NBER WORKING PAPER SERIES

IMPROVING NON-ACADEMIC STUDENT OUTCOMES USING ONLINE AND TEXT-MESSAGE COACHING

Philip Oreopoulos
Uros Petronijevic
Christine Logel
Graham Beattie

Working Paper 24992
http://www.nber.org/papers/w24992

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2018

We are indebted to the first-year economics instructors at the University of Toronto for their willingness to incorporate an experiment into their courses for a third consecutive year. We especially thank Aaron de Mello, our web developer, for his tireless commitment to designing and perfecting the experiment's website, as well as for his help with organizing and extracting the experimental data. Dana Britton, Holly Fenton, Ayush Gupta, Aamir Husain, Anthony Koundourakis-Soares, Julianna Lu, Sukhi Singh, and Kadachi Ye showed great enthusiasm and professionalism in their role as coaches. Seminar participants at the Canadian Institute for Advanced Research provided useful feedback. We also thank participants at the SOLE 2018 conference in Toronto for helpful comments and suggestions. All remaining errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Improving Non-Academic Student Outcomes Using Online and Text-Message Coaching
Philip Oreopoulos, Uros Petronijevic, Christine Logel, and Graham Beattie
NBER Working Paper No. 24992
September 2018
JEL No. I20,J24

ABSTRACT

We design and experimentally evaluate two low-cost, scalable interventions – an online preparatory module and a text-message coaching program – in a sample of over 3,000 undergraduate students at a large Canadian university. Supplementing administrative data on academic outcomes with a unique follow-up survey on student well-being and study habits, we estimate positive program effects on students' non-academic outcomes, despite estimating null effects on course grades and credit accumulation. Given the low costs associated with administering these programs, our results suggest that the positive impacts on student experiences may warrant program expansion even in the absence of impacts on academic outcomes.

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

Christine Logel
Renison University College
240 Westmount Road North
Waterloo, ON N2L 3G4
Canada
clogel@uwaterloo.ca

Graham Beattie
Loyola Marymount University
University Hall 4218
1 LMU Drive
Los Angeles, CA 90045
US
grb57@pitt.edu

A data appendix is available at http://www.nber.org/data-appendix/w24992
A randomized controlled trials registry entry is available at AEARCTR-0000810

1. Introduction

Designing and evaluating student support programs remains a priority on the public policy agenda in higher education, given a pervasive sense that many students underperform and acquire limited skills throughout post-secondary education (Arum and Roksa 2011). This sentiment is reflected in a growing literature that points to the promise of comprehensive support programs that bundle financial aid, coaching, tutoring, and group activities to induce behavioral change among students. Such programs have caused significant improvement in student grades and graduation rates, proving effective at both two-year (Scrivener and Weiss 2013; Evans et al. 2017) and four-year colleges (Bettinger and Baker 2014; Andrews, Lovenheim, and Imberman 2016; Page, Keho, Caslteman, and Sahadewo, forthcoming). Although these programs tend to improve student outcomes, they are often expensive and difficult to scale to large populations.

To expand student support in a cost-effective way, researchers and policymakers are increasingly turning toward briefer behavioral interventions that can be delivered to thousands of students simultaneously at low cost. These programs aim to improve academic outcomes by providing students with helpful information and encouraging them to focus more on the present, avoid relying on unsuccessful routines, and associate less with negative identities (Lavecchia, Liu, and Oreopoulos 2016, Damgaard and Nielsen 2018). Behavioral interventions are often built around timely information provision (Castleman and Page 2015, Castleman and Page 2016, Castleman and Meyer 2016; Oreopoulos and Petronijevic 2018), financial incentives (Angrist, Lang, and Oreopoulos 2009; Angrist, Oreopoulos, Williams 2014), personal assistance and coaching (Bettinger, Long, Oreopoulos, Sanbonmatsu 2012; Oreopoulos and Petronijevic 2018), and attempts to help students develop more adaptive mindsets (Yeager et al. 2016; Bettinger et al. 2018). Their efficacy often depends on whether the target outcome involves immediate concrete action (e.g., completing an application) or one that involves more continuous effort (e.g., GPA). For example, interventions that offer personal assistance or provide timely information have been shown to successfully affect relatively simple (often one-time) actions, such as completing a financial aid application (Bettinger et al. 2012), enrolling in college (Castleman and Page 2015), or renewing financial aid once enrolled (Castleman and Page 2016). In contrast, such interventions have been less successful at affecting outcomes such as grades or credit accumulation, which typically require continuous and sustained behavioral change from students (Angrist, Lang, and Oreopoulos 2009; Castleman and Meyer 2016; Oreopoulos and Petronijevic 2018).

While a large and still-growing literature finds comparatively small or null effects on academic performance, less attention has been devoted to investigating the effects on non-academic outcomes, such as mental health, school engagement, participation in extracurricular activities, and students' sense of belonging and support. These outcomes have important consequential effects on overall well-being, both during a program and beyond. A national survey of American college students, for example, found that fully 43% of students reported having felt very lonely in the previous 30 days and 42% reported having felt overwhelming anxiety (American College Health Association, 2018). Even among students with similar grades and persistence, those who endure greater levels of stress and depression may suffer lower levels of immediate and long-term utility. Non-academic outcomes are also becoming increasingly important as awareness of mental health problems and care about student experience increases.

Behavioral interventions may therefore make students feel happier, more engaged, or more supported, even if they fail to nudge them towards academic improvement. For example, new technologies like text messaging offer new means to converse with students conveniently and quickly at home or on weekends. More detailed and personalized data can be used to flag those struggling and reach out to offer help. These new methods for communicating offer fascinating possibilities for improving student experience and overall well-being.

We explore these issues in this paper, focusing on two behavioral interventions – a psychologically-informed and personalized online module and a text-message coaching program – and demonstrating that non-academic outcomes are a potentially important metric by which to evaluate program effectiveness. We design and experimentally evaluate the treatments in a sample of over 3,000 undergraduate students at the University of Toronto (U of T) during the 2016-17 academic year and supplement administrative grades data with a unique (and mandatory) follow-up survey designed to capture important non-academic outcomes.

The first treatment requires students to complete a 60-minute, one-time, online module during the first two weeks of class.[1] We designed the module based on the notion that a potential barrier to the effectiveness of many large-scale interventions is their inability to target the broad range of students' diverse academic needs. For example, some students benefit from interventions focused on study skills, while others have effective study skills and benefit more from interventions addressing low motivation. Other students have both skills and motivation but are hampered by stress in their personal lives, while students from historically underrepresented groups may underperform because of the pressure from contending with negative stereotypes (see Steele, 1997). To address the underlying heterogeneity among students, we incorporate a novel design that lets students personalize their experience according to their own academic needs through an online module we named 'Choose Your Own Challenge,' or CYOC.

[1] Interactive online modules are an increasingly common mode of delivering intervention materials, used, for example, in social psychology research on adaptive mindsets (Yeager and Walton 2011; Walton 2014; Bettinger et al. 2018), research on goal-setting interventions (Dobronyi, Oreopoulos, and Petronijevic 2017; Clark, Gill, Prowse, and Rush 2017), and research on timely information provision to graduating high school students (Oreopoulos and Ford 2016; Hastings, Neilson, and Zimmerman 2018).

The Choose Your Own Challenge (CYOC) treatment module teaches students helpful academic behaviors while guiding them to reflect on, and then overcome, behavioral and psychological barriers to implementing those behaviors. CYOC may also benefit future students by helping the university better understand student transitions. Part One of the module presents students with six broad factors critical to academic success,[2] with subsequent sections elaborating on each factor and taking students through tasks that draw on psychological 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.

The second treatment begins with both parts of the CYOC module and then enhances it with a text-message coaching program drawing on recent work in the economics of education (Castleman and Meyer 2016; Oreopoulos and Petronijevic 2018). Treated students are matched with senior undergraduate student coaches who offer advice and consultation via text message throughout the academic year. Students assigned to the online CYOC condition completed only that module, whereas students assigned to the enhanced text-message condition completed the same online module and were additionally invited to join the coaching program.[3]

[2] These include studying enough, studying effectively, seeking help, attending class, staying motivated, being patient and taking a long-term perspective.
[3] As we describe below, students in the control group were given a personality test measuring their relative ranking on each of the Big Five personality traits.

With respect to academic outcomes, our results indicate that neither treatment is effective at improving student grades or credit accumulation, both in the full sample and across student subgroups. Taking advantage of a unique follow-up survey that was conducted at the end of the fall semester, we then investigate treatment effects on non-academic outcomes. We follow the method of Kling, Leibman, and Katz (2007) to aggregate student responses to the survey into two main indices. The 'core well-being' index includes life satisfaction, belonging, confidence, and depression (reverse-coded) and the 'success strategies' index includes measures of study strategies and help-seeking, such as time management and frequency of meeting with instructors. We estimate that the enhanced text-message coaching treatment improves the core well-being index by 4 percent of a standard deviation and the success strategies index by 5 percent of a standard deviation, on average. A more disaggregated analysis shows that the effects on these non-academic outcomes are driven by treated students experiencing a greater sense of belonging and support and seeking help more often from course instructors and tutors. Using student responses to direct questions about the coaching program, we also show that the majority of treated students report feeling supported by their coaches, appreciating the messages they receive from their coaches, and having a better experience at U of T because of their coaches.

The effects of the online CYOC module on its own trend in a positive direction on both indices, at approximately 3 percent of a standard deviation, but are not statistically differentiable from zero. However, our experimental design does not let us tease apart the degree to which this online module contributed to the significant effects in the enhanced text-message coaching condition, as the effects of the CYOC module alone and of the module enhanced with text-message coaching are not statistically differentiable from each other. Given that our design does not include a condition testing the text-message coaching treatment on its own, we cannot determine whether treatment effects are additive, complementary, or entirely driven by the coaching treatment.

To our knowledge, our paper is the first to provide evidence for the effects of a brief behavioral intervention on non-academic outcomes in an education-based context.[4] Our findings suggest that such interventions can improve student experiences in college, despite not causing discernable improvement in course grades. Although the effects we estimate on student well-being and success strategies are modest, both the CYOC module and the text-message coaching intervention are relatively cheap to implement, suggesting that such programs are potentially worth administering even when effects on grades or credit accumulation are not present. The cost of administering the CYOC module together with the text-message coaching intervention is approximately $12 per student when factoring in setup costs and $2 per student when only considering the cost of sending and receiving text messages throughout the academic year. Our findings that treated students felt supported by their coaches, appreciated the messages they received, and thought coaches were in part responsible for their improved experience, suggest that these types of program can be a cost-effective way for colleges to improve student experience and support.

