

---

# Student Coaching: How Far Can Technology Go?

---

**Philip Oreopoulos**  
**Uros Petronijevic**

## ABSTRACT

*One-on-one coaching programs tend to have large effects on student outcomes, but they are costly to scale. In contrast, interventions that rely on technology to maintain contact with students can be scaled at low cost but may be less effective than one-on-one assistance. We randomly assign more than 4,000 students from a large Canadian university into control, online exercise, text messaging, and one-on-one coaching groups and find large effects on academic outcomes from the coaching program but no effects from either technology-based intervention. A comparison of key design features suggests that future technology-based interventions should aim to provide proactive, personalized, and regular support.*

## I. Introduction


Policymakers and academics share growing concerns about stagnating college completion rates and negative student experiences. Between 1970 and 1999, for example, while college enrollment rates of 23-year-old students rose substantially,

---

*Philip Oreopoulos is a professor of economics at the University of Toronto. Uros Petronijevic is an assistant professor of economics at York University. The authors are indebted to the first-year economics instructors at the University of Toronto for their willingness to try something different, and incorporate this experiment into their courses. They thank Aaron de Mello for tireless efforts to design, debug, and perfect the experiment's website, as well as for help with data extraction. Chantel Choi, Nabanita Nawar, Rachel Padillo, and Chadd Pirali showed great enthusiasm and professionalism in their role as coaches. The authors thank five anonymous reviewers for helpful comments and suggestions. Jean-William Laliberté and Shahin Hazrati provided outstanding research assistance. Seminar participants at CUNY, University of Toronto, and the Canadian Institute for Advanced Research provided useful feedback. Financial support for this research was provided by the Ontario Human Capital Research and Innovation Fund, a Social Sciences and Humanities Research Council Insight Grant (#435-2015-0180), and a JPAL Pilot Grant. Petronijevic also gratefully acknowledges support from Canada's Social Sciences and Humanities Research Council and Ontario's Graduate Scholarship. This RCT was registered in the American Economic Association Registry for randomized control trials under Trial number AEARCTR-0000810. Any omissions or errors are the responsibility of the authors. The data used in this article can be obtained beginning October 2018 through October 2021 from Philip Oreopoulos, 150 St. George St., Suite 30, Toronto, Ontario, M5S 3G7, Canada, philip.oreopoulos@utoronto.ca. [Submitted December 2016; accepted January 2017]; doi:10.3368/jhr.53.2.1216-8439R*

JEL Classification: I20, I23, J24

ISSN 0022-166X E-ISSN 1548-8004 © 2018 by the Board of Regents of the University of Wisconsin System

 Supplementary materials are freely available online at: <http://uwpress.wisc.edu/journals/journals/jhr-supplementary.html>

completion rates fell by 25 percent (Turner 2004). More recent figures suggest that only 56 percent of students who pursue a bachelor's degree complete it within six years (Symonds, Schwartz, and Ferguson 2011), and recent research questions whether students who attain degrees acquire meaningful new skills along the way (Arum and Roksa 2011). Students are increasingly entering college underprepared, with those who procrastinate, do not study enough, and have superficial attitudes about success performing particularly poorly (Beattie, Laliberté, and Oreopoulos 2016).

Much existing research focuses on lacking financial resources among both students and the institutions they attend as explanations for low completion rates and negative experiences. The impediments of student resource constraints are highlighted by youth from high-income families being more likely to attend college, even after accounting for cognitive achievement, family composition, race, and residence (Belley and Lochner 2007), and by student average work hours increasing during recent decades when college prices continued to rise but sources of financial aid did not follow suit (Scott-Clayton 2012). Financially constrained students are often forced to underinvest in higher education or to take on part-time employment, thereby reducing the time available for schoolwork.<sup>1</sup> Resource constraints among less-selective public universities and community colleges, where there are fewer resources per student, also contribute toward low completion rates and student dissatisfaction. Completion times have increased most among students who start college at these institutions (Bound, Lovenheim, and Turner 2012), where increases in student demand for higher education are not fully offset with increases in resources—something top-tier schools do by regulating enrollment size (Bound and Turner 2007).

While the economics of education literature has devoted much attention to the role of resource constraints, comparatively less attention has been given to understanding the role that students themselves play in the production of higher education. Yet, at both the high school and college levels, an emerging recent literature demonstrates the benefits of helping students foster motivation, effort, good study habits, and time-management skills through structured tutoring and coaching. Cook et al. (2014) find that cognitive behavioral therapy and tutoring generate large improvements in math scores and high school graduation rates for troubled youth in Chicago, while Oreopoulos, Lavecchia, and Brown (forthcoming) show that coaching, tutoring, and group activities lead to large increases in high school graduation and college enrollment among youth living in a Toronto public housing project.<sup>2</sup> Structured coaching has also recently been shown to improve outcomes among college students. Scrivener and Weiss (2013) find that the Accelerated Study in Associates Program (ASAP)—a bundle of coaching, tutoring, and student success workshops—in CUNY community colleges nearly doubled graduation rates, and Bettinger and Baker (2014) show that telephone coaching by *Inside Track* professionals boosts two-year college retention by 15 percent.

---

1. Simply providing access to financial aid may not be enough, as the application process can be prohibitively complex for some students. Bettinger et al. (2012) show that providing assistance with the Free Application for Federal Student Aid (FAFSA) increases both college entry and persistence, while Castleman and Page (2014b) demonstrate that reminding students about the steps to renew FAFSA aid also leads to higher renewal and persistence.

2. Other notable studies on effective coaching/tutoring interventions in K–12 education include Dobbie and Fryer (2013), Fryer (2014), Kraft (2015), Kosse et al. (2016), and Ander, Guryan, and Ludwig (2016).

While structured, one-on-one support services can have large effects on student outcomes, they are often costly to implement and difficult to scale up to the student population at large (Bloom 1984). In this paper, we build on recent advances in social-psychology and behavioral economics, investigating whether technology—specifically, online exercises and text and email messaging—can generate comparable benefits to one-on-one coaching interventions but at lower costs among first-year university students.

Several recent studies in social-psychology find that one-time, short interventions occurring at an appropriate time can have lasting effects on student outcomes (Yeager and Walton 2011; Cohen and Garcia 2014; Walton 2014). Relatively large improvements on academic performance have been documented as a result of several types of interventions, including those that help students define their long-run goals or purpose for learning (Morisano et al. 2010; Yeager et al. 2014a), teach the “growth mindset” idea that intelligence is malleable (Yeager et al. 2016), help students keep negative events in perspective by self-affirming their values (Cohen and Sherman 2014), and help teachers change the tone of feedback to students in order to build trust (Yeager et al. 2014b).<sup>3</sup> As a contrast to one-time interventions, other studies in education and behavioral economics attempt to maintain constant, low-touch contact with students or their parents at a low cost by using technology to provide consistent reminders aimed at improving outcomes. For example, several studies have shown that providing text, email, and phone call reminders to parents about their students’ progress in school boosts both parental engagement and student performance (Kraft and Dougherty 2013; Bergman 2016; Kraft and Rogers 2014; Mayer et al. 2015). Researchers have also used text messaging communication with college and university students directly to increase the likelihood of students enrolling in college (Castleman and Page 2014a) and renewing financial aid (Castleman and Page 2014b), and to attempt to improve students’ academic outcomes (Castleman and Meyer 2016).

