NBER WORKING PAPER SERIES

THE REMARKABLE UNRESPONSIVENESS OF COLLEGE STUDENTS TO NUDGING AND WHAT WE CAN LEARN FROM IT

Philip Oreopoulos
Uros Petronijevic

Working Paper 26059
http://www.nber.org/papers/w26059

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2019

This project would not have been possible without the enthusiastic support of the Department of Economics at the University of Toronto, especially the instructors of first-year economics courses, Lee Bailey, Tenzin Yindok, Bob Gazzale, Bill Wolfson, Kieran Furlong, Avi Cohen, Jack Parkinson, Michael Ho, Zachary Mahone, Gordon Cleveland, Paola Salardi, Loren Brandt, Eric Kam, Erica Rzepecki, and Peter Foltin. We are also indebted to Aaron DeMello and James Lahey for their tireless commitment to designing the project's well-run website and online exercises, as well as for their help with organizing and extracting the experimental data. Matthew Hendrickson helped extract the administrative data. We are very grateful to all the coaches who provided experienced assistance to participants, and to the Social Science Research Council of Canada (Insight Grant #435-2018-0268 and Insight Development Grant # 430-2017-00779), the Jamal Poverty Action Lab, and the University of Toronto's Learning and Education Advancement Fund for financial support. The paper was completed while Oreopoulos was visiting the Russell Sage Foundation as a Visiting Scholar, almost entirely without procrastination. We are grateful for seminar participant feedback from numerous institutions. All remaining errors are our own. The experiments in this paper were registered with the AEA RCT Registry (AEARCTR-0000810). The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Remarkable Unresponsiveness of College Students to Nudging And What We Can Learn from It
Philip Oreopoulos and Uros Petronijevic
NBER Working Paper No. 26059
July 2019
JEL No. I2,J24

ABSTRACT

We present results from a five-year effort to design promising online and text-message interventions to improve college achievement through several distinct channels. From a sample of nearly 25,000 students across three different campuses, we find some improvement from coaching-based interventions on mental health and study time, but none of the interventions we evaluate significantly influences academic outcomes (even for those students more at risk of dropping out). We interpret the results with our survey data and a model of student effort. Students study about five to eight hours fewer each week than they plan to, though our interventions do not alter this tendency. The coaching interventions make some students realize that more effort is needed to attain good grades but, rather than working harder, they settle by adjusting grade expectations downwards. Our study time impacts are not large enough for translating into significant academic benefits. More comprehensive but expensive programs appear more promising for helping college students outside the classroom.

Philip Oreopoulos
Department of Economics
University of Toronto
150 St. George Street
Toronto, ON M5S 3G7
CANADA
and NBER
philip.oreopoulos@utoronto.ca

Uros Petronijevic
York University
Department of Economics
Vari Hall
4700 Keele Street
Toronto, ON M3J 1P3
CANADA
upetroni@yorku.ca

A randomized controlled trials registry entry is available at AEARCTR-0000810
Appendices are available at http://www.nber.org/data-appendix/w26059

I. Introduction

In developed nations like the United States and Canada, higher education continues to be promoted as a key tool for improving productivity and wages, fostering innovation, reducing inequality, and encouraging economic growth (e.g., Deming, 2018; Acemoglu and Autor, 2012; Gurria, 2009; Psacharopoulos and Patrinos, 2018). College enrollment has steadily increased over the last several decades; 70 percent of young adults have at least some postsecondary education (U.S. Census Bureau, 2018, Statistics Canada, 2016). Many individuals are motivated to enroll by diminishing labor market opportunities for those with only a high school degree, or because they want to avoid working in low-skilled, low-paying jobs. Others desire to acquire new skills and compete for high-paying, intellectually stimulating, and satisfying occupations.

Although individuals with more education clearly realize better average outcomes than those with less, simply enrolling in college does not guarantee students will be better off. First, a substantial fraction of current enrollees fails to graduate. In the United States, the six-year completion rate among students beginning a four-year postsecondary program is 54.8 percent (Shapiro et al., 2019), while only about one in three students who enroll in two-year programs go on to graduate. Students who complete only some college education must incur large up-front costs but can expect to earn similar incomes as individuals with only a high school degree, especially among college enrollees who are in the middle or bottom of their entry class distribution (Oreopoulos and Petronijevic, 2013). Second, many students who do earn a college degree do so with weak grades and questionable human capital gains. Arum and Roksa's (2011) seminal research, for example, finds little evidence of improved skills and learning from attending college: two years after entry, students' scores on a test designed to measure abilities in critical thinking, complex reasoning, and writing increased (on average) by only 18 percent of the test's standard deviation, with no evidence of significant change in these skills for almost half of those surveyed. A leading explanation for low learning gains is that students invest little time into their studies. Past research shows that most college students spend fewer than 15 hours a week preparing outside of lecture for all of their courses, much less than the 25 to 40 hours per week usually recommended by university administrators (Brint and Cantwell, 2010; Babcock and Marks, 2011; Farkas et al., 2014). A need to spend time commuting or working for pay does not explain low levels of studying; rather, time-use surveys reveal that many students spend their time socializing or taking part in recreational activities (Arum and Roksa, 2014, Oreopoulos et al., 2019).

Even against the current backdrop of low completion rates and study effort, estimates of the *average* return to college remain significantly positive, including for students at the margin of admission (Oreopoulos and Petronijevic, 2013; Zimmerman, 2014; Ost et al., 2018). How well the investment pays off, however, depends on many factors, including financial cost (Lochner and Monge-Naranjo, 2012), incoming ability (Oreopoulos et al, 2018), college and teaching quality (Chetty et al., 2017; Hoxby and Stange, forthcoming), and field of study (Kirkeboen et al. 2016). The value of college investments also depends on the role students are willing and able to take in the development of their own human capital. Of particular importance are students' stocks of cognitive and non-cognitive skills when entering college, the information they have about how to study effectively, and their willingness to devote time to studying at the expense of other activities (e.g., Nyblom, 2015). The key question motivating this paper—and many education researchers—is whether low-touch interventions can affect these traits and behaviors and, in turn, cause improvement in academic outcomes and overall college experiences.

Research from behavioral economics suggests that psychological barriers may prevent students from realizing their preferred long-run outcomes (Lavecchia et al. 2016). Delaying studying, neglecting to take advantage of free tutoring services, or consistently getting distracted by social media are examples of how students' best intentions can go awry. Youth are particularly prone to these kinds of barriers because the brain's prefrontal cortex, which regulates forward-looking and critical thinking, does not fully develop until around the age of twenty-five (Giedd et al., 2012). Interventions that have proven effective at combating these barriers require students to complete a one-time action or a series of well-defined steps. For example, past research demonstrates the promise of nudging youth towards completing the college application process (Bettinger et al. 2012; Oreopoulos and Ford, 2019; Page et al. 2016), renewing financial aid (Castleman and Page, 2016), choosing selective colleges (Dynarski et al., 2018; Castleman and Sullivan, 2019), and choosing courses on time (Castleman and Page 2015). In contrast, nudging students toward improving study habits and attitudes has proven more challenging because it requires a sustained change in behavior over a prolonged period.

Prior studies find that offering structured, intensive and personalized support can help. One of the most successful experimentally-tested programs is the Accelerated Study in Associate Program (ASAP), which requires that college students enroll full-time, attend mandatory tutoring, receive regular counseling and career advising services, and are awarded free public transportation passes and funding for textbooks. ASAP doubled graduation rates at the City University of New York and had similarly large impacts on persistence in a replication attempt in Ohio (Scrivener et al., 2015; Sommo et al. 2018). Stay the Course (STC) is another college-based support system in which eligible students in Fort Worth, Texas, received regular intensive case management assistance along with emergency financial support. Completion rates for those eligible increased

3.7 percentage points, with this impact coming entirely from female students (Evans et al., 2017). A third comprehensive program is the Carolina Covenant aid program, in which eligible students receive financial aid and a variety of services, including career exploration workshops, peer mentoring, and support with navigating the university's wellness and academic programs. Eligibility increased credit accumulation through the first three years of college and suggestive evidence points to positive impacts on graduation rates (Clotfelter et al. 2018).

While encouraging, these programs cost thousands of dollars per student and are difficult to scale. We also know little about how they improve academic outcomes, and why they do not help even more students, as one might expect given their intensity. To explore whether offering 'lighter-touch' and less costly interventions might also benefit students, and to learn more about the mechanisms by which students can be assisted during college, we created a research environment in which we could conduct experiments with thousands of representative college students over many years. Teaming up with instructors of first-year economics courses at the University of Toronto (UofT)—who collectively teach about 5,000 students per year, including a quarter of all first-year students—we created the *Student Achievement Lab*. As one of the lab's key design features, instructors were asked to set aside a small grade requirement for each of their incoming classes to complete a one- to two-hour online 'warm-up exercise' within the first two weeks of the fall semester. Students registered an account, took a short introductory survey, and were then randomly assigned to treatment or control groups. Some treated students received follow-up communication through email, text messages, or face-to-face contact. We then linked our survey data to the university's administrative records to track academic outcomes and, in some cases, conducted follow-up surveys to collect non-academic outcomes such as study habits, aspirations, mental health, and perceptions of overall university experience.

Over a span of five years, we designed and tested several promising interventions based on past research and consultations with college administrators. In this paper, we group the interventions into six categories: (1) *Goal Setting*, in which students were asked to think about future aspirations and the importance of their current actions in relation to those aspirations; (2) *Mindset*, in which students were told about how adopting positive perspectives towards struggling with course material or feeling out of place on campus can lead to greater resilience and persistence; (3) *Online Coaching*, in which students were provided detailed advice about how to be a successful student; (4) online coaching with intensive follow-up communication through *One-Way Text Messages*, (5) *Two-Way Text Messages* between students and experienced upper-year student coaches, and (6) *Face-to-Face* regular meetings with coaches. Across all five years and six interventions, our total sample consists of approximately 25,000 students. While our previous studies (cited below) present separate results for some of these interventions (and include additional estimated effects from more subtle treatment variations and sub-analyses), presenting our five-year effort collectively in this paper facilitates a broader discussion of the overall potential for applying behavioral economics at the college level.

Our findings suggest that, at least for large four-year colleges like UofT, none of the interventions we test can generate a significant improvement in student grades or persistence. We can rule out treatment effects larger than 7 percent of a standard deviation and find precise null impacts even when focusing on students more at risk of performing poorly and those attending the two satellite campuses that are more representative of less-selective commuter colleges.

These results, however, belie more intermediate and subjective impacts. We find that our interventions using online coaching, text-based coaching, and especially face-to-face coaching improve study habits such as weekly study hours and the likelihood of meeting with a tutor or instructor. Study time increases, on average, by approximately two hours per week. But the estimated association between studying and grades (and causal estimates of these relationships from prior work) suggests that these improvements are not large enough to generate a significant change in aggregated academic outcomes. Turning to college experience, the coaching interventions improve subjective well-being, reduce stress, and make students feel more supported. Such impacts may be important in their own right, given the increase in attention by administrators to student experience and mental health.

To frame these results in greater context and explore the mechanisms driving our intermediate treatment effects, we develop a model of student effort decisions in which students choose study intensity based on their preferences, abilities, expected effort-to-grade relationships, and psychological barriers that lead actual effort to differ from target effort. We then use our unique sample and survey data to measure both changes in these factors between the start and end of the fall semester and the role our interventions played in affecting student behavior. We find clear evidence that actual hours of study are significantly lower than target hours, but that our interventions were not able to reduce these gaps. Our online and text-based coaching did reduce students' perceptions around the effectiveness of cramming for tests, and increased their motivation to attain higher grades, though the magnitude of these effects were small. Even if we were able to eliminate the behavioral gap between actual and target study time, in addition to the impacts we had on information updating and the perceived benefits of studying, we estimate that mean grades would have increased by only about 3.5 percentage points, or 27 percent of a standard deviation. This is not a large increase and suggests that more intensive interventions—along the lines of the comprehensive programs discussed above—are needed to meaningfully change students' performance and behavior.

Lastly, we find that students adjust effort and grade expectations when concluding less effort is required to attain the same grade, but not if they realize more effort is required. That is, when learning that it is easier to reach performance goals than originally believed, students respond by decreasing study time. However, students who realize more study time is required to attain good grades do not adjust much, but rather seem to reluctantly accept they will perform worse than originally believed. Their predicted and actual grades fall, consistent with our model of students adjusting their grade range targets downwards (no longer aiming for and expecting As, for example, but rather settling for Bs or Cs).

The remainder of the paper proceeds as follows: in Section II, we describe our *Student Achievement Lab* setup and review each field experiment. We also describe our data and methodology, and present descriptive statistics. Section III presents the overall results. In Section IV, we present a model of student effort to interpret our results and discuss rational and time-inconsistent explanations of poor performance, as well as opportunities for policies to help. We quantify the impact our experiments had on individuals' study-to-grade expectations and procrastination in Section V, and offer concluding remarks in Section VI.

II. The Student Achievement Lab: Setup, Interventions, and Data

A. Setup

The *Student Achievement Lab* (SAL) began in the fall of the 2014-15 academic year at the University of Toronto (UofT). In that first year, we conducted experiments only at the university's west-end satellite campus located in the city of Mississauga (we refer to this campus as UTM). UTM is primarily a commuter campus with approximately 12,500 undergraduate students. Roughly 80 percent of students live at home with their parent(s), slightly less than a quarter say that the campus was their first choice, and the majority say they work at least part-time while attending. Many of the students are immigrants or children of immigrants. Among undergraduates who entered in 2001, only 38 percent completed a degree in four years, while the six-year graduation rate was about 70 percent. SAL expanded in the following year (fall of 2015) to include UofT's two other campuses. The campus located in the east end of Toronto, the University of Toronto at Scarborough (UTSC), is similar to UTM, as it is primarily a commuter campus with completion rates of about 73 percent. UofT's St. George campus, UTSG, is located downtown and is more representative of a top four-year public college in the United States.[1] Students apply to each campus separately. Not surprisingly, UTSG is more selective and six-year completion rates are higher, at about 77 percent.[2]

During the fall semesters between 2014 and 2019, instructors of first-year economics courses at UofT incorporated into their course curriculum a small participation grade (usually 2 percent) for the completion of an online warm-up exercise lasting, on average, about an hour, with a deadline generally within the first two weeks of class. The grade requirement was highly effective in making almost all students participate (95 percent of all registered students at the start

[1] The St. George campus is ranked as one of the top universities in the world: https://cwur.org/2018-19.php.
[2] The St. George Arts & Science program is about twice as big as UTM and UTSC. In 2016-17, the full-time headcount at St. George, UTM, and UTSC was 25,056, 12,967, and 11,902 respectively (University of Toronto, 2018).

of the course, which at approximately 5,000 students per year constitutes 10 percent of the entire undergraduate student population).[3]

Students taking introductory economics courses are representative of the school's undergraduate student body. About a quarter of all first-year students at UofT enroll in a first-year economics course, half of which take the course as a requirement for their planned program of study. Students wanting to continue afterwards into one of the schools' competitive commerce or management programs must obtain a minimum grade (usually 67 percent) as part of that program's admissions requirements. Each year, about 30 percent of students drop their economics course before receiving an official grade. Of those who do complete, the 25th, 50th, and 75th percentiles in economics grades distribution are 58, 69.5, and 78 percent, respectively (using our baseline sample). Figures 1 and 2 depict our students' academic performance overall. Figure 1 displays the distribution of grades averaged across courses completed by the end of the first fall semester. The distribution is similar to that for economics alone, with the median grade being 70.5 percent and the 25th percentile being 62.0 percent. Figure 2 shows the histogram of credits completed at the end of the first school year for our sample. Many students initially enroll in five credits in order to try to complete their program in four years,[4] but by the end of the year, many drop some credits or fail to complete their courses. Only 30 percent of our sample received 5 or more credits by the end of the school year.

After logging in using their personal UofT account or creating and verifying a new account, students proceeded through the warm-up exercise by first taking a short initial survey to collect data not available administratively (such as parents' education, grade and study expectations,

---

[3] We restrict our sample to full-time students, defined as those paying full-time tuition, which permits them to enroll in at least 3.5 course credits over the school year.

[4] Students require 20 completed credits to earn a degree.

education aspirations, and subjective tendencies to cram for exams). They were then randomized into different groups, which we categorize and describe below.

During the last three years of the experiments, at the end of the fall semester or at the beginning of the next (winter) semester, we conducted a short follow-up survey also for a participation grade (usually worth 1 percent of the course's final grade for completion). We asked questions not available from administrative data, including questions about study habits, perceived learning outcomes, subjective well-being, attitudes towards grades, challenges with procrastination, and open-ended questions about first semester experiences, advice to other students, and feedback from treated students about the interventions.

B. Interventions[5]

1. Personality Test (Control Group)

Students assigned to the control group were given a set of questions about time preferences, non-cognitive abilities and interests. In order to make the exercise last as long as the treatment interventions, Control Group students were given two Big Five[6] personality tests: one based on an absolute score (e.g., Donnellan et al., 2006), making it possible for a student to score high in all five traits, and another based on a relative score (e.g., Hirsh and Peterson, 2008), indicating the extent to which one trait dominates a student's personality profile relative to other traits. The

[5] All surveys and interventions in their original form are available to peruse online at https://studentachievementlab.org. For additional operational details not all covered in this paper, readers may also refer to appendices provided in Beattie et al. (2016) for the Personality Test, Dobronyi et al. (2019) for the Goal Setting exercise, Oreopoulos and Petronijevic (2018), Oreopoulos et al. (2018), and Oreopoulos et al. (2019) for the online and follow-up coaching exercises, and Logel et al. (in progress) for the mindset exercises.

[6] The five traits are agreeableness, conscientiousness, extraversion, openness to experience, and emotional stability.

control group was also asked questions about risk tolerance (e.g., Dohmen et al., 2011), time preferences (e.g., Andersen et al., 2008), and grit (e.g., Duckworth and Quinn, 2009). The test took approximately 45 to 60 minutes to complete. Students were emailed a short report describing their relative Big Five scores and told that they might be interested in knowing which of their traits are most and least dominant.[7]

2. Goal Setting

Treated groups in the 2014-15 school year were assigned to an online warm-up exercise designed to focus students on their long-term goals. If students overemphasize the present, prompting them to think more carefully about their future may help remind them of the link between current behavior and long-term consequences (Locke and Latham, 2002; Locke et al., 1981; Smith, et al. 1990, Lavecchia, Liu, and Oreopoulos, 2016). Reflecting on the future may also lead to students discovering relevant knowledge and using more efficient strategies for achieving desired goals. Goal setting is also believed to decrease stress (Elovainio & Kivimäki, 1996) and increase working memory (see Morisano [2008] for an overview), making students with clear goals more likely to complete college (Braxton et al., 2004; Kirby & Sharpe, 2001).

[7] The Personality Test was not intended to affect subsequent academic performance or behavior, but data from respondents was used to explore which background and non-cognitive trait variables best predict the wide variance in first-year college performance. Beattie et al (2017) find that students who perform far below expectations also self-report greater tendency to procrastinate and being less conscientious ('gritty') than their peers. Those who perform unexpectedly and exceptionally well express purpose-driven goals and an intent to study more hours per week to obtain a high GPA. In a separate paper that uses follow-up survey data from SAL, Beattie et al (2019) examine the association between intermediate study inputs during a college semester and find that poor time management and lack of study hours are most associated with poor academic performance while large amounts of study time and regular use of student services are most associated with good academic performance. Worth noting as a prelude to this paper's discussion, both of these papers find that a student's high school grade used for admission is, by far, the most predictive variable for first-year performance, and that the additional non-cognitive variables examined do not improve predicted performance by much. A large variance remains even after accounting for observed differences in student background and study behavior.

Our Goal Setting intervention is motivated by the large effects found by Morisano et al. (2010). The authors conducted a randomized field experiment on 85 psychology undergraduates at McGill University with GPAs below 3.0. Students were offered financial remuneration to sit in a classroom and complete a written exercise lasting about two hours. In Part I, treated students were asked to write about their values and set goals that are meaningful, specific, challenging, and attainable. Students were asked to envision their ideal future social life, family life, and career, and to write about how to maintain a balanced life. In Part II, they were encouraged to identify seven or eight more specific goals and to examine each goal carefully, explaining why each was important and vividly describing potential obstacles and strategies for overcoming them and realizing their preferred future. The exercise took about two hours to complete. Relative to a control group receiving a personality test, the reported treatment effect was extremely large: an increase in end-of-semester GPA of about 70 percent of a standard deviation.

We designed a similar exercise to the one used by Morisano et al. (2010) from a version of the materials provided by one of the study's co-authors (which also closely relates to the exercise used in Morisano (2008) and in Schippers et al. (2015)). As discussed in Dobronyi et al. (2019), we view the differences between our exercise and the one by Morisano et al. (2010) as being very slight, with the possible exception that the McGill study had participants complete the task using paper in one sitting, and had adjudicators hand-check that the exercises were given sufficient consideration. All our interventions are delivered online. We instruct students to take their time to work through the exercise so that they might benefit and so that their responses might be used for research by the university to help other students. Minimum word count and time-on-page restrictions also were used, with students shown a pop-up window encouraging them to write in more detail or slow down to benefit from the task. Open-ended responses to the exercises suggest that a large majority of students took the exercise seriously and wrote inspiring and thoughtful responses.

To increase saliency of future goals, and to offer specific study tips to help during the school year, half of the Goal Setting treated students were offered the opportunity to receive one-way text-message reminders. Seventy-five percent opted in to receive text messages by providing their cell phone number. The remainder received email instead. The messages consisted mainly of academic tips and motivational support, sent three times a week during both the fall and winter semesters. Some reminders were personalized with goal-oriented messages, making explicit reference to the individual-specific goals each student provided during the completion of the online exercise. The full set of messages is recorded in Appendix D in Dobronyi et al. (2019). In this paper, we group all treated students who received the Goal Setting online exercise into the same category, regardless of whether they received the one-way follow-up communication.

3. Mindset

After targeting students' present-bias with Goal Setting experiments in the 2014-15 school year, we turned to another set of popular interventions from social-psychology called 'Mindset'. Mindset interventions attempt to encourage students to adopt a more positive perspective (a more positive mindset) when faced with setbacks or challenges in school. They originate from Carol Dweck and her colleagues' findings that differences in praise to children while working on abstract reasoning problems (e.g., "you must have worked hard" versus "you must be smart") affects subsequent motivation and performance (Mueller and Dweck, 1998). Dweck theorized that students inclined to view ability as innate—with a fixed mindset—were more likely to become discouraged by doing poorly on an initial test than students inclined to view their intelligence as malleable—with a growth mindset (Dweck, 1986). She and others suggested that positive mindsets could be encouraged by introducing students to relatable stories of resilience and reflective writing exercises (Yeager and Walton, 2011; Cohen and Garcia, 2014; Walton, 2014). Research in this area has examined ways to not only encourage growth mindsets (e.g., "I learn from my mistakes") but other positive perspectives about school such as social-belonging mindsets (e.g., "I will come to love this school and my classmates"), and academic mindsets (e.g., "What I'm learning in school relates to my life and affects my career"). Different interventions are used to target different challenges students face. Although the saliency of these short, often one-time exercises may fade over time, they may continue to have long-term impacts through encouraging a reinforcing pattern of positive habits and outcomes (Yeager & Walton, 2011).

We worked with three prominent researchers in this area, Christine Logel, Greg Walton, and David Yaeger, who provided us documentation of the Social-Belonging Mindset intervention used by Yaeger et al. (2016). In that study, incoming first-year students at a "high quality 4-year public university" were required to complete an online campus orientation. Those randomly assigned to the social-belonging intervention were asked to read short, descriptive stories from diverse upper-year students to help them understand that worries about belonging are normal in an academic transition and that they should expect these feelings to dissipate with time as they get more acquainted with their new environment, meet others like them, recognize support is available, and settle into a routine. They were then asked to write about how this message resonates with their own experiences so far and told that their response might be used to help future cohorts. First-generation and minority students combined were 3 percentage points more likely to earn 12 or more credits by the end of the first year compared to a control group (72 versus 69 percent) and significantly less likely to be deemed academically at risk based on a follow-up survey about subjective academic performance. We implemented a virtually identical exercise with first-year economics students at all three UofT campuses in the 2015 fall semester, adding additional video instructions and minimum word-count and time-on-page restrictions.

We also created two new mindset interventions to account for the unique characteristics of the students at our *Student Achievement Lab*. The first, which we call the International-Student Mindset intervention, was designed in collaboration with Christine Logel and Greg Walton, and focused on the fact that a large fraction of our study body consists of international students (35 percent). Not only must these students cope with transitioning into a new school environment, but they also must navigate living in a new country, away from parents and friends. We conducted focus groups and followed an iterative process outlined in Walton et al. (2017) to modify stories from our 2015 social-belonging mindset intervention and incorporate distinct themes highlighted by UofT international students. These included feeling homesick and isolated; having difficulty communicating with others in English, meeting friends, and understanding instructors during lectures; feeling shy to ask and answer questions; being intimidated by professors and staff; and feeling uneasiness from lack of guidance on how to study, and pressure from family to do well. Like the earlier mindset intervention, we created 1-2 paragraph stories by anonymous upper-year students describing their own experiences with these kinds of struggles and how they ultimately overcame them. Some stories were reported as being authored by domestic students to indicate that challenges around the university transition are normal and common. Participants were shown a summary from a "current student survey" that concluded, "most students worry at first about whether they belong at university but, after some time, they overcome these concerns and come to feel at home in their new environment." Students were then asked to read each of the seven created stories and describe their own thoughts and experiences about why students might feel initially unsure about their transition. They were told that their responses might be used anonymously to benefit students in future years.

The additional mindset intervention, which we call ‘Economics Mindset’, focused on the fact that all SAL participants were enrolled in an economics course. The intervention was designed together with David Yaeger, combining three messages: 1) sometimes it is difficult to see how ideas and tools from a first-year economics class can be applied in the real world, but in fact these courses are designed to train students to think more logically about a wide variety of important real-world problems and prepare them for tackling problems with more complexity later on; 2) some problems in the course are tricky and challenging, but working through new and difficult material is how we learn best, even when we make mistakes or struggle along the way; 3) professors give challenging questions not because they want some students to do poorly but because they want all students to push themselves to improve. Like other mindset interventions, students were told to consider these messages and related stories so that they might provide their own reactions and experiences to help future students. In writing to benefit others, students reinforce the same message to themselves (Aronson, 1999). At the end of displaying stories for each of the three messages, students were asked to provide their own thoughts about (respectively) 1) why students should care about learning to solve abstract multiple choice questions; 2) the advice they would provide a student doing poorly on their first economics test or no longer aiming to get a good grade; and 3) the purpose behind instructors assigning difficult questions.

We implemented both the International-Student Belonging Mindset and the Economics Mindset interventions in the 2017 fall semester. We assigned first-year international students who were randomly designated to be treated to the International-Student Belonging Mindset. We gave the Economics Mindset Intervention to students who were assigned to treatment but were not first-year international students. Because results were similar across these treatments, we combine them along with the Social-Belonging Mindset intervention from the 2015 fall semester into one category (called 'Mindset') when presenting our main results. Appendix A presents treatment effect estimates for these interventions separately (see Table A6).

4. Online Coaching Only

Based on results and feedback about the social-psychology interventions mentioned above, we also began to test interventions offering more direct coaching advice about how to perform well in university and have a successful experience. Several of the beneficial comprehensive college support programs mentioned in the introduction offer coaching and mandatory workshops about studying and performing well. Programs tested by Evans et al. (2018) and Bettinger and Baker (2014), in particular, have coaching as the main or key component. The personalized, ongoing, and proactive nature of these services, which we examine more below, may be important, but we intended to test as a baseline whether an inexpensive one-time online exercise providing similar advice to what a coach would offer could generate even a small impact on academic achievement.

We tested two online-only coaching programs at SAL. In the 2015 fall semester, those treated to the coaching exercise were asked to think about the future they envision and the steps they could take in the upcoming year at UofT to help make that future a reality. They were told that the exercise was designed for their benefit and to take their time while completing it. The online module lasted approximately 60 to 90 minutes and led students through a series of writing tasks about their ideal futures, both at work and at home, what they would like to accomplish in the current school year, how they intend to follow certain study strategies to meet their goals, and whether they wanted to get involved with extracurricular activities at the university. Varying minimum word-count and time restrictions were placed on several pages of the online exercise to ensure that students gave due consideration to each of their answers before proceeding. The exercise aimed to make students' distant goals salient in the present and to provide information on effective study strategies and how to deal with the inevitable setbacks that arise during an academic year. After the exercise, students were emailed a copy of the answers they provided to store for future reference.

Together with Christine Logel, we designed a second online-only coaching treatment the following school year (2016-17) that resembles a hybrid of the earlier interventions, incorporating elements of our goal-setting, mindset, and coaching treatments, while allowing each student to focus on the challenges they think are particularly important.[8] Part One of the exercise presents students with six broad factors critical to academic success,[9] with subsequent sections elaborating on each factor and taking students through tasks that draw on psychology research on attitude and behavior change. Part Two presents students with eight institutional barriers to success, most related to academic success factors, but also related to the implications of being part of a negatively stereotyped group, (i.e., "feeling that maybe 'people like them' are not especially welcome at U of T"), or of experiencing significant life challenges, (i.e., "dealing with a great deal of personal stress"). Students are invited to choose the two barriers most relevant to future students like them, identify and write about a reason why students might struggle with this problem, and identify and write about a potential solution. Figures A1 and A2 in Appendix A include sample screenshots of

[8] Oreopoulos, Petronijevic, Logel, and Beattie (2018) include more details.
[9] These include studying enough, studying effectively, seeking help, attending class, staying motivated, being patient and taking a long-term perspective.

the initial instructions and video, and one of Part One's modules about the importance of staying motivated while studying.

Both exercises mentioned in this section offer detailed and specific online coaching advice for performing well in university and having a good experience. We therefore group them both into the same category, which we call 'Online Coaching Only'. Table A6 in Appendix A includes baseline results with separate treatment effects (which are similar).

5. Online Plus One-Way Text Coaching

To help students stay motivated and remember study advice, a random subset of students finishing the online coaching exercise were also offered the opportunity to receive follow-up communication during the school year by text message or email. Students were told that they were selected by lottery to participate in a pilot project designed to help with their goals and provide extra support outside the classroom. About 85 percent of those invited provided a cell phone number. The remainder received emails with similar content. The initiative was called *You@UofT*—a name we chose to associate the program directly with the university and its effort to support students' individual goals.

During the 2015-16 school year, messages sent to students were mostly one-way, designed deliberately not to solicit a response—a design feature that allowed us to avoid having to hire, train, employ, and manage real coaches, making the marginal cost of the program almost zero (sending one text message costs about US$0.0075). Text messages were typically three to four lines in length while emails were longer and provided more detail (in 2015, students received both text and email messages). Our messages typically focused on three themes: academic and study preparation advice, information on the resources available at the university, and motivation and encouragement. Figure A3 in Appendix A shows a screen shot of the coaching manager we used to view outgoing (and incoming) messages for the one-way text message coaching treatment. Messages were signed from the 'You@UofT Support Team' rather than any individual person. Students were free to opt out of receiving email messages, text messages, or both at any time after the exercise, although few chose to do so.

In the 2016-17 school year at the UTM campus, we partnered with an existing for-profit company in the business of sending one-way text messages to college students with similar goals of improving academic achievement and persistence. Similar to our reasons for working with social psychologists to maintain high fidelity in the implementation of goal-setting and mindset experiments, our aim was to explore whether experienced commercial organizations might be better at designing text message coaching. Randomly selected students were still required to complete the online coaching exercise, but then were offered the outsourced one-way text message coaching (they did not know that the messages sent to them were from the outside organization). Students who did not provide cell phone numbers received our regular email messages with similar communication instead. The text message program remained labelled You@UofT and references were made to UTM student services. Some messages invited students to text back yes/no or numbered responses to receive automated replies.[10]

[10] For example, "Your profs recommend studying 3 hrs for every 1 hr spent in class. How many hours per week do you typically spend studying or working on school work?...(after any numbered response) Success requires a lot of studying but also studying well. If your effort isn't paying off get help from the Academic Skills Centre [LINK]".

Again, for the purpose of reporting results from a generalizable set of interventions from SAL, we estimate our main treatment effects below by grouping together the one-way-text coaching program administered by us and the one administered by the for-profit company. Similar but less precise results are presented separately in Table A6 in Appendix A.

6. Online Plus Two-Way Text Coaching

We investigated the impact of more intensive and personalized coaching by introducing two-way text message coaching, in which students were assigned to experienced upper-year undergraduate coaches whom they could message back with questions or simply check in about how their week was progressing and whether any challenges had arisen. Coaches were recruited based on their academic transcript and existing experience with mentoring, tutoring, and coaching students through other student services. They also received training from the university's Academic Skills Centre and from a one-day workshop discussing the You@UofT program and how to best communicate with students via text.

In the 2016-17 school year, a random subset of first-year economics students who completed the online coaching exercise were also offered an individual coach who would send them messages throughout the year and with whom they could communicate back. Approximately 90 percent of students chose to opt in by providing their phone numbers, and less than 3 percent later chose to opt out. Those who did not provide a number received weekly email messages of study advice and motivation instead. Coaches were instructed to initiate communication with each of their students at least once a week (often twice a week), which they typically did using suggested pre-programed batch messages designed to stimulate conversation. Batch messages could be sent at specific times of the day, personalized to mention the student's name, and sent to subgroups based on international or domestic status, first-year or non-first year status, and incoming high school grades. Coaches were also encouraged to follow up with individual students on specific issues they had recently discussed to make sure that students were effectively progressing. Once contact was established, conversations evolved organically, with coaches usually trying to determine how students were progressing throughout university, both academically and emotionally. Figure A4 in Appendix A provides a sample conversation using two-way coaching and our platform during the 2016-17 year. Coaches were trained to always respond as soon as possible to students who texted them and often used emojis, humor, and enthusiasm to engage.

We designed an online and coaching intervention the following year to emphasize the importance of sufficient study time. Many college administrators and faculty recommend two or three hours of study each week for each hour a student spends in class, implying at least 25 to 35 hours of effort outside of class for someone enrolled full-time. In contrast, many of our participants at SAL reported studying fewer than 15 hours per week for all their courses, with more than a quarter of our sample studying fewer than 10 hours we week. Poorly performing students who study such few hours are unlikely to benefit from any intervention that does not increase this variable.

SAL participants assigned to receive coaching at UTSG and UTM in 2017-18 and at all campuses in 2018-19 were introduced to an online warm-up exercise in two parts. In the first, students were told about UofT's recommendation for weekly study time (at least 4 to 6 hours per course, or at least 20 to 30 hours per week for a full course load) and shown several student stories about the importance of sufficient study time for academic performance and general life satisfaction. After reading through the stories, students wrote about how they could motivate themselves to stick to a regular study routine and identified the study strategies they thought would be the most helpful for doing so. In the second part of the online exercise, students were guided to make their own regular study schedule by building a weekly calendar. We made it easy to populate class times, then students were asked to indicate times in which they would likely be occupied working, sleeping, commuting, and socializing. Once they had accounted for items with little scheduling flexibility, students were asked to populate their calendars with sufficient weekly study times. Figure A5, in Appendix A, displays a screen shot of this part. Most students were able to upload their weekly schedules to their electronic phone or computer calendars.[11] All students receiving this planning exercise were also invited to receive follow-up communication with a virtual coach, who would send them a study tip and check in with them each week about their weekly study progress.[12] As with the earlier coaching interventions, the minority of students who did not provide a cell phone number message received similar email messages instead.

The planning interventions were designed to improve study time management through three key channels. First, by providing information about successful students' study habits, they make students aware of how much time is usually required to perform well in their courses. Second, by requiring that students create a weekly plan that details all their commitments, the intervention helps students better understand the time commitment required for all their other obligations outside of school. Third, the periodic reminders that students receive about their

[11] For the 2018-19 school year, students treated with this planning intervention could also indicate deadlines for particular tests, exams, and writing assignments. Based on these deadlines, we uploaded to their calendars suggested study strategies prior to these deadlines to prepare.

[12] Students who provided their cell phone numbers were assigned to a specific coach, and each coach was assigned a few time slots during the week to be the coach who was 'on call'. During each on-call time for a given coach, we sent a batch message to all students who were assigned to that coach to spur productive conversation. If students replied while their coach was still on call, that coach would continue the conversation. If students replied after their coach's shift ended, the coach who was currently on call or the team manager was responsible for closing the conversation.

planned study times help keep their goals salient throughout the academic year, and the personalized nature of the messages makes them more likely to engage and respond.

The virtual coaching programs were well received. Figure 3 charts text-back response rates from students who provided cell phone numbers. Combining samples over the three years that two-way text coaching was offered, we see that more than 65 percent replied at least once to their coach during the first semester. Weekly response rates were relatively high, especially during the first month, when a third to a half of eligible students replied every week. Students not responding still may have benefited from the advice and reminders we sent. As a quality check, we contacted some students who were not responding to any text messages. They mentioned that they felt too busy to reply but wanted to keep receiving them because they found them helpful. Figure 4 reinforces this conclusion, indicating feedback about the text-message coaching program from our follow-up surveys. A majority of students (whether they responded to the text messages or not) enjoyed the program and felt that they were doing better in university at least in part because of their coach. Seventy percent of respondents preferred that the coaching program continue into the following semester (should resources be available), and 87 percent said that it should be offered to the cohort of students next year. We also received several personal text and email messages at the end of the program expressing gratitude and appreciation from having participated.

7. Online and Face-to-Face Coaching

To compare the lower-cost, lower-touch interventions above with more intensive coaching efforts for helping students, we randomly offered a small sample of students during the 2015-16 and 2016-17 school years a coach to meet in person rather than communicating only through text.

The interventions were provided only at the UTM campus. After completing the online coaching exercise, 24 randomly selected students were offered one of four personal coaches in 2015-16 and 66 students were offered one of nine personal coaches the following year. Coaches were instructed to be proactive in arranging weekly 30-60 minutes meetings with their assigned students, and to reach out to reschedule when meetings did not occur. Coaches were also available in between meetings via Skype, email, or text. Students were sent messages of advice and motivation from their coaches, much like the other coaching programs described above.

Our coaches describe providing support to their students on a wide variety of issues, including questions about campus locations, booking appointments with counsellors, selecting majors, getting jobs on campus, specific questions about coursework, and feelings of nervousness, sadness, or anxiety. Approximately half of the interactions occurred face-to-face and half occurred via Skype or text messaging. Coaches were able to remember the issues each student was dealing with, proactively reach out to do regular status checks, and provide specific advice for dealing with each unique problem. The extra time afforded to coaches with low student-to-coach ratios allowed them to befriend their students, communicate informally and with humor, and slowly prompt students about their issues through a series of gentle, open-ended questions until students felt comfortable to open up about the details of their particular problems. Once trust was established between coaches and students, students felt more comfortable discussing challenging problems, making it easier for coaches to provide clear advice.

C. Data and Methodology

Our baseline sample includes all full-time UofT students between the 2014-15 and 2018-19 school years taking a fall semester introductory economics course and who at least started a SAL online warm-up exercise before October 1.[13] Students received a grade worth 1 to 2 percent for completion of the exercise, and about 95 percent of all initially registered students completed the exercise within the first few weeks of September. Random assignment was based on the last two digits of participants' student numbers, usually alternating assignments after each consecutive number. The Registrar's office assures us that these constructed variables are independent of other known information. Table 1 shows the number of students assigned to each intervention within each campus-year cluster. The table also compares the actual percentage of students assigned to each intervention relative to the percentage we should have expected had the assignment process been truly random. For example, in 2014-15 at UTM, students with identification numbers ending in 0, 1, 3, 5, 7, and 9 were assigned to the goal-setting exercise, while those with numbers ending in 2, 4, 6, and 8 were assigned to the personality test control group. The first column of Table 1 indicates that this process led to 59.7 percent of the 1,591 student participants to be assigned to the goal setting exercise while 40.3 percent were assigned to the control group, close to our expectations. In the following year, 19.7 percent of first year students participating in SAL were assigned to the mindset intervention, 19.6 percent were assigned to online coaching only, 30.7 percent were assigned to online and one-way text coaching, 0.4 percent to face-to-face coaching, and the remaining 29.6 percent to the control group. Upper-year students that year were not assigned to the mindset intervention, since that program specifically targeted transitioning freshmen.[14] The total sample size over all five years was 24,772 students.

[13] Full-time students are those registered to take at least 3.5 course credits over the school year. Typically, 5.0 credits each year are needed to complete a program in four years.
[14] Note that, as a precaution, we also changed the assignment rules each year so that students in the control group were not always associated with having an odd number for their last digit of their student number.

We estimate treatment effects by regressing outcome variables on treatment dummy variables plus fixed effects for each of the campus-year clusters listed in Table 1. The fixed effects are necessary because the interventions we designed changed over time, as did the sample populations. In addition, it was sometimes the case that some demographic groups of students (e.g., international or first-year students) or students at particular campuses were disproportionately assigned to certain interventions relative to other groups of students or students attending other campuses. We therefore include the cluster fixed effects to account for the mechanical correlation between treatment status and cohort, campus location, or background variables introduced by our assignment rules.[15] The treatment effects may be interpreted as average outcome differences between those from treatment groups and those from the control group within a given clustered sample. We do not condition these regressions on any additional background variables for ease of interpretation and because of missing high school admissions grades for some students (Table A1 in Appendix A shows that baseline results do not change when we do).

In addition to data collected through the warm-up exercise itself and follow-up surveys, we linked students to university administrative admissions records as well as follow-up academic outcomes such as course performance and GPA. Column 2 in Table 2 displays descriptive mean characteristics for the control group from our full sample. Column 3 indicates the corresponding standard deviation. A few observations from the table are particularly noteworthy. Most students self-report aspiring to pursue graduate studies after completing their undergraduate degree (65 percent). This widespread ambition suggests that good grade performance should matter to many.

[15] When estimating separate effects for each of the mindset treatments (see Table A6 in Appendix A), we expand the number of cluster fixed effects to account for only first-year international students being assigned to the International-Student Mindset intervention and only non-first-year international students being assigned to the Economics Mindset intervention. For the sake of readability, we do not explicitly present these clusters as columns in Table 1.

Indeed, the expected fall grade average is 80.6 percent.[16] Thirty-one percent of the students are international student, implying they pay larger tuition fees and have not lived in Canada until very recently. An even larger fraction does not speak English at home. Most students are admitted with very high grades—the average admissions grade (typically the top 6 high school courses) is 85.2 percent. Almost a third of students are first-generation (with both parents having less than a university education). Students expect to study in the fall semester an average of 17.9 hours a week, with a large standard deviation of 12.1 hours.

Columns 4, 6, 8, 10, 12, and 14 show estimated differences in mean characteristics between treatment groups and the control group (along with respective standard errors listed in columns 5, 7, 9, 11, 13, and 15). Estimates include fixed effects for the sample clusters mentioned above and listed in Table 1. Out of the 108 estimates, two are significant at the 1 percent level, six at the 5 percent level, and 10 at the 10 percent level, close to what would be expected by chance. Even these statistically significant differences are generally small, due to the large sample sizes. Together with Table 1 that shows each of the intervention groups appear to be in proportional size to what would be expected from random assignment, we take these results to suggest students were credibly randomly assigned in each experiment.

UofT administrative data allow us to track academic performance within the university until the start of the 2019 term. Table 3 shows means of the outcome variables for the control group that we can measure depending on the school year the experiment began. Credits earned and course grades for the fall semester during which the experiment began are observed for each of the five cohorts. These outcomes are generally worse for the 2014-15 cohort because that year

[16] Percentage grades that are 80 percent or above correspond to an A- or higher at UofT.

the SAL experiment was conducted only at the UTM satellite campus, where students tend to perform worse than those at the UTSG downtown campus. For the following four years that include students from all three UofT campuses, mean outcomes are similar. Many students do not end up completing the credits needed each year to graduate from their programs on time. For our earliest Goal Setting experiment conducted at the beginning of the 2014-15 school year, we observe that only 38 percent of the first-year students in the control group are recorded as graduating by the end of their fourth year. About 10 percent of first year students fail to persist into second year. By fourth year, about a quarter are no longer registered with the university.

III. Results

A. First Term Academic Performance

We first estimate effects from the six interventions described above on academic outcomes during the fall semester that each experiment began. Outcomes are regressed on intervention dummy variables plus fixed effects for the cluster groups listed in Table 1.[17] The second column of Table 4 shows estimated effects on missing grade data at the end of the first fall semester. Not having any grade data may indicate that a student dropped out of the program entirely. Another possibility, however, is that a student enrolled in only full-year courses and grades are not yet

[17] The same table in which the regressions also condition on background variables (showing similar results) is shown as Table A1 in Appendix A.

available.[18] The findings suggest that the interventions generally had no impact on the likelihood of missing grade data compared to the fraction of students missing data in the control group (13.1 percent). The exception is for students receiving Online Plus One-Way Text Coaching, with the estimated impact being positive, implying a counterintuitive increase in the likelihood of missing recorded grades. We discount the importance of this result given the fact that we do not find effects for the other, more intensive interventions, and that we do not find an impact for the same treatment on credits earned over the entire school year (results shown in Table 5).

Column 3 in Table 4 shows estimated treatment effects on non-missing average fall semester grades, measured in percentage points. The control group's mean average grade in the fall semester is 68.8 percent with a wide standard deviation of 13.5 percentage points. None of the six estimated effects are statistically significant at the 5 percent level. Not surprisingly, given our large sample size, most are precisely estimated close to zero. The largest effect is only 4.8 percent of a standard deviation. Columns 4 through 8 in Table 4 display results from testing for distributional treatment effects—that is, whether there exist impacts on receiving a fall semester grade average greater than 50, 60, 70, 80, or 90 percent respectively. For five of the six interventions, we also do not find more specific distributional effects on receiving a grade average above any threshold. Some of the thresholds tested for the Mindset intervention are marginally statistically significant. However, we fail to find corroborating evidence when looking at longer term outcomes (in Table 5), across the three campuses (Table A2 in Appendix A), and when looking at impacts on only math grades (arguably more objective evaluations of performance),

[18] Most courses at UofT are one-semester courses. Even courses that tend to last a school year, like economics, have been split across two semesters (e.g., micro and macro) to make course selection easier. An exception is at the UTM campus, where several large first-year courses remain defined as full-year.

only economics grades, and estimated impacts with controls (results shown in Appendix Tables A3, A4, and A1, respectively).

B. Persistent Academic Outcomes

Table 5 shows treatment effect estimates beyond the first term. The further out we look, the more we rely on earlier experiments. For outcomes after the first term, for example, we do not observe what happens to students who participated in the most recent Online Plus Two-Way Text Coaching program. Column 2 shows again the statistically insignificant intervention impacts on fall semester grades that were mentioned in reference to Table 4. Impacts on winter semester grades are also null. Column 4 indicates that, on average, students from the control group earn 3.1 credits by the end of the first school year of the experiment, and that this average is no different for the other intervention groups. The estimates are very close to zero with small 95 percent confidence regions. We can rule out effects larger than 8 percent of a standard deviation.

Eighty percent and 73 percent of students in our control group enroll in courses at UofT the second and third year, respectively, after taking the warm-up exercise. We find no significant differences between these persistence rates and those for students in any of our intervention groups. Only in the third year since taking a warm-up exercise do we find significant impacts on credits earned. These results are driven by the online and follow-up coaching programs given in the second year of SAL. Students receiving the online coaching intervention (with or without one-way coaching) earn about 7 percent of a standard deviation more credits than those in the control group. The impact on the 24 students selected to receive proactive face-to-face coaching is particularly large—73 percent of a standard deviation, though this result is only one out of sixty estimates in the table that is significant at the 1 percent level.

C. Academic Outcomes for Students at Greater Risk of Poor Performance

Certain types of students may be more affected by the Goal Setting, Mindset, and Coaching interventions than others. In previous work examining some of these interventions separately, we estimated treatment effects for dozens of different sub-groups after conducting a pre-analysis to focus on students thought to be more at risk of poor performance than others (Oreopoulos et al., 2018, 2019; Dobronyi et al, 2019; Oreopoulos and Petronijevic, 2018). We found no convincing evidence that the interventions improved first-year academic outcomes for any of the sub-groups examined.[19] In this paper, we summarize heterogeneous effects by focusing on students more or less at risk of performing poorly academically in their first term. Specifically, we first use the control group sample to estimate a students' propensity (probability) for receiving a grade less than 60 percent (or no grade at all), conditioning on a cubic function for high school admissions grade, mother's education, father's education, age, days since warm-up exercise introduced before registering, indicator variables for English as a second language and gender, and fixed effects for clustered sample group used for randomization. Each treated and control student for which we had such background information was then assigned a propensity score and ranked in order from

[19] Some subgroup examples include students who are male, first year, first generation, international students, live with their parents, working at least 8 hours per week, not sure about their program of study, self-report they tend to procrastinate, intend to complete their education with no more than an undergraduate degree, and expect to earn less than an A- grade average.

highest at risk for predicted poor performance to lowest.[20] We also tested for treatment effects among first-generation non-international students, international students, and students with English as a second language.

Table 6 shows estimated intervention effects on fall semester grades by the end of the same term that the warm-up exercise was introduced, for students with a non-missing high school admissions grade. This sample tends to omit some students who completed high school outside the province of Ontario, including outside of Canada. All treatment effect estimates are insignificant except for the combined Mindset interventions, where we estimate almost a full percentage point increase in average grades, about 7 percent of a standard deviation. The point estimates generally rise as we focus on students more at risk of receiving a low grade. However, we cannot reject that the estimated effects are the same for students above or below the propensity score median. We also find no corresponding statistically significant effect from the mindset interventions for those above median risk for winter semester grades, or for credits earned in first year, enrollment in second year, and grades in second year. We also show no statistically significant impacts on academic performance for two additional at-risk groups in Table A5 in Appendix A: first generation non-international students and students whose first language is not English. Overall, we find no evidence that our interventions improved undergraduate academic performance, even among students more at risk of poor academic performance.

D. Mental Health and Student Experience

[20] We estimate the propensity score using a leave-one-out procedure for students in the control group to avoid introducing bias in the subsequent analysis of treatment effects.

During the last three years of our SAL experiments, when we tested mostly different types of online and follow-up coaching interventions, we conducted follow-up surveys at the end of the fall semester. As with the initial warm-up exercise, students received a small grade for completion to encourage participation.[21] We use the follow-up surveys to investigate whether our interventions affected non-academic outcomes and intermediate outcomes that we cannot observe with administrative data. We asked a standard question about subjective well-being: "All things considered, how satisfied are you with your life as a whole these days?" Students responded on a 1 (Not at all satisfied) to 7 (Absolutely satisfied) scale. We also asked how satisfied they were with their university experience and whether they have felt stressed, sad, or depressed since the beginning of the academic year (0 (rarely or none of the time), 1 (some or a little of the time), 2 (occasionally or a moderate amount of the time), or 3 (most or all of the time)).

Table 7 shows estimated treatment effects on standardized measures of these variables (all converted to have mean zero and standard deviation one). We continue to include fixed effects for each group listed in Table 1. The table reveals large impacts for students assigned to receive face-to-face personal coaching. Self-reported university and life satisfaction are 20 and 23 percent of a standard deviation higher compared to the control group, respectively. The impacts on feeling stressed or depressed are also large but imprecisely estimated. If we create an overall mental health measure by averaging across these standardized variables, we estimate a 25.8 percent increase. There is also suggestive evidence that the interventions with text-message coaching improved overall mental health, though the impacts are smaller, with point estimates ranging from 3.7 to 8.5

[21] There are no significant treatment effects on starting or completing the follow-up surveys (the first row of Table 7 shows estimated impacts), though participation rates were lower in general (about 76 percent). This was largely due to students having already dropped the economics course and no longer being invited or required to take the survey.

percent of a standard deviation and significant only at the 10 percent level. Combining all online coaching treatments to increase statistical power, the treatments are estimated to raise overall mental health by 4.4. percent of a standard deviation, significant at the 5 percent level.

During one of the follow-up surveys, in 2016-17, we asked students about their university experience so far. Specifically, using a 1 to 6 scale, we asked whether students agree they feel like they belong at their university, whether being a university student is an important part of how they see themselves, whether they think their university wants them to be successful, and how confident they feel that they have the ability to succeed at their university. The bottom panel of Table 7 shows estimated coaching effects from our interventions on standardized versions of these measures. Again, for students offered face-to-face coaching, students indicate feeling much more supported and confident. The program seems to generate a clear sense that the university is trying to support their education. Students' sense of belonging and university support is 27.8 percent of a standard deviation higher with face-to-face coaching. They feel significantly more confident they will succeed as well. Those assigned to Two-Way Text Message Coaching also feel more supported, but less so than those with Face-to-Face Coaching. Overall feelings of university support are about 6 percent of a standard deviation higher than in the control group.

E. Study Behavior and Attitudes

We asked students at the end of the term about how much they studied during a typical week outside of midterms and finals. Table 8 shows estimated effects from our coaching interventions on this standardized outcome. Students assigned to Online and Two-Way Text Coaching studied

11.3 percent of standard deviation more, on average, than students in the control group, or about 1.3 hours. We find no significant effects for those assigned to receive face-to-face follow-up with a personal coach, though these estimates have wide confidence intervals (we cannot reject zero at the 95 percent significance level, but we cannot reject an effect size of 27.8 percent either). Those assigned to Online and Two-Way Text Coaching are also significantly less likely to report they cram for exams, less likely to miss class, and more likely to feel they manage their time well. We find some marginal significant effects on these outcomes for those assigned to only the online treatment, without follow-up coaching. Finally, we find some less precise but notably larger estimated effects on positive study strategies from face-to-face coaching, including rewriting course material in one's own words, seeking feedback, and managing time well. If we average over these standardized measures to create a summary measure of overall positive study behavior, we find a marginally significant effect from online coaching only (6 percent of a standard deviation), a larger impact from Online and Two-Way Text Coaching (13 percent) and an even larger effect from face-to-face coaching (19 percent). Overall, similar to the pattern of results found in Table 7 for the estimated treatment effects on mental health and student experience outcomes, we find small significant effects on study behaviors from the virtual coaching treatments, and large effects from Face-to-Face coaching.

One concern with these results is that treated students may feel more obliged to self-report more study hours than the control group, even though actual hours are the same. To address this, we asked multiple questions about study time during the last follow-up survey in the 2018-19 academic year (in which we only tested one intervention). As indicated in the bottom half of Table 8, we find significant effects from assigning students to receive Two-Way Text Coaching for all of our study time measures: self-reported weekly study time for all of a students' courses, for only their economics course, and the amount they plan to study the following (winter) semester. We also asked each student to create a brief time diary documenting what they did 'yesterday' (i.e., the day before the took the follow-up survey). Added up, students studied, on average, 3.3 hours per day with a standard deviation of 2.7 hours. We estimate students assigned to receive Online and Two-Way Text Coaching report studying an average of 0.3 hours more in the previous day, which averages to 2.1 hours over a week, similar to the estimated effects using the subjective weekly study time variables. The magnitudes of these impacts on study time are also similar, ranging from 10 to 20 percent of a standard deviation.

F. Reconciling Positive Study Effects and Null Grade Effects

Why do we find significant treatment effects from our virtual coaching interventions on study time but no evidence of effects on academic grade performance, credits earned, or persistence? One possibility is that our study habit measures are self-reported and treated students are more likely to over-report their study hour behavior, being primed to do so by the program's emphasis on the importance of studying in the warm-up exercise and the text messages sent to them from their coaches. We do not believe this is the case for four reasons. First, we estimate consistent and robust effects on study time when we ask students to report their study time across all courses, their study time for only their economics class, and the time they spent studying 'yesterday'. Second, our outcome data are collected several months after students complete the warm-up exercise, and we find similar effects for a smaller sample of students in which we collected data at the start of second semester. Third, we find effects on other study time outcomes, such as being less likely to cram for exams, miss class, and manage time poorly, while we find no effects on less-related study behavior such as tendencies to rewrite notes, visit a tutor, and attend office hours. Fourth, we also estimate similar impacts across UofT's three distinct campuses and find similar, largely significant effects over multiple years of the study, even though the coaching interventions were not exactly the same and did not always emphasize time management in each year.

If the positive effects on study time are real, why do we not find corresponding impacts on academic outcomes? Past research and common sense suggest that such a relationship should certainly exist, so we are left with somewhat of a puzzle. We argue that the most likely explanation is that the relationship between study time and academic achievement is too weak for us to observe an effect on grades, given the small change in study time that results from our experiments. Table 9 presents several estimates of the association between self-reported study time and mean fall grades for the 2018-19 cohort. In columns 1 to 4, we show the association of weekly study time with mean fall grades across all courses. The OLS relationships are small, implying that 13 additional hours of study each week (a one standard deviation change) is associated with a 1.04 percentage point higher mean grade (8 percent of a standard deviation).

While these associations are clearly not causal estimates of the return to study time, there are at least two other reasons for their relatively small magnitudes. First, study time is self-reported retrospectively, implying that measurement error may lead to downward-biased estimates. Second, grade inflation in some courses may weaken the link between study time and achievement. We address the measurement error concern by instrumenting for weekly study time with either variables from the baseline survey, or with 'yesterday's' study time from the follow-up survey, and we address grade inflation by focusing specifically—in columns 5 to 8—on math courses, where grading is likely to be more objective.[22] Our two instrumenting strategies appear to similarly mitigate measurement error, as the estimated coefficients increase (relative to the OLS estimates) by similar amounts in both cases. Further, as expected, the association between study time and grades is strongest in math courses, where grade inflation is least likely to be a problem. A one-standard deviation (13 hour) increase in weekly study time for math courses is associated with an 6.4 percentage point increase in mean math grades. If we assume this estimated association as the return to study time, our treatment-driven increase of 2.3 hours of studying per week is predicted to cause an increase in mean grades of 1.13 percentage points or 8.7 percent of a standard deviation. This is a small effect on achievement and one which we often cannot reject as being the treatment effect of our coaching interventions in Table A3 in Appendix A, where we estimate treatment effects on achievement in math courses specifically.[23]

We believe our treatment effects on study time are real and that study time generally affects achievement positively. Taken together, our results therefore suggest that our coaching interventions improved study behavior but not enough in magnitude to observe a significant and meaningful improvement in academic performance.

[22] When we instrument for weekly study time using data from the baseline survey, we use the following variables as instruments: study hours per week in high school, self-reported expected study hours per week this semester, tendency to regularly "cram" for exams, expected hours per week working for pay during the semester, and expected commuting time to campus.

[23] Although our estimate of the return to studying is adjusted for measurement error, it does not reflect a causal link between study time and grades. We attempt to account for this by performing a back-of-the-envelope calculation that uses the estimate of the causal return to study time from Stinebrickner and Stinebrickner (2008), who use time-diary data from Berea College together with an instrumental variables strategy to find that a one-standard deviation increase in studying per day increases student GPA by 90 percent of standard deviation. The standard deviation of daily study time in their data is 1.62 hours per day (or 11.34 hours per week) and the standard deviation of GPA is 0.686 points. Importing their estimate of the casual return to studying into our setting, we would expect to find that increasing studying time by 2.3 hours per week (our treatment effect) leads to an increase in GPA of 0.13 points—an effect that we should be able to marginally detect. It is possible that productivity of study time in our student population is lower or that the students who constitute the population driving the local average treatment effect in Stinebrickner and Stinebrickner (2008) are not well represented in our data.

IV. A Model of Student Effort

Why were the interventions we evaluated ineffective at improving overall academic achievement? Perhaps students already optimize when choosing how much they want to study and how to study efficiently relative to their abilities and preferences. Or perhaps these low-cost interventions are not intensive or personal enough to meaningfully change habits or goals. To explore these issues further, we describe a simple model of study effort to better understand why some students perform poorly and the mechanisms by which our interventions affected study behavior but not academic performance. We then map the model to our survey data gathered during the fall semester of our fifth experimental year (2018-19), track how students' beliefs about their academic abilities, study choices, and grade expectations changed over the semester, and measure the impact our interventions had on these objects.

Four main takeaways arise from the analysis: First, actual study behavior deviates from target study behavior. On average, students study five to eight fewer hours than they intended, suggesting procrastination or other behavioral barriers are at play. Our interventions, however, did not reduce the gap between actual and target study hours, which exists both for those with initially low and high study targets. Second, our interventions increased academic ambition, which we measure as a willingness to spend time studying to obtain higher grades and a self-reported enjoyment of studying. Third, our interventions also increased many students' beliefs about the need to study more for a good grade, but they did not respond by trying more, instead adjusting their grade expectations downwards (and subsequently indeed realizing lower grades). Fourth, more frequent, intensive and personalized interventions may be needed to address procrastination or motivate students more than online and text message nudges can, but even these expensive programs have limits because college students do not appear very responsive to support services offered outside the classroom.

A. The Education Production Function and Student Expectations

In the model, students take their expected abilities and preferences at the beginning of the semester as given and set goals around how much study effort to put forward. They then learn more about their abilities and preferences, revising their initial expectations, and update their study decisions and grade expectations accordingly. The difference between the time they report studying at the end of the semester and the time they expect to study at the beginning is a function of both the rational information update and a behavioral deviation from that update, which we refer to as procrastination tendencies. The difference could also arise from mistakes in time management, over-confidence with initial expectations, or lack of salience.

Let y_i denote the grade earned by student i at the end of the fall semester. We assume that the weekly study effort of each student, s_i, is mapped into grades according to the following linear production technology:

$$y_i = \alpha_i + \beta_i s_i + \epsilon_i, \qquad (1)$$

where α_i is the academic ability of student i—i.e., the grade she would expect to earn without any study effort—β_i is the return to each unit of additional studying for student i, and ϵ_i is an error term with mean zero.[24]

[24] We assume the simple linear specification for the production technology to keep the analysis tractable and to allow for an intuitive mapping of the theory to the survey data, where we ask students about their expected abilities and returns to studying during initial and follow-up surveys.

Students are uncertain about their academic abilities and returns to study effort at the beginning of the fall semester. We let $\hat{\alpha}_{i0}$ and $\hat{\beta}_{i0}$ denote, respectively, student i's expected ability and return to study effort at the start of the semester. Similarly, we let $\hat{\alpha}_{i1}$ and $\hat{\beta}_{i1}$ represent the updated values for these objects at the end of the semester. For a given amount of study intensity at time t (s_{it}), student i therefore expects to earn the following grade

$$\mathbb{E}_t(y_i|s_{it}) = \hat{\alpha}_{it} + \hat{\beta}_{it}s_{it}, \qquad (2)$$

where $t = 0$ and 1 denote the beginning and end of the semester, respectively.[25] With their grade expectations in mind, students then make study decisions according to their preferences over grades and the cost of study effort.

B. Student Preferences

We assume that students perceive the benefits of higher grades in discrete categories, defined by the grade cutoffs that correspond to the letter grades, B, and all other letter grades that are up to and below a C. Specifically, we let θ_{it}^{j} denote the utility benefit obtained by student i when she earns letter grade j and assume that $\theta_{it}^{A} > \theta_{it}^{B} > \theta_{it}^{C}$. Student i earns an A when $y_i > y^A$, earns a B when $y_i > y^B$, and earns a C when $y_i > y^C$, where $y^A > y^B > y^C$. At both the beginning and end of the semester, each student exerts of level of student intensity s_{it} to increase her expected grade, given by equation (2). The cost of study effort is given by the strictly increasing and convex function $c(s_{it})$.

[25] More precisely, in our data, the beginning and end of the semester are the times when students take the initial and follow-up surveys.

Note that we assume the benefit students derive from higher grades only changes (discretely) when they earn a grade that crosses a threshold for a higher letter grade: continuous changes in percentage grades within a given letter grade category do not give rise to any change in the benefit students derive from their study effort. We make this assumption because the patterns in our data suggest that students do indeed place much importance on attaining grades that correspond to certain thresholds. Some of this behavior is due to explicit thresholds determining whether students are admitted to specialized or honors programs. In Appendix B, we show that student percentage grade expectations bunch at multiples of ten, which indicate transitions between letter grades at UofT, and that only 30 percent of students report preparing for a test until they completely understand the material, with the remaining 70 percent preparing only enough to earn various letter grades. [26]

Note also that, for the purpose of our model, we deliberately group all grades up to and below a C into one category. We do so because (i) allowing for more grade categories does not add to the model's main insights and (ii) the data are consistent with students not differentiating much between letter grades that are a C or below. In Appendix B, we show that less than 2 percent of students expect to earn a grade that is a C or below, both across all courses and economics specifically, while only 9 percent of students report preparing for tests by studying enough to only earn a C or less.

[26] A model in which the benefit of higher grades is continuous (and increasing and concave) in the grade earned delivers similar predictions about student behavior but with some important differences. In particular, a model with a continuous benefit implies that students revise their expected grades upward when receiving a positive information update about their abilities and that they revise study time upward when receiving a negative information update. We do not find these patterns in the data. Instead, we see an asymmetric response, with students who receive a positive update leaving grade expectations relatively unchanged but decreasing study time choices and students who receive a negative update by downgrading grade expectations substantially but leaving study time relatively unchanged. These patterns are can be more easily generated by the threshold-based model we present here.

C. Student Decision-Making and the Interpretation of Observed Study Outcomes

The time students report studying each week at the end of the semester represents a combination of rational revisions to their initial expected study times, which reflect updated information about their academic abilities and preferences, and behavioral deviations from these rational revisions, which we conceptualize as procrastination.

Information-Driven Choice

Students' beliefs about their academic abilities and returns to studying ($\hat{\alpha}$ and $\hat{\beta}$) may change from the beginning and end of the semester. Using her beliefs in each time period and equation (2), student i determines the minimum amount of study effort that is required for her to expect to earn letter grade j in time period t as

$$s_{it}^{j} = \frac{y^{j} - \hat{\alpha}_{it}}{\hat{\beta}_{it}}, \qquad (3)$$

where $j = A, B,$ or C and $t = 0$ or 1. We assume that the distribution of student ability is such that the study time required for the lowest letter grade of C is non-negative for all students, implying that $y^{C} \geq \hat{\alpha}_{it}$ $\forall$ i and t.

When choosing between whether to exert enough study effort to expect an A or only enough to expect a B, student i compares the additional benefit of earning an A to the cost of additional studying, opting to aim for an A when

$$\theta_{it}^{A} - \theta_{it}^{B} \geq c\left(s_{it}^{A}\right) - c(s_{it}^{B}), \qquad (4)$$

where $s_{i}^{A,t}$ and $s_{i}^{B,t}$ are defined according to equation (3). Likewise, when choosing between aiming for a B or a C, student i studies enough to expect a B when

$$\theta_{it}^{B} - \theta_{it}^{C} \geq c(s_{it}^{B}) - c(s_{it}^{C}). \tag{5}$$

As discussed, the descriptive evidence suggests that few students approach their studies by aiming for a C or below. For ease of exposition, we assume that no student prefers to aim for a C over a B.[27] Formally, we normalize the benefit of obtaining a letter grade of C to zero for all students ($\theta_{it}^{C} = 0 \ \forall\, i, t$) and assume that the following condition holds

$$\underline{\theta}_{t}^{B} > c\left(\frac{y^{B} - \underline{\hat{\alpha}}_{t}}{\underline{\hat{\beta}}_{t}}\right) - c\left(\frac{y^{C} - \underline{\hat{\alpha}}_{t}}{\underline{\hat{\beta}}_{t}}\right) \tag{6}$$

for $t = 0$ and 1. Here, the underlined objects represent the minimum benefit of obtaining a B grade across all students, and the minimum values of perceived academic ability and the return to studying in each time period across all students. Because $c(\cdot)$ is strictly increasing and convex, the right-hand side of equation (6) is decreasing in both α and β, and is therefore maximized at the minimum values of both objects. Likewise, the left-hand side is lowest at $\underline{\theta}_{t}^{B}$, implying that condition (6) guarantees no student prefers to study only enough to expect a letter grade of C. In the model, all students will therefore study enough to expect to earn either an A or a B, which is consistent with the descriptive evidence in the data.

With this framework in hand, the optimal study choice of student i in time period t is written as

$$s_{it}^{*} = \begin{cases} s_{it}^{A} \text{ if } \theta_{it}^{A} - \theta_{it}^{B} \geq c(s_{it}^{A}) - c(s_{it}^{B}) \\ s_{it}^{B} \text{ if } \theta_{it}^{A} - \theta_{it}^{B} < c(s_{it}^{A}) - c(s_{it}^{B}) \end{cases}. \tag{7}$$

Behavioral Barriers

[27] As mentioned, we have only three grade thresholds in the model for ease of exposition, making the letter grade C our lower bound. The same intuition can be obtained from a model with more grade thresholds, in which no student prefers to attain a failing grading (F) over a D. However, the more nuanced model would add very little useful content at the expense of expositional clarity.

Equation (7) describes how students make rational decisions about study intensity at the beginning and end of the semester, given their preferences and beliefs about their academic abilities and returns to studying. We assume that observed study time at the end of the semester is given by the rational quantity implied by equation (7) and a behavioral deviation caused by procrastination. Specifically, we write observed study time at the end of the semester, $\tilde{s}_{i1}$, as

$$\tilde{s}_{i1} = \lambda_i^p + s_{it}^* + \nu_i, \tag{8}$$

where λ_i^p is a student-specific procrastination term and ν_i mean-zero noise. Seen this way, observed study outcomes are a function of rational behavior—based on preferences and expectations about academic ability—and behavioral challenges, such as procrastination tendencies and distractions. In Section V below, we demonstrate how multiple measures of study time from our survey data allow us to identify average procrastination behavior ($\bar{\lambda}^p$), while holding constant changes in study time that are driven by rational information updating.

D. Analyzing Changes in Study Time Over the Semester

We now describe how rational study choices and grade expectations change as students update beliefs about their abilities and returns to studying and also update their preferences. We then decompose the difference between actual and initially expected study time into a component driven by rational updating and a component driven by behavioral barriers like procrastination.

Considering students who plan to study enough to earn a letter grade of A, Proposition 1 describes how these students adjust their study effort and grade expectations when they receive new information.

Proposition 1: *Suppose student i is originally studying enough to expect to earn a letter grade of A. Hold fixed the difference between the perceived benefit of earning an A and the benefit of earning a B, $\theta_{it}^{A} - \theta_{it}^{B}$. If student i receives a positive update about her academic ability (α_i) or return to studying (β_i), she continues aiming for an A but with less study effort. If she receives a small negative update, she continues aiming for an A but with more study effort; if she receives an intermediate negative update, she lowers her expected grade to a B but decreases or does not change study effort; if she receives a sufficiently large negative update, she lowers her expected grade to a B and increases study effort.*

Proof: See Appendix B.

Intuitively, when students originally believe they are putting forth enough effort to earn top grades and effort exertion is costly, they reduce their effort upon learning that they are of higher academic ability or that each unit of effort is more productive. In contrast, when students learn it is more difficult to earn top grades than originally expected, they respond by trying harder and continuing to aim for an A as long as the additional benefit of doing so still exceeds the additional cost associated with studying enough for an A versus a B. When it no longer pays off to continue aiming for an A, students will revise their grade expectations down and reduce their study effort, as long as the downward revision to their beliefs is not too large. When students learn it is much more difficult to do well than they originally believed, they revise their expected grade down but increase study effort to ensure that even the lower expected grade is attainable. Proposition 2 establishes similar predictions for students who are originally aiming for a B.

Proposition 2: *Suppose student i is originally studying enough to expect to earn a letter grade of B. Hold fixed the difference between the perceived benefit of earning an A and the benefit of earning a B, $\theta_{it}^{A} - \theta_{it}^{B}$. If student i receives a negative update about her academic ability (α_i) or* *return to studying (β_i), she continues aiming for a B but with more study effort. If she receives a small positive update, she continues aiming for a B but with less study effort; if she receives an intermediate positive update, she increases her expected grade to an A and increases or does not change study effort; if she receives a large positive update, she raises her expected grade to an A but decreases study effort.*

Proof: See Appendix B.

Together, Propositions 1 and 2 imply that students modulate their study effort in response to updated information and in a potentially asymmetric way with respect to positive and negative information updates. That is, depending on their initial grade expectations and the size of the information update, the model outlines cases where students who realize it is *harder* to earn an A respond by not changing (or marginally decreasing) study time choices and decreasing grade expectations. Students are likely to make revisions of this nature when they are initially aiming for an A and receive an intermediate negative shock to their beliefs about their academic abilities.[28] The model also outlines cases where students who realize it is *easier* to earn an A respond by decreasing study time choices and not changing grade expectations. Revisions of this type are likely to occur when students are originally aiming for an A or aiming for a B and receive a relatively small information update. In the next section, we show the data are consistent with these predictions, as we find that students who realize it is harder to earn an A respond by revising grade expectations down but not increasing study time, while students who learn it is easier respond by significantly reducing study time and revising grade expectations far less.

Holding beliefs about academic ability constant, students' preferences may also change over the course of the semester, thus affecting their study time choices. We interpret a change in

[28] When they are initially aiming for an A and receive a large negative shock, they increase study time but still revise grade expectations down.

preferences that makes students value higher grades more as in increase in $\theta_{it}^{A} - \theta_{it}^{B}$. The next proposition establishes the intuitive idea that, for a given academic ability and return to studying, students are willing to work harder to earn higher grades when they value higher grades more.[29]

Proposition 3: *Holding* $\hat{\alpha}_i$ and $\hat{\beta}_i$ *fixed, the maximum amount of time a student is willing to study for an A is increasing in the difference between the perceived benefit of earning an A and the perceived benefit of earning a B,* $\theta_{it}^{A} - \theta_{it}^{B}$.

Proof: See Appendix B.

Rational revisions to study time and grade expectations in our model are therefore driven by changes to information about academic ability and changes in preferences for earning high grades. We also emphasize—both theoretically and empirically—that rational updates to study choices (because of changes in preferences and information) occur separately from procrastination behavior. That is, using equation (8), the difference between the actual number of hours per week a student reports studying at the end of the semester ($\tilde{s}_{i1}$), and her original expected study time, is

$$\tilde{s}_{i1} - s_{i0}^{*} = \underbrace{\lambda_i^p}_{\text{Procrastination}} + \underbrace{(s_{i1}^{*} - s_{i0}^{*})}_{\text{Rational Update}} + \nu_i. \tag{9}$$

Equation (9) makes clear that both students with high and low initial study expectations (s_{i0}^{*}) can procrastinate, as even students with low initial study goals may optimally desire to revise those goals up throughout the semester but fail to do so because they procrastinate. Indeed, in Section V below, we show that our average measure of procrastination does not differ between students with low and high initial study goals.

[29] We frame the proposition in terms of the maximum amount of time students are willing to study to earn *A* because we present evidence in Section V that our interventions cause treated students to report being willing to study more hours to earn higher grades than control students. We interpret this as suggestive evidence that our coaching interventions changed students' perceived benefits of higher grades.

In summary, we consider three main mechanisms through which our coaching interventions could affect study behavior: (i) changing the information students have about their academic abilities or returns to studying ($\hat{\alpha}_{it}$ and $\hat{\beta}_{it}$), (ii) changing the value students place on earning high grades, and (iii) helping them reduce procrastination behavior.[30] All are plausible channels through which the interventions could have caused students to increase study time. In particular, the interventions emphasized the importance of adequate study time for satisfactory performance, potentially causing students to revise their beliefs about their academic abilities ($\hat{\alpha}_i$), and provided students with effective study strategies and tips, possibly changing students' returns to study time ($\hat{\beta}_i$) and making each unit of study time more effective. The interventions also emphasized the long-term benefits of doing well in college and in first year particularly, potentially changing the value students place on earning high grades ($\theta_{it}^A - \theta_{it}^B$). In addition, in some cases, our programs attempted to mitigate students' tendencies to procrastinate (λ_i^p) by keeping their goals salient throughout the semester through consistent text or face-to-face contact.

V. Supporting Evidence for the Model and Decomposing Treatment Effects

In our model, four factors cause poor study effort: 1) high expected productivity from cramming ($\hat{\alpha}_{it}$), leading students who are better at cramming to study less; 2) low expected return to studying ($\hat{\beta}_{it}$); 3) low preferences for good grades ($\theta_{it}^A - \theta_{it}^B$), which we call low 'academic ambition'; and 4) actual study hours being less than target study hours (λ_i^p), which we label as 'procrastination'

[30] Our data do not allow us to decompose the proportion of the rational update due to information updating versus preferences, but we show that the residual between actual study time changes and how much of it we can explain with procrastination and information updating may be due to changes in preferences.

but could also reflect distraction and other behavioral barriers. Note also we use expected study time needed to earn an A as a summary measure for $\underline{\hat{\alpha}}_{it}$ and $\underline{\hat{\beta}}_{it}$, $\frac{80-\hat{\alpha}_{it}}{\hat{\beta}_{it}}$, which we label as 'academic savvy'. Our SAL interventions could influence all four factors. Getting students to think about future goals, for example, might make them more ambitious. Advice about how to study more effectively might improve expected returns to studying. Advice on time management, or reminders around goals and the need to study, may improve academic salience and help with procrastination. Cautioning students against cramming might lower its expected return, leading to an ambiguous reaction from students in choosing whether to study more for the same initially targeted grade, or studying less and settling for a lower grade. We test the extent to which each factor was affected below.

We use answers that students provided in the initial and follow-up surveys during the 2018-19 experiment to construct measures of these factors. In both surveys, students were asked to report the percentage grade they thought they would earn in their economics course if they studied 0, 1, 3, 7, 12, and 20 additional hours per week for the course, on top of any cramming two days before the midterm and final exam. Using the reported expected percentage grades as the dependent variable and the hours options as the independent variable, we estimate two (student-specific) regressions for each student—one with data from the initial survey and one with data from the follow-up—which allow us to construct estimates of each student's expected ability ($\hat{\alpha}_{it}$) and return to studying ($\hat{\beta}_{it}$) in economics at both the beginning and end of the semester. We measure expected ability at each time period as the estimated intercept from the relevant regression, while taking the estimated slope as the excepted return to studying. Students were also asked how much time they would be willing to study if guaranteed a grade of 70, 75, 80, and 85. Since students receive the grade with their stated amount of time, regardless of effort, the difference between their willingness to spend time for a guaranteed grade of 80 versus a 70 reflects preferences for higher grades, relative to their opportunity cost, whatever that might be. Finally, students were surveyed in November about their expected weekly study time in the winter term and their actual weekly study time in February. We use the difference between the two as our procrastination measure.[31]

A. Characteristics that Relate to Study Effort and Grade Outcomes

Table 10 documents the relationship between these variables and measures of study effort and achievement. On average, students believe they will obtain a grade of only 53.7 percent from cramming, and that their economics grade would increase, on average, by 1.8 percentage points for each additional hour per week of studying. This translates to 14.7 hours of study per week needed to obtain a grade of 80. Students also report a willingness to spend an average of 8.3 more hours of study to obtain a grade of 80 versus 70 over all their courses. Interestingly, in the top panel of Table 10, Columns 3 and 4 indicate that students' recorded study time does not systematically relate to their beliefs about cramming effectiveness or returns to studying. It is, instead, strongly correlated with academic ambition—students' willingness to spend time to guarantee a good grade (shown in columns 4 to 7). A 10-hour increase in a students' willingness to study for a guaranteed 'A' versus 'B' is associated with about a 5-hour increase in actual study time. Another proxy variable for preferences towards studying—how much a student agrees, on a 1 to 6 scale, whether they 'like to study'—is also highly correlated with actual study time. Conditioning on initial high school grades does not change these relationships. In short, the

[31] We use a simple linear model for tractability and ease of exposition. Regressing expected percentage grades on a quadratic function of hours and proceeding with estimates from that exercise does not meaningfully change the results.

estimates from the top panel of Table 10 suggest that motivation for good grades better predicts study behavior than students' expectations about their own study effectiveness.[32]

Our data also reveal that target study behavior does not line up with actual behavior. Figure 5, for example, displays kernel densities for target weekly study hours intended during the winter term, which students reported during late November and early December, and for actual weekly study hours during the winter term, which students reported in February. The median student intends to study 23.8 hours per week, and at least a quarter of students plan to study more than 35 hours per week. Relative to target hours, the distribution of actual hours shifts markedly to the left, with about a quarter of students studying more than 20 hours a week and about a quarter studying fewer than 6 hours. The median of actual reported study hours is 15.1 hours per week. The mean difference between target and actual study hours across all students is 8.7 hours.

Although many students falling short of their target study hours is suggestive of procrastination problems, a simple difference between target and actual hours does not isolate changes in study behavior that are driven by procrastination. The key challenge to identifying procrastination-driven changes separately from changes that are driven by information updating is that we do not observe the study time students would have rationally selected after updating without procrastinating. On this point, when comparing target hours for the winter term, measured in November, and actual reported hours, measured in February, the time lapsed may not be long enough to expect significant information updating.

As an alternative, however, we also measure weekly hours of procrastination by taking the difference between students' target hours in the winter term and actual hours reported for the fall

[32] The same conclusion arises when using target study hours stated at the start of the fall semester.

term, both of which are recorded *at the same time* in late November or early December. Because students record target hours for the next (winter) semester at the same as their actual weekly study times during the current (fall) semester, the information they have about their academic abilities is the same when answering both questions. However, procrastination does not determine how many hours they intend to study next semester, which instead reflects a choice based on preferences and present information, not behavioral barriers. We therefore treat expected weekly study time next semester as an observable proxy to the time students would have rationally chosen to study this semester based on just their preferences and updated information. Specifically, we denote expected study next semester as s_{i2}^* but assume that it is in fact equal to s_{i1}^* in our model. Then, using equation (9), the difference between expected study time next semester and actual study time this semester is written as

$$s_{i2}^* - \tilde{s}_{i1} = s_{i1}^* - \left(\lambda_i^p + s_{i1}^* + \nu_i\right) = -\lambda_i^p - \nu_i. \tag{10}$$

Equation (10) reflects the amount by which study time during the fall semester was affected by procrastination and mean-zero noise. Averaging equation (10) over all students therefore provides an average measure of procrastination, while holding constant (or removing) the effect of information updating. This alternative measure leads to a mean procrastination rate of 4.9 hours per week, compared to the 8.7 hours we estimate using both target hours and actual hours reported for the winter term.

As a further descriptive exercise, the bottom panel of Table 10 shows how expected academic savvy and ambition relate to actual grade performance. While these associations do not imply causal influences, they do reveal informative patterns. First, while beliefs about effectiveness of cramming and regular studying are not helpful for predicting actual study time, they strongly predict actual grades: a standard deviation increase in the grade that students expect from studying plus a standard deviation increase in the expected return to studying is together associated with a 6.3 percentage point increase in the actual fall term grade. Second, and in contrast, while academic ambition strongly predicts study time, it has little association with actual grades. Thus, the factors that relate to study behavior are not the same factors that relate to performance.

Surprisingly, our measure of procrastination—the deviation between actual and target study hours—does not predict grade performance. Some students who procrastinate much more can obtain similar grades (or better) than students who procrastinate less. They may be compensating through cramming or being more productive during the hours they do study. Conditioning on high school grade and reported weekly study time does not alter these relationships, though the high school grade itself is the strongest predictor of grade performance in college.[33] An interpretation of these results is that measures of academic ability—high school grade, performance from cramming, and return to study—matter more for academic performance but are perhaps less malleable than measures of ambition and procrastination. If ability determines college performance most, then the potential to help may be limited.

B. Treatment Effects on Academic Expectations, Ambition, and Procrastination

Students significantly deviate from their study intentions, but our interventions do not reduce this gap. Table 11 shows this lack of effect by reporting estimated treatment effects on multiple

[33] Note that in column 7 we find, similar to Table 9, that weekly study time correlates with grade performance, but not significantly: a one standard deviation increase in study time is associated with a 1.7 percentage point increase in fall grades, conditional on high school grade, academic savvy, ambition, and procrastination tendencies.

measures of procrastination and distraction. Our preferred measures—from the data used to produce Figure 5—indicated that students' winter target study times are 8.7 hours higher than the hours they actually report studying, on average, yet the estimated impact from our Online and Two-Way Coaching intervention in 2018-19 is insignificant (an increase in procrastination of 0.58 hours with a standard error of 1.2). We also estimate insignificant effects when measuring procrastination as the difference between winter target hours and actual fall term hours, and when considering either the population of students with initially high target study hours (above the median) or initially low target hours. We also find no impact on students' self-reported tendencies to feel distracted by social media and video screens (TV, Netflix, etc.).

Our coaching intervention did have a small impact on academic savvy, mostly by reducing the grade students believed they would get from minimal studying. In particular, the online and coaching intervention reduced the grade expected from only cramming by 1 percentage point, which our model suggests could cause students to increase their study time and continue to aim for a given grade threshold, or to reduce their study time and decide that trying to attain the higher threshold is not worth the effort.

Our coaching intervention had the largest impact on academic ambition, raising the number of hours students are willing to study for a grade of 80 versus 70 by 1, and causing a similar increase to the willingness to study for a grade of 85 versus 75. This represents about a 14 percent increase of a standard deviation. The small but significant impact is noteworthy because it is suggestive of the channel by which the coaching intervention affects study time (a little). We find supporting evidence from estimating significant treatment effects on students' agreement that earning good grades matters more than just ensuring program completion—a variable which likely reflects, in part, students' perceived benefits from higher grades.

C. Information Updating and Changes in Weekly Study Time and Grade Expectations

We close this section by exploring how students respond to updated beliefs about their study effectiveness and their abilities. The analysis reveals that students who realize their studying goals are easier to obtain than they initially thought respond by reducing effort, as predicted in our model. On the other hand, students who realize they are unable to achieve their grade targets with their initial study goals become discouraged. Rather than trying harder to reduce procrastination or study more, they begin to expect and, ultimately, accept a lower grade. This observation implies policy challenges for assisting struggling students.

To concisely use all the available information when tracking changes in students' beliefs about their academic abilities and effectiveness, we measure information updating for each student as the change in 'academic savvy'—that is, the change in the study hours that are required for students to expect to earn at least an A in economics—over the course of the semester:

$$\Delta s_i^A = \frac{80 - \hat{\alpha}_{i1}}{\hat{\beta}_{i1}} - \frac{80 - \hat{\alpha}_{i0}}{\hat{\beta}_{i0}}. \tag{11}$$

Equation (11) allows us to succinctly capture how changes in beliefs about both academic ability and the return to studying relate to eventual changes in study time and grade expectations. When Δs_i^A is positive, students receive a negative information update during the semester, learning that it is more difficult to earn an A than initially expected and that more study time is required to do so. The opposite is true when Δs_i^A is negative. This is our preferred measure of information updating because it has an intuitive connection to our model, where students consider the decision between studying enough to expect to earn an A or a B. In Appendix B, we show that our results remain very similar when considering reasonable alternative measures of information updating.[34]

We investigate how revisions in students' study times and grade expectations relate to our measure of information updating in panels (a) through (d) of Figure 6.[35] In panel (a), we plot the difference between actual and expected weekly study time for economics against the measure of information updating. There is a clear positive relationship, indicating that students revise their study time down when the change in their study gradients implies that they need to study fewer hours than initially expected to earn an A. Conversely, they study more hours than initially expected when their study gradients imply that they need to study more to earn an A. Note, however, that there is an asymmetric response, as the average change in study hours among students who learn they need to study more (those to the right of zero on the horizontal axis) is not statistically different from zero, while those who learn they need to study less (those to the left of zero) revise their study time in economics down by 1.34 hours per week.

Columns (1) and (2) in Table 12 report the estimated slope coefficient corresponding to the linear fit in panel (a) of Figure 6, with and without additional control variables, respectively. Among other demographic and background variables, our vector of controls includes flexible (cubic) functions of students' initially expected study times in economics, expected study times across all courses, and expected grades in economics. The dependent variables in the specifications in Table 12 are often changes relative to these initial expectations, making it important to flexibly control for systematic changes throughout the semester that are potentially correlated with

[34] See Table A7 and Figure A7 in Appendix A and the associated discussion in Appendix B.
[35] The panels in this figure are binned scatter plots. Each binned scatter plot is created by first grouping students into 20 equal-width bins (vingtiles) in the distribution of the variable on the x-axis and calculating the mean of both the y- and x-axis variables within each bin. The circles represent these means, while the lines represent the associated linear fit from the underlying student-level data.

information updating.[36] Reassuringly, the point estimates are similar across specifications with and without additional controls and are economically significant, implying that when students should expect to have to study 6.5 hours more per week to earn an *A*—a one standard-deviation change in the information update measure—they study 0.8 hours (16 percent of a standard deviation) more per week for economics than originally expected.

Panel (b) of Figure 6 shows a similar relationship when the dependent variable is the difference between actual and expected study time across all courses. Here, owing to data limitations, our measure of information updating is an imperfect proxy for beliefs about academic ability in all courses because it pertains specifically to academic ability in students' economic courses.[37] Nonetheless, we see very similar patterns: a one standard deviation decrease in required study time for an *A* in economics is associated with a decrease in weekly study time across all courses of 1.3 hours per week (approximately 9 percent of a standard deviation). Columns (3) and (4) in Table 12 show that the point estimates underlying these relationships remain qualitatively similar and statistically significant in specifications that include additional control variables. Consistent with the model, an asymmetric relationship again emerges as students who learn it is harder to earn an *A* do not significantly change study time, on average, while those who learn it is easier reduce study time by 2.1 hours per week.

In panel (c) of Figure 6, we show the relationship between students' expected percentage grade revisions in economics and information updating. Two points are worth making about this

[36] For example, suppose some students initially submit very high and unrealistic expectations for study time and grades. We would expect that these students mechanically revise down both study time and grade expectations, and, if such students are also more likely to be overly optimistic about their academic abilities, we would expect to find a correlation between these mechanical revisions and our measure of information updating. Flexibly controlling for the relationship between changes in expectations and initial expectations allows us to identify the effect of information updating conditional on this relationship.
[37] We do not have data on student beliefs about ability and the return to studying on a course-by-course basis.

figure. First, nearly all students revise their percentage grade expectations down, on average. Second, students who receive a positive information update revise their expected grade down only a little or not at all, while students who receive large negative updates revise their grade expectations down substantially. The average student who learns that less time is required to earn an A in economics revises their grade expectations down by 2.3 percentage points, while the average student who learns that more time is needed revises their grade expectation down by 11 percentage points.

Overall, a clear negative relationship prevails between information updating and grade expectation revisions. Columns (5) and (6) in Table 12 present the point estimates of the slope from the underlying linear fit, indicating that a one standard deviation increase in required study time for economics is associated with students expecting to earn grades that are approximately 6 percentage points lower than they originally believed. This is an economically significant magnitude, corresponding to approximately 32 percent of standard deviation in the dependent variable.

Panel (d) of Figure 6 and the point estimates in columns (7) and (8) of Table 12 show that students accurately revise their grade expectations upon learning new information. Specifically, the dependent variable in these specifications is the difference between students' *realized* economics grades and their expected grades at the start of the semester. Much like the differences in grade expectations, nearly all students earn grades that are lower than their initial expectations; students who received positive information updates earn grades that are closest to initial expectations, and students who received negative updates earn grades that are farthest away. The point estimates imply that a one standard deviation increase in the information update measure is associated with students scoring 4.5 percentage points lower than originally expected.

In summary, the descriptive associations indicate that students adjust study time and grade expectations in a way that is inversely related to the new information they learn about their academic abilities throughout the semester. However, students' responses to new information are asymmetric and depend on whether they believe it is easier or harder to do well than they originally expected. Students who update their beliefs such that they think earning high grades is easier than originally expected respond by studying less than their initial study goals and revise their grade expectations down only slightly or not at all. Students who learn it is more difficult to earn higher grades respond by not changing study time much but substantially revising grade expectations down. It appears that they revise their grade expectations correctly, as realized grades follow a similar profile to the profile of revised expectations.

VI. Conclusion

This paper summarizes a five-year effort to improve college performance through an inexpensive and scalable setup in which thousands of students complete a one- to two-hour online exercise for a small grade at the start of the academic year. After registering and completing a brief survey, students were randomly assigned to interventions that we group here into six categories: 1) Goal Setting, in which students were asked to think and write carefully about their future and contemplate how their actions today could help them tomorrow; 2) Mindset, in which students read encouraging stories about how others, like them, adopted positive perspectives to overcome challenges, and were asked to share their own related experiences to help future students; 3) Online Coaching, in which students were given helpful advice for academic success; 4) Online and One-Way Text Coaching and 5) Online and Two-Way Text Coaching, in which students also received follow-up text messages of tips, reminders and, in the latter case, an opportunity to communicate regularly with a personal coach; and 6) Online and Face-to-Face Coaching in which real coaches were assigned to students and proactively tried to meet regularly with them.

The fidelity of the experiments was very high. The grade requirement ensured a large representative sample of students from a large first-year economics course participated in the experiments at low cost. About 95 percent of those asked to complete the exercise did so. Feedback and open-ended responses suggested that students took the tasks seriously, thought carefully about the information provided, and were overall quite positive about the experience. Most of those who received follow-up virtual and face-to-face coaching wished the program would continue for them and would be offered to future students. The platform provides a unique way to collect a large set of quantitative and qualitative data over time. Other colleges and institutions can administer our exercises at their own institutions or modify them to ask other questions and try other interventions.[38]

Our Goal Setting and Mindset interventions were based largely on promising results from previous research from social psychology. The Goal Setting exercise was very similar to one tested by Morisano et al. (2010), who found grades increased by more than half a standard deviation for upper-year students at McGill University with GPAs less than 3.0. One of the Mindset interventions is very similar to one tested by Yeager et al. (2016), who found the program increased credits earned and continuation into second year for first-generation and minority students at the University of Texas at Austin. We also worked with three of the authors of these studies to design mindset interventions specific to our context and setting—one focused on encouraging international students to feel a greater sense of belonging at the University of Toronto

[38] Details of the interventions and assistance for designing similar experiments are available on this paper's online appendix and through the website studentachievementlab.org.

and in Canada, and one focused on encouraging students to see their first-year economics course as a stepping stone for learning more about complex and interesting world problems.

As reported, none of the social psychology interventions we tested improved academic performance, even for sub-groups more likely to be at risk of doing poorly. It is possible that our population is less responsive to these interventions than the students who participated in the earlier studies mentioned above. Recent research also suggests, however, that efforts to apply psychologically based theory to improve long-term outcomes using short, one-time exercises are often not robust. A meta-analysis of 100 publications in psychology journals finds that when effects are replicated with well-powered designs, the mean effect size is half that of the pilot study, on average (Open Science Collaboration, 2015). Furthermore, while 97 percent of pilot studies had statistically significant results, only 36 percent of replications had significant results. Social psychology interventions are no exception (Dee, 2015; Hanselman et al., 2016).

We also find that our coaching interventions had no discernable impact on grades or persistence. Students certainly appreciated receiving follow-up text messages and virtual coaching support after completing an online exercise with advice on how to have a successful year. Coaches made students feel more supported and even happier, but they did not significantly improve performance.

We developed a simple model of student effort and highlighted four main reasons for poor student performance: low ability, low expected return from studying, low preferences for good grades, and procrastination or other behavioral barriers—all of which (except incoming ability) may change during the semester. Our data suggest that many students begin the year believing that, with relatively little effort, they can attain high grades. Over the semester, they realize they need to work much harder to attain their grade goals. Students react by committing to study a small amount more than before, but not as much as would be needed to target their previous grade. This behavior is consistent with some students giving up on the goal of earning a higher grade and now settling for a lower one. In addition to rationally updating study effort from new expectations about the returns to studying, we also find evidence of considerable procrastination. We estimate that, on average, students in our sample study five fewer hours a week than they would prefer based on rational choice.

Our coaching interventions increased self-reported study time by about two hours a week. We also find effects on changes to students' perceived study-to-grade relationship and interest in studying, but not on our measure of procrastination. We interpret these findings as suggesting that our low-cost online and text efforts made students realize they must study more to attain their target grades, and increased the value students place on higher grades, as treated students report a greater willingness to study for several different grade categories than do control students.

Despite the documented increase in study time, we find no effect on academic outcomes. Our measurement error adjusted gradient between study time and grades suggests that students must increase their study time by more than 10 hours per week to notice more than a 5-percentage point improvement in grades. A likely explanation for why our coaching interventions improved study behavior but not grades, therefore, is that the study improvement was too small to observe a corresponding effect on performance. More comprehensive, but expensive, programs that offer more intensive support to students for significantly altering study behavior may be necessary to make progress in helping incoming college students succeed.

References

Acemoglu, Daron, and David Autor. 2012. "What does human capital do? A review of Goldin and Katz's The Race between Education and Technology." *Journal of Economic Literature*, 50(2): 426-463.

Andersen, Steffen, Glenn W Harrison, Morten I Lau, and E Elisabet Rutström. 2008. "Eliciting risk and time preferences." *Econometrica*, 76(3): 583-618.

Arum, Richard, and Jospia Roksa. 2011. "Academically Adrift: Limited Learning on College Campuses." Chicago, IL: University of Chicago Press.

Arum, Richard, and Josipa Roksa. 2014. "Aspiring Adults Adrift. Chicago, IL: University of Chicago Press.

Babcock, Philip, and Mindy Marks. 2011. "The falling time cost of college: Evidence from half a century of time use data." *Review of Economics and Statistics*, 93(2): 468-478.

Beattie, Graham, Jean-William P. Laliberte, and Philip Oreopoulos. 2018. "Thrivers and divers: Using non-academic measures to predict college success and failure." *Economics of Education Review*, 62: 170-182.

Beattie, Graham, Jean-William Laliberte, Catherine Michaud-Leclerc, and Philip Oreopoulos. 2019. "What sets college thrivers and divers apart? A contrast in study habits, attitudes, and mental health." *Economic Letters*, 178: 50-53.

Bettinger, Eric, and Rachel Baker. 2014. "The Effects of Student Coaching: An Evaluation of a Randomized Experiment in Student Advising." *Educational Evaluation and Policy Analysis*, 35(1): 3-19.

Bettinger, E. P., Long, T. B., Oreopoulos, P., & Sanbonmatsu, L. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment." *Quarterly Journal of Economics*, 127(3): 1205-1242.

Braxton, John, Amy Hirschy, and Shederick McClendon. 2004. "Understanding and Reducing College Student Departure." *ASHE-ERIC Higher Education Reports*, 30(3): 1-97.

Brint, Steven, and Allison M. Cantwell. 2010. "Undergraduate time use and academic outcomes: Results from the University of California Undergraduate Experience Survey 2006." *Teachers College Record*, 112: 2441-2470.

Castleman, B. L., & Page, L. C. 2015. "Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-Income High School Graduates?" *Journal of Economic Behavior & Organization*, 115: 144-160.

Castleman, B. L., & Page, L. C. 2016. "Freshman Year Financial Aid Nudges: An Experiment to Increase FAFSA Renewal and College Persistence." *The Journal of Human Resources*, 51(2): 389-415.

Castleman, B. L., Page, L. C., & Meyer, K. 2016. "Customized Nudging to Improve FAFSA Completion and Income Verification." SSRN Working Paper http://dx.doi.org/10.2139/ssrn.2854345.

Castleman, Benjamin L. and Zachary Sullivan. 2019. "Cash for college apps: The effects of conditional cash transfers on selective college enrollment." University of Virginia Working Paper.

Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." NBER Working Paper 23618. Cambridge, MA: National Bureau of Economic Research.

Clotfelter, Charles T., Steven W. Hemelt and Helen F. Ladd. 2018. "Multifaceted Aid for Low-Income Students and College Outcomes: Evidence from North Carolina." *Economic Inquiry*, 56(1): 278-303.

Deming, David. 2019. "What does education do?" Kershaw Award Lecture, APPAM Conference Fall 2018. https://www.dropbox.com/s/30zgwghqn88tdnq/David%20N.%20Kershaw%20Award%20What%20Does%20Education%20Do.mp4

Dobronyi, Christopher R., Philip Oreopoulos, and Uros Petronijevic. 2019. "Goal setting, academic reminders, and college success: A large-scale field experiment." *Journal of Research on Educational Effectiveness*, 12(1): 38-66.

Donnellan, M Brent, Frederick L Oswald, Brendan M Baird, and Richard E Lucas. 2006. "The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality." *Psychological Assessment*, 18(2): 192.

Duckworth, Angela Lee, and Patrick D Quinn. 2009. "Development and validation of the Short Grit Scale (GRIT–S)." *Journal of Personality Assessment*, 91(2): 166–174.

Dweck, Carol S. 1986. "Motivational processes affecting learning," *American Psychologist*, 41(10): 1040-1048.

Dynarski, Susan, C.J. Libassi, Katherine Michelmore, Stephanie Owen. 2018. "Closing the Gap: The Effect of a Targeted, Tuition-Free Promise on College Choices of High-Achieving, Low-Income Students." NBER Working Paper No. 25349. Cambridge, MA: National Bureau of Economic Research.

Elovainio, Marko, and Mika Kivimäki, 1996. "Occupational Stresses, Goal Clarity, Control, and Strain Among Nurses in the Finnish Health Care System." *Research in Nursing and Health*, 19: 517-524.

Evans, William N., Melissa S. Kearney, Brendan C. Perry, James X. Sullivan. 2017. "Increasing community college completion rates among low-income students: Evidence from a randomized controlled trial evaluation of a case management intervention" NBER Working Paper 24150. Cambridge, MA: National Bureau of Economic Research.

Farkas, Gary J., Ewa Mazurek, and Jane R. Marone. 2014. "Learning style versus time spent studying and career choice: Which is associated with success in a combined undergraduate anatomy and physiology course?" *Anatomical Sciences Education*, 9: 121-131.

Giedd, Jay N., Michael Stockman, Catherine Weddle, Maria Liverpool, Gregory L. Wallace, Nancy Reitano Lee, Francois Lalonde, and Rhloshel K. Lenroot. 2012. "Anatomic Magnetic Resonance Imaging of the Developing Child and Adolescent Brain." In *The Adolescent Brain: Learning, Reasoning, and Decision Making* edited by Valerie F. Reyna, Sandra B. Chapman, Michael R. Dougherty and Jere Confrey. Washington, D.C.: American Psychological Association.

Gurría, Angel. 2009. Remarks for the OECD launch of Education at a Glance 2009. OECD Conference Centre, Paris, 8 December 2009.

Hirsh, Jacob B, and Jordan B Peterson. 2008. "Predicting creativity and academic success with a "fake-proof" measure of the Big Five." *Journal of Research in Personality*, 42(5): 1323–1333.

Lavecchia, Adam, Heidi Liu, Philip Oreopoulos. 2016. "Behavioral Economics of Education: Progress and Possibilities." In *Handbook of Economics of Education Volume 5* edited by Eric A. Hanushek, Stephen J. Machin, Ludger Woessmann, 1-74. North Holland Press, Amsterdam.

Hoxby, Caroline M., and Kevin Stange. 2018. "Introduction to 'Productivity in Higher Education," Chapter in forthcoming NBER book *Productivity in Higher Education*, Caroline M. Hoxby and Kevin Stange, editors. University of Chicago Press.

Kirby, Dale, and Dennis Sharpe, 2001. "Student Attrition from Newfoundland and Labrador's Public College." *Alberta Journal of Educational Research*, 47: 353-368.

Kirkeboen, Lars. J., Edwin Leuven, and Magne Mogstad. 2016.. "Field of Study, Earnings, and Self-Selection." *Quarterly Journal of Economics*, 131(3): 1057-1111.

Lochner, Lance, and Alexander Monge-Naranjo. 2012. "Credit Constraints in Education." *Annual Review of Economics*, 4: 225-256.

Locke, Edwin, and Gary Latham, 1990. “A Theory of Goal Setting and Task Performance.” Englewood Cliffs, NJ: Prentice-Hall.

Locke, Edwin, and Gary Latham, 2002. “Building a practically useful theory of goal setting and task motivation: A 35-year odyssey.” *American Psychologist*, 57: 705-717.

Locke, Edwin, Karyll Shaw, Lise Saari, and Gary Latham. 1981. “Goal Setting and Task Performance: 1969-1980.” *Psychological Bulletin*, 90: 125-152.

Martin Nybom. 2017. “The Distribution of Lifetime Earnings Returns to College.” *Journal of Labor Economics*, 35(4): 903-952.

Morisano, Dominique, 2008. “Personal Goal-Setting in University Students: Effects on Academic Achievement, Executive Functioning, Personality, and Mood.” Department of Educational and Counseling Psychology: McGill University.

Morisano, Dominique, Jacob B. Hirsh, Jordan B. Peterson, Robert O. Pihl, and Bruce M. Shore. 2010. “Setting, Elaborating, and Reflecting on Personal Goals Improves Academic Performance.” *Journal of Applied Psychology*, 95(2): 255-264.

Mueller, Claudia M. and Carol S. Dweck. 1988. “Praise for intelligence can undermine children’s motivation and performance.” *Journal of Personality and Social Psychology*, 75(1): 33-52.

Oreopoulos, Philip, and Reuben Ford. 2019. “Keeping College Options Open: A Field Experiment to Help All High School Seniors Through the College Application Process.” *Journal of Policy Analysis and Management*, 38(2): 299-329.

Oreopoulos, Philip, Richard W. Patterson, Uros Petronijevic, and Nolan G. Pope. 2019. “When studying and nudging don’t go as planned: Unsuccessful attempts to help traditional and online college students.” NBER Working Paper 25036. Cambridge, MA: National Bureau of Economic Research.

Oreopoulos, Philip, and Uros Petronijevic. 2013. “Making college worth it: A review of the returns to higher education.” *Future of Children*, 23(1): 41-65.

Oreopoulos, Philip, Uros Petronijevic, Christine Logel, and Graham Beattie. 2018. “Improving non-academic student outcomes using online and text-message coaching,” NBER Working Paper 249992. Cambridge, MA: National Bureau of Economic Research.

Oreopoulos, Philip, and Uros Petronijevic, 2018. “Student Coaching: How Far Can Technology Go?” *Journal of Human Resources*, 53(2): 299-329.

Ost, Ben, Weixiang Pan, and Douglas Webber. 2018 “The Returns to College Persistence for Marginal Students: Regression Discontinuity Evidence from University Dismissal Policies." *Journal of Labor Economics*, 36 (3): 779-805.

Psacharopoulos, George, and Harry Anthony Patrinos. 2018. "*Returns to investment in education: a decennial review of the global literature.*" *Education Economics*, 26(5): 445-458.

Scrivener, Susan, Michael J. Weiss, Alyssa Ratledge, Timothy Rudd, Colleen Sommo, and Hannah Fresques. 2015. "Doubling Graduation Rates: Three-Year Effects of CUNY's Accelerated Study in Associate Programs (ASAP) for Developmental Education Students." Report, MDRC, New York, NY.

Schippers, Michaéla, Ad Scheepers, and Jordan Peterson. 2015. "A Scalable Goal-Setting Intervention Closes Both the Gender and Ethnic Minority Achievement Gap." *Palgrave Communications*, 1: Article Number 15014.

Shapiro, D., Dundar, A., Huie, F., Wakhungu, P., Bhimdiwala, A., and Wilson, S. E. 2019. Completing College: Eight Year Completion Outcomes for the Fall 2010 Cohort (Signature Report No. 12c). Herndon, VA: National Student Clearinghouse Research Center.

Smith, Ken, Edwin Locke, and David Barry. 1990. "Goal Setting, Planning and Organizational Performance: An Experimental Simulation." *Organizational Behavior and Human Decision Processes*, 46: 118-134.

Sommo, Colleen, Dan Cullinan, and Michelle S. Manno. 2018. "Doubling graduation rates in a new state: Two-year findings from the ASAP Ohio Demonstration," Policy Brief, MDRC, New York, NY. https://www.mdrc.org/publication/doubling-graduation-rates-new-state

Statistics Canada. 2016. Education Highlights Tables, 2016 Census. "Highest level of educational attainment (general) by selected age groups 25 to 34, both sexes, % distribution 2016, Canada, provinces and territories, 2016 Census – 25% Sample data". https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/edu-sco/index-eng.cfm

Stinebrickner, Ralph and Todd R. Stinebrickner. 2008. "The Causal Effect of Studying on Academic Performance." *The B.E. Journal of Economic Analysis & Policy*, 8(1): 1-55.

University of Toronto. 2018. "Enrolment Report 2017-18, February 16, 2018, Planning and Budget Office.

U.S. Census Bureau. 2018. Current Population Survey, 2018 Annual Social and Economic Supplement. "Detailed Years of School Completed by People 25 Years and Over by Sex, Age Groups, Race and Hispanic Origin: 2018" (Table 3). https://www2.census.gov/programs-surveys/demo/tables/educational-attainment/2018/cps-detailed-tables/table-03.xlsx

Walton, Gregory M., Mary C. Murphy, Christine Logel, and David S. Yeager. 2017. "The social-belonging intervention: A guide for use and customization." Unpublished manuscript, Stanford University

Yeager, David, Gregory Walton, 2011. “Social-Psychological Interventions in Education: They’re Not Magic.” *Review of Educational Research*, 81(2): 267-301.

Yeager, David S., Gregory M. Walton, Shannon T. Brady, Ezgi N. Akcinar, David Paunesku, Laura Keane, Donald Kamentz, Gretchen Ritter, Angela Lee Duckworth, Robert Urstein, Eric M. Gomez, Hazel Rose Markus, Geoffrey L. Cohen, and Carol S. Dweck. 2016. “Teaching a lay theory before college narrows achievement gaps at scale.” *Proceedings of the National Academy of Sciences*, 113(24): E3341-E3348.

Zimmerman, Seth D. 2014. “The Returns to College Admission for Academically Marginal Students.” *Journal of Labor Economics*, 32(4): 711-754.

Table 1
Random Assignment to Different Treatment and Control Groups Across Year and Campus

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Year	**2014**	**2015**	**2015**	**2016**	**2016**	**2016**	**2017**	**2017**	**2017**	**2018**
Campus	**UTM**	**All**	**All**	**UTSG**	**UTM**	**UTSC**	**UTSG**	**UTM**	**UTSC**	**All**
Sample	**All**	**1st Years**	**Upper Years**	**All**	**All**	**All**	**All**	**All**	**All**	**All**
Group				**Number Assigned to Group (Percent Assigned to Group) [Expected Percent Based on Random Assignment]**						
Goal Setting	**950**									
	(59.7) [60.0]									
Mindset		**986**					**1043**		**663**	
		(19.7) [20.0]					(35.5) [33.3]		(50.8) [50.0]	
Online Coaching Only		**962**	**334**	**1152**		**600**				
		(19.6) [19.6]	(28.4) [29.4]	(34.1) [33.3]		(50.2) [50.0]				
Online and One-Way Text Coaching		**1506**	**493**		**710**					
		(30.7) [30.0]	(41.9) [40.0]		(48.7) [47.5]					
Online and Two-Way Text Coaching				**1118**			**787**	**670**		**2723**
				(33.1) [33.3]			(26.8) [33.3]	(47.4) [50.0]		(49.7) [50.0]
Online and Face-to-Face Coaching		**17**	**7**		**66**					
		(0.4) [0.4]	(0.6) [0.6]		(4.5) [5.0]					
Control (Personality Test)	**641**	**1451**	**342**	**1112**	**681**	**595**	**1106**	**743**	**643**	**2689**
	(40.3) [40.0]	(29.6) [30.0]	(29.1) [30.0]	(32.9) [33.3]	(46.7) [47.5]	(49.8) [50.0]	(37.7) [33.3]	(52.6) [50.0]	(49.2) [50.0]	(50.3) [50.0]
Total Sample Size **24772**	**1591**	**4904**	**1176**	**3382**	**1457**	**1195**	**2936**	**1413**	**1306**	**5412**

Notes: The table displays the number of University of Toronto students enrolled in a first-year economics course assigned to each experiment category by year, campus and sample. Values in round brackets show the percent assigned to a group relative to each randomized sample. Values in square brackets show the expected percent assigned to each group based on the assignment rule. UTM = University of Toronto at Mississauga campus, UTSG = University of Toronto at St. George (downtown) campus, UTSC = University of Toronto at Scarborough campus.

Table 2
Descriptive Statistics and Balance Tests

			Difference between treatment and control group mean [standard error in square brackets]											
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	Control Mean [standard dev.]		**Goal Setting**		**Mindset**		**Online Coaching**		**Online and One-Way Text Coaching**		**Online and Two-Way Text Coaching**		**Online and Personal Coaching**	
Want Grad. Degree	0.65	[0.477]	-0.01	[0.024]	0.00	[0.012]	-0.01	[0.011]	0.01	[0.012]	-0.01	[0.009]	-0.01	[0.051]
Father's Education [0-8]	5	[2.29]	-0.11	[0.116]	0.01	[0.056]	-0.04	[0.053]	-0.02	[0.059]	-0.03	[0.042]	-0.21	[0.246]
Mother's Education [0-8]	4.7	[2.23]	-0.18	[0.113]	-0.01	[0.054]	0.00	[0.051]	0.04	[0.057]	-0.05	[0.041]	-0.20	[0.238]
First Generation Student	0.309	[0.462]	0.01	[0.023]	0.00	[0.011]	0.01	[0.011]	-0.01	[0.012]	0.01	[0.009]	0.02	[0.050]
Parent has Grad. Degree	0.33	[0.470]	-0.01	[0.024]	0.00	[0.012]	0.00	[0.011]	0.00	[0.012]	-0.02	[0.009]**	-0.01	[0.051]
First-Year Student	0.733	[0.442]			0.00	[0.008]	0.00	[0.008]	0.00	[0.009]	0.01	[0.006]**	0.06	[0.037]*
International Student	0.308	[0.462]			-0.01	[0.010]	-0.01	[0.010]	0.00	[0.011]	0.00	[0.008]	0.04	[0.045]
Tendency to Not Cram [1-7]	3.9	[1.50]			0.04	[0.037]	0.04	[0.035]	-0.01	[0.039]	0.03	[0.028]	-0.02	[0.162]
Exp. Avg. Weekly Study Hrs.	17.9	[12.06]			0.17	[0.292]	0.06	[0.276]	0.26	[0.305]	-0.07	[0.221]	-0.89	[1.282]
Exp. Avg. Weekly Work Hrs.	7	[9.49]			-0.26	[0.229]	-0.18	[0.216]	-0.31	[0.239]	0.03	[0.173]	0.04	[1.004]
Exp. Fall Grade [0-100]	80.6	[6.85]	-0.24	[0.330]	0.02	[0.158]	0.03	[0.149]	0.18	[0.165]	-0.01	[0.119]	0.59	[0.694]
# Days Since Sept 1 Began Exercise	10.6	[4.85]	0.47	[0.223]**	0.00	[0.107]	-0.04	[0.101]	0.16	[0.112]	-0.04	[0.081]	-1.94	[0.471]***
Grit Score: Finish What I Begin [1-5	3.8	[0.827]			0.04	[0.026]*	-0.02	[0.026]	-0.07	[0.044]	0.03	[0.015]**	0.17	[0.106]
English Mother Tongue	0.42	[0.493]	0.00	[0.026]	0.02	[0.012]*	0.00	[0.012]	0.00	[0.013]	0.00	[0.009]	-0.09	[0.054]*
Male	0.48	[0.500]	0.00	[0.026]	0.02	[0.013]	0.00	[0.012]	0.00	[0.013]	0.00	[0.010]	-0.01	[0.055]
Age	20.2	[2.09]	0.08	[0.089]	-0.02	[0.042]	0.01	[0.040]	0.07	[0.044]	-0.02	[0.032]	0.17	[0.186]
No High School Grade Data	0.27	[0.443]	-0.03	[0.023]	0.01	[0.011]	0.00	[0.010]	0.03	[0.011]***	0.00	[0.008]	0.05	[0.048]
HS Grade Admissions Avg [0-100]	85.2	[7.03]	0.02	[0.379]	-0.13	[0.197]	0.04	[0.184]	-0.02	[0.200]	0.09	[0.149]	-0.89	[0.830]

Notes: Column 1 lists each background variable (recorded prior to random assignment). Want Grad. Degree = highest expected education attainment is more than a Bachelor degree. Father and mother education categories range from none (0) to Doctorate degree (8). Exp. = Expected. Avg. = Average. HS = High School. Grad. = Graduate. Column 2 displays the mean of these variables among the control group, while column 3 shows the standard deviation. Columns 4, 6, 8, 10, 12, and 14 show the difference between the variable mean for the indicated treatment and control groups. Columns 5, 7, 9, 11, 13, and 15 show the estimated standard errors for these differences. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent levels respectively.

Table 3
Control Group Outcome Means and Standard Deviations by Cohort and Year Since Experiment

	Year Experiment Began (in Sept) [standard deviation in brackets]				
(1)	(2)	(3)	(4)	(5)	(6)
Relative Year Since Exp. Began	**2014**	**2015**	**2016**	**2017**	**2018**
Fall Grade Avg., Year 1	64.8	69.1	68.5	68.1	70.2
	[13.0]	[14.2]	[13.6]	[13.3]	[13.1]
Winter Grade Avg., Year 1	68.4	68.4	68.2	67.1	
	[14.0]	[15.7]	[15.7]	[13.8]	
Year 1 Grade Average	64.3	67.5	66.5		
	[13.9]	[13.6]	[14.3]		
Year 2 Grade Average	66.9	68.7	70.4		
	[13.0]	[13.0]	[12.8]		
Year 3 Grade Average	68.2	70.7			
	[13.8]	[12.9]			
Year 4 Grade Average	69.5				
	[13.0]				
Year 1 Total Credits Earned	2.7	3.7	3.6	3.5	
	[1.8]	[1.7]	[1.8]	[1.7]	
Year 2 Total Credits Earned	3	3.3	3.3		
	[2.1]	[2.0]	[2.2]		
Year 3 Total Credits Earned	3	3.2			
	[2.5]	[2.3]			
Year 4 Total Credits Earned	2.8				
	[2.6]				
Persistence Year 2	0.871	0.913	0.892	0.848	
Persistence Year 3	0.786	0.838	0.756		
Persistence Year 4	0.723	0.759			
Graduated by End of Year 4	0.38				

Notes: The table shows outcome means and, in square brackets, standard deviations for the control groups from each year of the experiment. Grade averages (Avg.) are listed as a percernt. Persistence variables show the fraction of first-year students in the first year of the experiment with any grade data in the following second, third, and fourth years. The graduation variable indicates the fraction officially gradating with any degree by the Fall Term of 2018 (after four years for first-years in the 2014 experiment).

Table 4
Estimated Treatment Effects on Initial Fall Term Grades [0-100]

	Outcome Variable						
(1)	(2) **Missing Fall Grade**	(3) **Fall Term Grade**	(4) **Grade>50**	(5) **Grade>60**	(6) **Grade>70**	(7) **Grade>80**	(8) **Grade>90**
Goal Setting	-0.008	0.254	0.012	0.005	0.015	0.012	0.000
	[0.018]	[0.781]	[0.015]	[0.023]	[0.029]	[0.023]	[0.010]
Mindset	0.008	0.655	0.01	0.022	0.013	0.018	0.01
	[0.008]	[0.353]*	[0.007]	[0.010]**	[0.013]	[0.011]*	[0.004]**
Online Coaching Only	0.000	0.072	-0.008	0.012	0.013	-0.002	0.002
	[0.008]	[0.337]	[0.007]	[0.010]	[0.012]	[0.010]	[0.004]
Online and One-Way Text Coaching	0.021	0.199	0.006	0.01	0.007	-0.001	0.003
	[0.009]**	[0.376]	[0.007]	[0.011]	[0.014]	[0.011]	[0.005]
Online and Two-Way Text Coaching	-0.001	-0.191	-0.001	-0.003	0.001	0.002	-0.001
	[0.006]	[0.269]	[0.005]	[0.008]	[0.010]	[0.008]	[0.003]
Online and Face-to-Face Coaching	-0.009	-0.456	-0.006	-0.053	-0.04	-0.019	-0.013
	[0.037]	[1.539]	[0.030]	[0.045]	[0.056]	[0.046]	[0.019]
Control Mean [& st.dev.]	0.131	68.8 [13.5]	0.924	0.795	0.517	0.2	0.025
Sample Size	24,772	21,305	21,305	21,305	21,305	21,305	21,305

Notes: The table shows coefficient estimates from regressing the indicated outcome variable on the different treatment categories plus fixed effects for each randomized group listed in Table 1. Grades are measured as a percent at the end of the fall term averaged over all courses completed in the first year of each experiment. Grade>X is an indicator variable for whether the Fall Term Grade exceeds X. Control means, standard deviations and sample sizes are also shown at the bottom. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent level respectively.

Table 5
Estimated Treatment Effects on Academic Performance and Persistence

	Outcome									
(1)	(2) **Fall Grade Year 1**	(3) **Winter Grade Year 1**	(4) **Credits Earned Year 1**	(5) **Final Grade Year 1**	(6) **Persisted Year 2**	(7) **Credits Earned Year 2**	(8) **Final Grade Year 2**	(9) **Persisted Year 3**	(10) **Credits Earned Year 3**	(11) **Final Grade Year 3**
Goal Setting	0.254	-0.602	-0.035	-0.594	-0.013	0.108	0.337	0.026	-0.001	0.519
	[0.781]	[0.920]	[0.083]	[0.747]	[0.020]	[0.096]	[0.750]	[0.022]	[0.104]	[0.734]
Mindset	0.655	0.354	0.068	0.545	-0.005	0.041	0.557	0.011	0.086	0.482
	[0.353]*	[0.406]	[0.040]*	[0.354]	[0.010]	[0.068]	[0.535]	[0.017]	[0.077]	[0.548]
Online Coaching Only	0.072	0.716	-0.012	0.409	0.011	0.006	0.493	0.013	0.152	0.371
	[0.337]	[0.388]*	[0.038]	[0.336]	[0.009]	[0.044]	[0.349]	[0.011]	[0.072]**	[0.509]
Online and One-Way Text	0.199	0.201	-0.004	0.103	-0.012	0.047	0.002	-0.012	0.128	0.419
Coaching	[0.376]	[0.430]	[0.041]	[0.367]	[0.010]	[0.049]	[0.382]	[0.012]	[0.064]**	[0.457]
Online and Two-Way Text	-0.191	-0.352	-0.044	-0.196	-0.001	-0.039	-0.097	0.01	Not available yet	
Coaching	[0.269]	[0.419]	[0.041]	[0.367]	[0.010]	[0.071]	[0.560]	[0.017]		
Online and Face-to-Face	-0.456	0.374	-0.043	0.086	0.043	0.29	1.176	0.053	1.421	4.302
Coaching	[1.539]	[1.760]	[0.174]	[1.535]	[0.042]	[0.197]	[1.535]	[0.047]	[0.392]***	[2.749]
Control Mean [& st.dev.]	68.8 [13.5]	68.3 [15.3]	3.1 [1.8]	67.6 [13.8]	0.804	3.0 [1.9]	69.2 [13.0]	0.727	2.9 [1.9]	71.0 [12.7]

Notes: The table shows coefficient estimates from regressing the indicated outcome variable on the different treatment categories plus fixed effects for each randomized group listed in Table 1. The year indicates the year since the experiment began. Control means and standard deviations are also shown at the bottom. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent level respectively.

Table 6
Estimated Treatment Effect on Fall Term Grade by Propensity Score Groups for Low Fall Grade

	Sample						
(1)	(2) **Full Sample [with HS grade]**	(3) **pscore> 90 pctile**	(4) **pscore> 80 pctile**	(5) **pscore> 70 pctile**	(6) **pscore> 60 pctile**	(7) **pscore> 50 pctile**	(8) **pscore<= 50 pctile**
Goal Setting	0.671	-1.495	0.991	1.521	0.608	0.952	1.375
	[0.855]	[3.231]	[2.007]	[1.594]	[1.347]	[1.188]	[1.104]
Mindset	0.941	2.395	1.949	1.665	1.591	1.281	0.684
	[0.417]**	[1.521]	[1.075]*	[0.860]*	[0.735]**	[0.651]**	[0.479]
Online Coaching Only	-0.194	-1.582	-1.249	-1.081	-0.923	-1.199	0.731
	[0.394]	[1.358]	[0.939]	[0.759]	[0.645]	[0.571]**	[0.484]
Online and One-Way Text	-0.027	-0.519	0.029	0.005	0.258	0.227	-0.279
Coaching	[0.433]	[1.294]	[0.930]	[0.757]	[0.649]	[0.584]	[0.583]
Online and Two-Way Text	-0.264	-1.907	-0.792	-1.535	-1.065	-1.158	0.216
Coaching	[0.315]	[1.238]	[0.874]	[0.693]**	[0.578]*	[0.504]**	[0.355]
Online and Face-to-Face	-0.221	-6.119	-3.284	0.599	0.944	1.480	-3.596
Coaching	[1.724]	[4.798]	[3.365]	[2.719]	[2.343]	[2.096]	[2.853]
Control Mean [& st.dev.]	68.0 [13.8]	59.7 [15.1]	61.0 [14.7]	61.7 [14.5]	62.7 [14.3]	63.5 [14.2]	72.2 [12.0]
Sample Size	16,154	1,539	3,067	4,633	6,212	7,844	8,310

Notes: The table shows coefficient estimates from regressing Fall Grades (in percent) from the experimental year on the different treatment categories plus fixed effects for each randomized group listed in Table 1. Except for Column 2, the samples include only those with non-missing high school grade data. Regression results are shown using different samples, restricted by the indicated percentile cut-offs of a propensity score for the likelihood of receiving a low grade (less than 60) based on background characteristics. See text for more details on the calculation of this score. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent level respectively.

Table 7
Estimated Treatment Effect Outcomes on Reported Mental Health and College Experience

	Outcome Coefficients [standard errors in brackets]					
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Outcome	**Online Coaching Only**	**Online and One-Way Text Coaching**	**Online and Two-Way Text Coaching**	**Online and/or Text Coaching**	**Online and Face-to-Face Coaching**	**Sample Size**
Completed Follow-up Survey (Cont.Mn=0.76)	-0.015 [0.010]	0.01 [0.016]	-0.01 [0.007]		-0.034 [0.039]	11,446
Subjective Mental Health Outcomes (standardized)						
Life Satisfaction	0.08 [0.038]**	0.078 [0.062]	0.042 [0.026]		0.317 [0.153]**	8,140
Univ. Satisfaction	0.021 [0.030]	0.066 [0.049]	0.023 [0.021]		0.208 [0.121]*	8,140
Feeling Less Stressed	0.013 [0.024]	0.029 [0.039]	0.017 [0.016]		0.136 [0.096]	8,140
Feeling Less Depressed	0.011 [0.041]	0.081 [0.062]	0.034 [0.044]		0.166 [0.153]	4,342
Overall Mental Health	0.041 [0.031]	0.085 [0.050]*	0.037 [0.021]*		0.279 [0.125]**	8,140
Overall Mental Health				0.044 [0.018]**	0.258 [0.122]**	8,140
Subjective Feelings of Support (standardized)						
Sense of Belonging	-0.061 [0.041]	0.103 [0.062]*	0.016 [0.045]		0.278 [0.153]*	4,276
University Wants Me to Succeed	0.056 [0.042]	0.06 [0.062]	0.09 [0.045]**		0.431 [0.154]***	4,276
University Supports Me	0.044 [0.042]	0.057 [0.062]	0.093 [0.045]**		0.278 [0.153]*	4,276
Confident I Can Succeed	0.047 [0.041]	0.048 [0.062]	0.009 [0.045]		0.288 [0.152]*	4,276
Overall Sense of Support	0.029 [0.041]	0.089 [0.062]	0.069 [0.045]		0.425 [0.152]***	4,276
Overall Sense of Support				0.057 [0.032]*	0.409 [0.150]***	4,276

Notes: The table shows coefficient estimates from regressing the indicated standardized outcome variable (with mean zero, standard deviation one) on the different treatment categories plus fixed effects for each randomized group listed in Table 1. Except for the first row, the sample is restricted to those responding to the follow-up surveys taken near or after the end of the first year fall term. See text for more details. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent level respectively.

Table 8
Estimated Treatment Effect Outcomes on Reported Study Behavior

	Outcome Coefficients [standard errors in brackets]					
(1)	(2) **Online Coaching Only**	(3) **Online and One-Way Text Coaching**	(4) **Online and Two-Way Text Coaching**	(5) **Online and/or Text Coaching**	(6) **Online and Face-to-Face Coaching**	(7) Sample Size
Study Behavior (standardized)						
Average Weekly Study Hours	0.067 [0.035]*	0.031 [0.053]	0.113 [0.023]***		0.006 [0.142]	9,662
Tend Not to Cram for Exams	0.038 [0.023]	0.003 [0.026]	0.054 [0.020]***		0.004 [0.107]	9,662
Number of Missed Classes	-0.205 [0.307]	-0.137 [0.319]	-0.135 [0.032]***		NA	3,995
Review Past Mistakes to Learn	0.021 [0.040]	-0.031 [0.056]	0.044 [0.042]		0.224 [0.152]	4,830
Rewrite Material in Own Words	0.013 [0.040]	-0.113 [0.056]**	-0.009 [0.042]		0.264 [0.152]*	4,830
Get Writing Feedback	0.002 [0.040]	0.029 [0.056]	0.054 [0.042]		0.287 [0.152]*	4,830
Meet with Tutor	-0.006 [0.040]	-0.149 [0.056]***	0.067 [0.042]		0.161 [0.152]	4,830
Manage Time Well	0.064 [0.037]*	-0.001 [0.056]	0.074 [0.025]***		0.332 [0.151]**	8,770
Overall Positive Study Behavior	0.06 [0.036]*	-0.051 [0.054]	0.127 [0.023]***		0.19 [0.145]	9,718
Overall Positive Study Behavior				0.091 [0.020]***	0.263 [0.143]*	9,718
Study Time (from 2018-19 data)	**Online Coaching Only**	**Online and One-Way Text Coaching**	**Online and Two-Way Text Coaching**	**Online and/or Text Coaching**	**Online and Face-to-Face Coaching**	Cont. Mn [Std Dev]
Weekly Study Time Fall Sem.			2.28 [0.409]***			14.4 [12.7]
Econ Weekly Study Time Fall Sem.			0.99 [0.172]***			5.0 [3.7]
Time Diary Daily Alone Study Time Fall Sem.			0.26 [0.085]***			2.6 [2.5]
Time Diary Daily Group Study Time Fall Sem.			0.06 [0.052]			0.7 [1.6]
Time Diary Daily Total Study Time Fall Sem.			0.32 [0.092]***			3.3 [2.7]
Current Weekly Study Time in Winter Sem.			2.18 [0.518]***			14.0 [10.6]

Notes: The table shows coefficient estimates from regressing the indicated standardized outcome variable (with mean zero, standard deviation one) on the different treatment categories plus fixed effects for each randomized group listed in Table 1. The sample is restricted to those responding to the follow-up surveys taken near or after the end of the first year fall term. See text for more details. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent level respectively.

Table 9

Estimates of the Gradient Between Weekly Study Time and Grades in the 2018-19 Cohort

	Mean Fall Grade				Mean Fall Math Grade			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	OLS	OLS	IV Using Baseline Survey	IV Using Daily Study Time	OLS	OLS	IV Using Baseline Survey	IV Using Daily Study Time
Study Time/Week	0.081***		0.225***	0.303***	0.150***		0.420***	0.493***
	[0.022]		[0.053]	[0.071]	[0.043]		[0.113]	[0.153]
Daily Study Time		0.442***				0.691***		
(first stage coeff.)		[0.101]				[0.230]		
Mean of Dep Var.			70.16				70.15	
[Standard Dev.]			[13.07]				[16.55]	
Observations	1,702	1,711	1,702	1,702	671	673	671	671

Notes: All regressions pool control group observations and include campus fixed effects. When we instrument for weekly study time using data from the baseline survey, we use the following variables as instruments: study hours per week in high school, self-reported expected study hours per week this semester, tendency to regularly "cram" for exams, expected hours per week working for pay during the semester, and expected commuting time to campus. When we instrument using daily study time we use the total study time students reported for "yesterday" -- i.e., the day before they took the follow-up survey. Robust standard errors appear in brackets. *** indicates significance at the 1 percent level.

Table 10
Characteristics that Relate to Study Effort and Grade Outcomes

	Ind. Var. Mean [Std. Dev]	**Dep. Var. = Reported Avg. Weekly Hrs of Study During Fall Term, mean 15.1, std. 11.1**				
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Alpha (exp. grade attainable with minimal cramming)	53.7	-0.045				
	[12.4]	[0.039]				
Beta (slope between grade and weekly study time)	1.8	-0.109				
	[0.71]	[0.667]				
Expected Weekly Hrs Needed to Get an A in Econ.	14.7		0.004	-0.016	0.002	0.031
	[7.1]		[0.042]	[0.040]	[0.039]	[0.046]
Extra Weekly Hrs Willing to Study to Guarantee A vs. B	8.3			0.441	0.437	0.487
	[7.2]			[0.050]***	[0.048]***	[0.058]***
Like to Study (1 - 6 scale)	3.6				1.69	1.22
	[1.2]				[0.300]***	[0.349]***
High School Grade	86.6					0.189
	[6.6]					[0.064]***
Observations		782	776	776	775	563
R-squared		0.002	0	0.093	0.13	0.152
		Dep. Var. = End of Fall Term Grade, mean 71.2, std. 11.9				
Alpha (exp. grade attainable with minimal cramming)	53.7	0.383	0.416	0.415	0.364	0.378
	[12.4]	[0.025]***	[0.042]***	[0.042]***	[0.044]***	[0.044]***
Beta (slope between grade and weekly study time)	1.8	1.757	2.26	2.282	2.363	2.311
	[0.71]	[0.446]***	[0.710]***	[0.712]***	[0.748]***	[0.740]***
Extra Weekly Hrs Willing to Study to Guarantee A vs. B	8.3		0.127	0.124	0.15	0.092
	[7.2]		[0.054]**	[0.054]**	[0.059]**	[0.061]
Hrs of Weekly Procrastination (Actual-Target Study Hrs)	8.4			-0.013	-0.029	0.01
	[15.7]			[0.027]	[0.028]	[0.030]
High School Grade	86.6				0.552	0.512
	[6.6]				[0.066]***	[0.067]***
Reported Avg. Weekly Hrs of Study During Fall Term	15.1					0.155
	[11.1]					[0.045]***
Observations		1739	704	704	534	534
R-squared		0.123	0.138	0.138	0.252	0.268

Notes: The table shows coefficient estimates from regressing the indicated dependent variable (Dep. Var.) on the indicated independent variables (Ind. Var). The sample is restricted to those responding to the 2018-19 follow-up surveys taken near or after the end of the first year fall term. exp = expected, Econ. = Economics. Avg. = Average. Hrs = Hours. Wkly = Weekly. std. = standard deviation. See text for more details. One, two, and three asterisks indicate statistical significance at the 10, 5, and 1 percent level respectively.

Table 11
Online Plus Two-Way Coaching Effects on Study Expectations, Ambition, and Procrastination

(1)	(2) Mean	(3) Std. Dev.	(4) Outcome Coef.	(5) Std. Error	(6) Sample Size
Alpha (exp. grade attainable with minimal cramming)	54.2	[12.3]	-1.153	[0.577]**	1,817
Beta (slope between grade and weekly study time)	1.7	[0.69]	-0.021	[0.033]	1,817
Expected Weekly Hrs Needed to Get an A in Econ.	16	[10.3]	0.841	[0.483]*	1,810
Extra Weekly Hrs Willing to Study to Guarantee A vs. B	8.3	[7.3]	1.008	[0.348]***	1,735
Extra Weekly Hrs Willing to Study to Guarantee A+ vs. B+	10.2	[8.6]	1.301	[0.410]***	1,735
Grades Don't Matter So Long As I Graduate (1-7 scale)	2.5	[1.2]	-0.107	[0.041]***	3,762
Procrastination (Winter Target Hrs - Winter Actual Hrs)	8.7	[15.7]	0.582	[1.151]	750
Procrastination (Winteter Target Hrs - Fall Actual Hrs)	4.9	[12.9]	0.036	[0.598]	1,875
Procrastination for Students with Low Initial Target Hrs	4.6	[12.9]	-0.749	[0.928]	730
Procrastination for Students with High Initial Target Hrs	5.1	[13.2]	0.533	[0.780]	1,145
Social Media, Screens Distract Me (standardized)	0	1	-0.026	[0.071]	806

Notes: The table reports estimated treatment effects from online and two-way coaching for the time-management program tested during the 2018-19 academic year. Sample sizes vary because some outcomes are collected from different surveys with different response rates (not correlated with treatment), and some variables were asked to a random subset only. exp = expected, hrs = hours. The social media variable is the standardized average of students' responses to their subjective agreement to the degree to which social media and video distract them. Students with low (high) initial target hours are those with stated target weekly study hours below (equal or above) the median (15 hours). *** indicates significance at the 1 percent level; ** indicates significance at the 5 percent level; * indicates significance at the 10 percent level.

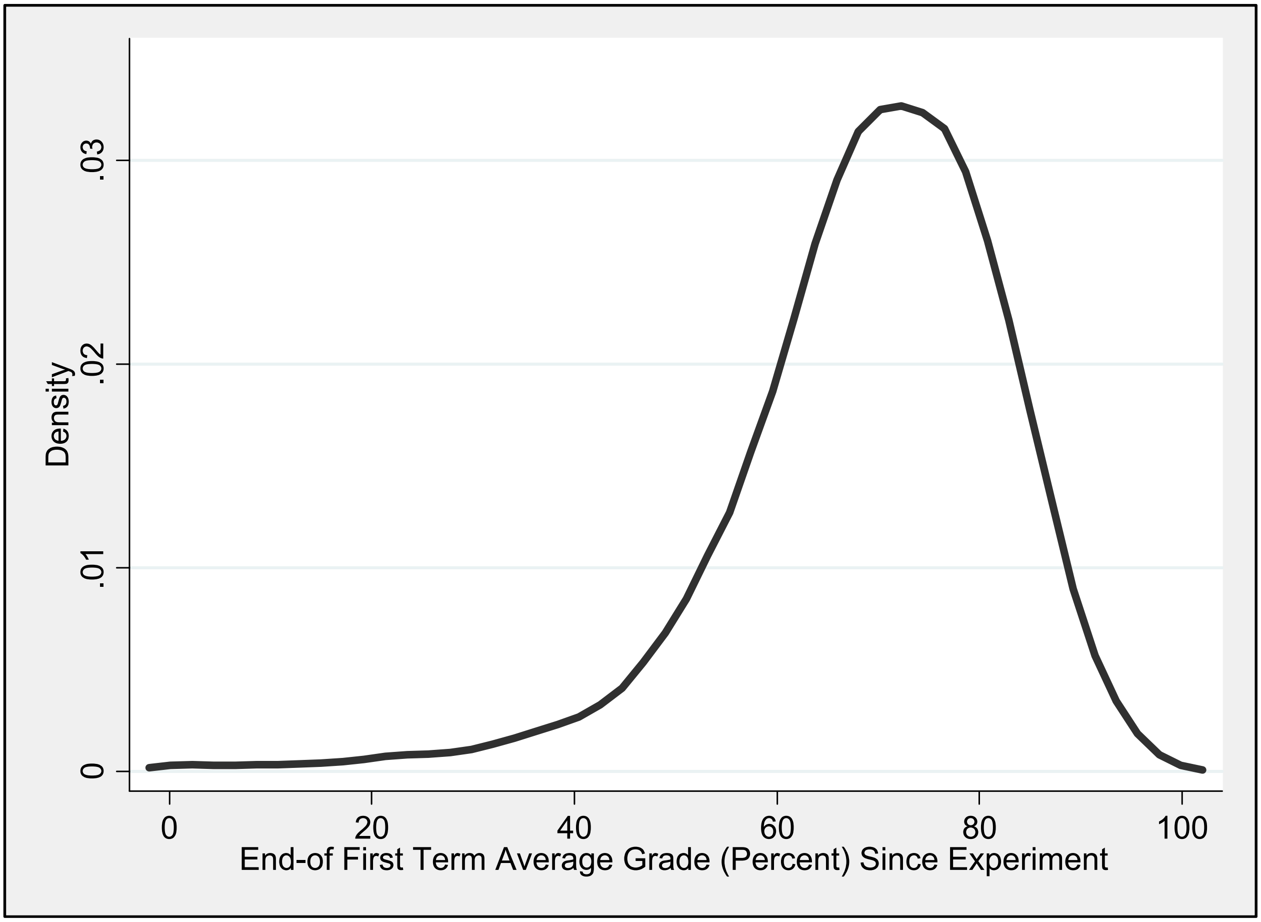
Table 12
Information Updating Revisions in Study Times and Grade Expectations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Actual - Expected Study Time in Economics		Actual - Expected Study Time in All Courses		Difference in Expected Econ Grades: At Follow-up – At Baseline		Actual Econ Grade - Expected Econ Grade at Baseline	
	Outcomes and Changes in Number of Hours Needed for an *A* Based on Changes in Student Study Gradients							
Δs_i^A	0.122***	0.108***	0.191***	0.094**	-0.987***	-0.861***	-0.716***	-0.631***
	[0.019]	[0.016]	[0.059]	[0.048]	[0.072]	[0.070]	[0.072]	[0.067]
Observations	1,765	1,664	1,765	1,664	1,765	1,664	915	915
Controls?	N	Y	N	Y	N	Y	N	Y

Notes: Each regression is estimated at the student level and the dependent variable indicated in the column headings. Control variables include age, expected weekly study time across all courses reported during the baseline survey, expected weekly study time in economics reported at during the baseline survey, the number of days it took for the student to start the online warmup exercise, campus fixed effects, commute time to campus (in minutes), cubic functions of students' initially expected economics grade, initially expected weekly study time in economics, and initially expected study time across all courses, indicators for expected performance categories, English as a second language, gender, first-year status, first-generation status, international student status, intending to earn more than a BA, self-reported enjoyment of studying, frequent use of a calendar, believing the first midterm in a course determines subsequent outcomes, the belief that grades do not matter as long as one graduates, managing time well, and having a strong tendency to study at the last minute. Robust standard errors are reported in brackets. *** indicates significance at the 1 percent level; ** indicates significance at the 5 percent level; and * indicates significance at the 10 percent level.

Figure 1

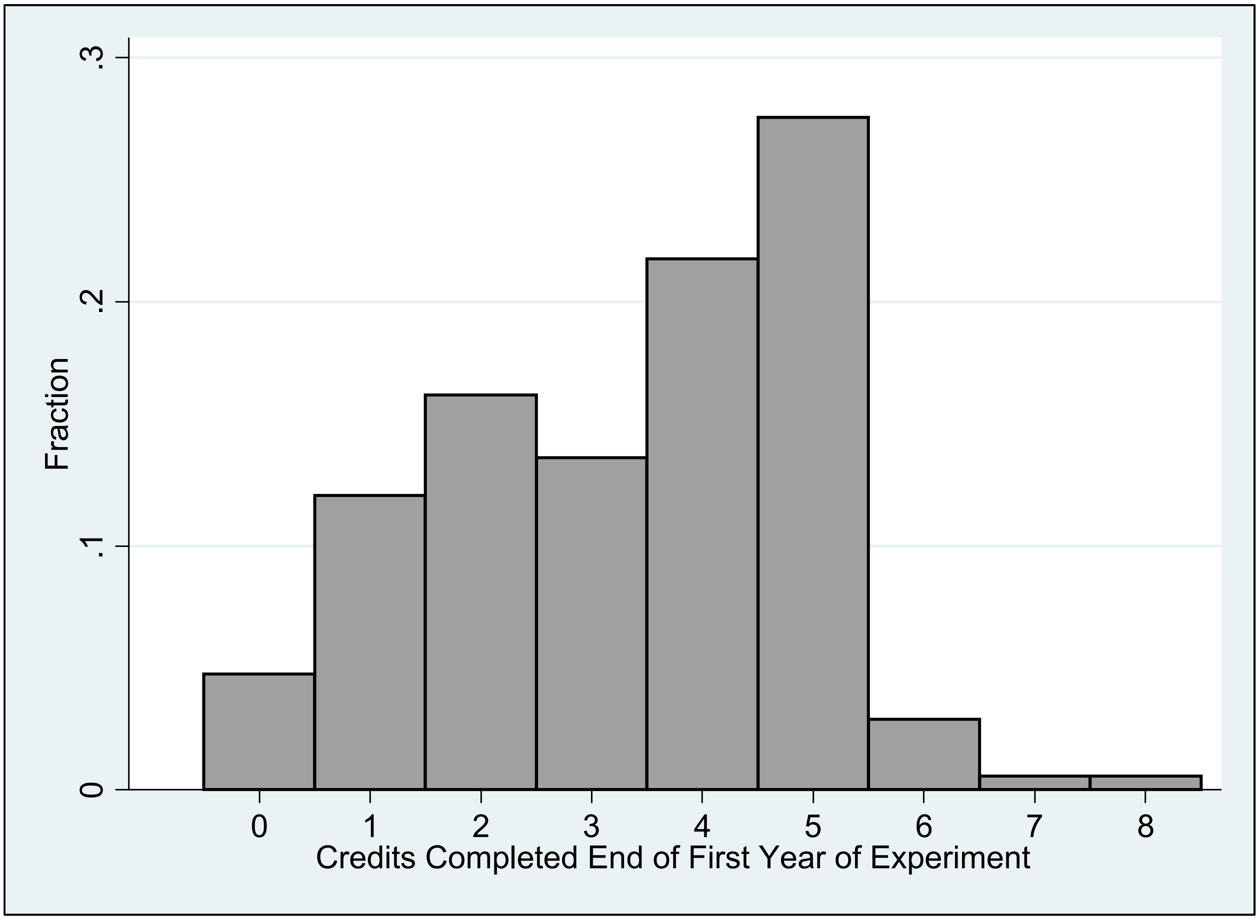
Fall Term Grade Distribution



Notes: Figure 1 graphs the kernel density estimate of all first year fall term grade averages for this paper's main sample of 2014-2018 first-year economics students. The density was calculated using a bandwidth of 2 and STATA's kdensity command. The median grade is 70.5, the 25th percentile is 62.0.

Figure 2

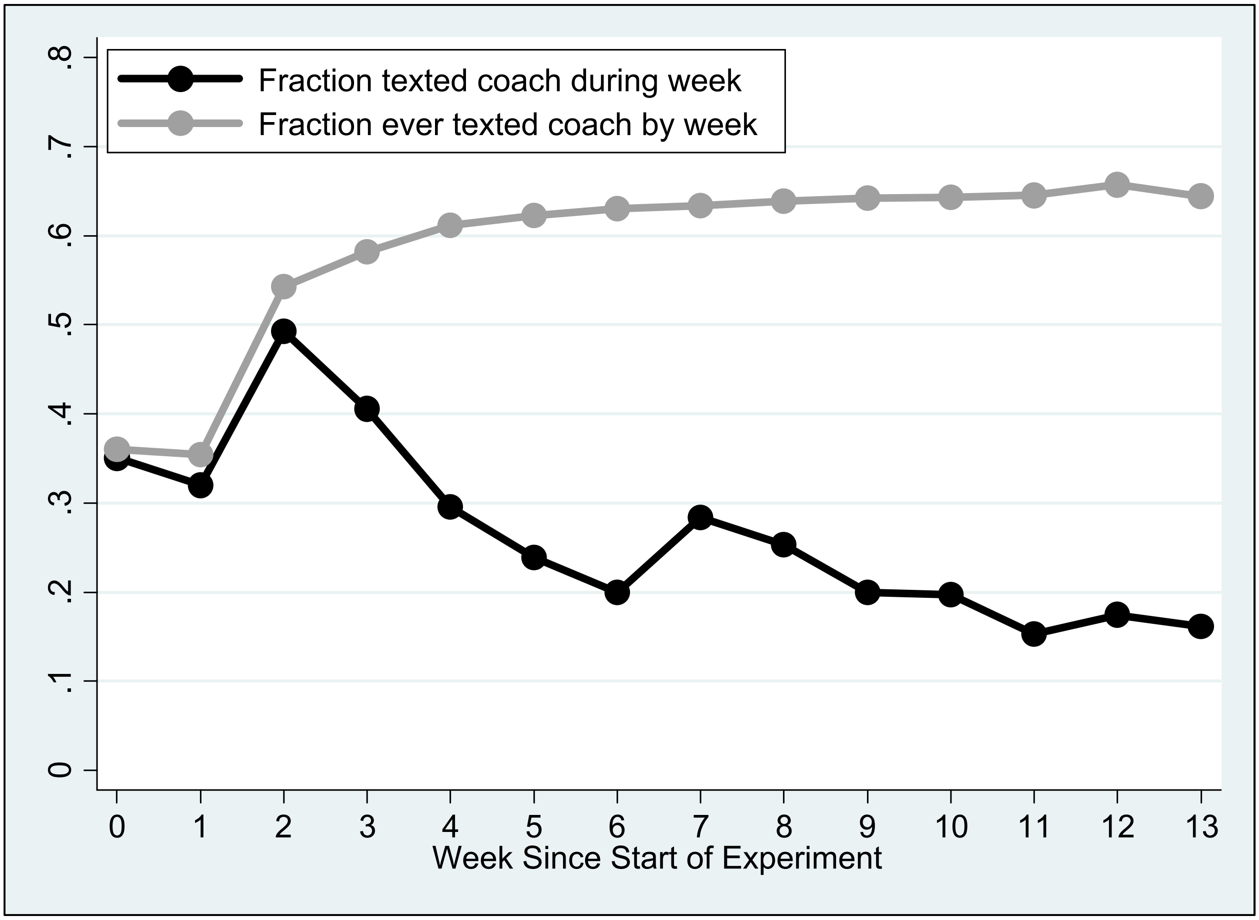
Total Credits Completed by End of First Year of Experiment (Sept-Aug)



Notes: Figure 1 displays the histogram of total credits completed by the end of the first year of the experiment. A full course load to graduate in four years with summers off would typically be 5 credits. The sample includes all first-year economics students in this paper's main sample (2014-2018).

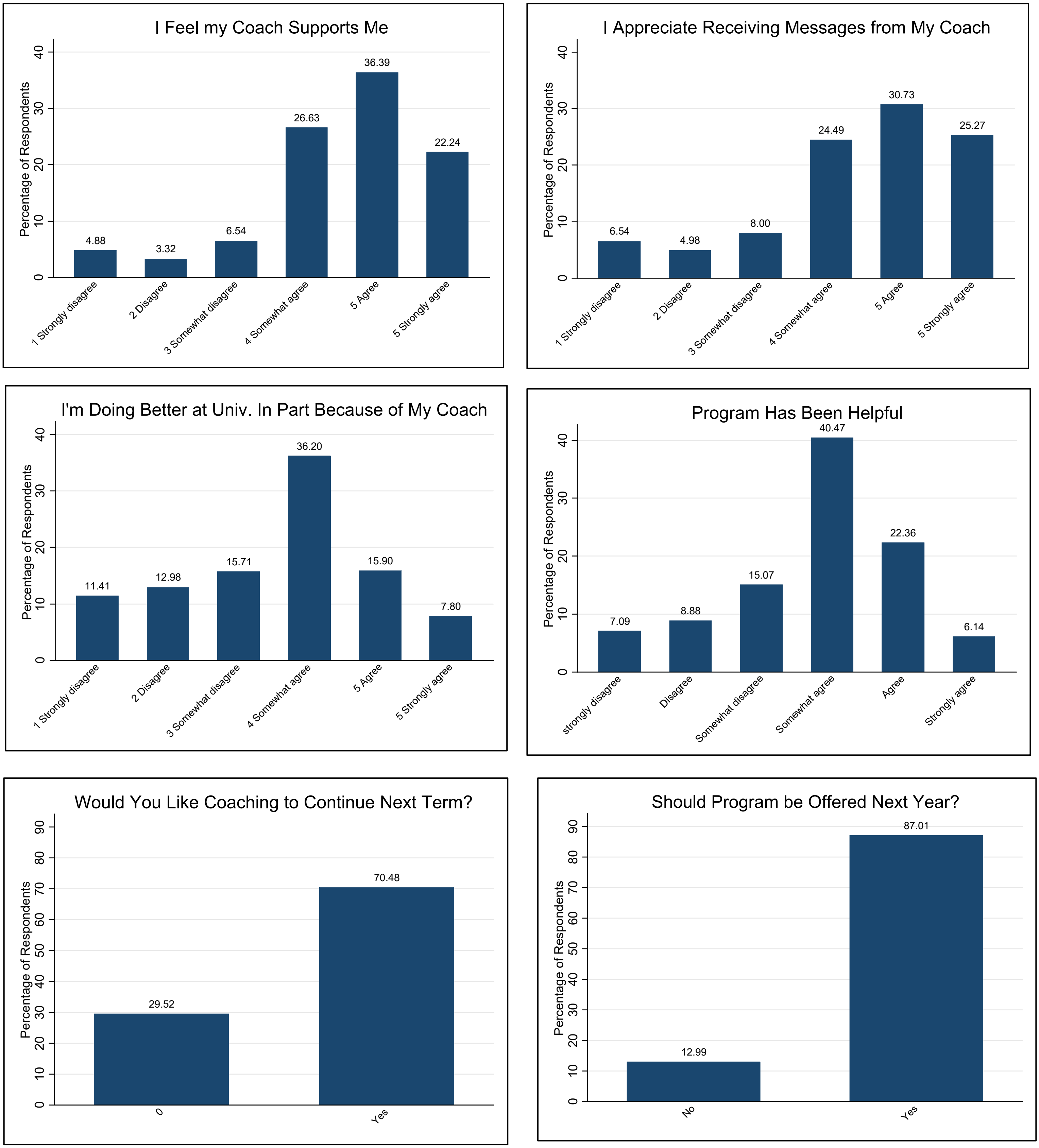
Figure 3

Fraction of Students Assigned to a Virtual Coach That Texted Back in a Given Week Since Start of Experiment And Fraction Ever Texted Back



Notes: The sample includes students agreeing to receive text-message coaching with Two-Way communication at the start of the 2016, 2017, and 2018 school years. The lighter line displays the fraction of this sample who ever texted back as of the indicated week during the first fall term of the experiment (with zero being the first Sunday after September 1). The darker line displays the fraction of this sample who texted anything back in a given week.

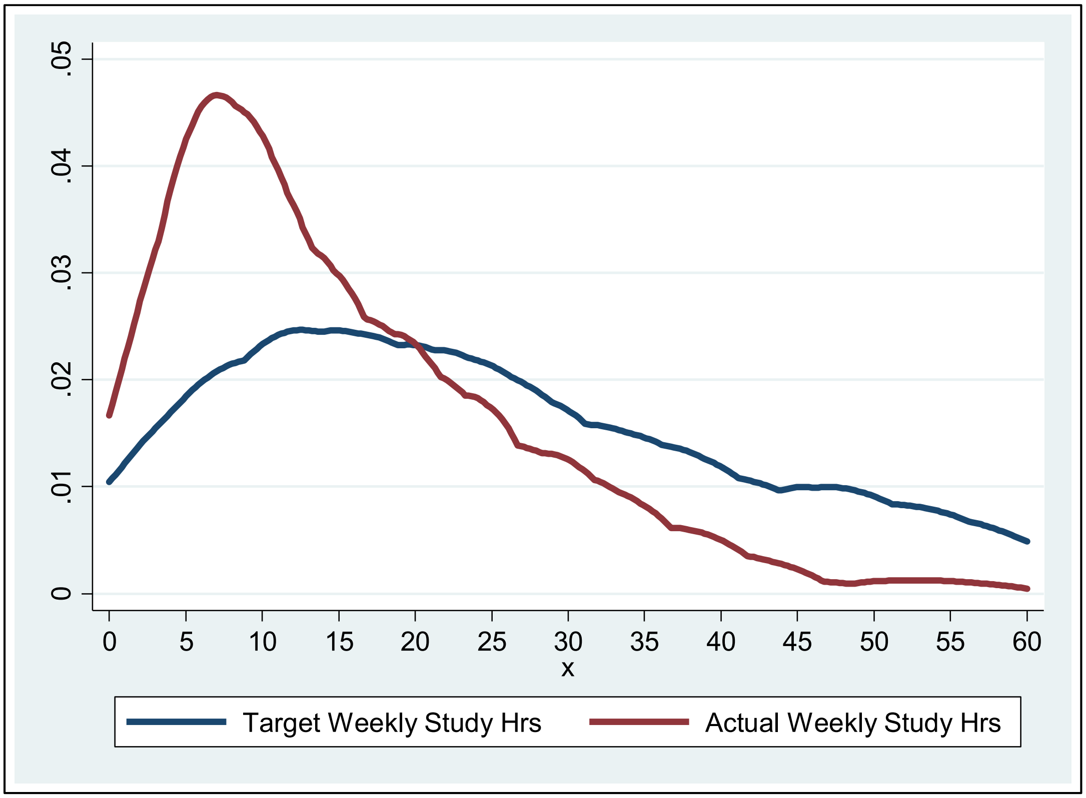
Figure 4: Student Feelings About the 2-Way Text-Message Coaching Program



Notes: The first three panels show the percentages of students in the text-message coaching program in 2016 who strongly disagree, disagree, somewhat disagree, somewhat agree, agree, and strongly agree with the statement that appears as the title of each panel. The last three panels show students responding about the 2018 coaching program.

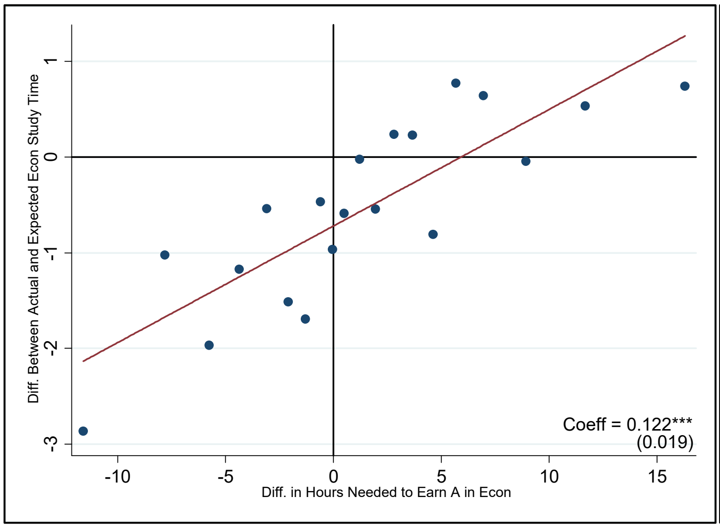
Figure 5

Target Versus Actual Reported Weekly Study Time

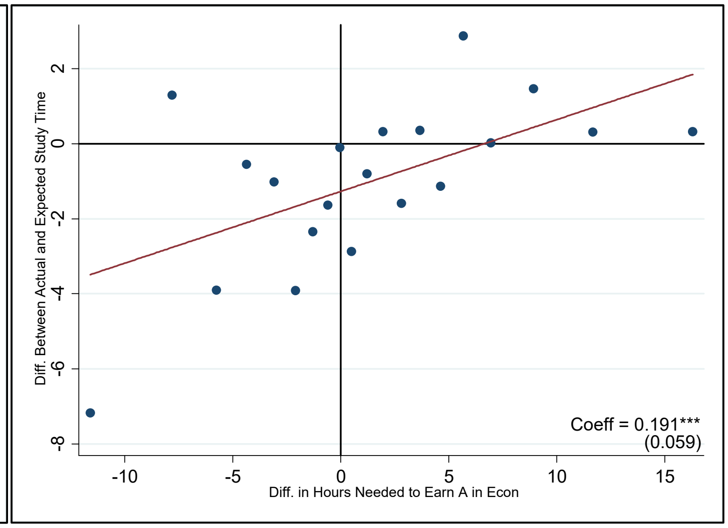


Notes: The sample is restricted to those responding to the 2018-19 follow-up surveys taken near or after the end of the first-year fall term. The blue line indicates the kernel density estimate of the reported target study hours in the next Winter semester, surveyed in late November 2018 to February 2019. The red line indicates the kernel density estimate of actual reported hours in a typical week so far in the winter term, surveyed in March 2019.

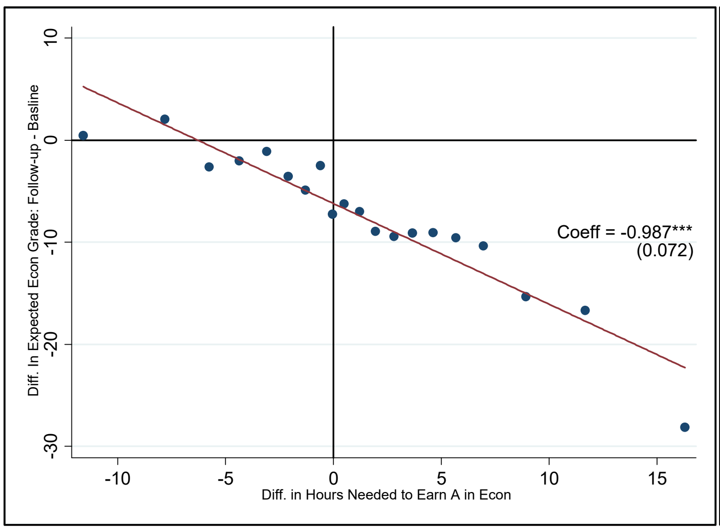
Figure 6: Study Time and Grade Expectation Revisions and Information Updating



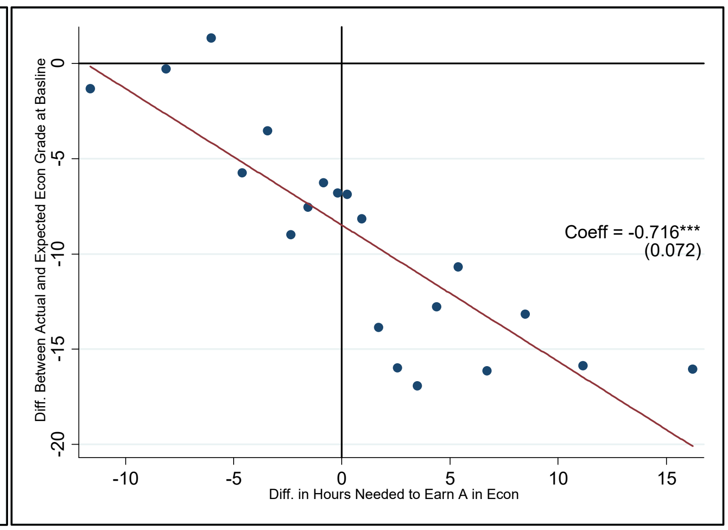
(a): Change in Econ Study Hours vs. Change in Hours for A



(b): Change in All Study Hours vs. Change in Hours for A



(c): Change in Econ Grade Expectation vs. Change in Hours for A



(d): Actual – Expected Econ Grade vs. Change in Hours for A

Notes: Panels (a) and (b) show the relationships between changes in students' study times and measures of changes in students' beliefs about their academic abilities. Panels (c) and (d) show the relationships between changes in students' expected and realized economics grades and measures of changes in students' beliefs about their academic abilities. Each binned scatter plot is created by first grouping students into 20 equal-width bins (vingtiles) in the distribution of the variable on the x-axis and calculating the mean of both the y- and x-axis variables within each bin. The circles represent these means, while the lines represent the associated linear fit from the underlying student-level data.