Our paper also contributes to the growing literature on text-message interventions in education contexts. These programs push helpful suggestions to students or their parents, with content focused on tips for learning, studying, or navigating through the environment at a given institution. They appear more effective when communicating with parents, as many studies estimate positive effects on student achievement from providing information to parents about the

[4] For reviews on the literature of behavioral interventions in education, see Lavecchia, Liu, and Oreopoulos (2016) and Damgaard and Nielsen (2018).

behavior and grades of their children and effective ways to help their children learn (Kraft and Dougherty 2013; Kraft and Rogers 2014; Mayer et al. 2015; Bergman 2017; Bergman and Chan 2017; Kraft and Monti-Nussbaum 2017). The evidence on programs that communicate with students directly is more nuanced. Sending messages to students with the goal of nudging them toward taking one-time, relatively simple actions has proven effective – for example, such interventions have been shown to increase the likelihood of students enrolling in college (Castleman and Page 2015) and renewing financial aid (Castleman and Page 2016) once enrolled. But texting campaigns have been relatively ineffective at changing outcomes for which improvement requires sustained changes in student behavior, such as test scores, course grades, or overall GPA (Fryer 2016; Castleman and Meyer 2016; Oreopoulos and Petronijevic 2018). Our estimated null effects on student academic outcomes are consistent with this finding. The treatment does improve non-academic outcomes, however, suggesting that future research on these interventions should consider and attempt to estimate broader treatment effects in order to obtain a more complete program evaluation.

The remainder of this paper is organized as follows. The next section provides a detailed description of our interventions and their implementation. Section 3 describes the experimental data and outlines our empirical strategy for estimating the treatment effects. Section 4 presents the results, while Section 5 provides concluding remarks.

2. Description of the Intervention

The setup of the interventions in our study is similar to the one used in Oreopoulos and Petronijevic (2018). Throughout the 2016-17 academic year, we ran an experiment at the main campus of the University of Toronto (St. George). We partnered with all first-year introductory economics instructors to make completion of our online 'warm-up' exercise worth 2 percent of students' final course grades. Students had to complete the exercise within the first two weeks of the fall semester to receive course credit. The type of exercise each student completed depended on whether he or she was randomly sorted to one of two treatment groups or the control group. We then administered a follow-up survey to all students in the final two weeks of the fall semester, approximately 12 weeks after the intervention exercise. The survey solicited students' feelings about non-grades outcomes, such as life satisfaction, feelings of support and belonging, and self-reported study habits. Completion of the follow-up survey was also worth 1 percent of students' final grade in their economics courses.

All students started the online warm-up exercise by creating an account and completing the same short introductory survey, in which they responded to background questions about their parents' education, their own expected educational attainment, first-year and international status, work and study plans, and tendencies to procrastinate or become distracted. After completion of the initial survey, students were randomly sorted to either a treatment group or the control group. Students sorted to the CYOC treatment group completed an online module. Students sorted to the text-message coaching treatment group also completed the online module but were additionally offered the opportunity to provide a cell phone number and participate in a text-message coaching program. Students assigned to the control group were given a personality test. We now describe the treatment and control modules, as well as the follow-up survey, in more detail.

2.1. Treatment One: Choose Your Own Challenge (CYOC) Online Module

We conceptualized the CYOC module based on the premise that students succeed and maintain well-being at least in part because of effective academic behaviors and adaptive perspectives. Teaching students effective academic behaviors is necessary but not sufficient for their success, as people can know effective behaviors but fail to follow through with them in nearly every domain.

Students also face a diverse array of barriers to implementing and following through with effective behaviors. These barriers include low motivation, personal life stress, and common identity threat concerns (i.e., concerns that one's social identities, such as race, gender, or socioeconomic status, could be devalued in a given context). Moreover, taking adaptive perspectives requires not just learning what those perspectives might be, but having the opportunity to reflect on them and incorporate them into one's way of thinking about the world.

As such, the CYOC intervention aimed to teach students effective behaviors and perspectives, increase their likelihood of following through on the behaviors and taking on the perspectives as their own, and address the diverse array of barriers to success, all while being cost-effective and implementable to large numbers of students. Consisting of two parts, the entire CYOC online module was designed to take 60 minutes to complete. An information page emphasized that the purpose of the exercise was to allow U of T and the economics department to learn more about students' perceptions of the transition to university, with the intent of later using this information to create helpful resources for future students. Stipulating that the information students provided would be used to help future generations follows the format of most studies on the belonging mindset (see, for example, Walton, Logel, et al. 2015). To underscore this framing, most segments included items asking students about the degree to which most students already understand each concept, or which one of a set of related statements they think is most relevant to most students.

In part one, students were asked to think about their own future and education, which we explained would help U of T better understand how students form strategies for achieving their goals.[5] An initial page listed six broad strategies critical to academic success: studying enough,

[5] There is a growing literature devoted to understanding whether encouraging students to focus on their goals can improve outcomes. See, for example, Dobronyi, Oreopoulos, and Petronijevic (2017) and Clark, Gill, Prowse, and Rush (2017).

studying effectively, getting help when you do not understand, keeping up and going to class, staying motivated, and being patient and taking a long-term perspective. Later subsections elaborated on each strategy, using elements designed to educate participants about the factor, change their attitude towards it, and then improve the intention-behavior link. For example, the subsection on ‘studying enough’ started with evidence of its efficacy (e.g., noting the strong relationship between studying and doing well) and provided detailed examples (e.g., finding at least 20 hours a week; treating it as a full-time job). Next, it used an ‘implementation intentions’ writing task, shown to increase the likelihood of behavior change (e.g., Gollwitzer and Schaal, 1998), in which students created a plan for studying enough and committed to it. Part one of the CYOC module also incorporated visualization techniques (e.g., “Let your mind imagine details of your environment” and “Where are you sitting? What does the desk look like?”).

Part two of the CYOC module was designed to address either barriers to carrying out the behaviors and perspective changes from part one or barriers that can remain even when those behaviors and perspectives are adopted (e.g., low motivation, personal life stress, and psychological threats experienced by students whose social identities are underrepresented or negatively stereotyped academically). It used a ‘self-persuasion’ design, in which students meaningfully engage to develop the material rather than passively reading through pre-determined content (see Canning and Harackiewicz 2015). To ensure that due consideration was given to each component of the module, we placed varying minimum word-count and time restrictions on several pages of the exercise.

The activity was framed in the context of attributing academic struggles to changeable situation factors rather than to unchangeable personal factors. To connect it to students’ identities, while keeping it focused on helping future students, the introductory page described how U of T only accepts students whose records show that they have the motivation, background knowledge, and skills to succeed, so each student is capable of doing well academically. It then presented a set of situations that could interfere with academic success, ostensibly for the purpose of getting students' advice about which problems are most common and how to solve them.

To prompt participants to connect this task to their own social identities, instructions describe how U of T accepts more than 15,000 students each year, so it is highly likely that the following year there will be at least some students with the same ethnic, religious, national background, age, and many of the same strengths and struggles as the participant. The module then presented a series of situational barriers to success, including those that could be related to social identity threat (e.g., "feeling that maybe 'people like them' are not especially welcome at U of T" and "Wondering if they just don't have what it takes to do well academically at U of T"), significant life challenges (e.g., "dealing with a great deal of personal stress and challenges along with classes"), and the academic success factors identified in part one (e.g., "Feeling unmotivated to devote time and energy to doing well in university" and "waiting too long to seek out help when class concepts are unclear"). Participants were asked to choose the two most important barriers for 'people like them.'

The survey program routed each student to content specific to the two barriers they chose. For each of those two barriers, students were provided with four potential reasons for that barrier and asked to choose one and write about why students might struggle with this problem. The four reasons focused on changeable situational factors and some included a subtle message to change students' attributions. For example, for the barrier of "feeling that maybe people like them were not especially welcome at U of T", the four reasons were designed to convey a message that belonging concerns are common (e.g., "Thinking they are the only ones wondering if they belong" and "Seeing other students and thinking that they seem to be completely comfortable at U of T").

Participants were then given a list of four solutions and asked to choose one and elaborate on it. In the above example, solutions include “giving it time – realizing that, in time, most students come to feel that they do belong at U of T.” Psychological interventions that convey to students experiencing identity threat that belonging concerns are common and pass with time have been shown to reduce belonging uncertainty and improve academic retention and success (see Walton and Cohen 2011).

Upon completion of the entire online module, students were emailed a printable poster of the tips for success presented in part one. Full documentation for the online exercise is available in Appendix A.

2.2. Treatment Two: Online Exercise with Follow-Up Text-Message Coaching

A second group of students was randomly offered the opportunity to participate in a text-message coaching program. These students also completed the CYOC module described above but, upon completion, were asked to provide their phone numbers to participate in a text-message coaching program. Branded *You@UofT*,[6] the coaching program was active throughout both the fall semester of 2016 and the winter semester of 2017. The experiment featured a total of 10 coaches, each being assigned between 70 and 185 students. The coaching team consisted of 8 senior undergraduate students and two of the paper’s coauthors, Oreopoulos and Petronijevic, making a team of 10 coaches. In the description of the data below, we provide details on the precise number of students assigned to each coach.

By assigning students to individual coaches, we attempted to combine the most promising features of the text-messaging and in-person coaching treatments that are evaluated in Oreopoulos

[6] As in Oreopoulos and Petronijevic (2018), we chose the name to emphasize that the program would help coach students toward their individual definitions of success.

and Petronijevic (2018). In that paper, we evaluated a mass two-way text-messaging campaign that facilitated communication with a large sample of students at low cost but was ineffective at improving outcomes. Students who participated in that text-messaging campaign were not assigned to individual coaches; instead, they were sent mass texts every week in which we invited them to share concerns and ask for help. In contrast, an in-person coaching treatment did improve academic outcomes, likely because coaches could proactively initiate discussions with students about challenges and establish relationships based on trust. The text-message coaching treatment in this paper is designed to continue to reach a large sample of students at low cost while integrating some of the features that made the in-person coaching program successful.

To that end, eight undergraduate students with previous student support experience were recruited to act as coaches.[7] Based on our results in Oreopoulos and Petronijevic (2018), coaches were trained to message their students regularly and to gently encourage those students to discuss the challenges they faced navigating through university. A web-based coaching platform we designed provided coaches with the ability to make notes about each individual student, allowing them to easily recall recently discussed topics with each student and follow-up regularly about specific issues. Coaches could also program batch messages to be sent at specific times of the day, further specifying subgroups of students to which each batch would be sent. Students could be differentiated based on international or domestic status, first-year or non-first year status, and incoming high-school grades. Coaches were also able categorize students into three distinct categories (red, yellow, and green), which indicated the degree of help or attention the coach deemed each student required.

[7] Most coaches participated through enrolling in a research opportunity course and received course credit rather than payment. Two coaches were not eligible to take the course and received a stipend of about $2,000.

Students who were randomly sorted to the text-message coaching treatment were offered the opportunity to enroll and had to make an active opt-in decision if they wanted to participate. Approximately 90 percent of students chose to opt-in, with less than 3 percent later choosing to opt out. As mentioned, 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 pre-programed batch messages designed to stimulate conversation. Coaches were also encouraged to follow-up with individual students on the specific issues they had recently discussed to make sure that students were effectively progressing.[8] Once contact was established, conversations evolved organically, with coaches usually trying to determine how students were progressing throughout university, both academically and emotionally.[9] Appendix B provides categorized summaries of the different types of message that coaches sent.

2.3. Control Group: Personality Test

As in Oreopoulos and Petronijevic (2018), students who were assigned to the control group at both campuses were given a personality test measuring the Big Five personality traits. The test could be completed in 45 to 60 minutes, and students were emailed a report describing their scores on each trait upon completion of the exercise. Beattie, Laliberté, and Oreopoulos (2018) use the data resulting from the personality test to explore non-academic predictors of performance in university. The appendix of that paper provides a full description of the personality test.

2.4. Follow-Up Survey

[8] Topics of conversation were also sometimes dictated by the events currently unfolding at the university. During midterm and exam periods, for example, coaches tended to guide conversations toward making study schedules and effective study strategies.

[9] Coaches did not act as tutors for course material; instead, they mentored students on effective study strategies, how to learn from past mistakes, and how to seek out campus resources when extra help was required.

During the last two weeks of the fall semester, approximately 12 weeks after the initial treatments, all students were again required to log into their accounts and complete a 15-minute follow-up survey for an additional 1 percent of their course grade. We used the follow-up survey to solicit students' answers to questions about non-academic outcomes, such as life satisfaction, confidence, feelings of belonging, and study habits, such as study strategies and the frequency with which they sought help from instructors, tutors, and advisors. Most questions could be answered using dropdown menus of prepopulated choices, but we also did ask all students to freely write responses to open-ended questions about their biggest challenges to academic success, how the university could better help them, and how they could better help themselves. Students who were in the text message coaching treatment were also asked about their satisfaction with the coaching program and open-ended questions about the how helpfulness of their coaches.

Beattie, Laliberté, Michaud-Leclrec, and Oreopoulos (2017) use the data from this follow-up survey to descriptively explore behavioral and mental health differences between high- and low-achieving university students. The follow-up survey is documented in full in the appendix of that paper. In Section 3.2 below, we discuss how we categorize the variables from this survey to use them in our analysis of treatment effects on non-grades outcomes.

3. Data Description and Empirical Strategy for Estimating Treatment Effects

In this section, we describe the experimental data and sample, the follow-up survey, and the interactions that took place between students and coaches in the text-message treatment. Having described the environment, we then outline our empirical strategy for estimating treatment effects on both academic and non-academic outcomes.

3.1. Experimental Randomization and Sample Description

We begin our data description by reporting the fractions of students sorted to treatment and control groups. Prior to the experiment, we intended to randomly sort one-third of the students to each group, based on the randomly-generated digits of their student numbers.[10] Table 1 shows that we successfully reached our randomization targets: the p-values for the tests of the difference between the realized and intended fraction are all well above conventional significance levels. Furthermore, the completion rates for both the personality test and the treatment modules are very high, each at 99 percent.

The full experimental sample consists of 3,395 students, from which 1,119 were assigned to the control group, 1,154 were assigned to only complete the online module, and 1,122 were assigned to the text-message coaching group. We could match 91 percent of the 3,395 students to the university's administrative data on grades on background characteristics, leaving us with a final analysis sample of 3,088 students for estimating treatment effects on course grades.[11]

Table 2 presents summary statistics and balancing tests. Across a rich set of student background variables – obtained from both our survey and the university's administrative data – almost no variables are statistically different, on average, between the control group and the two treatment groups. The only exceptions are that students assigned to complete only the online module are approximately 3.6 percentage points less likely to expect to earn an average grade of at least A- and to self-report checking their cell phones often, while students assigned to the text-

[10] Students provided their student numbers upon registering online for the experiment and had a strong incentive to provide the correct student number, as completion of the online exercise accounted for 2 percent of their final course grade.

[11] The matching rate is not significantly different across treatment and control groups. Regressing an indicator variable for whether a student cannot be matched to the grades data on treatment indicators results in coefficient estimates of 0.005 (se=0.012) and 0.019 (se=0.012) for the online only and the online with text message treatment groups, respectively.

message coaching treatment are 3.3 percentage points more likely to report expecting to work (for pay) more than 8 hours per week in the upcoming academic year. We show below that our main results are robust to controlling for these and many other background variables.

In terms of sample composition, half of our sample (55 percent) is female, and the average student starts university at 18.45 years of age and has an incoming high school average of 90.45 percent. Approximately 72 percent of our sample consists of first-year students, 45 percent speak English as their mother tongue, and 48 percent are international students. Approximately 43 percent of students live in residence and 29 percent of students are first generation (neither mother nor father attended university). Most students (68 percent) plan to earn more than a bachelor's degree and at least an A- average throughout their undergraduate studies, but the average student plans to study only 18.26 hours per week, less than the amount of time one would typically devote to a part-time job.

3.2. Follow-up Survey: Sample Description and Outcomes

We use twenty-four questions from the follow-up survey in our main analysis of treatment effects.[12] To address the issue of multiple hypothesis testing directly and to draw general conclusions about treatment effects on non-grades outcomes, we group student answers to these questions into two main indices. We construct each index using the method of Kling, Leibman and Katz (2007), by which we standardize a student's answer to each question relative to the control group mean and standard deviation and then take a simple (within-student) mean of the resulting standardized variables to construct the index. As shown in Table 3, we broadly define the two main

[12] Among the students with available grades data, 465 (15 percent) did not complete the follow-up survey at the end of the fall semester. The attrition rate, however, is not significantly different across treatment and control groups. Regressing an indicator variable for whether a student completed the follow-up survey on treatment indicators results in coefficient estimates of 0.016 (se=0.015) and 0.024 (se=0.016) for the online only and the online with text message treatment groups, respectively.

indices as the 'core well-being' index and the 'success strategies' index, with the former being an aggregation of 14 variables and latter being an aggregation of 10 variables. The last column of Table 3 lists the questions that contribute to the construction of each index.

These two indices are our main outcome variables from the follow-up survey. We provide a further disaggregation of the core well-being index by breaking it apart into four sub-indices, reflecting (i) overall satisfaction with life and the university, (ii) feelings of belonging at and support by U of T, (iii) confidence to succeed at U of T, and (iv) overall depression or stress. The success strategies index is also disaggregated into two sub-indices, reflecting (i) study strategies and (ii) help-seeking behavior. In our empirical analysis below, we report treatment effects on the two main indices and each of the sub-indices.[13]

3.3. Text Message Coaching Program and Data

Throughout the duration of the text-message coaching program, every text message sent by a coach or student was stored, allowing us to assemble a large dataset of the text dialogue between each student and his or her coach. To provide a fuller description of the text-message coaching program, we present summary statistics from this dataset in Table 4, focusing our discussion on the assignment of students to coaches and the resulting interactions that took place.

The average coach was assigned 100 students; however, there was wide variation in the number of students assigned to each coach: senior undergraduate coaches received between 70 to 95 students, while Oreopoulos and Petronijevic were assigned 185 students each. The variation in the number of students per coach occurred because we achieved a higher opt-in rate (90 percent) into the text-messaging program than we expected prior to launching the experiment. To avoiding

[13] Appendix C offers a completely disaggregated analysis of treatment effects on each of the twenty-four variables in Tables C1 and C2.

breaking the workload agreements that we made with senior undergraduate coaches, we allowed each coach to choose whether he or she would accept more than 70 students. Five coaches stopped at approximately 70 students, three coaches accepted 95 students, and the remaining students were assigned to Oreopoulos and Petronijevic.

Prior to the time we stopped sorting students to all coaches (when approximately 700 students had already completed the online exercise), students who were assigned to the text-message coaching treatment were also randomly assigned to a coach. After we stopped directing students to the coaches who did not wish to accept more than 70 students (and later, to coaches who did not wish to accept more than 95 students), the sorting of students to coaches was no longer random. Coaches who continued accepting students were less averse to a larger workload while the students who were being sorted to these coaches completed the warmup exercise relatively late. Using similar data from a prior experiment, Beattie, Laliberté, Oreopoulos (2018) show that these students tend to have a strong tendency to procrastinate and do not perform well in their courses. We address concerns over non-random sorting of students to coaches below by estimating specifications that control for three 'coach group' binary variables, each of which captures the sorting rule of students to coaches that applied during the (calendar) time when a student completed the online module.[14] Controlling for these binary variables creates conditional random assignment of students to coaches.

Among the 1,122 students who were randomly offered the coaching treatment, 114 opted not to provide a phone number and an additional 25 students asked to be removed from the program

[14] Students who completed during the initial time period were sorted to all coaches; students who completed during the second time period were sorted to Oreopoulos and Petronijevic and senior undergraduate coaches who were willing to accept between 70 and 95 students; and students who completed during the third time period were sorted only to Oreopoulos or Petronijevic.

during of the academic year, leaving us with a total of 983 students in our dataset of text message exchanges. A majority of students who were offered the coaching treatment (62 percent) sent a text message to their coach at least once. On average, students sent 17 text messages to their coaches in the fall semester and received a total 47 messages, with 32 of those messages being a batch message the coach sent to many students at once. Engagement fell in the winter semester, with students sending only 6 messages, on average, and receiving 23 from their coach, 18 of which were batch messages.

The average coach sent 1,521 non-batch messages and 3,122 batch messages to their students in the fall semester. If we do not double count batch messages that go out to multiple students at once, the average coach sent 47 unique batch messages in the fall semester. The decline in engagement in the winter semester is again reflected in the number of messages sent by coaches, as the average coach sent 482 non-batch messages, 1,734 batch messages, and 19 unique batch messages.

In Appendix B, we provide more detail about the batch text messages students received from their coaches, presenting summary statistics on text characteristics and categorized examples of select batch messages.

3.4. Empirical Strategy for Estimating Treatment Effects

Having successfully randomized students across treatment and control groups, we estimate the effects of the online-only and text-messaging treatments with a comparison of mean outcomes in a simple regression framework. The main specification we estimate is given by

$$y_i = \beta_0 + \beta_1 Online_i + \beta_2 Text_i + \rho' X_i + u_i, \tag{1}$$

where the outcome of student i is regressed on an indicator for the student being assigned to the online (CYOC) treatment only, an indicator for the student being assigned to complete the online module and receive text-message coaching and, in some specifications, additional student-level control variables.

The main parameters of interest are β_1 and β_2, the estimated average effects of the online module alone and online module enhanced with the text-message coaching. These parameters represent intent-to-treat effects, as students are included in the treatment groups if they are offered the *opportunity* to work through the online module or provide a cell phone number (based on the randomly-generated digits of their student numbers) without necessarily having to complete the module or opt-in to texting messaging. Given that our completion rates and opt-in rate are quite high, these estimates are likely close to the average treatment effect.

Our main outcomes of interest are course grades, overall grade point average (GPA), the number of credits earned, and the number of credits failed. In terms of non-grades outcomes, we explore treatment effects on our core well-being and success strategies indices as well as the effects on each of their sub-indices (described above). When the outcome of interest is course grades, we stack all course grades and run a regression at the student-course level. In these cases, we cluster standard errors at the student level. The effects on all other outcomes are estimated with regressions at the student level and robust standard errors are reported.

4. Experimental Results

In this section, we present the estimated effects of the CYOC condition and the enhanced text-messaging condition, followed by a series of robustness checks and an exploration of heterogeneous treatment effects.

4.1. Grades Outcomes

In Table 5, we estimate treatment effects on course grades using student-course level regressions and clustering standard errors at the student level. In columns (1) and (2), we report the estimated treatment effects from regressions without and with additional control variables. Importantly, the set of control variables includes ‘coach group’ binary variables for the date the student completed the online module, which preserves random assignment of students to coaches.[15] Treatment effects for the online module and text-messaging program across all course grades are small and statistically insignificant in both cases. Columns (3) and (4) show estimated effects on fall semester (September 2016 - December 2016) courses, while columns (6) and (7) show effects on winter semester (January 2017 – April 2017) courses.[16] Treatment effects are again statistically and economically insignificant in both cases. In the last columns, (8) and (9), the dependent variable is students’ first-year economics course grades, the class in which they completed the online module. Here, the treatment effects are larger than those estimated across all courses, but they remain small and insignificant.

In Table 6, we further present estimated treatment effects on student GPA and credit accumulation. Neither the online module nor the text-messaging treatment affected student GPA or credit accumulation, as we find precisely estimated null effects across all courses taken throughout the academic year, fall-semester courses, and winter-semester courses. Overall, the

[15] Additional control variables include student age, self-reported expected study hours per week, expected paid-work hours per week, expected average grade, tendency to finish what he or she starts, tendency to get discouraged by setbacks, tendency to study at the last minute, and indicator variables for first-year status, international student status, first-generation status, gender, English mother-tongue status, whether the student lives on residence, a self-reported desire to earn more than an undergraduate degree, and a self-reported tendency to check his or her cell phone often.

[16] Treatment effects in columns (1) and (2) were estimated using all available courses, which include those from the fall semester, those from the winter semester, and those that run across both semesters and conclude in the winter.

evidence clearly demonstrates that the online-module and text-message coaching treatment are ineffective at improving student grades and credit accumulation.

4.2. Non-Grades Outcomes

Having shown that neither treatment affected student grades, we now explore treatment effects on the indices created from our follow-up survey. Recall that the questions used to construct each index are reported in Table 3.

In Table 7, we report treatment effects on the core well-being index and its four sub-indices. The estimates in columns (1) and (2) show that students in the coaching treatment scored 4 percent of a standard deviation higher on the aggregate core well-being index. Breaking the index apart into four sub-indices, the treatment effect is driven by the effect on students' sense of belonging and support, as treated students score 5.5 percent of a standard deviation higher on the belonging and support index. Treatment effects on the three other sub-indices are also positive, although they are not statistically distinguishable from zero.

Although the estimated treatment effect of the online module on the core well-being index is not statistically different from zero, it also not statistically different from the effect of the text-message coaching treatment, as indicted by the p-values pertaining to the tests for treatment differences. Comparing the effects of the online module and the text-message coaching treatments on the sub-indices of the core well-being index indicates that, despite it not having an overall effect on the core well-being index, the point estimates for the effects of the online module are sometimes larger than the point estimates for the text-messaging treatment. In particular, students who only completed the online module reported higher overall satisfaction with life and the university and indicated great confidence to succeed in university. Taken together, we cannot rule out that the significant effect for the text-message coaching treatment is driven, at least in part, by students having completed the CYOC module.

For completeness and to further investigate mechanisms, we report treatment effects on each of the 14 variables used to construct the core well-being index in Appendix Table C1. The estimates imply that the effects of the text-message coaching program operate by causing students to feel more like being a student at U of T is a large part of their identity, like U of T wants them to succeed, and like they know where to get advice at U of T. Treatment effects on these variables range between 9.4 and 12.5 percent of a standard deviation and are statistically significant. Students in the text-message coaching treatment also felt more supported by U of T and were less likely to report having a tough transition to the university, although these effects are not statistically significant at conventional levels. Overall, the evidence is consistent with the text-message coaching program causing students to form stronger ties with the university and feel like they are better equipped to navigate the challenges in their environment.[17]

We now turn to treatment effects on the success strategies index, which are reported in Table 8. Students who were in the text-message coaching treatment scored 5 percent of a standard deviation higher on the aggregate success strategies index. Breaking the index apart into its two sub-indices in columns (3) to (6) shows that the effect is mainly driven by the help-seeking index, although the point estimates of the effects on the strategies index are approximately the same magnitude but estimated less precisely. Once again, we find that the effects of the text-message treatment and the online module are not statistically different from each other, despite the latter not having a statistically significant effect on student outcomes. The evidence is therefore

[17] As is the case for the main indices, despite the treatment effects for the online module not being statistically different from zero, they are also not statistically different from the effects of the coaching treatment, suggesting that the students in the coaching treatment may have benefited from completing the online module.

consistent with at least part of the text-message coaching treatment effect being driven by students having completion of the CYOC module, although we cannot determine whether treatment effects are additive, complementary, or driven entirely by the text-message coaching.

In Appendix Table C2, we break apart the treatment effects on the success strategies index and its two sub-indices by estimating treatment effects on each of the 10 variables used to construct these indices. The effect of the online module combined with the text-message coaching program on the aggregate indices is driven mainly by students meeting with instructors and free tutors more often. Relative to the control group mean, the text-message coaching treatment increased the number of times students met with instructors by 16.5 percent and the number of times they met with free tutors by 17 percent.

Given the evidence that the text-message program resulted in students meeting more often with tutors and instructors, it may appear somewhat surprising that we estimate no treatment effect on grades. There are two ways to reconcile these results. First, although the treatment increased the frequency of student meetings with instructors by 16.5 percent, students in the control group only met with an instructor 1.2 times, on average, during the fall semester. Therefore, despite visiting instructors more often, treated students still met with instructors relatively few times. Second, although we do not report these estimates here, we used the follow-up survey to estimate treatment effects on students' (self-reported) independent study hours outside of the classroom. The text-message coaching treatment did not cause students to spend more time devoted to independent weekly study throughout the fall semester.[18] Because students in the text-messaging

[18] These results are available upon request.

treatment did not study more on their own and visited instructors relatively little, it is perhaps not surprising that we find no treatment effect on course grades.

The evidence presented thus far shows that the text-messaging treatment caused students to feel a greater sense of belonging and support at university and to seek help more often. Students' answers to follow-up questions about the text-message coaching program demonstrate that the majority of students enjoyed the program and felt that their experience in first semester was better at least in part because of the program. During the follow-up survey, treated students were asked to express the extent to which they agree with the following statements: "I feel my coach supports me;" "I appreciate receiving messages from my coach;" and "I am doing better at U of T in part because of my coach." Figure1 shows the percentage of respondents who selected each of the possible categories for each statement.

Nearly 60 percent of students agree or strongly agree that their coach supports them, while 85 percent of students at least somewhat agree with that statement. Similarly, 54 percent of students agree or strongly agree with the statement that they appreciate receiving message from their coaches, while 79 percent of students at least somewhat agree. Relative to the first two statements, student support for the statement that they are doing better at U of T because of their coaches is a little lower, as only 20 percent of students agree or strongly agree with this statement, but 56 percent of students at least somewhat agree. Fewer students feeling that their coaches are at least partly responsible for them doing better at U of T than the number of students who appreciate their coach's messages or feel that their coach supports them is perhaps unsurprising, given that the coaching program had no effect on student grades and that many students likely interpreted the question as referring to doing better with respect to grades. It is clear, however, that nearly all students (80 percent) felt supported by their coaches and appreciated receiving text messages from them.

Overall, the analysis of treatment effects on non-grades outcomes suggests that the text-message coaching treatment caused students to feel a greater sense of belonging and support at U of T and encouraged them to seek help from course instructors slightly more often than students in the control group. Descriptive evidence from the follow-up survey also demonstrates that nearly all students felt supported by their coaches and appreciated receiving text messages from them. Despite the online module not having a statistically significant impact on non-grades outcomes, the evidence is consistent with at least part of the text-message coaching treatment effects being driven by students completing the online exercise.

4.3. Robustness Checks

In this subsection, we argue for the integrity of our experimental design and show that our results are robust to coach heterogeneity.

As in any field experiment, the validity of the results hinges on the control and treatment groups being balanced. As discussed above, the estimates in Table 2 show no significant differences between pre-determined student variables across the control and treatment groups, except for students in the text-message treatment group being slightly more likely to report expecting to work more than eight hours per week (for pay) during the upcoming academic year and students in the online-only group being slightly less likely to expect to earn at least an A-overall grade and to check their cell phones often. In Tables 5, 7, and 8 above, we demonstrated that our results are robust to controlling for these and many other pre-determined variables.

We also discussed above that students assigned to the text-message coaching treatment were initially sorted to all coaches randomly but that, after approximately 700 students were in the coaching program, we only kept sorting students to five coaches and then subsequently only kept sorting students to two coaches. Although this introduced non-random sorting of students to coaches (but still preserved random assignment to treatment), we can generate conditional random assignment to coaches by controlling for 'coach group' binary variables that capture the time a student completed the online module and was sorted to a coach. As shown above in Tables 5, 7, 8, our experimental results are robust to including these additional controls.

Apart from ensuring the treatment and control groups are balanced on pre-determined variables, one may also be concerned about the importance of coach heterogeneity. Specifically, with only 10 coaches, if coaches vary considerably in their quality, one may worry that the estimated treatment effects are quite sensitive to the presence of particular coaches. We address this concern by re-estimating many treatment effects for several outcomes of interest, with each estimated effect being obtained from a specification in which we drop the students of a given coach from the regression. The results from this exercise are presented in Appendix Table C3 for student grades (across all courses), the core well-being index, the belonging and support sub-index, the success strategies index, and the help-seeking sub-index. Each cell contains a treatment effect estimated from a separate regression and the column numbers indicate the number corresponding to the coach whose students we drop from the regression.

Regardless of which coach is dropped, and consistent with our main results above, the estimates in Appendix Table C3 indicate that the treatment was ineffective at improving student grades but was effective at improving students' non-grade outcomes. Further, all ten point estimates for a given outcome are very similar in magnitude and are not statistically differentiable from the main point estimates in Tables 5, 7, and 8. We therefore conclude that the estimated treatment effects are not driven by the effectiveness of any one particular coach.

In sum, there is robust evidence that the text-message coaching intervention did not improve student grades but it did modestly improve students' sense of belonging and support and caused students to seek out help more often from instructors and tutors.

4.4. Heterogeneous Treatment Effects

Having established that the text-message coaching intervention caused modest improvements in students' non-grades outcomes, we now explore potentially heterogeneous treatment effects across student subgroups. We focus attention on subgroups defined by gender, first-year student status, international student status, and whether the student reports (in the pre-randomization survey) experiencing below or above median difficulty in transitioning to university (on a 7-point scale). Potentially differential effects by gender are often of interest in education interventions. We further chose to explore the other three student subgroups because they each partition students into groups that differ in their familiarity or comfort with university and life in Toronto and both of our treatments are designed to help students form study strategies and adjust to university.

The panels of Table 9 report treatment effects across student subgroups on grades and non-grades outcomes. Turning first to heterogeneous treatment effects on course grades in panel (a), neither treatment was effective at improving student grades in any subgroup, except for the text-message coaching treatment having a positive effect among non-first-year students. Non-first-year students comprise only 25 percent of our sample. Given the lack of an effect on grades in the overall sample and in any other student subgroup, we believe it is likely that this result is due to chance.

Focusing next on non-grades outcomes in panels (b) to (e), text-message coaching treatment effects are larger for female students than for male students across all non-grades outcomes, but the point estimates are not statistically differentiable in any case. Among first-year and non-first-year students, treatment effects on the core well-being index and its belonging and support sub-index are larger among non-first year students but treatment effects on the success strategies index and its help-seeking sub-index are larger for first-year students. In neither case are the point estimates statistically differentiable. A similar theme emerges for the international-domestic student comparison, as treatment effects on well-being are larger for domestic students but treatment effects on success strategies are larger for international students. Again, however, the point estimates are usually not statistically differentiable, except for the effects on the main core well-being index. Splitting students by self-reported difficulty in transitioning to university reveals slightly larger treatment effects on the core well-being index and its belonging and support sub-index for students with less difficulty transitioning. Treatment effects on the success strategies index are larger for students who report having above-median transition difficulty, while treatment effects on the help-seeking sub-index are similar across both groups.

Although treatment effect estimates are rarely statistically differentiable across subgroups, the evidence is potentially suggestive of two conclusions. First, the text-message coaching treatment may have been more effective for women than for men. Second, treatment may have been more effective at improving well-being outcomes among students who are more familiar with university and life in Toronto (non-first year students, domestic students, and those with less difficulty transitioning to university) while it may have been more effective at improving success strategies among students who are less familiar with university and city life (first-year students, international students, and those with more difficulty transitioning). Ultimately, however, treatment effects in the overall sample and across all subgroups are modest, making it difficult to draw definitive conclusions about heterogeneous treatment effects.

5. Conclusion

In this paper, we designed and evaluated two large-scale behavioral interventions aiming to improve student experiences in college. Students assigned to the first intervention completed a novel online preparatory module while students assigned to the second intervention were also offered the opportunity to participate in a coaching program, in which they were matched with senior undergraduate coaches who mentored them via text message. We found that neither intervention improved student grades or credit accumulation but that the online module enhanced with the text-messaging program modestly increased students' sense of belonging and support at the university and caused students to seek out extra help from instructors and tutors more frequently. We also cannot rule out that at least part of the effect of the text-messaging treatment condition stems from students having completed the online preparatory module.

An exploration of heterogeneous treatment effects revealed that the effects of the text-messaging program on non-academic outcomes is driven by larger effects for female students, consistent with many studies on behavioral interventions in education (see, for example, Angrist, Lang, and Oreopoulos 2009 or Andrews, Lovenheim, and Imberman 2016). We also found suggestive evidence that effects on well-being outcomes are greater among students who are more familiar with life in the city and university while effects on success strategies are larger among students who are less familiar. Ultimately, however, we lack the statistical precision required to make definitive statements about heterogeneous treatment effects.

Although the treatment effects we estimate are modest, we note that the majority of students in the text-message coaching intervention report feeling supported by their coaches, appreciating their coach's messages, and feeling like they are doing better in university partly because of their coaches. Taken together, we view our results as the start of a conversation about the importance of non-academic outcomes and student perceptions in the evaluation of behavioral interventions in education contexts. Future evaluations should aim to consider treatment effects on non-academic outcomes, such as student well-being and success strategies, in addition to commonly explored outcomes like grades, credit accumulation, and persistence. As we demonstrate in this paper, an intervention can improve student experiences in college and also be perceived as beneficial by students while not necessarily affecting grades and credit accumulation.

Given the potential effects on non-academic outcomes and the low-cost of text-messaging programs, such interventions may be worth implementing even when they do not seem to affect student grades. The interventions we evaluated cost approximately $12 per student enrolled when we account for the setup costs of building the online platform. Conditional on having the platform, the costs of the text messages we sent are approximately $2 per student.[19] Given the lower costs of these programs than more intensive, in-person coaching models, the modest estimated effects we found on students' non-grades outcomes suggest that such programs are potentially worth expanding, as they may improve student experience. [20]

[19] This paper is part of a broader research agenda on low-cost, scalable interventions in education. To encourage further research and help reduce the costs of building an online platform, we have made all interventions available online and made them customizable for other researchers who are interested in implementation at their institutions. For details, see https://studentachievementlab.org/

[20] We studied the interventions in the context of the U of T, an institution with more than 40,000 undergraduate students and first year class sizes that are often in the hundreds. Interventions that help students become self-directed or provide personalized assistance may be particularly valuable at this type of larger institution.

Further research is needed to better understand how the effects we found generalize across other institutions and student populations. It is also worth exploring whether such interventions can be made even more targeted and cost-effective by experimentally varying the characteristics of the text messages students receive to develop a stronger understanding of the types of messages that generate student engagement and improve outcomes. We are currently exploring both avenues in ongoing work.

References

American College Health Association. American College Health Association-National College Health Assessment II: Reference Group Executive Summary Fall 2017. Hanover, MD: American College Health Association; 2018.

Andrews, Rodney J., Scott A. Imberman, and Michael F. Lovenheim. 2016. "Recruiting and Supporting Low-Income, High-Achieving Students at Flagship Universities." NBER Working Paper # 22260. Cambridge, Mass.: National Bureau of Economic Research.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. "Incentives and Services for College Achievement: Evidence from a Randomized Trial." *American Economic Journal: Applied Economics.* 1(1):136-163.

Angrist, Joshua, Philip Oreopoulos, and Tyler Williams. 2014. "When Opportunity Knocks, Who Answers? New Evidence on College Achievement Awards" *Journal of Human Resources.* 49(3): 572-610.

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

Beattie, Graham, Jean-William P. Laliberté 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 P. Laliberté, Catherine Michaud-Leclerc, and Philip Oreopoulos. 2018. " What Sets College Thrivers and Divers Apart? A Contrast in Study Habits, Attitudes, and Mental Health," NBER Working Paper # 23588. Cambridge, Mass.: National Bureau of Economic Research.

Bergman, Peter. 2017. “Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment.” *Columbia University Working Paper*.

Bergman, Peter and Eric Chan. 2017. “Leveraging Parents: The Impact of High-Frequency Information on Student Achievement.” *Columbia University Working Paper*.

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, Eric, Bridget Terry Long, Philip Oreopoulos, and Lisa Sanbonmatsu. 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.

Bettinger, Eric, Sten Ludvigsen, Mari Rege, Ingeborg F. Solli, and David Yeager. 2018. “Increasing Perseverance in Math: Evidence from a Field Experiment in Norway.” *Journal of Economic Behavior and Organization.* 146: 1-15.

Canning, Elizabeth A. and Judith M. Harackiewicz, 2015, "Teach It, Don't Preach It: The Differential Effects of Directly-Communicated and Self-Generated Utility-Value Information." *Motivation Science*, 1(1): 47-71

Castleman, Benjamin, and Lindsay Page. 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.

__________. 2016. “Freshman Year Financial Aid Nudges: An Experiment to Increase FAFSA Renewal and College Persistence.” *Journal of Human Resources*, 51(2): 389-415.

Castleman, Benjamin, and Katharine Meyer. 2016. “Can Text Message Nudges Improve Academic Outcomes in College? Evidence from a West Virginia Initiative.” Center for Education Policy and Workforce Competitiveness Working Paper # 43. Charlottesville, VA: University of Virginia.

Clark, Damon, David Gill, Victoria Prowse, and Mark Rush. 2017. “Using Goals to Motivate College Students: Theory and Evidence from Field Experiments.” NBER Working Paper # 23638. Cambridge, Mass.: National Bureau of Economic Research.

Damgaard, Mette Trier and Helena Skyt Nielsen. 2018. “Nudging in education.” *Economics of Education Review*. 64: 313-342.

Dobronyi, Christopher R., Philip Oreopoulos, and Uros Petronijevic. 2017. “Goal Setting, Academic Reminders, and College Success: A Large-Scale Field Experiment.” NBER Working Paper #23738. Cambridge, Mass.: National Bureau of Economic Research.

Evans, William N, Melissa S. Kearney, Brendan Perry, and 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, Mass.: National Bureau of Economic Research.

Fryer, Roland G. 2016. "Information, Non-Financial Incentives, and Student Achievement: Evidence from a Text Messaging Experiment.” *Journal of Public Economics*. 144:109-121.

Gollwitzer, Peter M. and Bernd Schaal. 1998. "Metacognition in Action: The Importance of Implementation Intentions." *Personality and Social Psychology Review*. 2(2): 124-136.

Hastings, J., Neilson, C. A., and Zimmerman, S. D. 2018. "The effects of earnings disclosure on college enrollment decisions." NBER Working Papers # 21300. Cambridge, Mass.: National Bureau of Economic Research.

Kling, Jeffrey R., Jeffrey B Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica*. 75(1): 83-119.

Kraft, Matthew A., and Shaun M. Doughtery. 2013. "The Effect of Teacher–Family Communication on Student Engagement: Evidence from a Randomized Field Experiment." *Journal of Research on Educational Effectiveness*. 6(3):199-222.

Kraft, Matthew A., and Manuel Monti-Nussbaum. 2017. "Can schools enable parents to prevent summer learning loss? A text messaging field experiment to promote literacy skills." *The ANNALS of the American Academy of Political and Social Science*, 674(1): 85-112

Kraft, Matthew A., and Todd Rogers. 2014. "The Underutilized Potential of Teacher-to-Parent Communication: Evidence from a Field Experiment." Harvard Kennedy School Faculty Research Working Paper Series, RWP-14-049. Cambridge, Mass.: Harvard University

Lavecchia, Adam, Heidi Liu and Philip Oreopoulos. 2016. "Behavioral Economics of Education: Progress and Possibilities" *Handbook of Economics of Education*. (5): 1-74.

Mayer, Susan E., Ariel Kalil, Philip Oreopoulos, and Sebastian Gallegos. 2015. "Using Behavioral Insights to Increase Parental Engagement: The Parents and Children Together

(PACT) Intervention.” NBER Working Paper #21602. Cambridge, Mass.: 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.

Oreopoulos, Philip and Reuben Ford. 2016. “Keeping College Options Open: A Field Experiment to Help All High School Seniors Through the College Application Process.” NBER Working Paper # 22320. Cambridge, Mass.: National Bureau of Economic Research.

Page, Lindsay C., Stacy S. Kehoe, Benjamin L. Castleman, Gumilang A. Sahadewo. “More than Dollars for Scholars: The Impact of the Dell Scholars Program on College Access, Persistence and Degree Attainment.” Journal of Human Resources, forthcoming.

Scrivener, Susan and Michael J. Weiss. 2013. “More Graduates: Two-Year Results from an Evaluation of Accelerated Study in Associate Programs (ASAP) for Developmental Education Students.” Policy Brief, MDRC.

Steele, C. M. 1997. A threat in the air: How stereotypes shape intellectual identity and performance. American Psychologist, 52(6), 613-629

Walton, Gregory. 2014. “The New Science of Wise Psychological Interventions.” *Current Directions in Psychological Science*. 23(1):73-82.

Walton, Gregory and Geoffrey Cohen. 2011. “A Brief Social-Belonging Intervention Improves Academic and Health Outcomes of Minority Students.” *Science*. 331: 1447-1451.

Walton, Gregory, Christine Logel, Jennifer Peach, Steven Spencer, and Mark Zanna. 2015. "Two Brief Interventions to Mitigate a "Chilly Climate" Transform Women's Experience, Relationships, and Achievement in Engineering." *Journal of Educational Psychology*. 107(2):468-485.

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

Yeager, David, Carissa Romero, Dave Paunesku, Christopher S. Hulleman, Barbara Schneider, Cintia Hinojosa, Hae Yeon Lee, Joseph O'Brien, Kate Flint, Alice Roberts, and Jill Trott. 2016. "Using Design Thinking to Improve Psychological Interventions: The Case of the Growth Mindset During the Transition to High School." *Journal of Educational Psychology*. 108(3):374–391.

Tables

Table 1: Treatment Randomization

	Control	Online Only	Text Messaging
Number of Students	1119	1154	1122
(i) Fraction of Total	0.329	0.339	0.330
(ii) Intended Fraction	0.33	0.33	0.33
p-value of (i) – (ii)	0.96	0.22	0.95
Completed Exercise	1110	1143	1113

Table 2: Summary Statistics and Balancing Tests

Student Characteristics	Control Sample Mean [Standard Deviation]	Online Only Difference [Standard Error]	Text Message Difference [Standard Error]	P-Value from F-Test of No Difference
Female	0.548	0.000	-0.003	0.987
	[0.498]	[0.022]	[0.022]	
Age in First Year	18.453	0.057	0.015	0.624
	[1.388]	[0.061]	[0.062]	
High School Average Grade	90.448	-0.233	-0.292	0.340
	[4.032]	[0.207]	[0.213]	
First-Year Student	0.715	0.003	0.012	0.797
	[0.452]	[0.019]	[0.019]	
English Mother Tongue	0.446	-0.012	0.001	0.816
	[0.497]	[0.022]	[0.022]	
Lives on Residence	0.431	-0.001	0.017	0.650
	[0.495]	[0.022]	[0.022]	
International Student	0.478	0.002	0.012	0.828
	[0.500]	[0.021]	[0.021]	
First-Generation Student	0.239	0.006	0.006	0.925
	[0.426]	[0.018]	[0.018]	
Expects to Earn more than BA	0.711	-0.015	0.009	0.463
	[0.453]	[0.019]	[0.019]	
Expects at least A- Average	0.682	-0.036*	-0.025	0.172
	[0.466]	[0.020]	[0.020]	
Checks Cell Phone Often	0.471	-0.037*	-0.034	0.147
	[0.499]	[0.021]	[0.021]	
Expected Study Hours/Week	18.265	-0.139	-0.317	0.820
	[12.095]	[0.500]	[0.505]	
Expected Work Hours/Week > 8	0.298	-0.002	0.033*	0.133
	[0.457]	[0.019]	[0.020]	
Not Discouraged by Setbacks (5-point Scale)	3.440	-0.048	-0.047	0.407
	[0.964]	[0.040]	[0.041]	
Finish what I start (5-point scale)	3.814	-0.021	0.007	0.695
	[0.837]	[0.035]	[0.035]	
Think about Future (7-point scale)	2.393	-0.025	-0.040	0.722
	[1.214]	[0.051]	[0.050]	
Tend to Cram for Exams (7-point scale)	3.941	0.001	0.055	0.602
	[1.521]	[0.064]	[0.063]	
Transition has been Challenging (7-point scale)	4.685	0.036	-0.029	0.628
	[1.615]	[0.068]	[0.068]	

Summary statistics and differences are calculated using the full sample of students. Robust standard errors are reported in brackets. *** indicates significance at the 1 percent level, ** at the 5 percent level, and * at the 10 percent level.

Table 3: Outcome Variables from Follow-Up Survey

Main Index	Sub Index	Survey Question
Core Well-Being	Satisfaction	All things considered, how satisfied are you with your life as a whole these days?
		All things considered, how satisfied are you with your experience at University of Toronto so far?
	Belonging and Support	I feel like I belong here at U of T
		Being a student at UofT is an important part of how I see myself
		UofT wants me to be successful here
		I know where to go if I need academic advice
		UofT does its best to help support me
		Other students understand more than I do about how things work here at UofT*
		I often remind myself of my goals and motivations for being here at UofT
		The transition to the University of Toronto has, so far, been challenging*
	Confidence	How important is it to you that you do well at U of T?
		How confident do you feel that you have the ability to do well at University of Toronto?
	Depression	Since the beginning of the academic year, I have felt sad or depressed*
		Since the beginning of the academic year, I have felt very stressed*
Success Strategies	Study Strategies	I manage my time well
		I try to learn from my mistakes on past tests and assignments
		I write thoughts and ideas down when I study to test my understanding
		I get feedback from my writing assignments before handing them in
	Help Seeking	Last term, how often did you meet with an instructor outside of class?
		Last term, how often did you meet with an academic advisor?
		Last term, how often did you use a free academic tutor?
		Last term, how often did you meet with a paid tutor?
		Last term, how often did you attend a workshop?
		Last term, how often did you participate in an informal study group?

*Responses to these variables are re-coded so that more beneficial outcomes have higher scores.

Table 4: Text Message Summary Statistics

	(1) Mean	(2) Standard Deviation	(3) Minimum	(4) Maximum	(5) Number of Observations
Coaches					
Number of Students Per Coach+	100.8	45.75	70	185	10
Number of Messages Sent by Coaches					
Fall Semester Non-Batch Messages	1,520.5	816.02	479	2,792	10
Fall Semester Batch Messages	3,121.7	2,729.71	834	9,355	10
Fall Semester Unique Batch Messages*	46.5	24.80	21	87	10
Winter Semester Non-Batch Messages	481.60	225.67	171	854	10
Winter Semester Batch Messages	1,733.7	1,908.55	488	6,859	10
Winter Semester Unique Batch Messages*	18.7	10.28	9	40	10
Students					
Fraction of Treated Students Sending At Least One Message	0.62	0.49	0	1	1,122
Number of Messages Sent by Students					
Fall Semester Messages	17.29	28.33	0	385	983
Winter Semester Messages	6.21	12.72	0	140	983
Number of Messages Received by Students					
Fall Semester Total Messages	47.22	25.62	0	287	983
Fall Semester Batch Messages	31.76	14.33	0	61	983
Winter Semester Total Messages	22.53	14.02	0	94	983
Winter Semester Batch Messages	17.63	10.93	0	39	983

+Multiplying the average number of students per coach by the number of coaches gives only 1,008 students. Of the 1,122 students who were assigned to the text-messaging treatment, 114 declined to participate and were thus not assigned a coach. An additional 25 were asked to be removed from the program, bringing the number of students who appear in the text-message data down to 983.
*We avoid double counting by counting each pre-programmed batch message that coaches sent to their students only once, despite such messages being received by many students of a given coach.

Table 5: Treatment Effects on Stacked Grades

	(1)	(2)	(3)	(4)	(6)	(7)	(8)	(9)
	All Course Grades		Fall Course Grades		Winter Course Grades		Economics Grade	
Online Only	-0.012	0.261	-0.687	-0.531	-0.163	0.085	0.452	0.823
	[0.575]	[0.532]	[0.624]	[0.574]	[0.761]	[0.707]	[0.761]	[0.730]
Text Messaging	0.169	0.240	-0.434	-0.410	-0.015	0.012	0.618	0.717
	[0.554]	[0.521]	[0.603]	[0.567]	[0.720]	[0.679]	[0.751]	[0.729]
Control Mean	70.851		72.999		70.993		67.957	
[Standard Deviation]	[15.881]		[14.120]		[16.526]		[15.995]	
Controls?	No	Yes	No	Yes	No	Yes	No	Yes
Observations	20,786	20,786	6,687	6,687	7,367	7,367	2,584	2,584

The dependent variable in each regression is indicated by the column headings. The unit of observation is a student-course. Additional control variables include student age, self-reported expected study hours per week, expected paid-work hours per week, expected average grade, tendency to finish what he or she starts, tendency to get discouraged by setbacks, tendency to study at the last minute, and indicator variables for first-year status, international student status, first-generation status, gender, English mother-tongue status, whether the student lives on residence, a self-reported desire to earn more than an undergraduate degree, and a self-reported tendency to check his or her cell phone often. Standard errors clustered at the student level are reported in brackets in columns (1) to (7). In columns (8) and (9), we report robust standard errors in brackets. *** indicates significance at the 1 percent level; ** indicates significance at the 5 percent level; and * indicates significance at the 10 percent level.

Table 6: Treatment Effects on GPA and Credit Accumulation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	All Courses			Fall Courses			Winter Courses		
	GPA	Credits Failed	Credits Earned	GPA	Credits Failed	Credits Earned	GPA	Credits Failed	Credits Earned
Online Only	0.041	-0.007	-0.036	0.023	0.013	-0.015	0.043	-0.002	-0.011
	[0.040]	[0.037]	[0.064]	[0.043]	[0.011]	[0.025]	[0.046]	[0.015]	[0.028]
Text Messaging	0.034	-0.039	-0.009	0.012	0.005	-0.006	0.007	-0.020	-0.025
	[0.041]	[0.035]	[0.063]	[0.044]	[0.010]	[0.025]	[0.047]	[0.014]	[0.028]
Control Mean	2.546	0.376	4.142	2.779	0.060	1.151	2.664	0.110	1.228
[Standard Deviation]	[0.979]	[0.862]	[1.461]	[0.995]	[0.222]	[0.562]	[1.067]	[0.359]	[0.631]
Observations	3,075	3,075	3,075	2,783	2,783	2,783	2,807	2,807	2,807

The dependent variable in each regression is indicated by the column headings. All regressions are run at the student level and control for student age, self-reported expected study hours per week, expected paid-work hours per week, expected average grade, tendency to finish what he or she starts, tendency to get discouraged by setbacks, tendency to study at the last minute, and indicator variables for first-year status, international student status, first-generation status, gender, English mother-tongue status, whether the student lives on residence, a self-reported desire to earn more than an undergraduate degree, and a self-reported tendency to check his or her cell phone often. 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.

Table 7: Treatment Effects on Core Well-Being Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
			Sub-Indices							
	Core Well-Being Index		Satisfaction Index		Belonging & Support Index		Confidence Index		Depression Index	
Online Only	0.018	0.028	0.047	0.061	0.012	0.018	0.047	0.066*	-0.016	-0.005
	[0.025]	[0.024]	[0.043]	[0.041]	[0.026]	[0.025]	[0.036]	[0.034]	[0.041]	[0.039]
Text Messaging	0.046*	0.041*	0.039	0.034	0.058**	0.053**	0.036	0.034	0.012	0.007
	[0.025]	[0.024]	[0.043]	[0.042]	[0.026]	[0.025]	[0.037]	[0.035]	[0.042]	[0.040]
P-Value for Test of Treatment Differences	0.275	0.577	0.859	0.517	0.0771	0.163	0.763	0.350	0.503	0.763
Controls?	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Observations	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623

The dependent variable in each regression is indicated by the column headings. Control variables include student age, self-reported expected study hours per week, expected paid-work hours per week, expected average grade, tendency to finish what he or she starts, tendency to get discouraged by setbacks, tendency to study at the last minute, and indicator variables for first-year status, international student status, first-generation status, gender, English mother-tongue status, whether the student lives on residence, a self-reported desire to earn more than an undergraduate degree, and a self-reported tendency to check his or her cell phone often. Robust standard errors are reported in brackets. * indicates is significance at the 10 percent level; ** indicates significance at the 5 percent level; *** indicates significance at the 1 percent level.

Table 8: Treatment Effects on Success Strategies Index

	(1)	(2)	(3)	(4)	(5)	(6)
			Sub-Indices			
	Success Strategies Index		Strategies Index		Help-Seeking Index	
Online Only	0.029	0.034	0.032	0.041	0.027	0.030
	[0.023]	[0.022]	[0.033]	[0.031]	[0.027]	[0.026]
Text Messaging	0.055**	0.051**	0.050	0.043	0.057**	0.057**
	[0.023]	[0.022]	[0.032]	[0.030]	[0.027]	[0.027]
P-Value for Test of Treatment Differences	0.271	0.431	0.655	0.966	0.263	0.306
Controls?	No	Yes	No	Yes	No	Yes
Observations	2,623	2,623	2,623	2,623	2,623	2,623

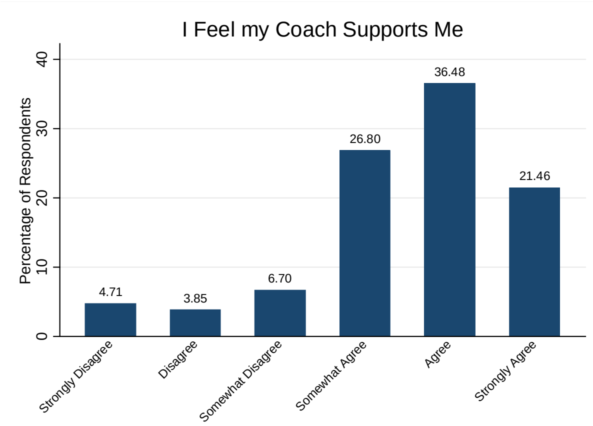
The dependent variable in each regression is indicated by the column headings. Control variables include student age, self-reported expected study hours per week, expected paid-work hours per week, expected average grade, tendency to finish what he or she starts, tendency to get discouraged by setbacks, tendency to study at the last minute, and indicator variables for first-year status, international student status, first-generation status, gender, English mother-tongue status, whether the student lives on residence, a self-reported desire to earn more than an undergraduate degree, and a self-reported tendency to check his or her cell phone often. Robust standard errors are reported in brackets. * indicates is significance at the 10 percent level; ** indicates significance at the 5 percent level; *** indicates significance at the 1 percent level.

Table 9: Treatment Effects Across Student Subgroups

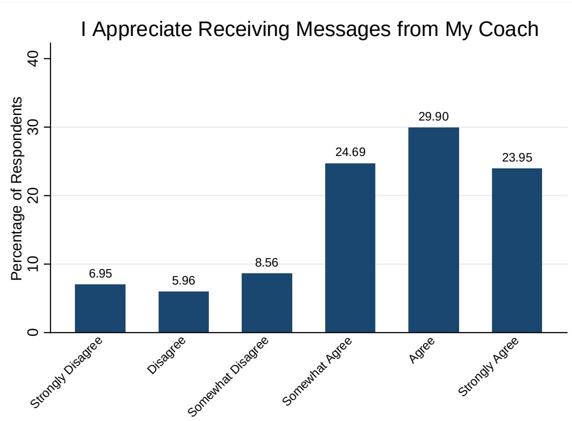
	Gender		First Year Status		International Student		Transition Difficulty (TD)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Female	Male	First Year	Not First Year	International	Domestic	TD > Median	TD < Median
				Panel (a): Grades				
Online Only	-0.435	1.121	0.439	-0.436	-0.221	0.848	0.272	0.129
	[0.657]	[0.863]	[0.540]	[1.405]	[0.816]	[0.683]	[0.779]	[0.712]
Text Messaging	-0.362	0.924	-0.395	2.686**	0.378	0.205	-0.112	0.287
	[0.662]	[0.826]	[0.543]	[1.336]	[0.783]	[0.693]	[0.744]	[0.721]
Control Mean	70.976	70.696	72.305	66.407	69.645	71.933	72.017	69.896
[Standard Deviation]	[15.388]	[16.472]	[14.756]	[18.206]	[16.603]	[15.124]	[15.125]	[16.414]
				Panel (b): Core Well-Being Index				
Online Only	0.022	0.029	0.027	0.024	-0.002	0.047	0.052	0.023
	[0.031]	[0.037]	[0.026]	[0.058]	[0.033]	[0.035]	[0.034]	[0.033]
Text Messaging	0.057*	0.016	0.030	0.101	-0.003	0.084**	0.042	0.052*
	[0.031]	[0.036]	[0.025]	[0.066]	[0.033]	[0.034]	[0.035]	[0.031]
				Panel (c): Belonging and Support Index				
Online Only	0.004	0.029	0.020	0.004	-0.008	0.034	0.037	0.016
	[0.032]	[0.039]	[0.027]	[0.062]	[0.034]	[0.036]	[0.036]	[0.034]
Text Messaging	0.075**	0.019	0.044*	0.107	0.015	0.089**	0.046	0.070**
	[0.032]	[0.038]	[0.026]	[0.072]	[0.033]	[0.036]	[0.036]	[0.033]
				Panel (d): Success Strategies Index				
Online Only	0.033	0.036	0.048**	-0.015	0.039	0.023	0.046	0.024
	[0.028]	[0.033]	[0.024]	[0.051]	[0.031]	[0.030]	[0.033]	[0.029]
Text Messaging	0.054*	0.044	0.057**	0.042	0.066**	0.041	0.073**	0.033
	[0.028]	[0.034]	[0.023]	[0.057]	[0.031]	[0.029]	[0.033]	[0.029]
				Panel (e): Help-Seeking Index				
Online Only	0.027	0.030	0.032	0.030	0.018	0.038	0.029	0.032
	[0.034]	[0.041]	[0.029]	[0.062]	[0.038]	[0.036]	[0.040]	[0.035]
Text Messaging	0.065*	0.042	0.069**	0.012	0.070*	0.049	0.056	0.061*
	[0.035]	[0.042]	[0.029]	[0.067]	[0.039]	[0.036]	[0.040]	[0.036]

The dependent variable in each regression is indicated by the panel headings. Control variables include student age, self-reported expected study hours per week, expected paid-work hours per week, expected average grade, tendency to finish what he or she starts, tendency to get discouraged by setbacks, tendency to study at the last minute, and indicator variables for first-year status, international student status, first-generation status, gender, English mother-tongue status, whether the student lives on residence, a self-reported desire to earn more than an undergraduate degree, and a self-reported tendency to check his or her cell phone often. Standard errors clustered at the student level are reported in brackets in panel (a). Robust standard errors are reported in brackets in panels (b) to (e). * indicates is significance at the 10 percent level; ** indicates significance at the 5 percent level; *** indicates significance at the 1 percent level.

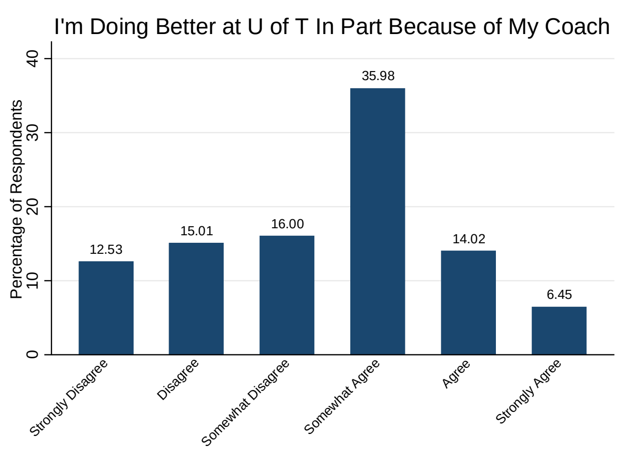
Figures



(a): I Feel my Coach Supports Me



(b): I Appreciate Receiving Messages from My Coach



(c): I am Doing Better at U of T In Part Because of My Coach

Figure 1: Student Feelings About the Text-Message Coaching Program

Notes: This figure shows the percentages of students in the text-message coaching program who strongly disagree, disagree, somewhat disagree, somewhat agree, agree, and strongly agree with the statement that appears as the title of each panel.

Appendix C: Supplemental Analyses

Table C1: Treatment Effects on Components of Core Well-Being Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7	8)	(9)	(10)	(11)	(12)	(13)	(14)
	Satisfaction Index		Belonging and Support Index								Confidence Index		Depression Index	
	Overall Life Satisfaction	Satisfaction with UofT	Belonging at UofT	Identify with UofT	UofT Wants me to Succeed	Know where Advice	UofT Supports Me	Others Know More	Often Think of Goals	Tough Transition	Important to do Well at UofT	Confident to do Well at UofT	Feelings of Depression	Feelings of Stress
Online Only	0.147**	0.016	-0.054	0.040	0.071	0.043	0.055	0.034	-0.015	0.006	0.073	0.104	0.006	-0.014
	[0.063]	[0.060]	[0.051]	[0.057]	[0.059]	[0.054]	[0.053]	[0.054]	[0.054]	[0.059]	[0.053]	[0.066]	[0.040]	[0.038]
Text Messaging	0.089	0.002	0.027	0.125**	0.104*	0.094*	0.081	0.025	0.004	0.050	0.075	0.009	0.029	-0.016
	[0.064]	[0.061]	[0.051]	[0.056]	[0.058]	[0.054]	[0.053]	[0.053]	[0.054]	[0.060]	[0.054]	[0.067]	[0.041]	[0.039]
P-Value for Test of Treatment Differences	0.361	0.816	0.114	0.129	0.561	0.340	0.620	0.862	0.715	0.463	0.965	0.156	0.583	0.971
Control Mean	4.456	4.410	4.219	4.289	3.938	4.225	3.923	3.113	4.474	2.692	5.894	4.309	2.432	2.108
[Standard Deviation]	[1.349]	[1.333]	[1.119]	[1.237]	[1.312]	[1.178]	[1.175]	[1.155]	[1.151]	[1.232]	[1.203]	[1.473]	[0.861]	[0.830]
Observations	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623

The dependent variable in each regression is indicated by the column headings. Control variables include student age, self-reported expected study hours per week, expected paid-work hours per week, expected average grade, tendency to finish what he or she starts, tendency to get discouraged by setbacks, tendency to study at the last minute, and indicator variables for first-year status, international student status, first-generation status, gender, English mother-tongue status, whether the student lives on residence, a self-reported desire to earn more than an undergraduate degree, and a self-reported tendency to check his or her cell phone often. Robust standard errors are reported in brackets. * indicates is significance at the 10 percent level; ** indicates significance at the 5 percent level; *** indicates significance at the 1 percent level.

Table C2: Treatment Effects on Components of Success Strategies Index

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Study Strategies Index					Help-Seeking Index			
	Manage Time Well	Learn from Mistakes	Study with Pen	Review Writing	Times Met with Instructor	Times Met with Advisor	Times Met with Free Tutor	Times Met with Paid Tutor	Times Went to Workshop	Times Used Study Group
Online Only	0.097*	0.019	0.020	0.059	0.059	0.098*	0.081	0.007	-0.069	0.173
	[0.052]	[0.046]	[0.057]	[0.061]	[0.081]	[0.058]	[0.100]	[0.093]	[0.084]	[0.130]
Text Messaging	0.035	0.053	0.018	0.100	0.198**	0.051	0.259**	0.077	-0.032	0.120
	[0.052]	[0.046]	[0.056]	[0.061]	[0.086]	[0.060]	[0.105]	[0.095]	[0.082]	[0.133]
P-Value for Test of Treatment Differences	0.243	0.455	0.982	0.503	0.104	0.440	0.086	0.465	0.646	0.688
Control Mean	3.589	4.644	4.095	3.597	1.200	0.752	1.533	0.911	1.102	2.529
[Standard Deviation]	[1.171]	[1.021]	[1.243]	[1.320]	[1.724]	[1.224]	[2.165]	[1.967]	[1.814]	[2.774]
Observations	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623	2,623

The dependent variable in each regression is indicated by the column headings. Control variables include student age, self-reported expected study hours per week, expected paid-work hours per week, expected average grade, tendency to finish what he or she starts, tendency to get discouraged by setbacks, tendency to study at the last minute, and indicator variables for first-year status, international student status, first-generation status, gender, English mother-tongue status, whether the student lives on residence, a self-reported desire to earn more than an undergraduate degree, and a self-reported tendency to check his or her cell phone often. Robust standard errors are reported in brackets. * indicates is significance at the 10 percent level; ** indicates significance at the 5 percent level; *** indicates significance at the 1 percent level.

Table C3: Leave-One-Coach-Out Text-Message Coaching Treatment Effects

Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Grades (All Courses)	0.312	0.483	0.236	0.152	0.343	0.343	0.079	0.402	0.225	0.204
	[0.531]	[0.521]	[0.530]	[0.535]	[0.525]	[0.538]	[0.535]	[0.534]	[0.533]	[0.546]
Core Well-Being Index	0.046*	0.040*	0.040*	0.044*	0.037	0.045*	0.037	0.024	0.058**	0.033
	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]	[0.024]	[0.025]	[0.024]	[0.025]
Belonging and Support Index	0.061**	0.051**	0.051**	0.058**	0.049*	0.056**	0.051**	0.034	0.069***	0.048*
	[0.025]	[0.025]	[0.025]	[0.025]	[0.025]	[0.025]	[0.025]	[0.026]	[0.025]	[0.026]
Success Strategies Index	0.052**	0.052**	0.052**	0.055**	0.052**	0.053**	0.043**	0.044*	0.053**	0.048**
	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]	[0.022]	[0.023]
Help-Seeking Index	0.056**	0.054**	0.061**	0.059**	0.058**	0.057**	0.047*	0.052*	0.057**	0.056**
	[0.027]	[0.027]	[0.027]	[0.028]	[0.027]	[0.027]	[0.027]	[0.028]	[0.027]	[0.028]
Controls?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Each cell contains the estimated treatment effect of the text-messaging intervention and the associated standard error from a separate regression. In a given row, the dependent variable is the same across all specifications and indicated by the description in the first column. The specifications in each column differ according to which coach's students are excluded from the regression. The excluded coach number is given by the column number. The control variables include student age, self-reported expected study hours per week, expected paid-work hours per week, expected average grade, tendency to finish what he or she starts, tendency to get discouraged by setbacks, tendency to study at the last minute, and indicator variables for first-year status, international student status, first-generation status, gender, English mother-tongue status, whether the student lives on residence, a self-reported desire to earn more than an undergraduate degree, and a self-reported tendency to check his or her cell phone often. Robust standard errors are reported in brackets. * indicates is significance at the 10 percent level; ** indicates significance at the 5 percent level; *** indicates significance at the 1 percent level.