We contribute to these literatures by examining whether benefits comparable to those obtained from one-on-one coaching can be achieved at lower costs by either a one-time, online intervention designed to affirm students’ goals and purpose for attending university or a full-year text and email messaging campaign that provides weekly reminders of academic advice and motivation to students. We work with a sample of more than 4,000 undergraduate students who are enrolled in introductory economics courses across all three campuses of the University of Toronto, randomly assigning students to one of three treatment groups or a control group. The treatment groups consist of (i) a one-time, online exercise completed during the first two weeks of class in the fall, (ii) the online intervention plus text and email messaging throughout the full academic year, and (iii) the online intervention plus one-on-one coaching in which students are assigned to upper-year undergraduate coaches. Students in the control group are given a personality test measuring the Big Five personality traits.

We find large positive effects from the one-year coaching service, amounting to approximately a 5 percentage-point increase in average course grades and a 0.35 standard deviation increase in grade point average (GPA). In contrast, we find no effects on academic outcomes from either the online exercise or the text messaging campaign,

---

3. While the studies cited above all carefully explain the settings and times in which the interventions are likely to be effective, there is growing skepticism about the generalizability of some interventions due to recent failed replication attempts (see, for example, Kost-Smith et al. 2012; Dee 2015; Harackiewicz et al. 2016).

even after investigating potentially heterogeneous treatment effects across several student characteristics, including gender, age, incoming high school average, international-student status, and whether students live on residence. Our results suggest that the benefits of personal coaching are not easily replicated by low-cost interventions using technology. As we describe below, coaches were instructed to be proactive by regularly initiating contact with their students and, whenever possible, to provide concrete actionable steps for solving a given problem. Our text messaging approach was not able to replicate this proactive approach. Students had to initiate contact, and our team was unable to “dig deep” into each problem to the same degree as the coaches. We discuss the key challenges—and potential solutions—to using technology to implement coaching-type support at large scale in our discussion of the results.

While our main contribution is assessing the scope for technological interventions to reproduce the benefits of one-on-one coaching, our paper also makes two more general contributions. To our knowledge, we provide the first causal analysis of the effects of a large-scale text messaging campaign on the academic outcomes of college students. The most closely related to work to ours in this respect is Castleman and Meyer (2016), who analyze the effects of text message reminders on academic outcomes such as GPA, the number of credits attempted, and persistence. The authors work with a sample of rural, low-income college students in West Virginia and find that text campaign participants attempted more credits than nonparticipants, although they are unable to make causal claims about the program’s effectiveness because students were not randomly assigned to participation. In contrast, we randomly assign students into the messaging treatment, estimating no effects on academic outcomes. We also show that assigning students to upper-year undergraduate coaches can lead to potentially large academic improvements without the need for professionally trained coaches, as were used in Bettinger and Baker (2014). Instead, a consistent characteristic across a variety of effective coaching studies appears to be proactive coaches or mentors who regularly contact students to provide support (Cook et al. 2014; Oreopoulos, Lavecchia, and Brown, forthcoming).<sup>4</sup>

The remainder of this paper is organized as follows: In the next section, we describe the intervention in greater detail, explaining how each treatment and control group exercise was implemented. Section III describes the data and our empirical strategy for estimating the effects of the intervention, while Section IV presents the results. We discuss the results in Section V and provide concluding remarks in Section VI.

## II. Description of the Intervention

We implemented our intervention across all three University of Toronto (U of T) campuses, working with a sample of all students registered for first-year economics classes in the fall of 2015. We cooperated with the instructors of each of these classes, having them agree to make completion of our online “warm-up” exercise worth 2 percent of students’ final grade. Students had to complete the exercise in the first two weeks of the fall semester to receive credit.<sup>5</sup> The type of online exercise each student had

---

4. Having our coaches be proactive is a key difference between our coaching program and that which resulted in negligible effects in Angrist, Lang, and Oreopoulos (2009).

5. We describe our sample, randomization strategy, and balancing tests in the next section.

to complete depended on whether the student was randomly assigned to one of the three treatment groups or the control group. Each student created an account and completed the same introductory survey, in which we asked several background questions, including the highest level of education obtained by students' parents, the amount of education they expect to obtain, whether they are first-year or international students, and their work and study time plans for the upcoming year.

### *A. Treatment 1: Online Exercise*

Students in the first treatment group then worked through an online exercise designed to get them thinking about the future they envision and the steps they could take in the upcoming year at U of T to help make that future a reality. They were told that the exercise was designed for their benefit and to take their time while completing it. The online module lasted approximately 60 to 90 minutes and led students through a series of writing exercises in which they wrote about their ideal futures, both at work and at home, what they would like to accomplish in the current year at U of T, how they intend to follow certain study strategies to meet their goals, and whether they want to get involved with extracurricular activities at the university. Varying minimum word count and time restrictions were placed on several pages of the online exercise to ensure that students gave due consideration to each of their answers before proceeding.<sup>6</sup> The exercise aimed to make students' distant goals salient in the present and to provide information on effective study strategies and how to deal with the inevitable setbacks that arise during the course of an academic year. After the exercise, students were emailed a copy of the answers they provided to store for future reference. A full description of the online exercise is available in Online Appendix A.<sup>7</sup>

The online exercise is related to the one students completed in Morisano et al. (2010). In that study, 85 struggling students (those with GPAs below 3.0) at McGill University were recruited to take part in an experiment. Approximately half of the students were randomly sorted into a one-time, intensive, online goal-setting exercise that required 2.5 hours to complete, while other students were sorted to online control activities. The goal-setting exercise consisted of eight segments, which together engaged students to imagine their ideal futures, think about goals that could be pursued to realize that future, rank the importance of those goals, discuss the significance of achieving each goal, detail specific plans for the pursuit of each goal, and evaluate the degree of their commitment to these pursuits. The authors found that students who participated in the goal-setting exercise experienced 70 percent of a standard deviation increase in GPA over a four-month period, while the control group experienced no change.

During the 2014 to 2015 academic year, we ran a pilot study in which we randomly assigned students registered for first-year economics classes to the goal-setting exercise in Morisano et al. (2010). We did not find any discernable effects on students' academic outcomes for the goal-setting exercise, leading us to adopt a different model for the online intervention in the current paper. Based on feedback from students, instructors,

---

6. Nearly all students took the exercise seriously, writing coherent statements that served as logical answers to the relevant questions. There were very few instances of students writing random words to hit the word-count minimum, and the students who did only did so on some questions, not throughout the entire exercise.

7. Online appendixes can be found at <http://jhr.uwpress.org/>.

and administrative support staff, our current exercise is shorter and designed to get students thinking specifically about how attending U of T can help them realize their ideal futures. To that end, the second half of the module involves answering questions about study habits, effective study strategies, using university resources, and coping with stress.

### ***B. Treatment 2: Text Messaging Campaign***

As an extension of the one-time, online reflection of the academic environment at U of T, students in the second treatment group completed the same online exercise but were additionally offered the opportunity to provide their phone numbers and participate in a text and email messaging campaign lasting throughout both the fall semester in 2015 and the winter semester in 2016. The messaging campaign was called You@UofT—a name we chose to emphasize that the program was geared to provide personalized assistance and help students reach their individual definitions of success. Students had the opportunity to choose the frequency with which they received text and email messages, with choices including once a week, two to three times per week, and three or more times per week. All students who were randomly sorted into this treatment received email messages, while only those students who provided their cell phone numbers received text messages throughout the year.<sup>8</sup> Students were free to opt out of receiving email messages, text messages, or both at any time after the exercise, although few chose to do so.

A full documentation of all the text and email messages we sent throughout the experiment is available in Online Appendix B. Our messages typically focused on three themes: academic and study preparation advice, information on the resources available at the university, and motivation and encouragement. Students always received both a text and email message. Text messages were typically three to four lines in length, while emails were longer and provided more detailed information with which students could follow up. The You@UofT program offered personalized two-way communication, as both text and email messages regularly encouraged students to look further into the content and to reach out to us if they had specific or general questions. Approximately 25 percent of students engaged in two-way communication with our team via text messages, compared to only 3 percent of students responding via email.

There was wide variation in the types of response we received from students. For example, some students asked for locations of certain facilities on campus or how to stay on residence during the holiday break, while others said they needed help with English skills or specific courses. Some students expressed relatively deep emotions, such as feeling anxious about family pressure to succeed in school or from doing poorly on an evaluation. We also received messages of thanks for our appropriately timed advice or motivation, and several students messaged us to tell us how well they were doing in their courses and how much they appreciated the communication. No matter the type of message received and when, we attempted to provide a personalized support service, typically responding to the inquiry within a few hours (and usually within less than one hour). The You@UofT program served as a virtual coach from whom students could

---

8. A total of 2,024 students were randomly sorted into the messaging campaign treatment and 1,540 (76 percent) provided their phone numbers.

expect a rapid response at any time. In this sense, the program leveraged technology to provide a personalized coaching service at large scale to all students while keeping the cost per student lower than what is typically incurred with one-on-one coaching.

The design of our text messaging intervention is motivated by previous studies in which researchers directly communicate with students (rather than the parents of students) with the goal of improving student outcomes. Our intervention is most related to Castleman and Page (2014a, 2014b) and Castleman and Meyer (2016). Castleman and Page (2014a) use an automated and personalized text messaging campaign to help students enroll in college and avoid “summer melt” by providing them with ten text message reminders of the steps required for college matriculation after high school graduation, while Castleman and Page (2014b) send college freshman a series of 12 personalized text message reminders to encourage them to renew their financial aid by refile the Free Application for Federal Student Aid (FAFSA) form. The text messaging campaign designed to combat “summer melt” led to large increases in college enrollment among students with limited access to college-planning support, and the financial aid renewal campaign led to a 19 percent increase in sophomore year persistence in community college. These studies attempt to nudge students toward fulfilling a one-time action—enrolling in college or submitting a renewal form—while our text messaging campaign is geared toward improving students’ academic performance outcomes, which involves altering student behavior over a prolonged period and perhaps even creating new lifestyle habits.

As mentioned above, Castleman and Meyer (2016) attempt to do the same among low-income college students in West Virginia, although students are not randomly assigned to treatment in their sample. Freshman student participants in that study were sent three to four messages per month encouraging them to use campus resources, register for courses early, and reapply for financial aid. Students also received messages of general encouragement and affirmation during the transition to college. The authors find that students who received texts attempted more credits, but they do not find any discernable effects on GPA.

Our text messages are similar in spirit, although we send two to three messages per week and allow for two-way communication, encouraging students to either respond to our messages or initiate contact about a topic of their choice. Influencing academic outcomes likely involves influencing students’ behavior and helping them sustain the new habits over the course of many months. We therefore chose to send texts more frequently than prior studies to increase the salience of our messages. The opportunity for two-way communication is also an important distinguishing feature of our design, as it affords us the opportunity to mimic the interactions of a student and an in-person coach but at much lower cost.

### ***C. Treatment 3: In-Person Coaching***

To test how the effects of the You@UofT program compare to those of in-person, one-on-one coaching, a third group of students also completed the online exercise and was offered the opportunity to participate in a pilot project in which they would be assigned to an upper-year undergraduate student acting as a personal coach. Coaches were available to meet with students to answer any questions via Skype, phone, or in person,

and would send their students regular text and email messages of advice, encouragement, and motivation, much like the You@UofT program described above. In contrast to the messaging program, however, coaches were instructed to be proactive and regularly monitor their students' progress. Whereas the You@UofT program attempts to "nudge" students in the right direction with academic advice, coaches play a greater "support" role, sensitively guiding students through problems.

The coaching program was offered only to students at one of the university's satellite campuses, the University of Toronto at Mississauga campus. Our coaching treatment group was established by randomly drawing 24 students from the group of students that were randomly assigned into the text message campaign treatment. At the conclusion of the online exercise, instead of being invited to provide a phone number for the purpose of receiving text messages, these 24 students were given the opportunity to participate in a pilot coaching program. A total of 17 students agreed to participate in the coaching program, while seven students declined. These 17 students were assigned to a team of four upper-year undergraduate coaches, who participated in our program as part of a research opportunity program. Each coach originally agreed to coach six students throughout the academic year but was eventually responsible for only four or five students as result of seven students declining participation.

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

#### ***D. Control Group: Personality Test***

Students assigned to the control group were given a personality test measuring the Big Five personality traits. They were told the exercise is based on current research and gives a unique opportunity to learn more about personality traits. The test could be completed in approximately 45 to 60 minutes and, after the exercise, students were emailed their scores in a report describing how they fair on each of the Big Five traits.<sup>9</sup> Students were instructed that the report reveals rank-orders of the five traits, which may be interesting for knowing which are their most and least dominant traits.

---

9. Beattie, Laliberté, and Oreopoulos (2016) use the data resulting from the personality test exercise to explore nonacademic predictors of performance in university. That paper's appendix describes the exercise in more detail.



### III. Data Description and Empirical Strategy

In this section, we describe the data we collected from the experiment and how we estimate the effects of the three treatments.

#### A. Data Description

Our experiment is registered with the American Economic Association's registry for randomized controlled trials. Prior to the experiment, we intended to sort 30 percent of students into the control group, 20 percent into the online-exercise-only treatment, and 30 percent into the treatment group that received the text messaging campaign in addition to the online exercise. The remaining 20 percent of students were sorted into an online belonging exercise similar to the exercise that appeared in Walton et al. (2015). This intervention and its ineffectiveness will be the topic of a separate, standalone paper, and we therefore do not discuss this treatment in detail throughout the remainder of this paper. While students in this treatment are dropped from the main analysis, we do mention some of the key estimates of the effects of the belonging treatment in the results section below.

Students were sorted into one of the treatment groups or the control group according to the randomly generated last digits of their student numbers, which they provided upon registering online for our experiment.<sup>10</sup> As mentioned, we established the personal coach treatment group by drawing 24 students at random from the group of students that was intended to be a part of the text messaging campaign. Table 1 shows some basic statistics about our randomization strategy among first-year students, which indicate that we successfully reached each of our randomization targets.<sup>11</sup> Furthermore, high fractions of students completed each exercise, with completion rates ranging from 95 to 99 percent.

Our experimental sample consists of 5,179 students, where 1,820 were sorted into the control group, 1,311 were given the online exercise only, 2,024 were offered text message reminders, and 24 were matched a senior undergraduate coach. Most of our sample consists of freshman students, as 3,941 are in their first year of studies. We can match 4,926 students from our experimental data to the university's administrative records, although some students are missing the relevant GPA outcomes and others are missing campus and year-of-study indicators,<sup>12</sup> which leaves us with a final analysis sample consisting of 4,840 students (93 percent of the original experimental sample). Students with missing outcome data are excluded from the main analysis, but, as we show in Subsection IV.C, our results are robust to setting course-specific grades equal to zero for students with missing values.

Table 2 shows summary statistics for baseline characteristics among students in the control group along with differences between each treatment group mean and control

---

10. Since completing the exercise was a course requirement worth 2 percent of the final grade in introductory economics, students had a high incentive to provide their real student numbers and complete the exercise.

11. The table conditions on first-year students because the online belonging exercise mentioned above was only offered to students in their first year of study. Thus, students registered in second year or above are more likely (than the intended prerandomization fractions) to be in one of the other three treatment groups. We account for this in our empirical strategy below by including first-year fixed effects in every regression.

12. As we explain below, our randomization design requires that we control for campus location and year of study in our empirical strategy.

**Table 1**  
*Treatment Randomization among First-Year Students*

	Treatment Group			
	Control	Goal-Setting Alone	Goal-Setting with Reminders	Belonging
Number of students	1,455	964	1,522	972
(i) Fraction of total	0.30	0.20	0.30	0.20
(ii) Intended fraction	0.30	0.20	0.30	0.20
<i>p</i> -value of (i) = (ii)	0.55	0.50	0.14	0.70
Completed the exercise	1,440	922	1,466	960
Fraction selecting SMS frequency				
1 time a week			0.40	
2 times a week			0.19	
3+ per week			0.42	

Note: The intended fraction of students in the goal-setting with reminders group comprises the 24 students who were assigned to the coaching program.

group mean for each baseline characteristic. The treatment indicators are never jointly significant in explaining variation in any student characteristic. The only individual exceptions are that students in the online-only group are slightly more likely to live on residence and to be first-generation students while students in the personal coaching group report having a slightly less difficult time transitioning to university. We explore whether our main results are sensitive to the imperfect balancing by investigating heterogeneous treatment effects across several student subgroups (including those defined by these three variables) and additionally controlling for these variables as a robustness check.

In terms of the descriptive statistics, approximately half of our sample is female, and the average student 18.5 years old. Approximately half of the students are nonnative English speakers and half are not Canadian citizens. Only 30 percent of students live on residence, but this fraction is pulled down by the two satellite campuses of U of T, the Mississauga and Scarborough campuses, which are both commuter campuses. At the main campus, St. George, 40 percent of students live on residence. The main campus also has students with higher incoming high school average grades: while the average is 87 percent across all students, it is 90 percent at the St. George campus. Approximately 24 percent of students are first generation, and 43 percent have international status.

### ***B. Empirical Strategy***

Since we successfully randomized students into various treatment groups, we estimate the effects of each treatment by simply comparing mean outcomes in a regression framework. These estimates are “Intent to Treat” effects, each representing the average impact from being invited to complete the exercise, regardless of whether students completed or not.

**Table 2**  
*Summary Statistics and Balancing Tests*

Student Characteristics	Treatment Status				P-Value from F-Test of No Difference
	Control Sample	Online Only	Text Messaging	Personal Coaching	
		Mean [Standard Deviation]	Difference [Standard Error]	Difference [Standard Error]	
Female	0.524 [0.500]	0.011 [0.018]	0.013 [0.017]	0.098 [0.105]	0.700
Age in first year	18.511 [1.428]	0.025 [0.058]	0.064 [0.052]	-0.533 [0.332]	0.216
High school average grade	87.284 [5.122]	-0.093 [0.182]	-0.208 [0.164]	0.120 [1.004]	0.642
Non-English mother tongue	0.551 [0.498]	-0.002 [0.018]	-0.005 [0.016]	-0.064 [0.105]	0.932
Non-Canadian citizenship	0.541 [0.498]	0.002 [0.018]	-0.001 [0.017]	-0.069 [0.105]	0.926
Living in residence	0.300 [0.458]	0.035** [0.016]	0.019 [0.014]	-0.041 [0.090]	0.143

(continued)

**Table 2** (continued)

Student Characteristics	Treatment Status				P-Value from F-Test of No Difference
	Control Sample	Online Only	Text Messaging	Personal Coaching	
	Mean [Standard Deviation]	Difference [Standard Error]	Difference [Standard Error]	Difference [Standard Error]	
Expects to get more than undergraduate degree	0.657 [0.475]	0.008 [0.017]	0.007 [0.015]	0.069 [0.097]	0.87
First generation student	0.237 [0.425]	0.027* [0.016]	0.011 [0.014]	-0.036 [0.092]	0.375
Expect to receive A-average or more	0.659 [0.474]	0.008 [0.017]	0.022 [0.015]	-0.082 [0.097]	0.39
Hours expected to study	18.243 [10.890]	-0.323 [0.398]	-0.029 [0.355]	-0.175 [2.257]	0.855
Expect to work more than 8 hours/week	0.376 [0.485]	-0.002 [0.017]	-0.013 [0.016]	0.019 [0.099]	0.847
Economics is a required course	0.577 [0.494]	-0.001 [0.017]	-0.003 [0.015]	0.084 [0.096]	0.842

International student	0.436 [0.496]	-0.024 [0.018]	-0.016 [0.016]	-0.008 [0.102]	0.571
Have a clear vision for future (1-7)	4.842 [1.467]	0.023 [0.054]	-0.028 [0.048]	-0.299 [0.308]	0.603
Think about the future (1-7)	5.559 [1.226]	0.041 [0.045]	-0.003 [0.040]	-0.199 [0.256]	0.605
Tend to cram for exams (1-7)	4.105 [1.521]	-0.035 [0.055]	-0.008 [0.049]	-0.002 [0.312]	0.936
Identify with university (1-7)	5.646 [1.427]	-0.028 [0.053]	0.000 [0.047]	0.145 [0.300]	0.893
Transition has been so far challenging (1-7)	4.774 [1.670]	0.001 [0.061]	0.015 [0.054]	-0.782** [0.344]	0.146

Notes: Summary statistics and differences are calculated using the full sample of students. All differences are conditional on campus location and first-year fixed effects. Significant at \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Given that almost all students finished, however, the estimates are likely close to “Average Treatment” effects, measuring the average effect from completing the exercise for the entire sample. More formally, we estimate the following equation:

$$(1) \quad y_{ij} = \alpha + \beta_1 \text{Online}_i + \beta_2 \text{Text}_i + \beta_3 \text{Coach}_i + \delta_j + \mu \text{FY}_i + \epsilon_{ij},$$

where the outcome of student  $i$  who attends campus  $j$  is regressed on indicators for each of the three treatment exercises students were given, campus fixed effects, and a first-year student indicator. We include campus fixed effects because the coaching treatment was only offered at the Mississauga campus, which accepts students with lower high school averages who tend to perform worse in university than students who attend the main campus, St. George. We include the first-year indicator to account for students who are enrolled in second year and above being more likely to be in one of the three treatment groups than students in first year, as only first-year students were randomly assigned to the online belonging exercise that is not analyzed in this paper. The main parameters of interest are  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ , which represent the effect of the online treatment, the online plus messaging treatment, and the online plus coaching treatment, respectively. As mentioned, we include all students in the analysis, irrespective of whether they completed the online exercise, provided a cell phone number, or agreed to participate in the coaching program, implying that our parameter estimates all represent intent to treat effects.

Our main outcomes of interest are course grades, GPA, number of credits earned, and number of credits failed. When the outcome is course grades, we stack all of the reported course grades for a given student and run a regression at the student-course level in which we cluster the standard errors by student. For all other outcomes, we run the regression at the student level and report robust standard errors.

## IV. Results

### A. Main Results

Table 3 presents the results from stacked regressions at the student–course level, in which the dependent variable is a student’s course grade outcome, and we consider all courses, fall semester courses, and winter semester courses separately. Standard errors are clustered by student. The results in Column 1 use grades from all courses as the dependent variable and show that neither the online exercise on its own nor the online exercise and texting messaging treatment had any effect on course grades. The insignificant effects are not due to statistical imprecision. We can rule out impacts above 6 percent of a standard deviation using a 95 percent confidence interval. In contrast, the personal coaching treatment had relatively large effects, boosting the average course grade by 4.92 percentage points, which amounts to 30 percent of the control group standard deviation. Reassuringly, including student age and gender as additional control variables in Column 2 does not change the result. Columns 3 and 4 use a student’s course-specific grade point relative to the average course grade point as the dependent variable.<sup>13</sup> While the coaching effects

13. The number of observations differs from that in Columns 1 and 2 because we are missing course averages for some courses in the administrative data.

**Table 3**  
*Effects on Stacked Grades*

	All Course Grades				Fall Course Grades				Winter Course Grades			
	Grades		Grade Relative to Class Average		Grades		Grade Relative to Class Average		Grades		Grade Relative to Class Average	
	No Controls (1)	With Controls (2)	No Controls (3)	With Controls (4)	No Controls (5)	With Controls (6)	No Controls (7)	With Controls (8)	No Controls (9)	With Controls (10)	No Controls (11)	With Controls (12)
Online only	0.469 [0.471]	0.462 [0.472]	0.063 [0.041]	0.063 [0.041]	0.143 [0.575]	0.143 [0.575]	0.024 [0.043]	0.024 [0.043]	0.596 [0.607]	0.581 [0.607]	0.111* [0.057]	0.111* [0.057]
Text messaging	-0.116 [0.424]	-0.132 [0.423]	-0.002 [0.038]	-0.003 [0.038]	0.073 [0.505]	0.075 [0.506]	0.016 [0.038]	0.016 [0.038]	-0.163 [0.549]	-0.197 [0.549]	-0.020 [0.053]	-0.021 [0.053]
Coaching	4.920** [1.930]	4.876** [1.927]	0.270 [0.175]	0.273 [0.174]	4.897*** [1.874]	4.935*** [1.880]	0.269 [0.167]	0.267 [0.168]	6.659*** [2.107]	6.525*** [2.094]	0.431*** [0.207]	0.431*** [0.207]
Control mean	68.905 [16.166]	68.905 [16.166]	0.077 [1.181]	0.077 [1.181]	69.428 [15.744]	69.428 [15.744]	0.066 [1.174]	0.066 [1.174]	68.826 [16.820]	68.826 [16.820]	0.073 [1.176]	0.073 [1.176]
Observations	29,591	29,591	18,162	18,162	10,106	10,106	10,065	10,065	10,554	10,554	5,864	5,864
R-squared	0.021	0.022	0.013	0.014	0.029	0.029	0.017	0.017	0.015	0.016	0.014	0.014
F-stat	2.691	2.675	1.760	1.792	2.290	2.311	0.910	0.895	3.918	3.865	3.352	3.360
Prob > p	0.0447	0.0456	0.153	0.146	0.0763	0.0742	0.435	0.443	0.00833	0.00897	0.0183	0.0181

Notes: The dependent variable in each regression is indicated by the column headings. The unit of observation is a student-course. All regressions control for campus and first-year fixed effects. Additional control variables include age in first year and gender. Standard errors clustered at the student level are reported in brackets. Significant at \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

are not statistically significant, they are substantially larger than the effects of the online exercise or the messaging campaign, each of which are zero.

Columns 5 to 8 consider grades from courses taken only in the fall semester as the outcome of interest. The coaching effects on fall grades are slightly weaker than those on grades from all courses, but coached students still earn higher grades, on average. Columns 9–12 show treatment effects on grades from courses taken only in the winter semester. Here, the coaching treatment effects are even stronger, as coaching boosts the average grade by 6.6 percentage points (or 39 percent of the control group standard deviation). Students in the coaching treatment also tend to earn higher relative grades in their winter courses, scoring approximately 0.43 grade points higher than the average student in their courses. It thus appears that the effects of the coaching treatment strengthened over time. It may be the case that students developed more trust with their coaches as the academic year progressed and that they learned how to use resources more effectively.

Figures 1 to 3 show graphically the effects of the coaching treatment strengthening over time. Each figure shows the treatment-group-specific distributions of residual grades, after campus and first-year effects are removed. Figure 1 reports the residual-grade distributions for all courses (full-year, winter semester, and fall semester) and clearly shows that the coaching distribution is shifted rightward relative to the control group distribution and the distributions for the online and texting treatments. Indeed, a Kolmogorov–Smirnov test rejects that the coaching distribution is the same as the control and texting distributions at the 1 percent level and the online-only distribution at the 5 percent level. Contrasting Figures 2 and 3 shows that the strongest coaching effects emerge in the winter semester, as the coaching distribution's rightward shift relative to the other distributions is much more pronounced in the winter semester in Figure 3 than in the fall semester in Figure 2.<sup>14</sup>

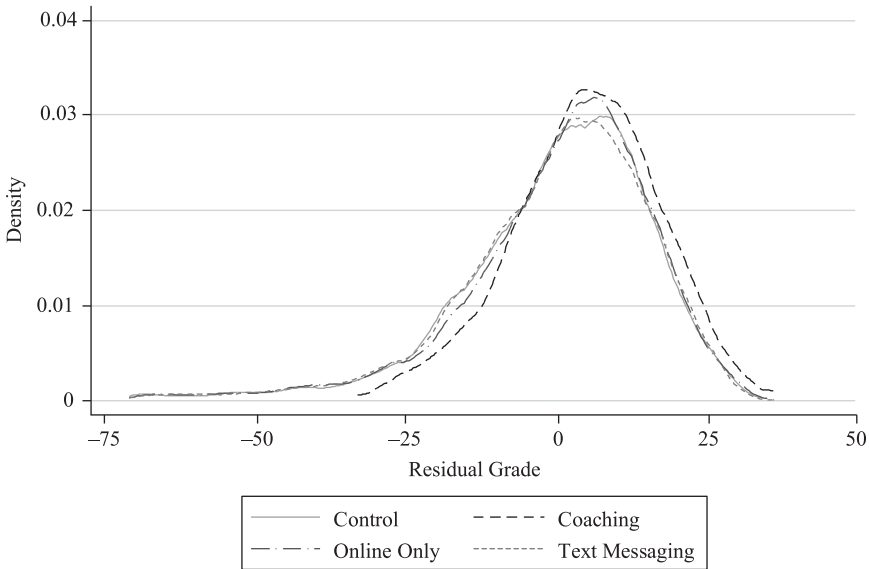
In Oreopoulos and Petronijevic (2016), we show with linear regressions that the coaching treatment decreases the likelihood of students earning extremely low grades and that these effects strengthen over time.<sup>15</sup> In particular, coached students are 8 percentage points less likely to earn a grade below 60 percent across all courses and are 12 percentage points less likely to earn a grade below 60 percent in winter semester courses. In the winter semester, coached students are also 16 percentage points more likely to earn a grade above 75 percent.

Table 4 shows treatment effects on other academic outcomes with one observation per student and student-level regressions. The dependent variables are constructed using outcomes from all courses. The coaching treatment causes a 0.35 grade point increase in student GPA, equivalent to approximately 35 percent of the control group standard deviation. Coached students failed fewer credits and earned more credits, on average, than students in the control group. As with stacked grade outcomes, there are no detectable effects on GPA or the number of credits failed or earned from the online exercise treatment or the text messaging campaign. Although we do not report these results

14. Note that the coaching group density for fall grades in Figure 2 does not have overlapping mass with the other densities in the left tail of the grade distributions. Thus, while the coaching program does not cause a pronounced shift of the grade distribution in the fall, it does appear to prevent students from earning extremely low grades, as the grade distribution is truncated at a residual grade of  $-14.6$ .

15. These results can be found in Tables 3 to 5 of Oreopoulos and Petronijevic (2016).





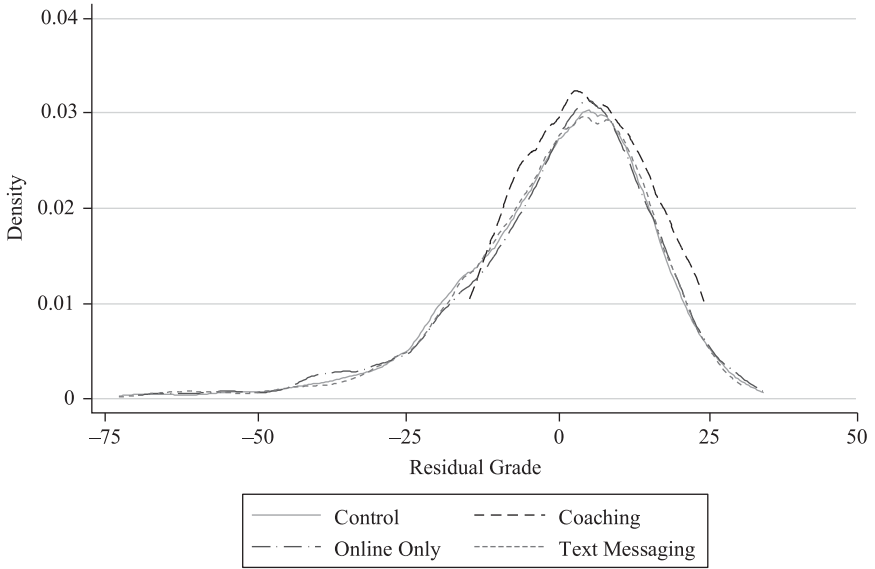
**Figure 1**  
*Grade Distributions across All Courses by Treatment Status*

This figure presents residual grade distributions using grade outcomes from all (full-year, winter, and fall) courses. To construct the figure, we stack course grade outcomes for all students from all courses, regress course grades on campus and first-year fixed effects, and obtain the residuals from this regression. The figure shows the density of these residuals for each of the three treatment groups and the control group.

separately, the effects of the coaching treatment on these outcomes are again stronger in the winter semester than in the fall semester.

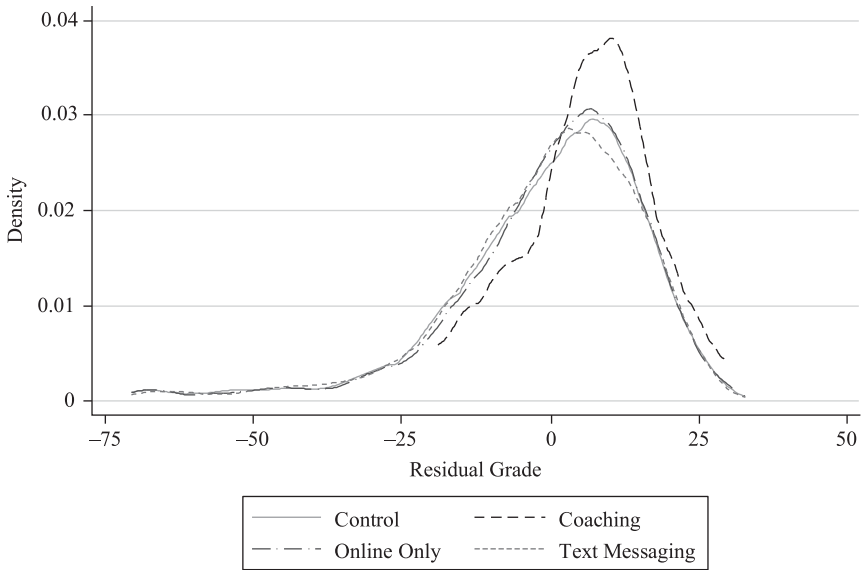
As mentioned above, 20 percent of our initial experimental sample (972 students) was sorted into an online belonging exercise similar to that in Walton et al. (2015). Although we intend to discuss this treatment and its results in a separate paper, we briefly mention some of the key estimates of its effects. Much like the online exercise and the text messaging campaign, the belonging treatment appears to have had no effect on student outcomes.<sup>16</sup> The effects on course grades, averaged from all courses, fall courses, and winter courses are 0.5 grade points, 1.02 grade points, and 0.4 grade points, respectively (with a standard deviation of about 16 points). Referring to Table 3, the estimates for all courses and winter courses are nearly identical to those for the online exercise and are indistinguishable from zero. The effect of 1.02 grade points during the fall semester is significant at the 10 percent level and corresponds to approximately 6 percent of a standard deviation. The same patterns emerge when the outcome is student GPA, as the

16. We explore heterogeneous treatment effects for the online-only, text messaging, and coaching interventions below. Although the results are not reported, the belonging treatment also has no effect on most student subgroups, with possible exceptions being on the fall course grades of students who are less than 20 years old and those who are in their first year of studies.



**Figure 2**  
*Grade Distributions across Fall Courses by Treatment Status*

This figure presents residual grade distributions using grade outcomes from fall courses only. To construct the figure, we stack course grade outcomes for all students from all fall courses, regress course grades on campus and first-year fixed effects, and obtain the residuals from this regression. The figure shows the density of these residuals for each of the three treatment groups and the control group.



**Figure 3**  
*Grade Distributions across Winter Courses by Treatment Status*

This figure presents residual grade distributions using grade outcomes from winter courses only. To construct the figure, we stack course grade outcomes for all students from all winter courses, regress course grades on campus and first-year fixed effects, and obtain the residuals from this regression. The figure shows the density of these residuals for each of the three treatment groups and the control group.

**Table 4**  
*Effects on All Courses Outcomes*

	GPA		Credits Failed		Credits Earned	
	No Controls (1)	With Controls (2)	No Controls (3)	With Controls (4)	No Controls (5)	With Controls (6)
Online only	0.022 [0.037]	0.021 [0.037]	-0.012 [0.027]	-0.012 [0.026]	-0.039 [0.050]	-0.037 [0.050]
Text messaging	-0.019 [0.033]	-0.021 [0.033]	0.008 [0.024]	0.009 [0.024]	-0.042 [0.045]	-0.040 [0.044]
Coaching	0.358** [0.182]	0.345* [0.183]	-0.143 [0.152]	-0.136 [0.117]	0.501* [0.283]	0.429* [0.247]
Control mean [Standard deviation]	2.380 [0.990]		0.346 [0.717]		3.684 [1.360]	
Observations	4,840	4,840	4,840	4,840	4,840	4,840
R-squared	0.039	0.042	0.010	0.012	0.094	0.114
F-stat	1.750	1.660	0.496	0.670	1.520	1.461
Prob > p	0.155	0.173	0.685	0.570	0.207	0.223

Notes: The dependent variable in each regression is indicated by the column headings. All regressions are run at the student level and control for campus and first-year fixed effects. Additional control variables include age in first year and gender. Robust standard errors are reported in brackets. Significant at \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

effects of the belonging exercise are indistinguishable from those of the online exercise for all courses and winter courses. In the fall semester, the belonging treatment appears to have improved GPA by 7 points (6 percent of a standard deviation), on average, but the effect is insignificant at conventional levels. Given the negligible effects of the belonging exercise, we focus the remainder of the paper on the coaching results and contrasting them with the online exercise and text messaging campaign.

### **B. Heterogeneous Treatment Effects**

In this section, we explore heterogeneous treatment effects across different student subgroups and across the three U of T campuses. As mentioned above, only 24 students are in the coaching treatment. We therefore investigate the heterogeneous effects of coaching along with the other treatments only for completeness; with a sample size of only 24 students, we lack the necessary power to meaningfully distinguish potential differences in the effects of coaching across different subgroups.

Table 5 shows treatment effects on all course grades across a variety of student subgroups. The effects of the online-only and text messaging treatments are not statistically significant for any type of student, with the lone exception being a small

**Table 5**  
*Effects on All Grades across Student Subgroups*

	Gender		Age		High School Grade		Mother Tongue	
	Female (1)	Male (2)	20 or Older (3)	19 or Younger (4)	Above Campus Median (5)	Below Campus Median (6)	English (7)	Other (8)
Online only	-0.083 [0.587]	1.120 [0.757]	-1.518 [1.357]	0.804 [0.499]	-0.618 [0.659]	0.664 [0.796]	1.245* [0.660]	-0.185 [0.665]
Text messaging	-0.493 [0.539]	0.339 [0.668]	-0.366 [1.177]	-0.123 [0.451]	-0.518 [0.614]	-0.235 [0.687]	0.150 [0.613]	-0.364 [0.584]
Coaching	3.176 [2.208]	6.728* [3.455]	7.036 [5.859]	4.755** [2.132]	2.242 [2.611]	4.829* [2.646]	5.104 [3.269]	4.834*** [1.799]
Control mean [Standard deviation]	69.804 [14.958]	68.070 [17.264]	67.208 [17.952]	69.186 [15.836]	72.492 [14.564]	63.942 [16.617]	68.642 [16.119]	69.121 [16.203]
Obs.	16,016	13,575	4,500	25,091	11,441	11,090	13,293	16,298
R-squared	0.017	0.027	0.018	0.024	0.031	0.011	0.027	0.019
F-stat	1.113	1.863	1.016	2.858	0.702	1.519	2.087	2.786
Prob > p	0.343	0.134	0.385	0.0357	0.551	0.208	0.0999	0.0394

Notes: The dependent variable in each regression is course grades. The unit of observation is a student-course. All regressions control for campus fixed effects, first-year status, age in first year, and gender. Standard errors clustered at the student level are reported in brackets. Significant at \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

positive effect of the online-only treatment on students whose mother tongue is English. We find that coaching effects are stronger for men, students who are 20 years of age or older, first-generation students, and students who are not in first year. Given the small coaching treatment sample size, however, we are hesitant to push these results further without investigation on a bigger a sample of students.

We also investigate whether there are heterogeneous treatment effects across the three U of T campuses. Table 6 shows the effects of the three treatments on grade outcomes from all courses at the Mississauga, Scarborough, and St. George campuses, respectively. The effects of the coaching treatment are only reported among students attending the Mississauga campus, as only these students were randomly offered the coaching service. Neither the online exercise nor the text messaging campaign had any effect on student grade outcomes at the main campus, St. George, or at the Scarborough campus.<sup>17</sup>

The estimated effects of the online exercise and the texting campaign at the Mississauga campus are larger than those found in the pooled sample and at the other two

17. Table 6 shows treatment effects on all course grades. While we do not report the results, there are also virtually no effects on grades from full-year courses, winter courses, and fall courses at both the St. George and Scarborough campuses.

International Student		First Generation		First-Year Student		Lives on Residence		Transition Difficulty	
Yes (9)	No (10)	Yes (11)	No (12)	Yes (13)	No (14)	Yes (15)	No (16)	Above Median (17)	Below Median (18)
0.194 [0.770]	0.638 [0.594]	0.720 [1.045]	0.515 [0.560]	0.793 [0.504]	-1.203 [1.201]	0.158 [0.726]	0.444 [0.601]	0.051 [0.804]	0.757 [0.570]
-0.756 [0.688]	0.333 [0.534]	0.606 [0.867]	-0.380 [0.512]	-0.013 [0.462]	-0.762 [1.032]	-0.442 [0.693]	-0.064 [0.525]	0.700 [0.732]	-0.460 [0.515]
5.665*** [2.135]	4.486 [2.877]	7.504*** [1.623]	3.273 [2.770]	3.061 [2.086]	8.579** [3.965]	6.533*** [1.869]	4.726** [2.182]	5.075 [3.128]	4.433* [2.349]
68.998 [16.524]	68.998 [15.766]	68.038 [15.977]	69.470 [16.170]	69.506 [15.733]	66.192 [17.744]	71.748 [14.730]	67.650 [16.608]	67.299 [16.891]	69.916 [15.570]
12,321 0.021 3.234 0.0215	17,270 0.026 1.092 0.351	6,631 0.020 7.430 6.29e-05	20,702 0.023 1.324 0.265	23,539 0.013 1.701 0.165	6,052 0.021 2.235 0.0826	9,373 0.020 4.677 0.00295	20,218 0.014 1.801 0.145	10,574 0.041 1.173 0.319	19,017 0.015 2.713 0.0434

campuses separately. Columns 1 and 2 of Table 6 show that the online exercise boosts course grades by 1.93 percentage points, on average. The estimate is significant at the 10 percent level and implies that the online exercise increases grades by 11 percent of the control group standard deviation. This is a relatively small effect when compared to the coaching treatment, which increases grades by 5.95 percentage points, or 35 percent of a standard deviation. In Oreopoulos and Petronijevic (2016), we show that the online exercise also boosts the likelihood that students earn a grade above 80 percent and decreases the likelihood of earning a grade below 60 percent, but the effects are again smaller than those from the coaching treatment.<sup>18</sup>

### C. Robustness

In this subsection, we conduct sensitivity analysis, showing our results are robust to the inclusion of additional control variables, alternative parametrization of the treatment indicators, and missing outcome data.

18. See Table 10 of Oreopoulos and Petronijevic (2016).

**Table 6**  
*Effects on All Course Grades by Campus*

	Mississauga Campus				Scarborough Campus				St. George (Main) Campus			
	Grades		Grade Relative to Class Average		Grades		Grade Relative to Class Average		Grades		Grade Relative to Class Average	
	No Controls (1)	With Controls (2)	No Controls (3)	With Controls (4)	No Controls (5)	With Controls (6)	No Controls (7)	With Controls (8)	No Controls (9)	With Controls (10)	No Controls (11)	With Controls (12)
Online only	1.905* [0.989]	1.958** [0.984]	0.143** [0.068]	0.146** [0.068]	0.950 [0.957]	0.936 [0.959]	0.074 [0.072]	0.077 [0.072]	-0.386 [0.645]	-0.387 [0.645]	-0.069 [0.066]	-0.069 [0.065]
Text messaging	1.453 [0.931]	1.486 [0.924]	0.079 [0.065]	0.081 [0.064]	-0.957 [0.912]	-1.008 [0.913]	-0.065 [0.068]	-0.063 [0.068]	-0.362 [0.554]	-0.362 [0.554]	-0.017 [0.054]	-0.019 [0.054]
Coaching	5.974*** [2.013]	5.930*** [2.024]	0.331* [0.180]	0.330* [0.179]	—	—	—	—	—	—	—	—
Control mean	64.762 [17.069]	-0.009 [1.202]	68.894 [15.988]	0.162 [1.178]	70.694 [15.439]	0.081 [1.140]	—	—	—	—	—	—
Observations	6,583	6,583	6,563	6,563	7,162	7,162	7,089	7,089	15,846	15,846	4,510	4,510
R-squared	0.016	0.025	0.022	0.028	0.005	0.006	0.009	0.010	0.009	0.009	0.009	0.011
F-stat	3.507	3.512	2.230	2.296	2.008	2.088	1.813	1.862	0.273	0.274	0.563	0.571
Prob > p	0.0149	0.0148	0.0830	0.0761	0.135	0.124	0.164	0.156	0.761	0.761	0.569	0.565

Notes: The dependent variable in each regression is indicated by the column headings. The unit of observation is a student—course. All regressions control for first-year fixed effects. Additional control variables include age in first year and gender. Standard errors clustered at the student level are reported in brackets. Significant at \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

The summary statistics in Table 2 indicate that students in the online-only group are slightly more likely to live on residence and to be first-generation students while students in the personal coaching group report having a slightly less difficult time transitioning to university. We include these three variables in our control vector and rerun the analysis on grades from all courses, winter courses, and fall courses. The results are presented in Columns 1, 3, and 5 of Table 7 and are very similar to the main results. Students who only completed the online exercise or participated in the text messaging campaign did not experience higher grades, while the effect of the personal coaching treatment remains large and significant.

In our main specifications, the online-only treatment is considered separately from the messaging and coaching treatments. However, all three treatment groups participated in the online exercise, implying that it may be warranted to consider all three groups as having received the online exercise in the empirical specification. Since the effects of the online exercise are often positive (although rarely significant, except for students at the Mississauga campus) estimating the effects of the other treatments as the gain that occurs above the online exercise may render the estimates of the coaching effects statistically insignificant. We test whether this is the case by rerunning the analysis after setting the online-only treatment indicator equal to one for students in the text messaging and personal coaching groups. The estimated treatment effects on all course grades, winter course grades, and fall course grades are presented in Columns 2, 4, and 6 of Table 7. The coaching effects remain significant at the 10 percent level for all courses and at the 5 percent level for winter and fall courses.

As a final robustness check, we show that missing outcome data do not drive our results. Since we have grade outcome data for 93 percent of the original experimental sample, it would be surprising if our conclusions were sensitive to our decision to simply exclude missing student observations from the main analysis. We verify that this is the case by setting course-specific grades equal to zero when they are missing for a given student and then using these student–course observations in the regression analysis. The estimated effects on all course grades, winter course grades, and fall course grades are presented in Columns 7 to 12 of Table 7. The online exercise and text messaging campaign are still estimated to have no effect on student outcomes, while the effects of the coaching treatment are large and significant (except on grades from fall courses). The coaching effects on all courses and winter courses are larger than our main estimates, although they are not statistically distinguishable.

We also estimated the effects of each treatment on students' grades in their first-year economics courses—the courses through which they were introduced to our online platform and assigned to one of the treatment arms. Although we do not report the results,<sup>19</sup> treatment effects in economics courses are virtually identical to those estimated with all course grades: students in the online and texting groups did not experience any gains, while coached students experienced a 5.23 percentage-point boost to their economics grade, equivalent to 31 percent of a standard deviation.

As mentioned, our coaching treatment has a sample size of only 24 students, which is small, especially relative to the control group. Angrist and Pischke (2009) and Imbens and Kolesar (2016) discuss how these factors may cause bias in conventional standard

---

19. Additional results are available upon request.

**Table 7**  
*Robustness Checks*

	All Grades		Winter Grades		Fall Grades		All Grades (Zero if Missing)		Winter Grades (Zero if Missing)		Fall Grades (Zero if Missing)	
	Main Param. (1)	Alt. Param. (2)	Main Param. (3)	Alt. Param. (4)	Main Param. (5)	Alt. Param. (6)	Main Param. (7)	Alt. Param. (8)	Main Param. (9)	Alt. Param. (10)	Main Param. (11)	Alt. Param. (12)
Online	0.390 [0.491]	0.390 [0.491]	0.476 [0.636]	0.476 [0.636]	-0.018 [0.598]	-0.018 [0.598]	0.063 [0.579]	0.063 [0.579]	0.045 [0.770]	0.045 [0.770]	0.325 [0.716]	0.325 [0.716]
Text	-0.253 [0.439]	-0.644 [0.480]	-0.271 [0.572]	-0.747 [0.609]	-0.001 [0.520]	0.018 [0.593]	-0.496 [0.504]	-0.559 [0.568]	-0.762 [0.673]	-0.807 [0.754]	-0.133 [0.622]	-0.457 [0.707]
Coaching	4.228** [1.966]	3.838* [1.972]	5.532** [2.190]	5.055** [2.193]	3.852** [1.828]	3.870** [1.848]	7.248** [2.560]	7.185*** [2.571]	8.785*** [3.212]	8.740*** [3.223]	4.713 [3.153]	4.388 [3.164]
Control Mean	68.905	68.905	68.826	68.826	69.428	69.428	68.905	68.905	68.826	68.826	69.428	69.428
[Standard Deviation]	[16.166]	[16.166]	[16.820]	[16.820]	[15.744]	[15.744]	[16.166]	[16.166]	[16.820]	[16.820]	[15.744]	[15.744]
Observations	27,333	27,333	9,744	9,744	9,311	9,311	29,563	29,563	10,262	10,262	9,788	9,788
R-squared	0.033	0.033	0.025	0.025	0.048	0.048	0.022	0.022	0.031	0.031	0.039	0.039
F-stat	2.159	2.159	2.685	2.685	1.523	1.523	3.302	3.302	3.292	3.292	0.882	0.882
Prob > p	0.0908	0.0908	0.0450	0.0450	0.207	0.207	0.0195	0.0195	0.0197	0.0197	0.450	0.450

Notes: All regressions control for campus fixed effects, first-year fixed effects, age in first year, gender, an indicator for whether a student lives on residence, an indicator for whether a student is first generation, and a variable indicating (on a seven-point scale) the difficulty of transitioning to university. Standard errors clustered at the student level are reported in brackets. Significant at \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .



error estimates, both suggesting the “HC2” and “HC3” corrections as potential solutions. In addition to the robust standard errors reported in the paper, we have calculated conventional (unadjusted) standard errors, HC2 corrected standard errors, and HC3 corrected standard errors. Following Angrist and Pischke (2009), and using the maximum of all four standard error estimates, we continue to estimate statistically significant effects of the coaching treatment on GPA from all courses and winter courses at the 10 percent level, suggesting that our results are not biased by the small sample size.<sup>20</sup>

In sum, there is robust evidence that neither the online exercise nor the text messaging campaign were effective at improving students’ academic outcomes, both in the general population and across various student subgroups. The lone exception may be the positive effects of the online exercise among students at the Mississauga campus, although these effects are small relative to the coaching treatment. They are also larger in magnitude than the effects of the text messaging treatment, calling into question whether their statistical significance is due to real treatment effectiveness or random chance. In contrast to the one-time online intervention and the consistent-contact text messaging campaign, we find economically and statistically significant effects of the personal coaching treatment and a wide variety of academic outcomes. We discuss the potential reasons why the coaching treatment was more effective than the other two treatments and how the text messaging campaign can be adjusted to improve its effectiveness in the following section.

## V. Discussion

We find that neither the one-time online intervention nor the text messaging campaign have significant effects on student outcomes, while personal coaching boosts students’ grades and GPA by approximately 35 percent of a standard deviation. The key disadvantage of our coaching programs—and others like it—is that it is costlier to implement and scale up than one-time online interventions or interventions that rely heavily on technology for constant contact with students.

Although our upper-year coaches participated in the experiment as part of a research opportunity program (for course credit), such students would typically require at least \$20 per hour from the university to provide coaching services. With each of our coaches devoting approximately seven total hours per week to coaching, this conservative wage rate implies that the coaching program would regularly cost over \$13,000 to service 17 student participants. In contrast, after the initial setup costs, the online intervention is done at no additional cost and the total cost of the messaging campaign that serviced more the 1500 student participants was approximately \$1,200 for the entire academic year. Given the large differences in relative costs, it is worth discussing the key differences between the coaching treatment and the text messaging campaign, with the goal of learning how to modify the texting initiative to increase its effectiveness.<sup>21</sup>

---

20. The effects on GPA from fall courses are only significant at the 13 percent level when the most conservative standard error estimate is used. For the student-course level regressions, in which we estimate treatment effects on students’ course grades, we continue to use standard errors clustered at the student level, as these are likely most appropriate, given the repeated observations per student.

21. An alternative way to reduce the costs of one-on-one coaching is to recruit upper-year undergraduates to volunteer their time as coaches, with the promise of gathering valuable experience to place on a resume.

A common characteristic across many successful coaching programs is regular student–coach interaction facilitated either by mandatory meetings between coaches and students or proactive coaches who regularly initiate contact (Scrivener and Weiss 2013; Bettinger and Baker 2014; Cook et al. 2014; Oreopoulos, Lavecchia, and Brown forthcoming). Indeed, having proactive coaches is the key difference between our coaching service and that offered in Angrist, