was only slightly positive. This is because, relative to the control group, treatment students received term GPAs that were just slightly above their prior cumulative GPAs, while a minority of other treatment students got term GPAs that fell significantly below their prior cumulative GPAs. This first suggests that the achieving students increased effort just enough in order to reach the threshold for the goal. Meanwhile, the minority of other treatment students perhaps became discouraged by the goal as the term progressed, highlighting a potentially perverse effect of goal-setting: once an agent believes a set goal is unattainable, they decrease effort relative to had they never set the goal in the first place.

Finally, through a series of field experiments with additional treatment arms, this paper finds that commitment, defined as the student putting their own money on the line, is a principle contributor to the effectiveness of the mechanism. For each experiment, a random subset of applicants were selected to a treatment arm where they 1) were refunded their wager and 2) received the exact same net monetary benefit as those in the regular *CollegeBetter* pool for the term if they achieved the commitment challenge. These students underperformed relative to the students in the regular *CollegeBetter* pool, and performed similarly to the control group, suggesting that consistent with loss aversion, an important component of the success of the mechanism is having students back their goals by placing their own money on the line.

This paper offers several important contributions to the literature. To my knowledge, this project is the first to introduce a formal commitment device (i.e. a contract with monetary punishments) to education. In other settings, commitment contracts have helped agents battle problems of present bias, and so a natural extension is to investigate how formal commitment devices could benefit procrastinating college students. *CollegeBetter* is also arguably the first program in educa-

tion to appeal to agents' loss aversion (Kahneman and Tversky, 1979). In previous studies (e.g. Levitt et al., 2016; Fryer Jr et al., 2012, see Section 2), subjects were first endowed with the monetary award, and then were threatened to lose the endowment for failing to achieve. One cannot be certain if these framing effects truly shifted the subjects' reference points, or in other words, it is questionable whether the students/teachers in these studies actually felt like they had something to lose for failing. With *CollegeBetter*, since students had to put their own money at risk, they could actually finish the program with less money than had they not participated in the first place. Finally, and perhaps most importantly, *CollegeBetter* is extremely cost-effective to implement since the rewards given to goal-achievers are completely funded by the wagers foregone by those who do not achieve the goal. Hence, even if concern is drawn on the mechanism effectiveness, the program is very unlikely to produce a worse cost-benefit outcome compared to the interventions from previous studies due to the zero-sum nature of payments. Put broadly, this paper helps to shape the knowledge of how behaviorally-inspired incentive schemes can be utilized to boost student performance.

The remainder of this paper proceeds as follows. In the next section, I provide some background of the relevant literature. Section 3 describes the mechanism. Section 4 introduces the lab and field experiments that were offered at the University of California, Davis. Section 4 presents the results. Section 5 concludes. Section 6 discusses concurrent and future experiments of interest.

2 Background

The mechanism outlined in this paper primarily draws inspiration from two branches of research: 1) the commitment device literature from behavioral economics, and 2) the monetary incentives and student motivation literature from education. Each subsection below provides a brief review from these fields of study.

2.1 Commitment Devices

Whether it is to eat healthier, exercise more frequently, or to save money, people are always setting goals. These goals can be set with a wide range of procedure, from a promise to oneself to a formal wager with a friend. People also often fall short of their goals. Why do people set, and fail to accomplish, their goals? Bryan et al. (2010) broadly define a commitment device as an arrangement a person makes to formalize and facilitate their goals. The purpose of the commitment device is to assist an agent in fulfilling an intention for future behavior that would otherwise be challenging due to problems of self-control and present bias. Commitment devices work by restricting the agent's future choice set by making certain choices more costly. So long as the agent is at least partially

aware that their future self deviates from what their present-self defines as optimal future behavior (i.e. a partial sophisticate), then the agent will, on some margin, be willing to pay some cost for the commitment device (O'Donoghue and Rabin, 1999). Commitment devices need not only involve economic penalties; for instance, when an individual makes an internal commitment to be healthy, the individual may feel shame or guilt any time he or she eats junk food.

Examples of informal, ad hoc commitments can be found in many aspects of life, and these arrangements can be viewed as a type of demand for commitment. The majority of the time, these informal commitments fail to help the agent achieve their desired outcome. From Gallup (2015), approximately 58% of smokers would like to quit smoking, and 88% of smokers have made at least one unsuccessful attempt at quitting. A study from Wiseman (2007) followed 3,000 New Year's resolution makers and found that though 52% of the subjects were confident of success, only 12% successfully accomplished their resolution. Over 60% of adults in the U.S. report having tried to lose weight at some point in their life, and nearly 20% of these adults report never succeeding in their attempt(s) to lose weight (Gallup 2011). In education, Ariely and Wertenbroch (2002) found that students were willing to self-impose deadlines when writing essays, and that these students performed better than those who did not set deadlines.

The prevalence of these informal arrangements has motivated researchers to better understand the potential to create markets for more formal commitment contracts. An early study from Thaler and Benartzi (2004) designed a savings commitment device called Save More Tomorrow (SMarT) where employees could voluntarily have a portion of their future salary contributed to retirement savings. Ashraf et al. (2006) introduced a program called SEED (Save, Earn, Enjoy Deposits) via a bank in the Philippines that offered one of two types of a commitment devices: 1) a savings account where deposits would only become available on a future date (with no interest), and 2) an account where deposits would only become available once the account holder reached a certain level of savings; those who were offered these commitment accounts increased their savings by 80% compared to the control group. Similar to SMarT, Breman (2011) tested a commitment device across two field experiments that let donors commit to making a future donation; donation sizes were largest for those offered the commitment device relative to requests for increasing current donations, and these increases persisted after the commitment period. Giné et al. (2010) introduced a commitment banking account called CARES (Committed Action to Reduce and End Smoking) which let smokers recover their account if they passed a urine test for smoking cessation after six months; over 11% of smokers opened an account, and smokers randomly offered CARES were 3 percentage points more likely to pass the smoking cessation test than the control group. Lastly, Royer et al. (2015) found long-lasting behavioral changes in exercise behavior when commitment contracts were combined with financial incentives to attend the gym.

2.2 Monetary Incentives in Education

There are many reasons researchers in education have been interested in utilizing monetary incentives to improve student outcomes. Whether it's because they have time-inconsistent preferences or are underestimating the returns to school, students may be investing too little effort into their own education. Through monetary incentives, students can receive an immediate return that provides an extra motivation to increase effort. Incentives can also be used to motivate parents or teachers to put more effort into their children (e.g. [Glewwe et al., 2010](#)). Overall, incentives have been shown to work well with improving attendance, with mixed results on effort and achievement. Unfortunately, even in cases where monetary incentives lead to positive effects, researchers draw concern about the cost-effectiveness of the program.³ Presented below are just a few studies on monetary incentives in education.

Studies using incentives to promote attendance have overall been positive. These studies typically focus on poorer communities due to the higher opportunity costs of attending school. Studies from [Schultz \(2004\)](#) and [Behrman et al. \(2005\)](#) evaluated a program called PROGRESA in Mexico which paid families an average of \$55 a month for sending their children to school. [Angrist et al. \(2002\)](#) and [Angrist et al. \(2006\)](#) evaluate two programs in Columbia called PACES and Familias en Accion which randomly assigned vouchers to cover part of the cost of attending school for adequate amounts of attendance. For these studies, not only were enrollment and attendance boosted, but so were pass rates and scores on achievement tests. Further evaluation of the Columbia programs from [Barrera-Orsorio et al. \(2011\)](#) found increased enrollment and graduation rates in post-secondary institutions, as well as positive spillovers onto siblings not receiving the subsidy. On the other hand, a recent NBER working paper from [Visaria et al. \(2016\)](#) found that attendance and test scores were negatively affected in an experiment conducted in rural India in response to the introduction and subsequent removal of a reward scheme for school attendance.

The results from studies on the effects of monetary incentives for academic performance are more mixed, and often depend on the type of task. A study from [Bettinger \(2012\)](#) examined monetary incentives for better grades in primary schools in Ohio, where students could receive up to \$100; the program had positive effects for math scores, but not for other subjects. [Fryer \(2011\)](#) conducted a study that distributed \$6.3 million in payments to roughly 38,000 students in over 250 schools to find positive returns when incentivizing educational inputs (e.g. attendance, good behavior, wearing uniforms), while incentives for educational outputs (e.g. grades) were largely ineffective. [Rodriguez-Planas \(2012\)](#) looked at the effects of the Quantum Opportunity Program,

³Other opponents of monetary incentives in education highlight that monetary incentives could “crowd out” the underlying, intrinsic motivations that guide educational decisions. While extrinsic incentives have been shown to occasionally have perverse effects in other prosocial domains ([Frey and Jegen, 2001](#)), there is little evidence of negative crowding-out effects in education.

which offered a combination of monetary rewards and mentoring to high school students for graduating and attending postsecondary school. Angrist et al. (2009) similarly looked at treatments which combined monetary incentives with mentoring for college students. These programs were most efficient when incentives were combined with mentoring, perhaps suggesting that mentoring helps students make goal setting and achievement more concrete. Finally, Levitt et al. (2011) conducted a series of field experiments in a low-performing school district in Chicago which varied the recipients of the monetary reward (students vs. parents) and the structure of the reward (piece rate vs. lottery). If the student met the requirements of the program every month, then they earned an expected value of \$400. Average effects of the program were modest, with the largest effects coming from students on the threshold of meeting the achievement requirements; these marginal students also saw continued success in school after the removal of the monetary incentives.

More recent studies have drawn inspiration from behavioral economics to improve the efficiency of monetary incentives in education. For example, Levitt et al. (2016) conducted a series of field experiments across thousands of primary and secondary school students to find 1) both monetary and non-monetary rewards had positive effects on performance, but only when delivered without a delay, consistent with hyperbolic discounting, 2) suggestive evidence that rewards framed as losses were more effective, and 3) non-monetary incentives were considerably more cost-effective for younger students. Fryer Jr et al. (2012) conducted an across-school experiment to find improvements in teacher quality when teachers were paid in advance and asked to give the money back if their students did not succeed, relative to teachers who received a deferred reward.

Overall, there are many promising results presented from these studies to warrant further investigation on the usage of monetary incentives in education. Perhaps the largest concern, however, is to not only understand how to make these interventions effective, but also to address cost-effectiveness. As Gneezy et al. (2011) conclude in their literature review of incentives in education, “The results of these experiments are somewhat disappointing: given the relatively small effect sizes it is not clear that these programs represent the best return on investment. Yet these are pioneering attempts that did achieve some changes in the investment in education. We see this as encouraging future research which should concentrate on making the incentives more cost effective.”

3 The *CollegeBetter* Mechanism

The mechanism at the center of this paper, called *CollegeBetter*, is loosely based off the parimutuel betting market. These betting markets were first introduced in the early 1900s at race-tracks, and are often used today in betting on sporting events and office pools. With parimutuel betting systems, bets of a particular type are placed together in a pool, and payoff odds are calcu-

lated by sharing the pool among winning bettors. With *CollegeBetter*, students place parimutuel wagers on their own academic success. The mechanism works by first setting up a betting pool defined by a “commitment challenge”. A student i can join the pool by placing a wager b_i on themselves to successfully accomplish the commitment challenge. Wagers are collected to create a pot size of $B = \sum b_i$. After the defined commitment period is over, university transcripts are used to verify which students achieve the commitment challenge. Those who do not succeed do not recover their wager b_i ; those who do succeed get to split the pot B proportionally according to their wager size. Setting $W = \sum_j b_j$ where j is an index for students who completed the commitment challenge, successful student i ’s revenue is $(\frac{b_i}{W} * B)$ for a net return of $(\frac{b_i}{W} * B - b_i)$. An unsuccessful student’s revenue is zero for a net return of $-b_i$. Note that if all students successfully complete the commitment challenge (i.e. $W = B$), then each student i would simply recover their wager b_i for a net return of 0. Thus, the baseline for this mechanism is the classic formal commitment device, where an agent’s account is returned for successful commitment.⁴

For the *CollegeBetter* experiments discussed in this paper, a pool was set up each term with a two-part commitment challenge. First, the student had to successfully raise their cumulative grade point average (GPA).⁵ In other words, the student had to get a term GPA that was higher than the student’s cumulative GPA entering the term. If the student had no cumulative GPA entering the term, perhaps because it was the student’s first term at UC Davis, then the student was not eligible to participate. Secondly, the student had to enroll in a minimum number of units for a letter grade. This requirement reduces the concern of creating an incentive for the student to enroll in an “easy” schedule of courses. For simplicity, the wager amount allowed for each pool was fixed across all participants such that each successful student would receive the same net return.

4 Experiments

Starting with the Winter 2015 quarter, an experimental pool was set up each term at the University of California, Davis. In this paper, a total of five experimental pools will be discussed: Winter 2015, Spring 2015, Fall 2015, Winter 2016, and Spring 2016.⁶ Experiments across terms differed in various ways, as described below, in order to 1) measure selection into *CollegeBetter*,

⁴In parimutuel betting markets involving sporting events, the manager usually takes a cut of the pool, called the vigorish, before redistributing the winnings. In all *CollegeBetter* pools, no vigorish was charged, thus guaranteeing that students who succeeded would, at a minimum, get their buy-in fee back.

⁵Grade Point Average (GPA) is the primary evaluation measure utilized by U.S. universities for student success. GPA is calculated by dividing the total amount of grade points earned by the total amount of enrolled units attempted. More grade points are awarded for higher letter grades (e.g. $A = \frac{12}{3} = 4.0$, $A- = \frac{11}{3} \approx 3.7$, $F = 0$).

⁶Experimental pools continue to run today through the website www.CollegeBetter.com. Section 6 considers alternative variations to the *CollegeBetter* experiments to identify other questions of importance on student motivation and success. The agreement with UC Davis is to continue running an experimental pool every term through the 2017-2018 school year.

2) estimate causal effectiveness, and 3) explore the degree to which the monetary commitment can explain the effectiveness of the mechanism. In all experiments, information about student participation was kept private.⁷ The experiments are introduced in subsections below, presented in chronological order, and are also briefly summarized in [Table 1](#).⁸

4.1 Winter 2015 - Experiment in Economics courses

To start, one cannot simply let students sign up for a *CollegeBetter* pool, then compare the academic performances of students who sign up against those who do not, as this estimate will certainly suffer from endogeneity bias; those who sign up for *CollegeBetter* are different in unobservable ways from those who do not sign up, and these unobserved factors likely correlate with academic success. Hence, one needs to consider an experiment in order to generate a proper reference group to compare to the students who do participate in *CollegeBetter*.

The first *CollegeBetter* experiment was conducted in January of 2015. Along with a team of research assistants, I went into several undergraduate Economics courses at UC Davis, described the *CollegeBetter* program, and asked students to complete a short survey.⁹ Students who completed the survey were entered into a raffle to win one of several cash prizes. All survey-takers agreed to release their academic transcripts at the end of the quarter. The survey included an application to join the *CollegeBetter* pool for the term. Recall that the commitment challenge for all experimental pools required 1) the student to raise his/her cumulative GPA by the end of the term, and 2) to take a minimum number of units for a letter grade for the term. Students interested in signing up for the *CollegeBetter* pool would complete the application and submit a \$10 wager on themselves to achieve the commitment challenge.

Of the 412 students who took the survey, 169 completed an application to join the *CollegeBetter* Winter 2015 pool. Then, approximately half of the 169 applicants were randomly selected to participate in the *CollegeBetter* Winter 2015 pool, while the excluded half were immediately refunded their \$10 wager. The randomization was carried out after the last day students could add or drop classes, thereby ensuring that students were not changing the composition of their classes after observing whether they were randomly selected to participate or not. By comparing the performance of students who entered the *CollegeBetter* Winter 2015 pool against those who wished to enter the pool but were not randomly selected to do so, this experiment can generate an

⁷Selection into and the effectiveness of the *CollegeBetter* mechanism would likely be influenced were participation in the program made publicly available, as prior studies have linked academic success to social interactions and competitions ([Babcock, 2008](#)).

⁸All experiments received approval from the Institutional Review Board at UC Davis.

⁹While I am the sole author of this research project, the experiments certainly would not have been possible without the assistance of many university administrators and undergraduate research assistants. Thus, at various points in this paper, I will use “we” and “us” to simply refer to the full team of individuals who helped conduct the experiments.

unbiased point estimate for the causal effectiveness of *CollegeBetter* for those who sign up for the mechanism. We can also contrast survey responses of those who completed the application against those who did not to predict selection into *CollegeBetter*. See [Appendix A.1](#) for a complete list of survey questions for the Winter 2015 experiment.

4.2 Spring 2015 - Experiment brought online

Following the Winter 2015 experiment, the website www.CollegeBetter.com was launched primarily to improve the efficiency of the signup process. Through the website, students could complete an application to join the *CollegeBetter* Spring 2015 pool and submit their \$10 wager electronically via credit card, PayPal, or Venmo. Students completing the application were also required to email us a receipt of their preordered university transcript to be sent at the end of the Spring 2015 term. Applications were accepted until the last day students could add or drop classes. Then, similar to the Winter 2015 experiment, approximately half of the applicants were randomly selected to participate, while the remaining half were excluded from the pool and were refunded their \$10 wager. For this pool, www.CollegeBetter.com was promoted through emails sent to students enrolled in Economics courses for the term. Again, this experiment helps to generate an unbiased estimate for the effectiveness of *CollegeBetter*. A total of 77 students fully completed the online application.¹⁰

4.3 Fall 2015 - Time-preference elicitation, commitment, & loss aversion

For the Fall 2015 *CollegeBetter* experimental pool, signups continued online in similar manner to the Spring 2015 experiment, except a few small edits were made. First, the wager amount was increased to \$20. Second, student applicants would, after reading a terms and conditions form, provide an online signature agreeing to release their university transcripts; this substantially simplified the overall application process. Third, recruitment was no longer restricted to emailing students enrolled in Economics courses. After the success of the Winter 2015 pool in Economics courses, the UC Davis Institutional Review Board granted us permission to send emails to faculty and advisers across all UC Davis departments asking them to forward information about www.CollegeBetter.com to their students.

Furthermore, two major additions were included in order to (1) predict selection into *CollegeBetter* as a function of elicited discount rates and present-bias (δ , β) and beliefs about the monetary

¹⁰The Spring 2015 experimental pool had the fewest number of subjects primarily due to incomplete applications, where many students failed to send us a receipt of their pre-ordered university transcript. This component of the application was altered in the subsequent Fall 2015 term.

returns to accomplishing the commitment challenge, and (2) identify the importance of commitment and loss aversion for *CollegeBetter* participants.

For (1), we recruited 120 undergraduates at UC Davis to participate in a laboratory experiment. The lab experiment was advertised as a decision-making experiment where subjects would earn money based on decisions made on a computer screen in a lab. In this lab experiment, similar to the experiments conducted in the Economics courses in Winter 2015, students were first given information about the *CollegeBetter* pool for the term. After accepting or rejecting the invitation to sign up at www.CollegeBetter.com, students completed a survey of questions intended to elicit beliefs about the monetary returns to participating in *CollegeBetter*, as well as the student’s time preferences and propensity to be present biased.

Students’ time preferences were elicited through the Convex Time Budget (CTB) method introduced by [Andreoni and Sprenger \(2012\)](#). With the CTB method, one can easily identify time discounting and utility function curvature at the individual level, given simple structural assumptions. Importantly, the design also tests for individual present bias and hyperbolic discounting, and as discussed in the [Section 2.1](#), these are factors that economists typically link to procrastinating behavior and demand for commitment. During the survey, students faced a series of allocation problems, where each allocation problem was characterized by dividing a number of cash-valued tokens m between period t and period $t + k$. The relative cash-value of the token in period $t + k$ versus period t was determined by an interest rate r . Each allocation problem differed by t , $t + k$, and r . Variations in k and r help identify discounting and utility function curvature, while variations in t identify present bias. In short, for each allocation problem, subjects are maximizing some assumed utility function $U(c_t, c_{t+k})$ subject to the convex budget set $(1 + r)c_t + c_{t+k} = m$. At the end of the lab experiment, one allocation problem was executed at random to determine each subjects payment.¹¹ [Table 2](#) captures the 45 different allocation problems each subject faced.

Eliciting a student’s expected monetary returns to participating in *CollegeBetter* requires estimating two parameters: the student’s belief about the monetary return to succeeding, and the likelihood the student him or herself succeeds. To estimate the former, we simply asked subjects “How much money do you predict each winner of *CollegeBetter* will earn, net of the \$20 buy-in?” and incentivized the question by awarding \$15 to the subject for guessing within \$0.15 of the true net earnings of participants at the end of the term. There are several ways to proxy for a student’s beliefs about their own likelihood of success. Objective measures include the number of units the

¹¹Other measures were taken in order to equalize the subjects’ costs of receiving payments in period t versus period $t + k$ (See [Andreoni and Sprenger, 2012](#), for further discussion). For example, we sent two small participation payments of \$2.50 in both period t and period $t + k$, regardless of the subjects choices, and added the experimental earnings from the CTB mechanism to these two participation payments. This eliminates any convenience the subject may gain by concentrating tokens into one period, since two payments would be sent regardless. Payments were instantly wired to students’ account via PayPal or Venmo.

student enrolled in for the term. We also included questions in the survey such as “How difficult is your course load this quarter compared to previous quarters? More/As/Less difficult.” See [Appendix A.2](#) for a complete list of survey questions from the Fall 2015 lab experiment.

For (2), to identify the importance of commitment and loss aversion for *CollegeBetter* participants, we included an additional treatment group to the baseline design. So, for the Fall 2015 term, applicants were randomly assigned to one of three groups:

- (a) Control group: student refunded \$20 wager and excluded from the *CollegeBetter* pool
- (b) Treatment group: student entered into the regular *CollegeBetter*
- (c) Treatment group: student refunded \$20 wager and told they would receive same net winnings as (b) for successful commitment

Thus, those in treatment group (c) received the same net monetary benefits, conditional on accomplishing the goal, as those in treatment group (b), but did not have to commit their own money into the pool.¹² During the application period, students were not made aware of treatment group (c); this is important in order to keep focus on the same population of students who were solely interested in *CollegeBetter*, and not pick up a subset of students who were motivated by the possibility of making money while avoiding the \$20 wager.¹³ So long as the effect of formal commitment, defined as losing money for failure to achieve the goal, is additively separable from all other components entering the *CollegeBetter* production function (e.g. monetary incentives), the difference in performance between the two treatment groups (b) and (c) captures the effect of the monetary commitment; this is because (c) encapsulates everything captured in (b) with the exception of the student having their own money on the line to lose. Furthermore, the difference between the treatment group (c) and the control group (a) can reveal the effect of monetary incentives on student performance, and can provide results that are easily comparable to the previous interventions seen in education ([Gneezy et al., 2011](#)). A total of 247 students completed an application, and these students were split evenly across the three groups - 83 in (a), 82 in (b), and 82 in (c).

4.4 Winter & Spring 2016 - Further identification of commitment

Finally, the Winter 2016 and Spring 2016 experimental pools were used to further identify the importance of having *CollegeBetter* participants putting their own money on the line. Applications

¹²For example, suppose 100 students were selected into treatment group (b). This would create a pot size of $\$20 \times 100$, or \$2,000. If 60 of the 100 students successfully committed to the challenge, then each student in treatment group (b) would take home $\$2,000/60$, or \$33.33, for a net gain of $\$33.33 - \$20 = \$13.33$. Each student in treatment group (c) then would also take home \$13.33 if they successfully accomplished the commitment challenge.

¹³Applicants who were randomly selected for treatment group (c) for this experimental pool were told that the group was a one-time special event that was funded for by the Russell Sage Foundation.

continued to be offered online through www.CollegeBetter.com. For the Winter 2016 pool, a total of 247 students fully completed the application, 82 of which were randomly assigned to (a), 82 to (b), and the remaining 83 to (c).

For the Spring 2016 pool, the control group (a) was removed, so applicants were randomly selected to either (b) the regular *CollegeBetter* pool or (c) the treatment group where the \$20 wager was refunded and the student was told they would receive the same conditional net winnings as group (b) for accomplishing the commitment challenge. At the start of the Spring 2016 term, an unexpected email policy change hamstrung our primary recruitment effort. As a consequence, the decision to remove control group (a) from the Spring 2016 experimental pool was driven by a desire to maximize power in identifying the importance of commitment and loss aversion. A total of 141 students completed an application, 70 of which were assigned to (b) and 71 to (c).

5 Results

This paper seeks to answer three questions related to the *CollegeBetter* mechanism: (1) What type of students sign up for the mechanism? (2) What is the effectiveness of the mechanism? And (3) to what degree can the monetary commitment explain the effectiveness of the mechanism? Results from the experiments are divided according to these three questions in subsections below. Refer to [Table 1](#) for a summary of which of the three research questions each experiment seeks to answer. Since experimental pools were set up across multiple terms, students could participate in multiple experiments. For simplicity, all results presented below only focus on the first experimental pool each student participated in.¹⁴

5.1 Selection into *CollegeBetter*

Two of the five experiments help to identify selection into the *CollegeBetter* mechanism: [Winter 2015](#) and [Fall 2015](#). For the Winter 2015 experiment, a total of 412 students from Economics courses were invited to apply for a spot in the *CollegeBetter* experimental pool. Of these 412 students, 169 expressed interest in joining the *CollegeBetter* pool.

The first two columns of [Table 3](#) display summary statistics by whether the student applied for a spot in the [Winter 2015](#) pool (“Yes” vs. “No” under “Interested in *CollegeBetter*?”), and the third column calculates whether the difference between interested and uninterested students is sta-

¹⁴Understanding the potential long run impacts of *CollegeBetter* on student selection and success is an interesting research question. Unfortunately, the university did not agree to provide transcripts for students beyond the term they participated in the experiment. Thus, I only observe grades across terms for students who signed up for multiple experiments. Focusing on the first experimental pool for each student removes any potential long run effects confounding the estimates, and thus, results can be interpreted as the contemporaneous effect of *CollegeBetter* on student outcomes.

tistically significant across each covariate. For instance, students interested in *CollegeBetter* had significantly lower cumulative GPAs entering the Winter 2015 term (2.947) than students who did not apply for a spot in the *CollegeBetter* pool (3.244). This directly implies that students interested in the mechanism are lower-performing, and perhaps underachieving relative to their belief about their true ability. Gender is also a significant predictor of selection, where 58.6% of *CollegeBetter* applicants were male compared to just 41.7% of non-applicants. Applicants were also significantly more likely to agree with the statement “I often start studying for exams too close to the exam date such that I feel regret,” providing some suggestive evidence of sophistication and demand for commitment against present bias (O’Donoghue and Rabin, 1999). Finally, *CollegeBetter* applicants shared beliefs that the fraction of *CollegeBetter* participants that would successfully commit to the goal was relatively high.¹⁵ At the bottom of Table 3, I regress an indicator for whether the student applied for *CollegeBetter* on the full set of covariates to find that the estimated coefficients are jointly significant, and that there is little concern for multiple hypothesis testing.

In the Fall 2015 term, 120 students in a lab experiment were invited to join the *CollegeBetter* program for the term, 57 of which completed the application online. Then, the students completed a survey of questions, including a series of 45 Convex Time Budget (CTB) allocation decisions (Andreoni and Sprenger, 2012). Given simple structural assumptions on the utility function and the nature of discounting, each student’s time preferences, present bias, and utility function curvature were estimated using the student’s responses to the CTB allocation decisions. Similar to the majority of experimental studies on time preferences, my primary analysis assumes a time-separable constant relative risk aversion (CRRA) utility function with quasi-hyperbolic discounting (Strotz, 1955; Phelps and Pollak, 1968; Laibson, 1997),

$$U(c_t, c_{t+k}) = \frac{1}{\alpha}(c_t)^\alpha + \beta\delta^k \frac{1}{\alpha}(c_{t+k})^\alpha, \quad (1)$$

where α is the CRRA curvature parameter, δ is the one-day discount factor, β is the present bias parameter, and c_t and c_{t+k} correspond to the subject’s experimental token allocations. Nested within this quasi-hyperbolic function is the classic exponential discounting model when $\beta = 1$. The individual is said to be present-biased when $\beta < 1$, and β does not enter the utility function for $t > 0$. Present-biased agents of the quasi-hyperbolic model are often used by economists to rationalize settings where individuals continually delay activities with short-run costs and long-run benefits, such as with procrastinating college students. Maximizing (1) subject to the convex

¹⁵Students’ beliefs were elicited with the following survey question: “Make a guess! What fraction of students who participate in *CollegeBetter* will accomplish the goal of raising their cumulative GPA? Whoever guesses closest to the true success rate will win \$50 at the end of the quarter.”

budget set $(1 + r)c_t + c_{t+k} = m$ yields the demand function

$$c_t = \begin{cases} \left[\frac{(\beta\delta^k(1+r))^{\left(\frac{1}{\alpha-1}\right)}}{1+(1+r)(\beta\delta^k(1+r))^{\left(\frac{1}{\alpha-1}\right)}} \right] \times m & \text{if } t = 0 \\ \left[\frac{(\delta^k(1+r))^{\left(\frac{1}{\alpha-1}\right)}}{1+(1+r)(\delta^k(1+r))^{\left(\frac{1}{\alpha-1}\right)}} \right] \times m & \text{if } t > 0 \end{cases} . \quad (2)$$

Each individual's α , β , and δ can be extrapolated when estimating (2) using non-linear least squares (see [Appendix A.3](#) for more detail).¹⁶

[Table 4](#) presents the results from the [Fall 2015](#) lab experiment, and is laid out in similar fashion to [Table 3](#). Of all the covariates, only two come up as statistically significant. Roughly 39% of those who applied for *CollegeBetter* agreed with the statement that the classes they were enrolled in for Fall 2015 were more difficult than average, which is significantly lower than the 67% of non-applicants. *CollegeBetter* applicants also shared a higher belief that their GPAs would increase compared to non-applicants. Other measured differences in covariates, though statistically insignificant at levels below 10 percent, reveal interesting aspects about selection. Both measures of time-preferences (β - δ) are very weak predictors of selection. Applicants shared beliefs that the monetary returns, conditional on succeeding, for participating in *CollegeBetter* were higher than non-applicants. Looking at the two survey questions that were significant in the Winter 2015 experiment, we again find that students interested in *CollegeBetter* were more likely to agree with the statement that they procrastinate too much when studying for exams, but the sign on gender reverses; in the Fall 2015 experiment, females were more interested in *CollegeBetter* than males.

Coupling the results from these two experiments, one can make some fairly definitive conclusions about the channels of selection into the mechanism. Interested students are poorer performers overall (lower GPAs) who are confident that they can raise their grades. This suggests that interested students are those who believe that their prior poor performance cannot be explained by something “permanent” that is inherent to the student, such as a lack of ability. Instead, these students perhaps lacked motivation in prior terms, or had low cumulative GPAs due to “bad luck” in prior terms. Note, however, that for both the [Winter 2015](#) and [Fall 2015](#) experiments, years spent at UC Davis do not predict selection; if “bad luck” were the primary driver of selection, then one would assume younger students (e.g. Freshmen) would be more likely to apply since

¹⁶One drawback with using NLS is that without additional distributional assumptions, it cannot account for censored data issues built into corner solutions. Additional analyses considered that are omitted from this paper include 1) alternative utility function assumptions including CARA and CRRA with various Stone-Geary parameters and 2) estimation of the derived tangency condition from maximizing equation (1). [Appendix Table A.1](#) presents distributional statistics on elicited parameter estimates for this study and the [Andreoni and Sprenger \(2012\)](#) study. [Appendix Table A.2](#) reports elicited time preferences and present bias parameters for each individual subject.

their cumulative GPAs would be more adversely affected by a negative shock, while older students (e.g. Seniors) would not sign up since their cumulative GPAs are weighted across many terms of enrollment and thus would not be influenced by a bad term.

Meanwhile, the evidence of selection due to commitment is mixed. On one hand, applicants are more likely to be self-identified procrastinators, but on the other, lab elicited present bias β does not predict selection, which either suggests commitment is not a driver of selection, or that the students are underestimating the extent to which they need commitment to battle their self-control problems (i.e. naïfs from O'Donoghue and Rabin, 1999). Alternatively, as the literature on time preference elicitation is still growing, there could simply be a disconnect between lab-elicited present bias via Convex Time Budgets (Andreoni and Sprenger, 2012) and actual procrastination in real-life effort (e.g. Augenblick et al., 2015).¹⁷

5.2 Effectiveness of *CollegeBetter*

In order to identify the causal effectiveness of *CollegeBetter*, we can simply compare the performance of students who applied for a spot in the *CollegeBetter* pool and were randomly selected to participate against those who applied for a spot but were randomly excluded from the pool. Each of the Winter 2015, Spring 2015, Fall 2015, and Winter 2016 experiments have a group of students who participated in *CollegeBetter* and a control group, allowing for the opportunity to measure the effectiveness of *CollegeBetter*.

Balance test

To start, I want to first ensure that the experiment is balanced across covariates. That is, I test whether there are any statistically significant differences in observable characteristics between the treatment groups (applicants randomly selected to participate) and control groups (applicants randomly excluded from the *CollegeBetter* pools) in order to address any potential concerns of unobservables that correlate with assignment and student outcomes. To do this, I utilize the Winter 2015 experiment since each student who completed an application also completed the survey of questions administered in Economics courses.¹⁸ The final three columns of Table 3 display average student survey responses by group and calculate whether the difference between the treatment and control group is statistically significant for each survey response. Two of the 14 characteristics

¹⁷A recent NBER working paper provides a thorough discussion on time preference elicitation with monetary decisions versus time-dated consumption and effort, and explains why variations in monetary decision experiments could be driven by factors other than underlying time preferences (Cohen et al., 2016).

¹⁸For the Spring 2015, Fall 2015, and Winter 2016 pools, the only student characteristics I observe correspond to information on the students' transcripts (e.g. class-standing, previous GPA). On these covariates, I find no statistically significant differences between the treatment and control groups.

come up as significant at the 90% level; applicants randomly accepted into the *CollegeBetter* pool were less likely to be an international student and more likely to agree with the statement “I wait until the last second to buy gifts (birthday, Christmas, etc.).” Given the testing is done across a multitude of variables, these two variables could have come up as significant merely by chance. At the bottom of the final column of [Table 3](#), the F -statistic and p -value are reported from a test for whether the coefficients are jointly equal to zero from a regression of an indicator for whether the applicant was randomly selected to participate on the full set of student observables. With a p -value of 0.873, I fail to reject the null hypothesis that the coefficients are jointly equal to zero and conclude that the Winter 2015 experiment is well balanced.

Main results - commitment challenge completion rates

Next, I combine the results from the [Winter 2015](#) experiment with the remaining experimental pools ([Spring 2015](#), [Fall 2015](#), and [Winter 2016](#)) into one figure to evaluate the effectiveness of *CollegeBetter*. Recall that for these latter three terms, the experimental pools were carried out online via www.CollegeBetter.com, where students could apply for the term’s pool by providing identifying information and their parimutuel wager. For the Winter 2015 and Spring 2015 pool, the wager amount was \$10, while for Fall 2015 and Winter 2016, the wager amount was \$20. [Figure 1](#) presents mean commitment challenge completion rates by experiment and by treatment group. I find that for all four experiments, students randomly selected to participate were more likely to achieve the commitment challenge (raise cumulative GPA while enrolling in sufficient number of units) than students who applied for a spot in *CollegeBetter* but were randomly excluded. When the results across all four experiments are pooled together, such that each student gets equal weight irrespective of term, I find that treatment students were nearly nine percentage points more likely to achieve the commitment challenge, and this estimate is statistically significant at the five percent level.

Taking a step back from evaluating differences in relative performance, and looking strictly at the levels of performance by treatment group, I find some additionally interesting results. For instance, the mean commitment challenge completion rate of the control group (students who applied for a spot but were randomly excluded) was *under* 50%. That is, less than half of the students who wanted to place a wager on themselves, but were not randomly selected to do so, actually raise their grades.¹⁹ Couple this result with the selection story from [Table 4](#) in the [previous results section](#), where *CollegeBetter* applicants were very confident about their chances of raising their grades for the term, and what is left is a severe disconnect between the students’ percep-

¹⁹In separate work I’ve done for UC Davis looking at time-to-degree, I find that students’ grades tend to trend upward through their time at Davis (e.g. Seniors tend to get higher grades than Freshmen). This trend is also true for the population of students who select into *CollegeBetter*.

tion of their likelihood of increased future performance and their actual subsequent performance. Simply put, students who sign up for the mechanism are, on average, overconfident, and tend to overestimate their own likelihood of succeeding.

Turning to the treatment group, the mean commitment challenge completion rate was roughly 56%, so successful committers pocketed nearly an 80% net return on their wager.²⁰ Still, for the \$20 buy-in group, this amounts to a mere \$16 of net profit. Consequently, the mechanism becomes a \$20 loss for failure and a \$16 gain, in expectation, for successful commitment. Overall, the estimated positive treatment effects are extremely promising, particularly when considering, ex post, how the stakes correspond to an under-\$40 changing of hands.²¹

Figure 2 shows that the increases in commitment cannot be explained by treatment students taking fewer units and, consequently, enrolling in an easier course load. Though I carried out the randomization after the last day students could add or drop courses, students could feasibly partake in other extreme methods to drop their courses.²² I find that treatment students enrolled in fewer units in only two of the four experiments, and that for these two experiments, the differences in mean number of units are statistically insignificant (p -values of 0.77 and 0.90 for Spring 2015 and Fall 2015, respectively). Treatment students in the Winter 2016 experiment enrolled in statistically significantly *more* units on average than the control students (p -value=0.09), suggesting, if anything, that treatment students maintain a more difficult course load in order to ensure they meet the minimum unit requirement for the commitment challenge. When pooling the students across all four experiments, I find no statistically distinguishable difference between treatment and control students in mean number of enrolled units.

Similar to Figure 1, Figure 3 presents mean commitment challenge completion rates by term and by treatment group for the sample of students who had cumulative GPAs of 2.50 and lower entering the term. Students in this GPA are at a heightened risk of being placed on academic probation since they average more letter grades of C's, D's, and F's than B's and A's. I find that the treatment effects are largely driven by this subsample of students. That is, the poorest performing students who were randomly selected to participate significantly outperform the poorest performing students who applied for a spot but were randomly excluded (65.71% vs. 44.83% success rate). This result is perhaps unsurprising: those with poorer past performance have mathematically more margin to achieve the commitment challenge. On the other hand, there are reasons why these stu-

²⁰In pools with equal wagers, a successful student's net return is equivalent to $(\frac{1}{p} - 1) \times 100\%$, where p is the proportion of participants who successfully committed (i.e. $p = W/B$ from section 3.)

²¹The relatively large net returns for goal-achievers are also promising with regard to the long run implications of *CollegeBetter*. Businesses similar to *CollegeBetter* that focus on other domains (e.g. exercise) have documented poor long-run participation rates since nearly all participants succeed, meaning the majority of goal-achievers simply earn their buy-in back.

²²For example, students can receive an "Incomplete" grade at the end of the term if they provide a valid excuse for why they missed an integral class assignment (e.g. the final exam).

dents have poor GPAs, one of which could be the disconnect between their perceived likelihood of positive future performance and actual performance. Furthermore, over half of those in the control group, though they have a large margin to improve, see their GPAs drop. In total, from [Table 3](#) I found that poorer performing students were more likely to select into the mechanism, and conditional on selection, [Figure 3](#) shows that lower performing students respond to the treatment, and so, the mechanism is overall serving those students who are perhaps in the most need of increased motivation and improved performance.

Average term GPAs and distributional effects

Next, I investigate changes in the magnitude of student GPA in response to participation in the mechanism. In [Table 5](#), I consider a regression of term GPA on an indicator for whether the student was randomly selected to participate, while also controlling each student's pre-term cumulative GPA, number of term units, and number of pre-term cumulative units. Each column corresponds to a single regression for an experiment of interest. The top row presents the coefficient on the indicator for whether the applicant was randomly selected to participate. These regressions suggest little to no differences in average term performance between treatment and control students. For the Winter 2015 experiment, treatment students saw a minimal boost in term GPA (0.087), but for the remaining three experimental pools, treatment students actually experienced slightly *lower* average term GPAs than control students. Across all four experiments, estimates are statistically indistinguishable from zero.²³

The results from [Figure 1](#) and [Table 5](#) may, at first, seem conflicting: How can treatment students both (1) be more likely to achieve the commitment challenge of improving their grades and (2) attain term GPAs that are, on average, only slightly larger than the control group? [Figure 4](#) reveals an interesting story to help reconcile these findings, where I plot smoothed histograms of the distributions of GPA differential by treatment and control group. That is, for each student, I subtract the student's pre-term cumulative GPA from their term GPA, and then plot the estimated kernel density. First, note that significant bunching of GPA differentials occur just above zero for treatment students; in other words, a significant amount treatment students received term GPAs that were just slightly above their prior cumulative GPAs. This suggests that achieving treatment students increased effort just enough to reach the threshold for the commitment challenge. Meanwhile, a minority of other treatment students got term GPAs significantly below their prior cumulative GPAs relative to the control group. The bunching of GPA differentials around -1 for treatment students comes from leftward movement in the distribution, where these students would

²³[Appendix Figure A.1](#) plots the corresponding mean term GPAs by treatment group and by term. [Table A.4](#) and [Table A.5](#) present results from regressions where I consider an indicator for whether the student completed the commitment challenge and the number of units the student enrolled in for the term, respectively, as outcome variables.

have otherwise had GPA differentials between 0 and -1, as implied by the distribution for the control group.²⁴

To test for whether the difference in the distributions is statistically significant at different parts of the distribution, I present a series of results from quantile regressions. First, [Table 6](#) shows that median treatment students had 0.081 larger GPA differentials than median control students, while the 10th percentile of treatment students saw their GPA differential decrease by 0.131 grade point units relative to control students. [Figure 5](#) plots the coefficients on a treatment indicator from a sequential-quantile regression estimated at each decile with 90% confidence intervals. This graph again shows that a minority of treatment students received sharp, but statistically insignificant, decreases in their GPAs relative to control students. Meanwhile, the largest treatment gains are experienced by the third, fourth, and fifth deciles of the distribution, each of which exhibit statistical significance at least at the 10% level.

These results highlight several potentially limiting effects of *CollegeBetter* and of goal-setting in general. Students in the treatment group, upon realizing that the goal was no longer attainable, possibly decided to decrease effort relative to had the goal not been set in the first place. Furthermore, successful treatment students did not exert additional effort to go above and beyond their goal, suggesting that the commitment device no longer provided any motivation once the students reached the threshold.

5.3 Importance of commitment for *CollegeBetter* participants

Lastly, I turn to the [Fall 2015](#), [Winter 2016](#), and [Spring 2016](#) experimental pools in order to identify the importance of commitment and loss aversion for students who participate in the mechanism. Recall that each of these terms included an additional treatment group where applicants were refunded their initial \$20 wager, but were given the same monetary incentive as those who participated in the regular *CollegeBetter* program, conditional on achieving the commitment challenge. If students in this additional treatment group perform significantly worse than those in the regular *CollegeBetter* mechanism, then one can conclude that commitment and loss aversion, defined as the student incurring monetary losses for unsuccessful goal attainment, are primary drivers of student success in the mechanism. If students across these two groups perform roughly the same, then there must be some other component of the mechanism that is driving the student success (e.g. monetary gains).

[Figure 6](#) presents the results from these three terms. Recall that students in the [Spring 2016](#) experimental pool were randomly selected to either the regular *CollegeBetter* group or the non-monetary commitment group (i.e. there was no randomly excluded group). Across all three ex-

²⁴The p -value from a two-sample Kolmogorov-Smirnov test (see [Conover, 1999](#)) for equality of distribution functions between the treatment and control groups is 0.110.

perimental pools, I find that those who participated in the regular *CollegeBetter* group were more likely to achieve the commitment challenge of raising their GPA than those in the non-monetary commitment group. In other words, students who put their own money at risk, versus those whose buy-in was covered by the experimenters, were more likely to succeed. When pooled together, the mean performance difference between these two groups is seven percentage points (58.9% versus 51.9% commitment challenge completion rate). Note, however, that this difference is statistically insignificant at levels below 10 percent (p -value of 0.154). From the [Fall 2015](#) and [Winter 2016](#) experiments, we see that students in the non-monetary commitment group performed only slightly better than the group of applicants who were excluded from the mechanism entirely (51.9% versus 49.6% commitment challenge completion rate); this difference is statistically indistinguishable from zero (p -value of 0.678). Thus, the commitment rates of the non-monetary commitment group were closer to the control group than the regular *CollegeBetter* group. In total, these results suggest that commitment and loss aversion are the principle drivers of the success of the mechanism, while further highlighting only modest effects when solely offering rewards for student achievement.²⁵

6 Conclusions

Procrastination is a pervasive issue for college students ([Steel, 2007](#); [Knezevic, 2012](#)). This paper centers on a novel mechanism called *CollegeBetter.com* which draws inspiration from the commitment device and monetary incentive literatures to help college students battle problems of procrastination and present bias. The mechanism works by setting up pools which students can join by placing a monetary wager on themselves to achieve the pool’s “commitment challenge.” Students who successfully commit to the pool’s challenge recover their wagers and split losing wagers proportionally. This paper presents results from a series of lab and field experiments to (1) measure selection into the mechanism as a function of elicited preferences and beliefs, (2) draw inferences on the effectiveness of the mechanism, and (3) identify channels through which the mechanism operates.

Demand for the *CollegeBetter* mechanism is driven by lower performing students who tend to overestimate their own ability and probability of improving their grade point averages (GPAs). Meanwhile, the evidence of demand due to commitment against present bias is more mixed: structurally estimated present bias elicited in the lab via Convex Time Budgets ([Andreoni and Sprenger, 2012](#)) weakly predicts selection, while students who sign up are more likely to self-identify as procrastinators. The mechanism overall effective: students randomly selected to participate (treatment students) were more likely to raise their GPAs than students who applied for a spot but were ran-

²⁵[Table 7](#) presents results from a series of regressions to evaluate differences in term GPA from these experimental pools. Much like the results from [Table 5](#), I find little changes in average GPA across groups.

domly excluded (control students). This positive treatment effect is driven by the lowest performing applicants who had pre-term average letter grades of C and lower. One adverse effect can be seen in the distributions of GPA differentials: while many treatment students allocate just enough effort to raise their GPAs, a small fraction of treatment students see significant drops in their GPAs compared to the control group. Through additional treatment arms, I find that commitment, defined as the student putting their own money on the line, is a key contributor to effectiveness of the mechanism.

Findings from this study make several important contributions to the literature. While commitment contracts have helped present biased agents in a variety of other settings, this paper is the first to introduce a market for commitment contracts with monetary losses in education. Furthermore, to the best of my knowledge, *CollegeBetter* is the first intervention in education to allow for the possibility that agents lose money for failing to achieve. Prior interventions solely utilized frames to create loss domains (i.e. agents were first given an endowment, then lost the endowment for failure to achieve; Fryer Jr et al., 2012; Levitt et al., 2016), while participants in *CollegeBetter* could actually leave the program with less money than had they not participated entirely. Hence, *CollegeBetter* is arguably the first intervention in education to appeal to agents' loss aversion, and the results from the experiments are the first to test the importance of monetary loss aversion on student outcomes. Lastly, and perhaps most importantly, due to the zero-sum nature of payments, *CollegeBetter* represents a program that produces more cost-efficient outcomes compared to earlier monetary incentive interventions in education.

Finally, the findings from this paper offer key insights and possibilities from a policy standpoint. In general, the results provide promising evidence that programs that appeal to behaviorally-inspired concepts can be utilized to significantly improve student performance in a cost-efficient manner. More specifically, the results suggest that incentive schemes are more likely to be effective if students have something at risk for failure. Though less-obvious, there are certainly margins in current programs and policies that can be shifted to include loss components. For instance, in higher education, programs could alter the delivery of merit-based awards to include sanctions for poor performance, much like Wisconsin's Learnfare program for secondary schoolers (Dee, 2011). Additionally, the way course grades or test scores are framed could plausibly improve student effort; for example, students could lose points for insufficient performance on various questions.²⁶

The *CollegeBetter* program itself could easily be expanded to numerous universities, as students could directly supply their academic transcripts for verification. Ideally, more experiments would be conducted before considering an expansion of *CollegeBetter*, particularly on the long run, potentially distorting general equilibrium consequences of the program. For example, an es-

²⁶An example of this, though likely not implemented to appeal to loss averse preferences, can be seen with the SAT exam. SAT takers lose points for incorrect responses, while blank answers result in no points lost or gained.

established presence of *CollegeBetter* could encourage inefficient shifts in how students allocate their courses across terms.²⁷ To identify general equilibrium effects, information must be collected at the student population level, and not be solely restricted to *CollegeBetter* participants. Given the experiments found that lower achieving students were more likely to sign up, and conditional on selection, lower achieving students were more likely to succeed, *CollegeBetter* would likely generate larger positive treatment effects at universities with lower performing students compared to the UC Davis population.

7 Future Directions

There are many potential extensions of interest for this line of research. As I continue to present this research to various audiences, I have received numerous suggestions on alternative commitment challenges, experiments, and mechanism structures, all of which provide interesting and exciting future avenues to consider. Below I list just a couple additional experiments that I am concurrently running, or am planning on implementing in the near future.

7.1 Inputs versus outputs

One consideration is to shift the focus of the commitment challenges from educational “outputs” (e.g. grades) to “inputs” (e.g. studying). Under simple assumptions, traditional economic theory suggests that incentivizing output is socially optimal. While incentives for inputs reduce the costs of engaging in those particular inputs, incentives for output also act as a price subsidy while simultaneously allowing the agent to flexibly decide which input from their production function to engage in. Since students are assumed to choose their inputs (e.g. studying behavior) optimally to maximize their outputs (e.g. grades), incentivizing outputs will produce an outcome that is at least as socially optimal as incentivizing inputs.

Contrary to standard economic theory, [Fryer \(2011\)](#) found that incentives for student outputs were largely ineffective, while incentives for inputs led to both increased engagement in the inputs and better grades. While [Fryer \(2011\)](#) provides a list of potential explanations (e.g. noisy output, students have vague idea of how to produce output), evidence from a job market paper by [Hirshleifer \(2015\)](#) offers some insight: [Hirshleifer \(2015\)](#) conducted an experiment in India where individuals with incentivized inputs significantly outperformed those with incentivized outputs, and found that the treatment effect can be linked to present bias. This result suggests that incen-

²⁷In an extreme case where a student solely cares about maximizing returns from *CollegeBetter*, the student would front-load all of his/her difficult classes and save the easier classes for later terms. Such shifts in course load could lead to overall decreases in student performance.

tivizing inputs for students can be relatively more efficient since the incentive helps to alleviate the sharp short run discount factor β that leads to time inconsistency and procrastination.

Hence, setting up *CollegeBetter* pools with commitment challenges focused on inputs could prove to be even more effective than those focused on outputs. There are two primary obstacles to implementing pools with input commitment challenges. First, which inputs should be selected? In the two aforementioned studies, the inputs of interest were fairly obvious; in [Hirshleifer \(2015\)](#), students partook in a math computer program that included interactive learning modules (inputs) and exams (outputs), while in [Fryer \(2011\)](#), students were incentivized to read books (inputs) or for their performance on reading comprehension exams (outputs). For the general college student population, the mapping of inputs into outputs is less clear, and so it is not completely obvious which inputs should be the subject of the commitment challenge. The second difficulty with input commitment challenges deals with the impracticality of measuring and verifying inputs. For example, how can we confirm number of hours a student spent studying? If we introduce or utilize a setting where we can observe student studying, then already we are influencing the natural environment in such a way that could merely lead to a substitution effect. For instance, if the commitment challenge is defined by hours spent in a library, then, on some margin, students would shift their studying location from their homes to the library.

To investigate the plausibility of input commitment challenges leading to positive outcomes, I conducted a pilot experiment in a single economics course (the Economics of Education, coincidentally) offered during the 2016 summer session at UC Davis. As the instructor for the course, not only was I able to observe course outputs (e.g. grades), I could also measure and verify inputs (e.g. homework submission). The students' grades were determined by their performance on a midterm, a final, and the completion of three out of four homework assignments. I set up a *CollegeBetter* pool in this class with the commitment challenge "Complete all four homework assignments." Since the course grade was a function of completing only three homework assignments, students who joined this pool committed to exerting effort on a homework assignment that would have no direct influence on their final grade. Thus, participating students were committing to an input that would in expectation, all else equal, boost their performance on the final exam. Still, if participation in the pool distorts behavior away from more efficient inputs (e.g. hours spent on the homework assignment are less efficient than spending the hours studying at the library), then *CollegeBetter* could lead to worse outcomes.

A total of 129 students enrolled in the course, which lasted six weeks. Homework assignments were due on Fridays during the second, third, fourth, and fifth weeks of instruction. Students had until the Friday of the first week of the term to complete their application, which was provided through a Google Form on the course website. Similar the baseline design, those who applied for a spot were either randomly selected to participate in the pool or randomly excluded and refunded

their \$20 wager. With this experiment, I can test whether an input commitment challenge causally improves both student inputs (homework completion) and outputs (final grades). Note, however, that this experiment does not offer concrete evidence on whether a commitment challenge on inputs is more efficient than one on outputs. I could compare the relative differences in student grades between this pilot and the previous experimental pools, but one might hesitate to do this considering there are unobservable differences in the student population during summer courses versus the regular school year.

The results from this pilot study are very promising. Students randomly selected to participate in the pool were substantially more likely to complete all four homework assignments compared to the control group (83.3% vs. 29.9%). Treatment students also scored statistically significantly higher grades in the course (raw scores of 72.5% vs. 60.8%). These pilot results, though conducted on a very specific and small population of students, provides evidence that *CollegeBetter* can also work to increase student investment into inputs.

7.2 Formal versus informal commitments

One of the major contributions of this research project is that it is the first to introduce a formal commitment device (i.e. a contract with monetary punishment) to an educational setting. For the majority of the previous experimental pools, the punishment for failure to commit was \$20. From the the results [Section 5.2](#), successful committers netted around an 80% return. It is perhaps surprising that college students are responding so substantively to what amounts to, ex post, not very large monetary stakes; these students, after all, are spending thousands of dollars to attend UC Davis.²⁸

On the other hand, there are potentially positive interaction effects between the formal commitment with informal commitments the students are making. As discussed in [Section 2.1](#), informal commitments, which do not involve economic penalties, are formed by agents in a variety of settings; any time an individual sets a goal and makes a promise to oneself, the individual may feel shame or guilt for failing to achieve the goal. When an individual signs onto a formal commitment contract, they are likely to endure additional costs beyond the monetary punishment if they fail to achieve. With *CollegeBetter*, students are perhaps not only motivated by the potential loss of money (formal commitment) and any internal costs (e.g. shame) that come with failure, but also by potential interactive effects (e.g. student is ashamed they signed up for a costly program and still failed to commit). Indeed, I find some suggestive evidence of this in the form of participant testimonials. For instance, one student exclaimed to me that *CollegeBetter* undoubtedly motivated

²⁸For comparison, the size of the monetary incentives seen in the previous literature ranges to as low as in the tens to as high as in the hundreds of dollars.

her to perform because she did not want to go home during Spring break and tell her parents that not only did she get bad grades, but that she had lost money in the process of doing so.²⁹

Separately identifying these channels is obviously difficult, as the formal commitment and any interaction effects are defined by the presence of the monetary punishment. Beyond the student testimonials, alternative experiments to provide additional suggestive evidence center on simply changing the wager amount and structure. For instance, I could let students endogenously select their wager amount, and reward them proportionally to their wager amount as described in [Section 3](#). Then, with the same baseline experimental design described in the beginning of [Section 4](#), I can compare the treatment effects of those with high monetary stakes against those with smaller monetary stakes. If the treatment effect for those with high stakes is relatively larger, then I cannot conclude anything about the presence of interaction effects, since the increased performance could be driven solely by increased monetary loss aversion, or by both monetary loss aversion and an interaction effect. However, if the treatment effect for those with high stakes is similar to those with small stakes, then I can conclude that some interaction effects are at play. This is because, conditional on estimating positive treatment effects regardless of stakes, the results from [Section 5.3](#) suggest that at least part of the treatment effect is driven by monetary loss aversion. If the treatment effect does not monotonically increase in the size of the wager, then it must be that the mere presence of the monetary wager is driving the commitment effect.

I have reached an agreement with the UC Davis IRB to allow for flexible wagers between \$20 and \$100 for all experimental pools during the 2016-2017 school year. As these experiments are still ongoing, results are not yet available.

7.3 Single versus multiple threshold commitment challenges

All the presented experiments contained single threshold commitment challenges, where students needed to earn a term GPA that was higher than their previous cumulative GPA while maintaining a certain level of course units. The most attractive feature of a dichotomous commitment challenge is its simplicity, where the participants form a concrete goal with a relatively clear understanding of what determines their winnings. A potential drawback to setting a single threshold commitment challenge is captured in [Figure 4](#), where students who believe they can no longer reach the threshold may subsequently reduce their effort; furthermore, students who are certain to reach the threshold will not receive any additional motivation from participating in *CollegeBetter*. One can consider including multiple thresholds to the commitment challenge, with corresponding sizes punishments and rewards, in order to mitigate decreased motivations.

For instance, a pool can be defined by two separate commitment thresholds, “Receive a GPA

²⁹See www.CollegeBetter.com/testimonials for more student testimonials about their experience with *CollegeBetter*.

of 2.0 or higher”³⁰ and “Raise your cumulative GPA,” where students who complete the first challenge earn their wager back, and students who complete the second challenge split losing wagers proportionally. Any student with a pre-term GPA higher than 2.0 who, at some point during the term, no longer believes they can raise their cumulative GPA will cease to be completely demotivated by *CollegeBetter*. Unfortunately, even with this simple example, one can identify a major tradeoff with introducing alternative commitment challenges: any commitment challenge beyond “Raise your cumulative GPA” will favor students of different inherent ability. In this example, all else equal, higher ability students will have an easier time achieving the 2.0 GPA. Commitment challenges that are defined by increases in GPA (e.g. raise your cumulative GPA by x units) could mathematically be unattainable for higher achieving students, thus favoring lower achieving students. Still, it is an empirical question to investigate whether these alternative commitment challenges with or without multiple thresholds improve performance. I have an agreement with the university and the IRB to continue experiments through 2018, and are seeking IRB approval to flexibly alter the commitment challenges we offer across pools.

³⁰In other words, don’t flunk your classes this quarter.

References

- AMBEC, S. AND N. TREICH (2007): “Roscas as financial agreements to cope with self-control problems,” *Journal of development economics*, 82, 120–137.
- ANDREONI, J. AND C. SPRENGER (2012): “Estimating Time Preferences from Convex Budgets,” *The American Economic Review*, 3333–3356.
- ANGRIST, J., E. BETTINGER, E. BLOOM, M. KREMER, AND E. KING (2002): “The effects of school vouchers on students: evidence from Colombia,” *American Economic Review*, 92, 1535–1558.
- ANGRIST, J., E. BETTINGER, AND M. KREMER (2006): “Long-Term Educational Consequences of Secondary School Vouchers: Evidence from Administrative Records in Colombia,” *The American Economic Review*, 96, 847–862.
- ANGRIST, J., D. LANG, AND P. OREOPOULOS (2009): “Incentives and services for college achievement: Evidence from a randomized trial,” *American Economic Journal: Applied Economics*, 136–163.
- ARIELY, D. AND K. WERTENBROCH (2002): “Procrastination, deadlines, and performance: Self-control by precommitment,” *Psychological science*, 13, 219–224.
- ASHRAF, N., D. KARLAN, AND W. YIN (2006): “Deposit collectors,” *Advances in Economic Analysis & Policy*, 5.
- AUGENBLICK, N., M. NIEDERLE, AND C. SPRENGER (2015): “Working over time: Dynamic inconsistency in real effort tasks,” *The Quarterly Journal of Economics*, 1067, 1115.
- BABCOCK, P. (2008): “From ties to gains? Evidence on connectedness and human capital acquisition,” *Journal of Human Capital*, 2, 379–409.
- BARRERA-OSORIO, F., M. BERTRAND, L. L. LINDEN, AND F. PEREZ-CALLE (2011): “Improving the design of conditional transfer programs: Evidence from a randomized education experiment in Colombia,” *American Economic Journal: Applied Economics*, 167–195.
- BASU, K. (2011): “Hyperbolic discounting and the sustainability of rotational savings arrangements,” *American Economic Journal: Microeconomics*, 143–171.
- BEHRMAN, J. R., P. SENGUPTA, AND P. TODD (2005): “Progressing through PROGRESA: An impact assessment of a school subsidy experiment in rural Mexico,” *Economic development and cultural change*, 54, 237–275.
- BESHEARS, J., J. J. CHOI, D. LAIBSON, AND B. C. MADRIAN (2013): “Simplification and saving,” *Journal of economic behavior & organization*, 95, 130–145.
- BETTINGER, E. P. (2012): “Paying to learn: The effect of financial incentives on elementary school test scores,” *Review of Economics and Statistics*, 94, 686–698.

- BETTINGER, E. P. AND R. B. BAKER (2014): “The Effects of Student Coaching: An Evaluation of a Randomized Experiment in Student Advising,” *Educational Evaluation and Policy Analysis*, 36, 3–19.
- BOUND, J., M. F. LOVENHEIM, AND S. TURNER (2010): “Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources,” *American Economic Journal: Applied Economics*, 2, 129–57.
- (2012): “Increasing Time to Baccalaureate Degree in the United States,” *Education Finance and Policy*, 7, 375–424.
- BREMAN, A. (2011): “Give more tomorrow: Two field experiments on altruism and intertemporal choice,” *Journal of Public Economics*, 95, 1349–1357.
- BRYAN, G., D. KARLAN, AND S. NELSON (2010): “Commitment devices,” *Annu. Rev. Econ.*, 2, 671–698.
- CARRELL, S. E., M. E. PAGE, AND J. E. WEST (2010): “Sex and Science: How Professor Gender Perpetuates the Gender Gap,” *The Quarterly Journal of Economics*, 125, 1101–1144.
- CARRELL, S. E. AND J. E. WEST (2010): “Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors,” *Journal of Political Economy*, 118, pp. 409–432.
- COHEN, J., K. M. ERICSON, D. LAIBSON, AND J. M. WHITE (2016): “Measuring Time Preferences,” Tech. rep., National Bureau of Economic Research.
- CONOVER, W. (1999): “Statistics of the Kolmogorov-Smirnov type,” *Practical nonparametric statistics*, 428–473.
- DEE, T. S. (2011): “Conditional cash penalties in education: Evidence from the Learnfare experiment,” *Economics of Education Review*, 30, 924–937.
- FAIRLIE, R. W., F. HOFFMANN, AND P. OREOPOULOS (2014): “A Community College Instructor Like Me: Race and Ethnicity Interactions in the Classroom,” *American Economic Review*, 104, 2567–91.
- FREY, B. S. AND R. JEGEN (2001): “Motivation crowding theory,” *Journal of economic surveys*, 15, 589–611.
- FRYER, R. G. (2011): “Financial Incentives and Student Achievement: Evidence From Randomized Trials,” *The Quarterly Journal of Economics*, 1755–1798.
- FRYER JR, R. G., S. D. LEVITT, J. LIST, AND S. SADOFF (2012): “Enhancing the efficacy of teacher incentives through loss aversion: a field experiment,” Tech. rep., National Bureau of Economic Research.
- GINÉ, X., D. KARLAN, AND J. ZINMAN (2010): “Put your money where your butt is: a commitment contract for smoking cessation,” *American Economic Journal: Applied Economics*, 213–235.

- GLEWWE, P., N. ILIAS, AND M. KREMER (2010): “Teacher Incentives,” *American Economic Journal: Applied Economics*, 205–227.
- GNEEZY, U., S. MEIER, AND P. REY-BIEL (2011): “When and Why Incentives (don’t) Work to Modify Behavior,” *The Journal of Economic Perspectives*, 191–209.
- GUGERTY, M. K. (2007): “You cant save alone: Commitment in rotating savings and credit associations in Kenya,” *Economic Development and cultural change*, 55, 251–282.
- HIRSHLEIFER, S. R. (2015): “Incentives for Effort or Outputs? A Field Experiment to Improve Student Performance,” .
- HOFFMANN, F. AND P. OREOPOULOS (2009a): “A Professor Like Me: The Influence of Instructor Gender on College Achievement,” *Journal of Human Resources*, 44, 479–494.
- (2009b): “Professor Qualities and Student Achievement,” *The Review of Economics and Statistics*, 91, 83–92.
- KAHNEMAN, D. AND A. TVERSKY (1979): “Prospect theory: An analysis of decision under risk,” *Econometrica: Journal of the Econometric Society*, 263–291.
- KNEZEVIC, M. (2012): “Procrastination: a student’s worst enemy?” Retrived from <http://www.theguardian.com/education/mortarboard/2012/may/09/students-procrastinating-exams>.
- LAIBSON, D. (1997): “Golden eggs and hyperbolic discounting,” *The Quarterly Journal of Economics*, 443–477.
- LEVITT, S. D., J. A. LIST, S. NECKERMANN, AND S. SADOFF (2011): “The impact of short-term incentives on student performance,” *Unpublished mimeo, University of Chicago*.
- (2016): “The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance,” *American Economic Journal: Economic Policy*.
- LINDO, J. M., N. J. SANDERS, AND P. OREOPOULOS (2010): “Ability, Gender, and Performance Standards: Evidence from Academic Probation,” *American Economic Journal: Applied Economics*, 2, 95–117.
- LUSHER, L., D. CAMPBELL, AND S. CARRELL (2015): “TAs Like Me: Racial Interactions between Graduate Teaching Assistants and Undergraduates,” Tech. rep., National Bureau of Economic Research.
- MARTORELL, P. AND I. MCFARLIN JR (2011): “Help or hindrance? The effects of college remediation on academic and labor market outcomes,” *The Review of Economics and Statistics*, 93, 436–454.
- O’DONOGHUE, T. AND M. RABIN (1999): “Doing it now or later,” *American Economic Review*, 103–124.

- OREOPOULOS, P. AND U. PETRONIJEVIC (2013): “Making college worth it: A review of the returns to higher education,” *The Future of Children*, 23, 41–65.
- PAGE, L. C. AND J. SCOTT-CLAYTON (2015): “Improving College Access in the United States: Barriers and Policy Responses,” Tech. rep., National Bureau of Economic Research.
- PHELPS, E. S. AND R. A. POLLAK (1968): “On second-best national saving and game-equilibrium growth,” *The Review of Economic Studies*, 35, 185–199.
- RODRIGUEZ-PLANAS, N. (2012): “Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the United States,” *American Economic Journal: Applied Economics*, 121–139.
- ROYER, H., M. STEHR, AND J. SYDNOR (2015): “Incentives, Commitments, and Habit Formation in Exercise: Evidence from a Field Experiment with Workers at a Fortune-500 Company,” *American Economic Journal: Applied Economics*, 7, 51–84.
- SCHOUWENBURG, H. C. (2004): “Procrastination in Academic Settings: General Introduction,” *American Psychological Association*, 3–17.
- SCHULTZ, T. P. (2004): “School subsidies for the poor: evaluating the Mexican Progresa poverty program,” *Journal of development Economics*, 74, 199–250.
- STEEL, P. (2007): “The nature of procrastination: a meta-analytic and theoretical review of quintessential self-regulatory failure,” *Psychological bulletin*, 133, 65.
- STROTZ, R. H. (1955): “Myopia and inconsistency in dynamic utility maximization,” *The Review of Economic Studies*, 23, 165–180.
- THALER, R. H. AND S. BENARTZI (2004): “Save more tomorrow: Using behavioral economics to increase employee saving,” *Journal of political Economy*, 112, S164–S187.
- VISARIA, S., R. DEHEJIA, M. M. CHAO, AND A. MUKHOPADHYAY (2016): “Unintended consequences of rewards for student attendance: Results from a field experiment in Indian classrooms,” Tech. rep., National Bureau of Economic Research.
- WISEMAN, R. (2007): “New Years Resolution Experiment,” *Published online at http://www.quirkology.com/UK/Experiment_resolution.shtml*.

Main Tables and Figures

Table 1: Summary of *CollegeBetter* experimental pools

	Research question of interest		
	Selection into mechanism	Mechanism effectiveness	Importance of commitment
Winter 2015	Survey in Economics courses included application, compare applicants against non-applicants	Randomly exclude or enter applicants into pool	—
Spring 2015	—	Applications online, randomly exclude or enter applicants into pool	—
Fall 2015	Lab experiment with survey, Convex Time Budget allocation decisions	Applications online, randomly exclude or enter applicants into pool	Additional treatment group: Buy-in covered, same conditional reward as regular participants
Winter 2016	—	Applications online, randomly exclude or enter applicants into pool	Additional treatment group: Buy-in covered, same conditional reward as regular participants
Spring 2016	—	—	Applications online, randomly select participants to enter mechanism or additional treatment group

Notes: Each row corresponds to a single experimental pool conducted at the University of California, Davis (See [Section 4](#) for further detail). Experiments differed in various ways in order to 1) measure selection into *CollegeBetter*, 2) estimate causal effectiveness, and 3) explore the degree to which the monetary commitment can explain the effectiveness of the mechanism.

Table 2: Allocation problems for Convex Time Budget mechanism

t (start date)	k (delay)	Token budget	a_t	a_{t+k}	$(1+r)$	Daily rate (percent)	Annual rate (percent)
0	35	50	\$0.19	\$0.20	1.05	0.147	65.3
0	35	50	\$0.18	\$0.20	1.11	0.301	164.4
0	35	50	\$0.16	\$0.20	1.25	0.64	528.9
0	35	50	\$0.14	\$0.20	1.43	1.024	1,300.9
0	35	50	\$0.20	\$0.25	1.25	0.64	528.9
0	70	50	\$0.19	\$0.20	1.05	0.073	29.6
0	70	50	\$0.18	\$0.20	1.11	0.151	67.4
0	70	50	\$0.16	\$0.20	1.25	0.319	178.1
0	70	50	\$0.14	\$0.20	1.43	0.511	362.1
0	70	50	\$0.20	\$0.25	1.25	0.319	178.1
0	98	50	\$0.19	\$0.20	1.05	0.052	20.5
0	98	50	\$0.16	\$0.20	1.25	0.228	113
0	98	50	\$0.13	\$0.20	1.54	0.441	286.4
0	98	50	\$0.10	\$0.20	2	0.71	637.1
0	98	50	\$0.20	\$0.25	1.25	0.228	113
7	35	50	\$0.19	\$0.20	1.05	0.147	65.3
7	35	50	\$0.18	\$0.20	1.11	0.151	67.4
7	35	50	\$0.16	\$0.20	1.25	0.64	528.9
7	35	50	\$0.14	\$0.20	1.43	1.024	1,300.9
7	35	50	\$0.20	\$0.25	1.25	0.64	528.9
7	70	50	\$0.20	\$0.20	1	0	0
7	70	50	\$0.19	\$0.20	1.05	0.073	29.6
7	70	50	\$0.18	\$0.20	1.11	0.151	67.4
7	70	50	\$0.16	\$0.20	1.25	0.319	178.1
7	70	50	\$0.14	\$0.20	1.43	0.511	362.1
7	98	50	\$0.19	\$0.20	1.05	0.052	20.5
7	98	50	\$0.16	\$0.20	1.25	0.228	113
7	98	50	\$0.13	\$0.20	1.54	0.441	286.4
7	98	50	\$0.10	\$0.20	2	0.71	637.1
7	98	50	\$0.20	\$0.25	1.25	0.228	113
35	35	50	\$0.19	\$0.20	1.05	0.147	65.3
35	35	50	\$0.18	\$0.20	1.11	0.301	164.4
35	35	50	\$0.16	\$0.20	1.25	0.62	528.9
35	35	50	\$0.14	\$0.20	1.43	1.024	1,300.9
35	35	50	\$0.20	\$0.25	1.25	0.64	528.9
35	70	50	\$0.19	\$0.20	1.05	0.073	29.6
35	70	50	\$0.18	\$0.20	1.11	0.151	67.4
35	70	50	\$0.16	\$0.20	1.25	0.319	178.1
35	70	50	\$0.14	\$0.20	1.43	0.511	362.1
35	70	50	\$0.20	\$0.25	1.25	0.319	178.1
35	98	50	\$0.19	\$0.20	1.05	0.052	20.5
35	98	50	\$0.16	\$0.20	1.25	0.228	113
35	98	50	\$0.13	\$0.20	1.54	0.441	286.4
35	98	50	\$0.10	\$0.20	2	0.71	637.1
35	98	50	\$0.20	\$0.25	1.25	0.228	113

Notes: Each row represents an allocation problem from the Convex Time Budget mechanism used to elicit individual (β, δ) preferences (Andreoni and Sprenger, 2012). a_t and a_{t+k} represent the value of a token on the earlier and later date, respectively. Variations in k and r help identify discounting and utility function curvature, while variations in t identify present bias.

Table 3: Summary statistics from Winter 2015 experiment

	Interested in <i>CollegeBetter</i> ?			Randomly Entered?		
	Yes	No	Difference	Yes	No	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Cumulative GPA entering term	2.947 (0.577)	3.244 (0.530)	-0.297*** (0.058)	2.925 (0.536)	2.973 (0.626)	-0.048 (0.090)
Cumulative # units entering term	62.554 (54.034)	60.984 (50.531)	1.569 (5.515)	61.137 (55.243)	64.250 (52.865)	-3.113 (8.418)
Male	0.586 (0.494)	0.417 (0.494)	0.169*** (0.050)	0.604 (0.492)	0.564 (0.499)	0.040 (0.076)
Years at U.C. Davis	1.922 (1.094)	1.884 (1.001)	0.034 (0.034)	1.901 (1.146)	1.936 (1.036)	-0.035 (0.169)
Asian	0.615 (0.488)	0.608 (0.489)	0.008 (0.049)	0.626 (0.486)	0.603 (0.493)	0.024 (0.075)
Employed (e.g. full-time, part-time, summer job)	0.369 (0.484)	0.411 (0.493)	0.042 (0.050)	0.330 (0.473)	0.416 (0.496)	-0.086 (0.075)
International	0.173 (0.379)	0.163 (0.370)	0.009 (0.038)	0.121 (0.328)	0.231 (0.424)	0.110* (0.058)
Major in College of Letters and Science	0.767 (0.424)	0.767 (0.424)	0.001 (0.044)	0.727 (0.448)	0.813 (0.392)	-0.086 (0.067)
Made New Years Resolution	0.443 (0.498)	0.443 (0.498)	0.000 (0.051)	0.407 (0.494)	0.487 (0.503)	-0.080 (0.077)
Pays bills late	0.743 (0.439)	0.746 (0.436)	-0.003 (0.045)	0.769 (0.424)	0.711 (0.457)	0.059 (0.068)
Gets easily distracted on internet (e.g. Facebook, Twitter)	0.882 (0.326)	0.842 (0.365)	0.040 (0.036)	0.912 (0.284)	0.842 (0.367)	0.070 (0.050)
Waits until last second to buy gifts (e.g. Christmas, birthday)	0.712 (0.454)	0.732 (0.444)	-0.020 (0.046)	0.769 (0.424)	0.645 (0.482)	0.124* (0.070)
Procrastinates too much when studying for exams	0.772 (0.421)	0.623 (0.486)	0.147*** (0.047)	0.791 (0.409)	0.750 (0.436)	0.041 (0.065)
Belief about fraction of <i>CollegeBetter</i> participants that will achieve goal	58.938 (19.937)	52.238 (22.109)	6.700*** (2.164)	60.044 (19.118)	57.659 (20.897)	2.384 (3.107)
F-statistic [p-value] from regression of indicator for “Yes” on all of the above observables	2.599*** [0.003]			0.873 [0.575]		
Total observations	169	243		91	78	

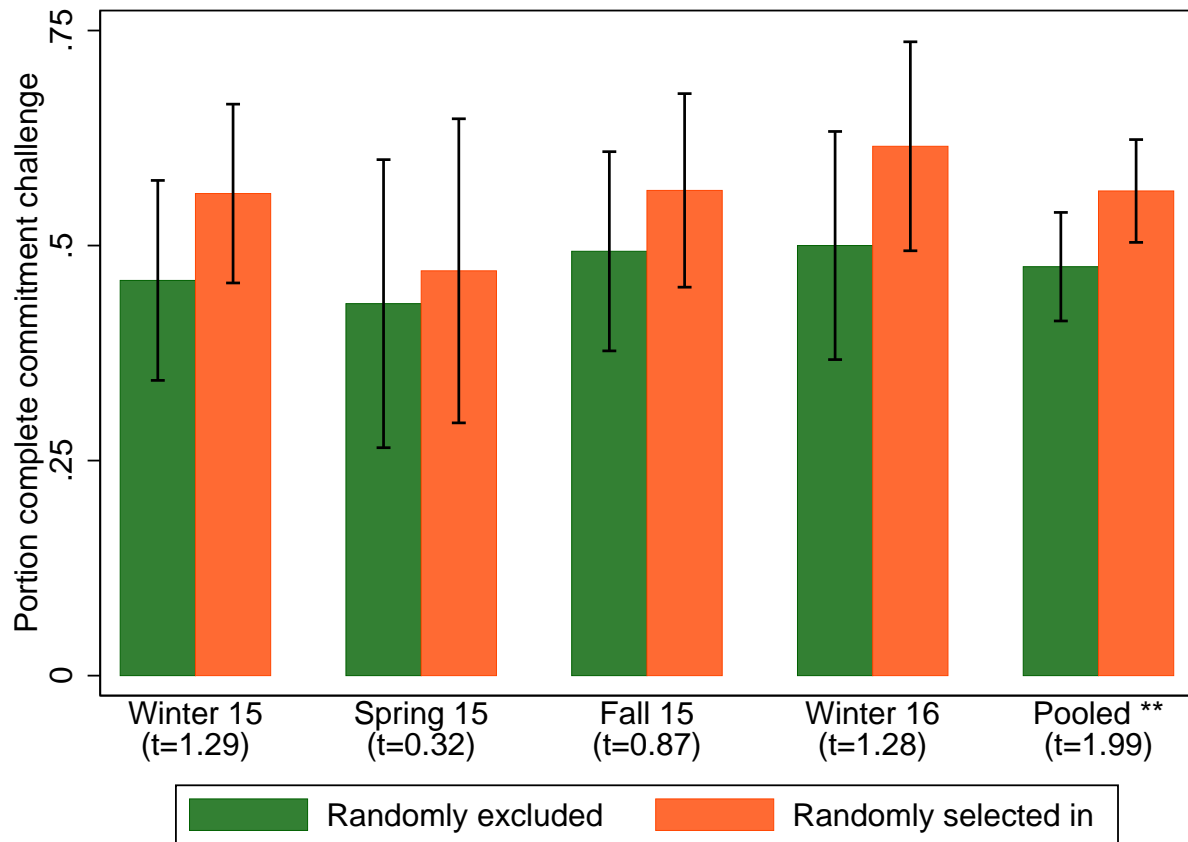
Notes: See Appendix A.1 for a complete list of survey questions. Standard deviations in parentheses in columns headed “Yes” and “No”, standard errors in parentheses in columns headed “Difference”. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Table 4: Summary statistics from Fall 2015 experiment

	Interested in <i>CollegeBetter</i> ?			Interested in <i>CollegeBetter</i> ?		
	Yes	No	Difference	Yes	No	Difference
Male	0.386 (0.491)	0.444 (0.501)	-0.058 (0.091)	25.612 (12.420)	23.713 (11.959)	1.899 (2.227)
CRRA curvature: α	0.877 (0.218)	0.876 (0.157)	0.001 (0.038)	0.386 (0.491)	0.667 (0.475)	-0.281*** (0.088)
Present bias: β	1.057 (0.143)	1.028 (0.154)	0.029 (0.030)	0.333 (0.476)	0.254 (0.439)	0.079 (0.083)
Discount factor: δ	1.004 (0.015)	1.002 (0.011)	0.002 (0.003)	0.825 (0.384)	0.762 (0.429)	0.063 (0.075)
Monthly rent	519.546 (190.653)	471.188 (198.159)	48.357 (35.579)	0.719 (0.453)	0.683 (0.469)	0.037 (0.084)
Smokes regularly	0.018 (0.132)	0.016 (0.126)	0.002 (0.024)	0.404 (0.495)	0.429 (0.499)	-0.025 (0.091)
Years at UC Davis	3.333 (1.123)	3.333 (1.032)	0.000 (0.197)	5.421 (1.017)	4.794 (1.461)	0.627*** (0.232)
International Student	0.140 (0.350)	0.127 (0.336)	0.013 (0.063)	0.246 (0.434)	0.317 (0.469)	-0.072 (0.083)
Pays in-state tuition	0.842 (0.368)	0.889 (0.317)	-0.047 (0.063)	0.842 (0.368)	0.905 (0.296)	-0.063 (0.061)
Owens a credit card	0.614 (0.491)	0.667 (0.475)	-0.053 (0.088)	0.965 (0.186)	0.937 (0.246)	0.028 (0.040)
Employed	0.614 (0.491)	0.524 (0.503)	0.090 (0.091)	0.351 (0.481)	0.317 (0.469)	0.033 (0.087)
Average weekly expenses	91.535 (53.930)	74.646 (70.046)	16.890 (11.501)	0.333 (0.476)	0.286 (0.455)	0.048 (0.085)
Major in Letters and Science	0.368 (0.487)	0.381 (0.490)	-0.013 (0.089)	0.123 (0.331)	0.190 (0.396)	-0.068 (0.067)
Exercises regularly	0.614 (0.491)	0.540 (0.502)	0.074 (0.091)	0.070 (0.258)	0.016 (0.126)	0.054 (0.036)
Made New Years Resolution	0.316 (0.469)	0.429 (0.499)	-0.113 (0.089)	0.316 (0.469)	0.238 (0.429)	0.078 (0.082)
Self-stated patience (1 - impatient to 7 - patient)	4.877 (1.402)	4.921 (1.548)	-0.043 (0.271)	4.860 (1.187)	4.810 (1.134)	0.050 (0.212)
F-statistic [p-value] from regression of indicator for “Yes” on all of the above observables				4.440*** [1E-8]		
Total observations				57	63	120

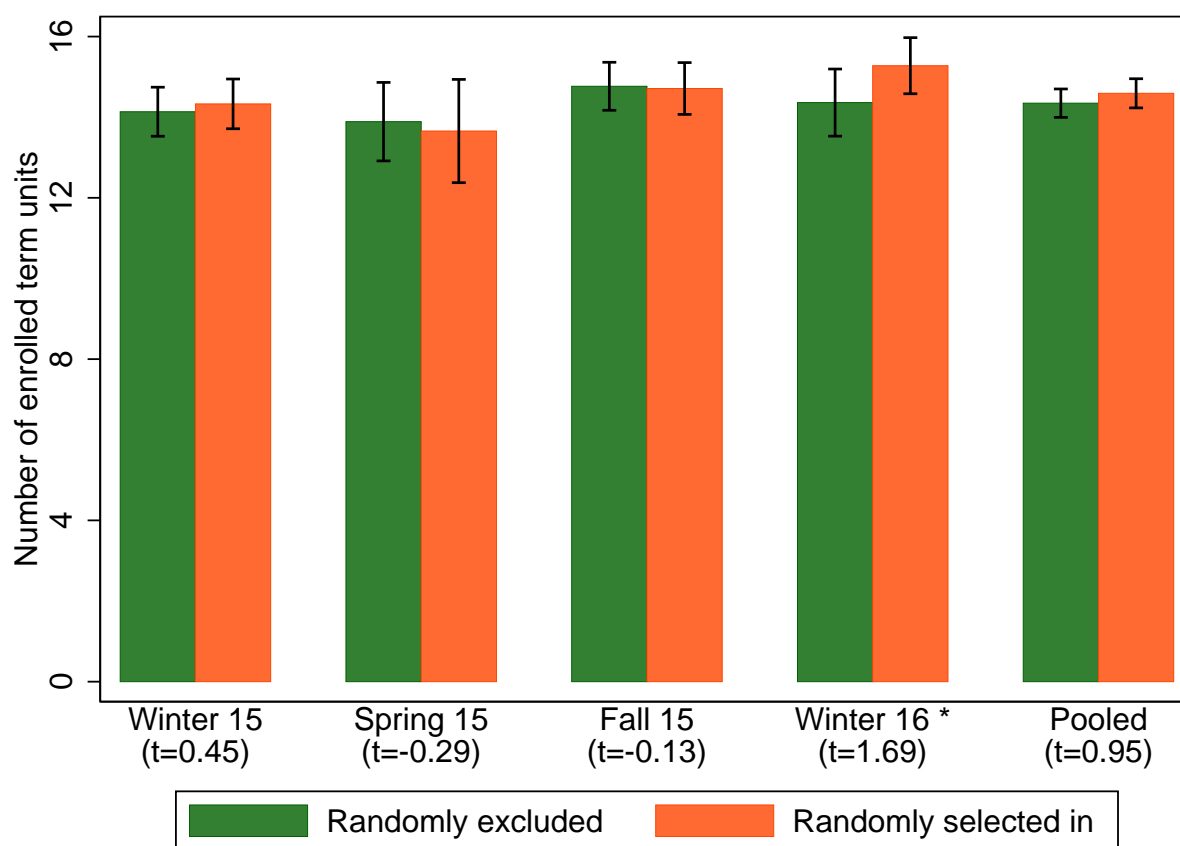
Notes: See Appendix A.3 for a complete list of survey questions. Standard deviations in parentheses in columns headed “Yes” and “No”, standard errors in parentheses in columns headed “Difference”. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Figure 1: Effectiveness of *CollegeBetter* - Mean commitment challenge completion rates



Notes: Each bar reports the mean commitment challenge completion rate by experimental pool and by treatment group. 95% confidence intervals are included for each mean completion rate.

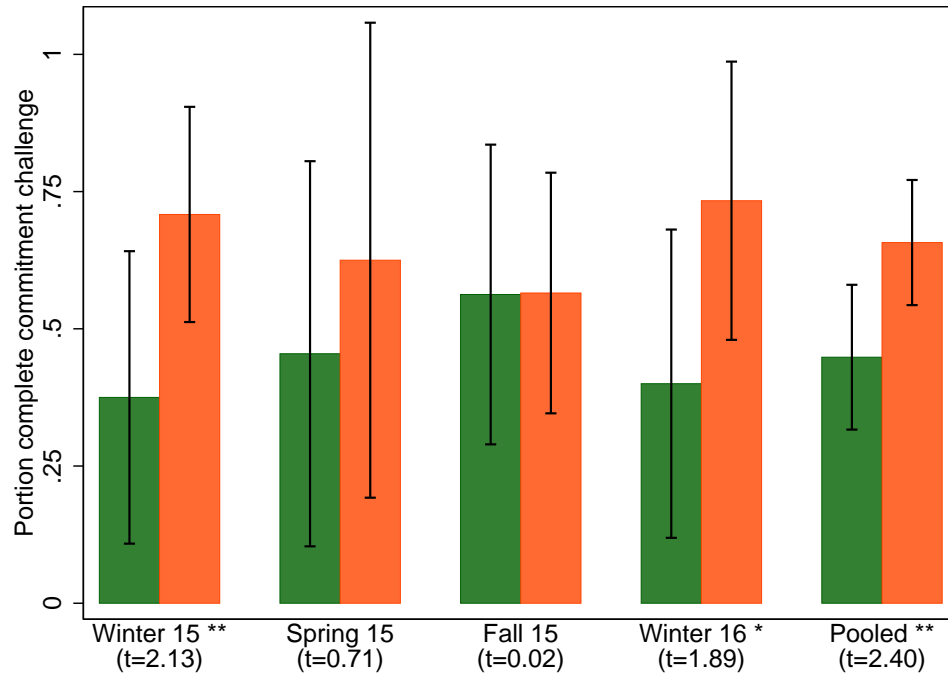
Figure 2: Effectiveness of *CollegeBetter* - Mean number of enrolled units



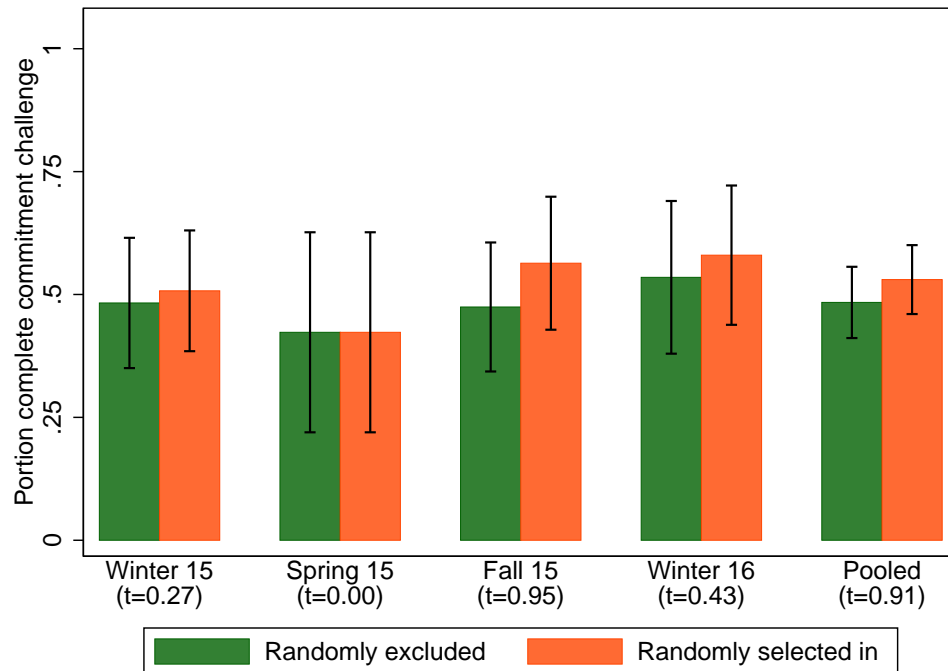
Notes: Each bar reports the mean number of units enrolled for the term by experimental pool and by treatment group. 95% confidence intervals are included for each mean number of units.

Figure 3: Effectiveness of *CollegeBetter* - Mean commitment challenge completion rates

(a) Students with ≤ 2.50 cumulative GPA entering term



(b) Students with > 2.50 cumulative GPA entering term



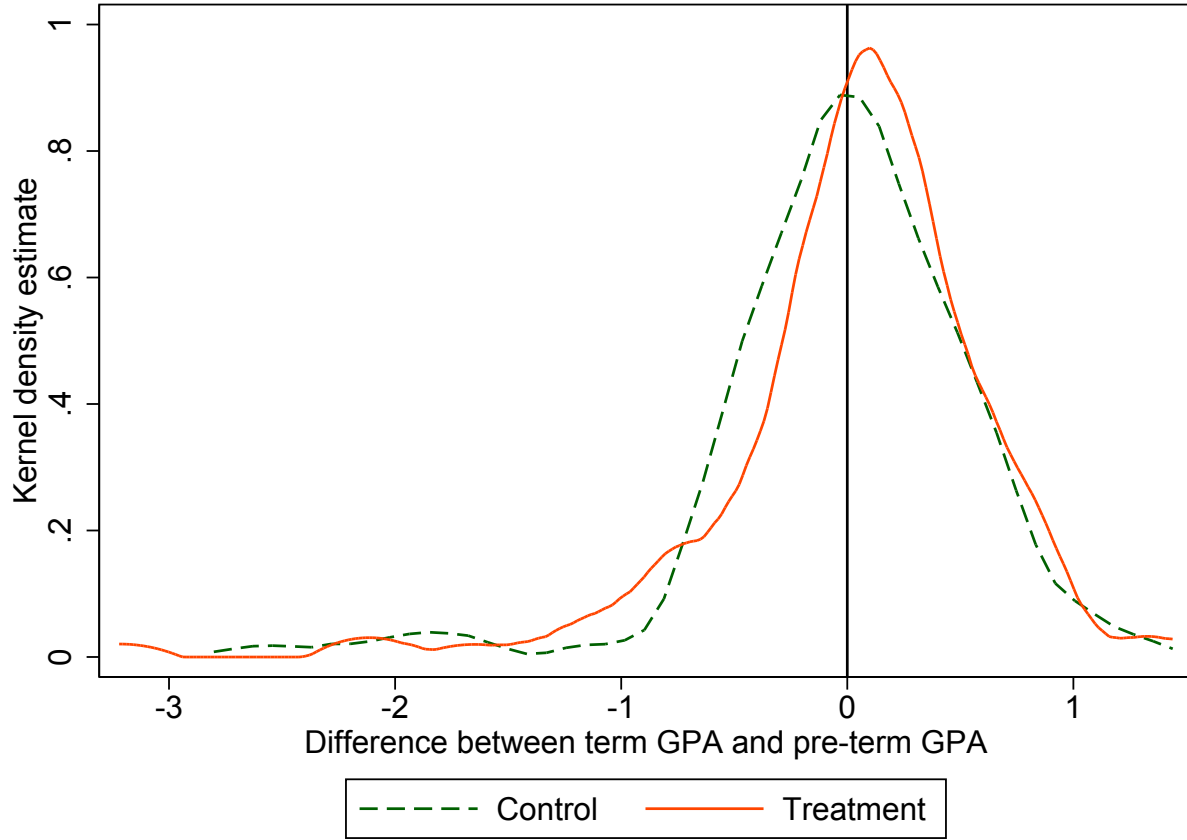
Notes: Each bar reports the mean number of units enrolled for the term by experimental pool and by treatment group. 95% confidence intervals are included for each mean number of units.

Table 5: Effectiveness of *CollegeBetter* - Term Grade Point Average (GPA)

	Full sample					Students w/ ≤ 2.50 Pre-term GPA				
	W15	S15	F15	W16	Pooled	W15	S15	F15	W16	Pooled
Randomly Entered	0.087 (0.083)	-0.050 (0.165)	-0.001 (0.079)	-0.007 (0.109)	0.012 (0.050)	0.386** (0.184)	-0.497 (0.285)	-0.136 (0.138)	0.264 (0.193)	0.055 (0.104)
Pre-term Cumulative GPA	0.834*** (0.084)	0.758*** (0.201)	0.992*** (0.089)	0.957*** (0.108)	0.892*** (0.053)	0.382* (0.209)	2.607*** (0.656)	2.954*** (0.438)	0.979*** (0.301)	1.024*** (0.248)
Term Units Enrolled	0.063*** (0.022)	0.101* (0.051)	0.075*** (0.022)	0.034 (0.035)	0.062*** (0.016)	0.118*** (0.029)	0.150 (0.124)	0.099*** (0.018)	0.077 (0.051)	0.116*** (0.018)
Pre-term Cumulative Units	0.001 (0.001)	0.002 (0.001)	0.002* (0.001)	0.002** (0.001)	0.001** (0.000)	-0.001 (0.002)	0.000 (0.004)	0.001 (0.002)	0.002 (0.002)	-0.001 (0.001)
Observations	165	64	153	123	505	40	17	39	30	126
Mean of outcome	2.956	2.948	2.911	2.864	2.919	2.293	2.401	2.344	2.296	2.324

Notes: Each column presents results from a single regression. Term GPA is the student's GPA (on a scale from 0 to 4) for the term. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Figure 4: Distribution of Grade Point Average differential



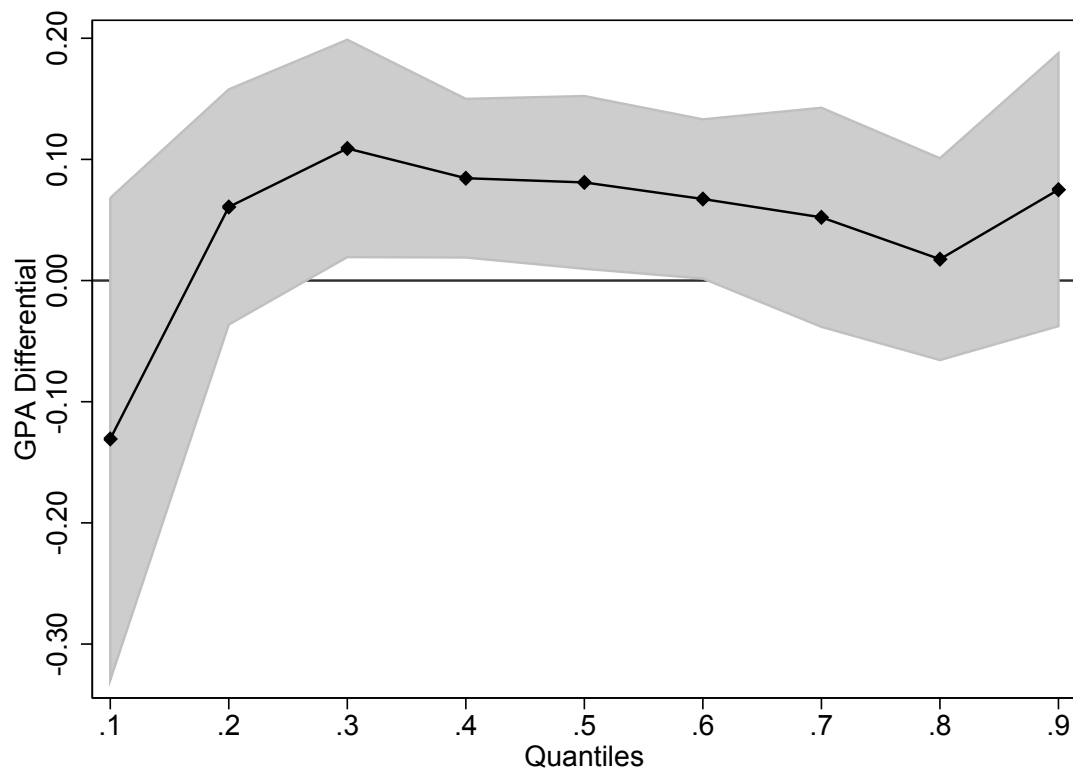
Notes: The density function of GPA differential (term GPA minus pre-term cumulative GPA) is estimated using kernel density estimation, and plotted separately by treatment (students randomly selected to participate) and control (students randomly excluded) groups. Epanechnikov kernel used to calculate kernel density estimates.

Table 6: Quantile regressions - Grade Point Average (GPA) differential

	GPA Differential					
	Mean	P10	P25	Median	P75	P90
Randomly Entered	0.035 (0.052)	-0.131 (0.108)	0.077 (0.063)	0.081* (0.045)	0.035 (0.058)	0.075 (0.072)
Constant	-0.013 (0.036)	-0.548*** (0.037)	-0.278*** (0.045)	0.020 (0.033)	0.327*** (0.045)	0.589*** (0.047)
Observations	509	509	509	509	509	509
Mean of outcome	0.005	0.005	0.005	0.005	0.005	0.005

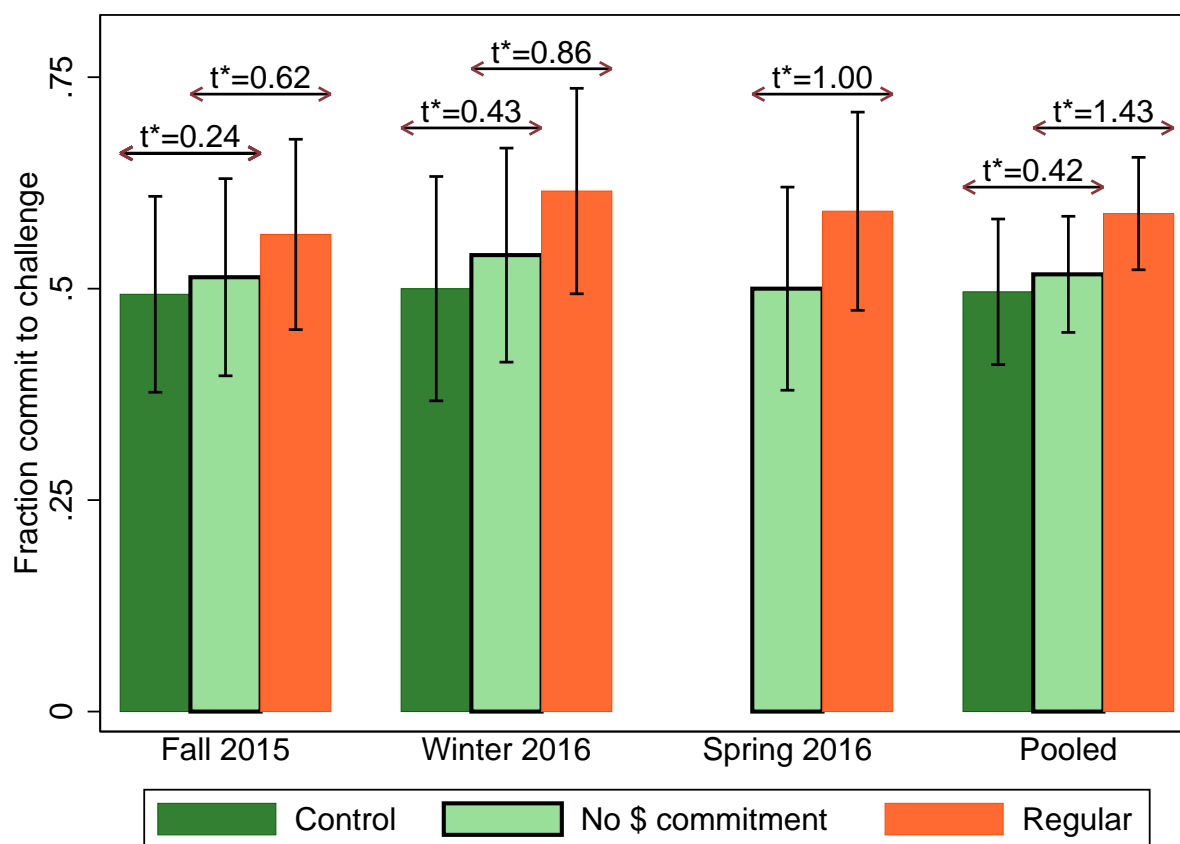
Notes: Each column presents results from a single regression. The dependent variable for each regression is GPA differential (term GPA minus pre-term cumulative GPA). Columns two through six correspond to quantile regressions to test for effects at the 10th, 25th, 50th, 75th, and 90th percentiles, respectively. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Figure 5: Decile regressions - Grade Point Average (GPA) differential



Notes: Each point corresponds to the coefficient on an indicator for whether the student was randomly entered into the pool from a sequential-quantile regression, estimated at each decile. 90% confidence intervals shaded in gray. The variance-covariance matrix is estimated via bootstrapping, and includes between-quantile blocks.

Figure 6: Importance of monetary commitment - Mean commitment challenge completion rates



Notes: Each bar reports the mean commitment challenge completion rate by experimental pool and by treatment group. 95% confidence intervals are included for each mean completion rate.

Table 7: Importance of monetary commitment - Term Grade Point Average (GPA)

	Full sample				Students w/ ≤ 2.50 Pre-term GPA			
	F15	W16	S16	All	F15	W16	S16	All
<i>CollegeBetter</i> participants	-0.008 (0.079)	0.002 (0.109)	-0.001 (0.098)	-0.005 (0.059)	-0.128 (0.128)	0.387** (0.173)	-0.458 (0.369)	0.023 (0.126)
No Monetary Commitment	0.038 (0.096)	0.001 (0.096)	— —	0.011 (0.062)	-0.130 (0.171)	-0.013 (0.208)	— —	0.006 (0.135)
Pre-term Cumulative GPA	0.920*** (0.089)	1.084*** (0.092)	1.000*** (0.119)	1.002*** (0.059)	2.525*** (0.391)	1.130*** (0.411)	1.503* (0.719)	1.578*** (0.259)
Term Units Enrolled	0.072*** (0.026)	0.015 (0.029)	0.028 (0.032)	0.042** (0.018)	0.093*** (0.021)	0.026 (0.060)	0.037 (0.050)	0.065** (0.031)
Pre-term Cumulative Units	0.001 (0.001)	0.003*** (0.001)	0.003** (0.001)	0.002*** (0.001)	0.001 (0.002)	0.002 (0.002)	0.008** (0.003)	0.002 (0.001)
Observations	226	186	140	552	63	51	24	138
Mean of outcome	2.941	2.841	2.972	2.915	2.440	2.162	2.114	2.281

Notes: Each column presents results from a single regression. Term GPA is the student's GPA (on a scale from 0 to 4) for the term. The first two rows correspond to indicators for which treatment group the student was randomly selected into. For the Fall 2015 and Winter 2016 experiments, students randomly excluded (control students) are the omitted category to identify the two indicators; for the Spring 2016 experiment, no students were randomly excluded, so the omitted category is the No Monetary Commitment group. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Results Appendix

Table A.1: Individual discounting, present bias, and curvature parameter estimates

Panel A: Results from Lusher (2016)

	N	Median	5 th Percentile	95 th Percentile	Min	Max
Daily Discount Factor: $\hat{\delta}$	99	0.9999	0.9992	1.0162	0.9982	1.0975
Present Bias: $\hat{\beta}$	99	1.0003	0.9484	1.2871	0.5604	1.8549
CRRA Curvature: $\hat{\alpha}$	99	0.9470	0.5509	0.9986	-0.1226	0.9987

Panel B: Results from Andreoni and Sprenger (2012)

	N	Median	5 th Percentile	95 th Percentile	Min	Max
Daily discount factor: $\hat{\delta}$	86	0.9991	0.9948	1.0005	0.9902	1.0146
Present bias: $\hat{\beta}$	86	1.0011	0.9121	1.1075	0.7681	1.3241
CRRA Curvature: $\hat{\alpha}$	86	0.9665	0.7076	0.9997	-0.1331	0.9998

Notes: NLS estimators from Fall 2015 Convex Time Budgets experiment. Parameter estimation was not achieved for 20 of Lusher (2016) subjects and 11 of Andreoni and Sprenger (2012) subjects.

Table A.2: Individual estimates from Fall 2015 CTB experiment (Part 1)

Subject #	Annual Rate	$\hat{\beta}$	$\hat{\alpha}$	Interior	Proportions of Response	
					Zero Tokens Sooner	All Tokens Sooner
1	.108	.984	.976	.07	.8	.13
2	.305	.973	.902	0	.51	.49
3	.007	.988	.959	.22	.73	.04
4	0	1.003	.999	0	.98	.02
5	0	1.003	.999	0	1	0
6	0	1.003	.999	0	.98	.02
7	0	1.003	.999	0	.98	.02
8	-.076	.975	.947	.38	.6	.02
9	-.076	.975	.947	0	1	0
10	-.076	.975	.947	0	1	0
11	.026	1.216	.98	0	.96	.04
12	0	1.003	.999	0	.98	.02
13	-1	1.014	.808	.13	.87	0
14	.166	.997	.879	0	.53	.47
15	0	1.003	.999	0	.98	.02
16	-.723	1.051	.673	1	0	0
17	.182	1.007	.982	0	.84	.16
18	-.181	.948	.901	.64	.22	.13
19	.076	1.855	.963	0	.89	.11
20	-.917	1.063	.551	1	0	0
21	-.39	1.026	.887	.33	.49	.18
22	-.779	1.073	.738	1	0	0
23	.299	1.091	.812	.16	.31	.53
24	-.701	.91	.821	.64	.27	.09
25	-.701	.91	.821	0	1	0
26	-.019	1.053	.997	.02	.98	0
27	-.998	1.801	-.123	.98	.02	0
28	-.998	1.801	-.123	0	1	0
29	-1	.001	-6.966	.58	.42	0
30	.134	1.003	.967	0	.84	.16
31	.134	1.003	.967	0	1	0
32	.942	1.323	.665	.2	.18	.62
33	0	1	0	0	0	1
34	0	1.003	.999	0	.98	.02
35	-.369	1.004	.869	1	0	0
36	-1	1.069	.895	.02	.98	0
37	.094	.981	.928	.58	.2	.22
38	.189	.974	.939	.16	.6	.24
39	.219	1.01	.927	.33	.44	.22
40	-.906	.999	.47	.93	0	.07
41	.206	.999	.978	0	.84	.16
42	.036	.997	.966	.2	.78	.02
43	.321	1.052	.92	.31	.47	.22
44	.213	1.024	.926	0	.71	.29
45	0	1.003	.999	0	.98	.02
46	-.019	1.053	.997	.02	.98	0
47	-.019	1.053	.997	0	1	0
48	-.019	1.053	.997	0	1	0
49	.178	1.315	.551	.84	0	.16
50	-.688	1.041	.733	1	0	0
51	.305	.973	.902	0	.51	.49
52	-.704	1.01	.761	1	0	0
53	0	1.003	.999	0	.98	.02
54	-.665	1.079	.8	.84	.16	0
55	-.571	.885	.841	.56	.36	.09
56	.219	.978	.924	.04	.6	.36
57	.219	.978	.924	0	1	0
58	-.03	1.034	.95	0	.89	.11
59	-.818	1.101	.583	1	0	0
60	0	1.003	.999	0	.98	.02

Table A.2: Individual estimates from Fall 2015 CTB experiment (Part 2)

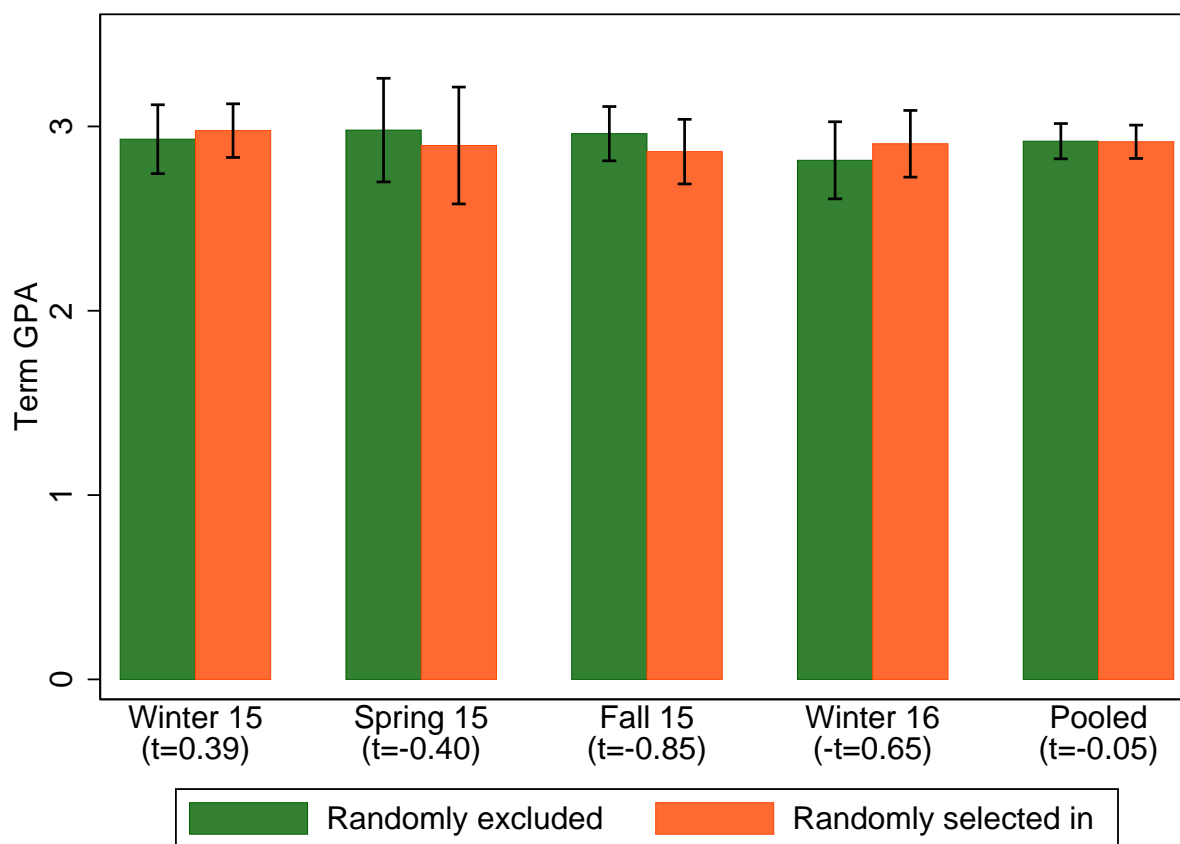
Subject #	Annual Rate	$\hat{\beta}$	$\hat{\alpha}$	Interior	Proportions of Response	
					Zero Tokens Sooner	All Tokens Sooner
61	.304	.99	.915	0	.58	.42
62	0	1.003	.999	0	.98	.02
63	.125	1.056	.947	0	.8	.2
64	-.007	.95	.98	0	.89	.11
65	-.216	1.022	.921	.56	.42	.02
66	.164	.991	.914	.49	.38	.13
67	.206	1.03	.97	0	.84	.16
68	-.403	1.193	.78	.91	.02	.07
69	-1	.56	.239	.96	.04	0
70	0	1.003	.999	0	.98	.02
71	0	1.003	.999	0	.98	.02
72	-.213	.961	.912	.31	.53	.16
73	-.187	1.032	.848	.84	.11	.04
74	-.019	1.053	.997	.02	.98	0
75	0	1.003	.999	0	.98	.02
76	-1	.339	-7.557	.13	.87	0
77	.161	1.015	.964	.24	.69	.07
78	-.583	1.287	.793	.98	.02	0
79	.198	1.03	.969	.16	.71	.13
80	-.019	1.053	.997	.02	.98	0
81	-.89	1.147	.645	1	0	0
8207	.93	0
83	.	.	.	0	1	0
84	.208	1.005	.971	0	.82	.18
85	-.383	1	.99	.16	.84	0
86	.218	.996	.937	.11	.6	.29
87	-.085	1.147	.916	1	0	0
88	.054	.965	.97	0	.84	.16
89	.054	.965	.97	0	1	0
90	-.354	.968	.942	.02	.89	.09
91	0	1.003	.999	0	.98	.02
92	0	1.003	.999	0	.98	.02
93	.336	1.083	.924	0	.69	.31
94	.608	1.086	.61	0	.11	.89
95	-.394	1.276	.989	.04	.96	0
96	0	1.001	.999	0	.93	.07
97	-.019	1.053	.997	.02	.98	0
98	-.067	1.035	.83	.91	0	.09
99	-.067	1.035	.83	0	1	0
100	-.297	.998	.822	.96	0	.04
101	0	1.003	.999	0	.98	.02
102	.113	.966	.97	0	.82	.18
103	.113	.966	.97	0	1	0
104	0	1.003	.999	0	.98	.02
105	0	1	0	0	0	1
106	0	1.003	.999	0	.98	.02
107	.244	.997	.893	.56	.2	.24
108	.21	1.017	.978	0	.87	.13
109	.21	1.017	.978	0	1	0
110	-.749	.968	.804	.96	.02	.02
111	0	1.003	.999	0	.98	.02
112	0	1.003	.999	0	1	0
113	-.997	1.282	.149	.82	.07	.11
114	.111	.965	.969	.04	.78	.18
115	.673	1.235	.67	0	.22	.78
116	0	1.003	.999	0	.98	.02
117	0	1.003	.999	0	1	0
118	.004	.952	.999	0	.93	.07
119	-.995	.829	.568	1	0	0
120	.219	1.013	.964	0	.8	.2

Table A.3: Main results from Winter 2015 experiment

	Achieve Goal			Term GPA		Units Enrolled	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Randomly Entered	0.101 (0.078)	0.119 (0.078)	0.038 (0.093)	0.131 (0.083)	0.021 (0.092)	0.283 (0.469)	0.412 (0.547)
$X \leq 2.5$ Pre-term GPA			0.337* (0.176)		0.457** (0.211)		-0.523 (1.092)
Observations	165	165	165	165	165	165	165
Mean of outcome	0.515	0.515	0.515	2.956	2.956	14.242	14.242
Controls		X	X	X	X	X	X

Notes: Each cell reports the coefficient on an indicator for whether the student was randomly entered to participate in the program from a regression of one of three outcome variables. “Achieve Goal” is an indicator equal to one for students who successfully raised their cumulative GPAs by the end of the Winter 2015 term. “Term GPA” is the student’s Winter 2015 term GPA. “Units Enrolled” is equal to the number of course units the student enrolled in for the Winter 2015 term. Controls include the full set of student characteristics from Table 1. Columns (3), (5), and (7) include an interaction between the treatment indicator with an indicator for whether the student had a pre-Winter 2015 cumulative GPA of 2.5 or lower. Robust standard errors presented in parentheses. Four transcripts from the pool of 169 interested students failed to process. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Figure A.1: Effectiveness of *CollegeBetter* - Mean term Grade Point Average (GPA)



Notes: Each bar reports the mean term grade point average (GPA) by experimental pool and by treatment group. 95% confidence intervals are included for each mean term GPA.

Table A.4: Effectiveness of *CollegeBetter* - Completed commitment challenge

	Full sample						Students w/ ≤ 2.50 Pre-term GPA					
	W15	S15	F15	W16	Pooled		W15	S15	F15	W16	Pooled	
Randomly Entered	0.088 (0.076)	0.048 (0.116)	0.066 (0.079)	0.101 (0.090)	0.076* (0.043)		0.360** (0.159)	-0.111 (0.198)	0.073 (0.138)	0.360* (0.188)	0.189** (0.087)	
Pre-term Cumulative GPA	-0.121* (0.063)	-0.144 (0.125)	-0.047 (0.093)	0.062 (0.105)	-0.078* (0.043)		-0.367** (0.164)	1.354** (0.473)	1.483*** (0.426)	0.129 (0.438)	0.091 (0.191)	
Term Units Enrolled	0.047*** (0.012)	0.053*** (0.017)	0.047*** (0.011)	-0.004 (0.016)	0.035*** (0.007)		0.040*** (0.014)	0.034 (0.052)	0.025 (0.018)	0.008 (0.021)	0.037*** (0.008)	
Pre-term Cumulative Units	0.001 (0.001)	0.003*** (0.001)	0.002** (0.001)	0.002*** (0.001)	0.001*** (0.000)		-0.001 (0.002)	0.002 (0.003)	0.002 (0.002)	0.002 (0.002)	0.000 (0.001)	
Observations	165	64	153	123	505		40	17	39	30	126	
Mean of outcome	0.515	0.453	0.529	0.561	0.523		0.575	0.529	0.564	0.567	0.563	

Notes: Each column presents results from a single regression. The outcome variable is an indicator for whether the student completed the commitment challenge for the term. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

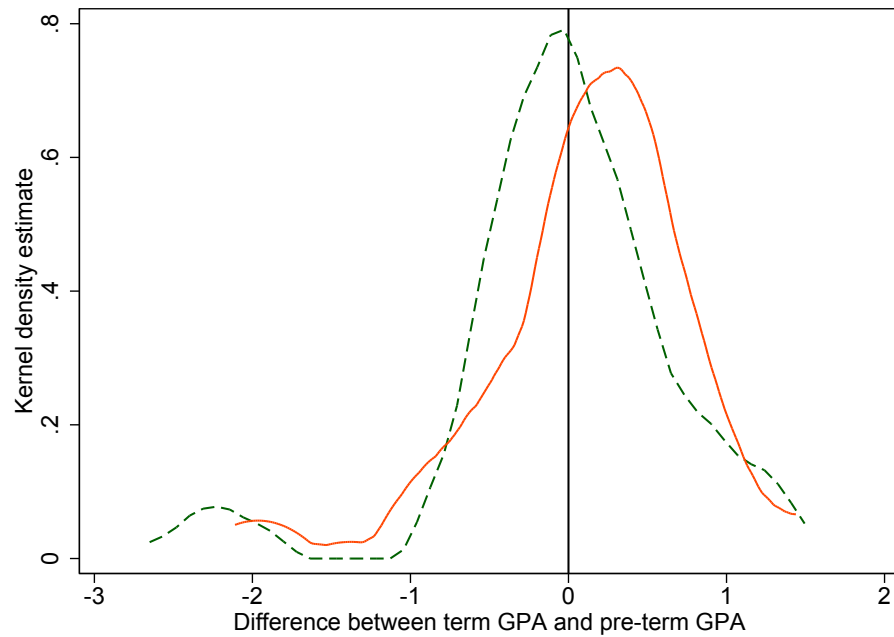
Table A.5: Effectiveness of *CollegeBetter* - Term units enrolled

	Full sample						Students w/ ≤ 2.50 Pre-term GPA					
	W15	S15	F15	W16	Pooled		W15	S15	F15	W16	Pooled	
Randomly Entered	0.209 (0.430)	-0.338 (0.823)	0.111 (0.439)	0.844 (0.521)	0.279 (0.254)		-0.126 (1.098)	2.284* (1.202)	0.402 (1.263)	1.315 (1.279)	0.795 (0.614)	
Pre-term Cumulative GPA	0.091 (0.447)	1.691** (0.748)	1.591** (0.683)	1.554*** (0.549)	0.921*** (0.294)		0.350 (1.154)	-0.197 (1.993)	6.819 (4.291)	-0.640 (2.376)	0.788 (1.072)	
Pre-term Cumulative Units	0.003 (0.004)	0.004 (0.007)	-0.007* (0.004)	-0.002 (0.006)	0.002 (0.002)		-0.007 (0.010)	0.017 (0.011)	-0.011 (0.011)	0.012 (0.019)	-0.001 (0.006)	
Observations	165	64	153	123	505		40	17	39	30	126	
Mean of outcome	14.242	13.750	14.739	14.846	14.477		13.950	13.118	13.603	13.667	13.663	

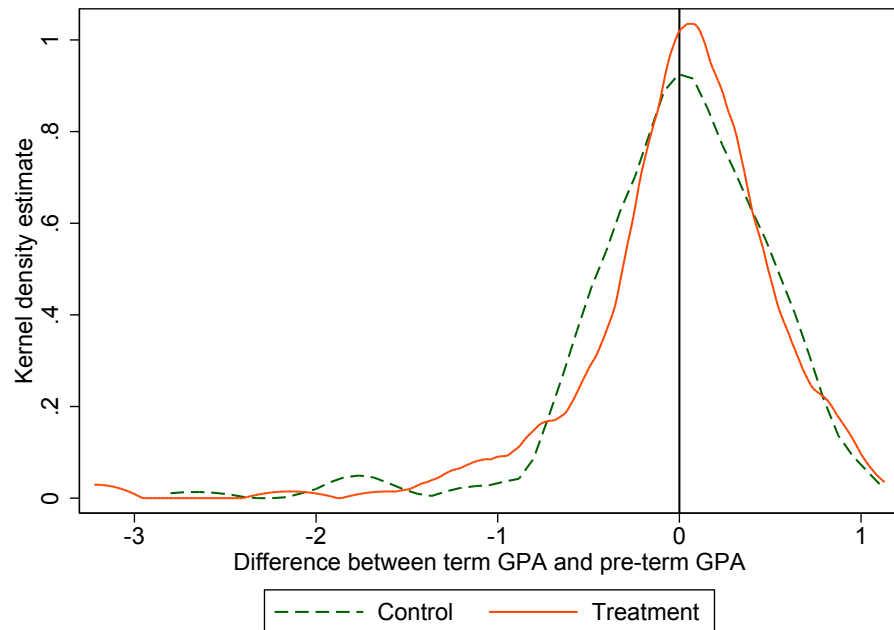
Notes: Each column presents results from a single regression. The outcome variable is the number of units the student enrolled in for the term. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels, respectively.

Figure A.2: Distribution of Grade Point Average differential

(a) Students with ≤ 2.50 cumulative GPA entering term



(b) Students with > 2.50 cumulative GPA entering term



Notes: The density function of GPA differential (term GPA minus pre-term cumulative GPA) is estimated using kernel density estimation, and plotted separately by treatment (students randomly selected to participate) and control (students randomly excluded) groups. Epanechnikov kernel used to calculate kernel density estimates.

Appendix A.1 - Questions from Winter 2015 survey

1. This is the most important question! Would you like to participate in College Better?

-Remember, you must submit \$10 of your own cash at the end of class.

-If you do not raise your cumulative GPA by the end of the Winter 2015 quarter, you will lose your \$10.

-In order to raise your cumulative GPA, you must get a GPA in the Winter 2014 quarter that is higher than your cumulative GPA prior to the Winter 2015 quarter.

-If you do achieve this goal, you will split the pot of money with everyone else who also succeeded and raised his or her cumulative GPA.

-We will email you instructions about how to claim your winnings in April.

-Note that you will not lose money so long as you raise your cumulative GPA.

• Yes, I want to participate in College Better for the Winter 2015 quarter

• No, I do not want to participate in College Better for the Winter 2015 quarter

2. What is your gender? • Male • Female

3. What is your race? Check all that apply. • Caucasian • Black • Chinese • Latino/Hispanic • Japanese • Korean • Other Asian • Hawaiian/Pacific Islander • Other

4. What is your major(s)/minor(s)?

5. In what year were you born?

6. Select the number of years you've attended UC Davis • 1 • 2 • 3 • 4 • 5 • 6 • 7 • 8+

7. If you answered "Yes" in question 1 and chose to participate in College Better, do you think your participation will influence which courses you choose to enroll in for the remainder of this quarter? • Yes • No • I do not want to participate in College Better

8. Have you participated in an economics experiment before? • Yes • No

9. Do you have a job? Check all that apply

• Yes-Full Time • Yes-Part Time • Yes-Summer Job • No

10. Do you pay in-state or out-of-state tuition? • I pay in-state tuition • I pay out-of-state tuition

11. Are you an international student? • Yes • No

12. Make a guess! What fraction of students who participate in College Better will accomplish the goal of raising their cumulative GPA? Whoever guesses closest to the true success rate will win \$50 at the end of the quarter. You do not have to participate in College Better to be eligible for the \$50 prize.

13. Suppose we allowed the buy-in fee to be determined by the student, where the minimum buy-in would be \$5 and the maximum buy-in would be \$100, and where payouts are proportional to the amount of money the student contributed. Students who contributed a larger buy-in would take a larger cut of the pot, so long as they completed the challenge. Would you participate in this group, where the goal is to raise your cumulative GPA? If so, how much would you buy-in for?

• Yes • If Yes, how much? • No

14. College Better is currently in the process of being developed into an online program. Are you interested in receiving emails with more information about College Better in the future? • Yes • No

15. If you are not participating, please explain why (check all that apply): • I am participating! • Too expensive. I don't have \$10 to commit to College Better • I don't need it • Even if I do raise my cumulative GPA, I doubt I'll make that much money • I don't think that I will successfully raise my cumulative GPA • Other:

For each of the following statements, select the response that best matches you:

16. I often pay bills right at the deadline • Not me • Like me

17. I often get distracted on the internet (Facebook, Twitter, Youtube) when trying to study or be productive • Not me • Like me

18. I wait until the last second to buy gifts (birthday, Christmas, etc.) • Not me • Like me

19. I often start studying for exams too close to the exam date such that I feel regret • Not me • Like me

20. Did you make a New Year's Resolution(s) for 2015? • Yes • No

Appendix A.2 - Questions from Fall 2015 lab experiment survey

Welcome! This experiment consists of two parts. In the first part, the experimental administrator Lester Lusher and his assistants will divulge information about a program founded by the Department of Economics at UC Davis called *CollegeBetter*. Afterward, if you are interested in the *CollegeBetter* program, we will give you an opportunity to join for the Fall quarter. If you are uninterested in *CollegeBetter* (or signed up for *CollegeBetter* before coming to today's experiment), then you will simply wait for the second part of the experiment. At this point, as you wait for Lester to describe the *CollegeBetter* program, click the following link to learn more about *CollegeBetter*: www.collegebetter.com

Now that Lester has described the *CollegeBetter* program, and you have decided whether or not to join *CollegeBetter*, please fill out the information below about yourself. Once you complete all the questions, click "Continue" and read the instructions for the second part of the experiment. You may begin the second part of the experiment on your own, or you can wait until Lester describes the experiment to everyone in the session.

1. Did you sign up for the *CollegeBetter* Fall 2015 pool? • Yes! I did so just now. • Yes, and I did so before this experiment! • No, I am uninterested.
2. Here's your first chance to win money! How much money do you think those who succeed in the *CollegeBetter* Fall 2015 pool will take home, net of the \$20 buy-in?
-If your guess is within \$0.15 of the actual amount won by each student at the end of the Fall 2015 term, then you will earn \$15! Even those who do not sign up for *CollegeBetter* are eligible for this prize.
3. What is the likelihood that your Fall 2015 term GPA will exceed your current cumulative GPA? (On a scale from 1 - very unlikely to 7 - very likely) • 1 • 2 • 3 • 4 • 5 • 6 • 7
4. How difficult are the courses you are taking this Fall 2015 quarter? • More difficult than average • Average difficulty • Less difficult than average
5. What is your gender? • Male • Female

6. Do you have a job/internship? • Yes • No
7. How many hours/week to you spend at your job/internship?
8. If you have a job/internship that pays, how much do you earn per hour?
9. Is English your first language? • Yes • No
10. What is your year in school? • Freshman • Sophomore • Junior • Senior • Fifth year+
11. What is/are your major(s)?
12. What College(s) is/are your major(s) in? (Check all that apply) • Letters and Science • Agriculture and Environmental Sciences • Biological Sciences • Engineering
13. What best describes your study habits? • I keep up on some classes but cram before exams for others • I study every day and keep up on classes • I generally don't keep up on my classes and tend to cram for all my exams
14. Have you ever pulled an all-night to study for an exam? • Yes - Often • Yes - Not often • Never
15. If you took the SAT, what was your math score?
16. If you took the SAT, what was your verbal score?
17. If you took the SAT, what was your writing score?
18. If you took the ACT, what was your Math score?
19. If you took the ACT, what was your English score?
20. If you took the ACT, what was your Reading score?
21. If you took the ACT, what was your Science score?
22. Do you smoke cigarettes? • Yes • No

23. Do you exercise regularly (three or more times per week)? • Yes • No
24. Do you find that you are able to maintain the body weight that you like? • No I struggle with my weight • Yes I'm basically able to control my weight
25. Do you have a credit card? • No and I don't want one • No but I want one • Yes
26. If yes, what is the estimated balance on all of your credit cards combined?
27. If you pay your credit card bill yourself, each month do you normally: • Pay the balance in full • Pay the minimum payment • Pay more than the minimum but less than the full balance
28. Do you have a bank account? • Yes • No
29. If you have a bank account, what interest rate do you earn on your bank balances?
30. Do you have a debit card? • Yes • No
31. How much money would you say you spend in total in an average week?
32. How much money do you spend monthly on rent?
33. How willing are you to take risks in general? (On a scale from 1, unwilling to 7, very willing)
• 1 • 2 • 3 • 4 • 5 • 6 • 7
34. How patient are you in general? (On a scale from 1, extremely impatient, to 7, extremely patient) • 1 • 2 • 3 • 4 • 5 • 6 • 7
35. Have you participated in an economics experiment before? • Yes • No
36. Do you pay in-state or out-of-state tuition? • I pay in-state tuition • I pay out-of-state tuition
37. Are you an international student? • Yes • No
38. Which of the following reflects this statement as it applies to you

-“I wait until the last second, or am often late, when paying bills” • Not like me • Like me

39. Which of the following reflects this statement as it applies to you

-“I often get distracted on the internet (e.g. facebook, youtube) when trying to study or be productive” • Not like me • Like me

40. Which of the following reflects this statement as it applies to you

-“I wait until the last second to buy gifts (birthday, Christmas, etc)” • Not like me • Like me

41. Which of the following reflects this statement as it applies to you

-“I often start studying for exams too close to the exam date such that I feel regret” • Not like me • Like me

42. Which of the following reflects this statement as it applies to you

-“I have trouble getting up in the morning” • Not like me • Like me

43. Did you make a New Year’s Resolution for 2015? • Yes • No

44. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? (Please give your answers in days.)

45. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets? (Please give your answer in minutes.)

46. If a baseball and a bat cost \$1.10 together, and the bat costs \$1.00 more than the ball, how much does the ball cost?

Appendix A.3 Nonlinear Least Squares (NLS) estimation using Convex Time Budgets

In the Fall 2015 lab experiment, 120 subjects each completed 45 Convex Time Budget ([Andreoni and Sprenger, 2012](#)) allocation decisions. Assume that each subject j makes decisions $c_{t_{ij}}$, $i = 1, 2, \dots, 45$, decisions according to [equation \(2\)](#) with some mean-zero, serially correlated error. More

specifically, let

$$f(m, r, k, t; \beta, \delta, \alpha) = \begin{cases} \left[\frac{(\beta \delta^k (1+r))^{\left(\frac{1}{\alpha-1}\right)}}{1+(1+r)(\beta \delta^k (1+r))^{\left(\frac{1}{\alpha-1}\right)}} \right] \times m & \text{if } t = 0 \\ \left[\frac{(\delta^k (1+r))^{\left(\frac{1}{\alpha-1}\right)}}{1+(1+r)(\delta^k (1+r))^{\left(\frac{1}{\alpha-1}\right)}} \right] \times m & \text{if } t > 0 \end{cases},$$

such that

$$c_{tij} = f(m, r, k, t; \beta, \delta, \alpha) + e_{ij}.$$

Stacking the 45 allocation decisions for individual j ,

$$\mathbf{c}_{tj} = \mathbf{f}(m, r, k, t; \beta, \delta, \alpha) + \mathbf{e}_j.$$

The vector \mathbf{e}_j is zero in expectation with (45×45) variance-covariance matrix \mathbf{V}_j . Then, stacking over all 120 subjects,

$$\mathbf{c}_t = \mathbf{f}(m, r, k, t; \beta, \delta, \alpha) + \mathbf{e}.$$

Assuming that errors are uncorrelated across individuals (but allowing for e_{ij} to be correlated within individuals across allocation decisions), $E(\mathbf{e}'_j \mathbf{e}_g) = 0$ for $j \neq g$. Then, \mathbf{e} is zero in expectation with covariance matrix $\mathbf{\Omega}$, a block diagonal (5400×5400) matrix, with individual covariance matrices \mathbf{V}_j . Defining the criterion function $S(m, r, k; \beta, \delta, \alpha)$ as the sum of squared residuals,

$$S(m, r, k, t, \beta, \delta, \alpha) = \sum_{j=1}^{120} \sum_{i=1}^{45} (c_{tij} - f(m, r, k, t; \beta, \delta, \alpha))^2,$$

and minimize $S(\cdot)$ using non-linear least squares with standard errors clustered on the individual level to obtain $\hat{\beta}$, $\hat{\delta}$, $\hat{\alpha}$. *Stata* has a built-in econometrics package that permits the estimation of individual-level parameters using NLS. $\hat{\mathbf{\Omega}}$ is estimated as the individual-level clustered error covariance matrix.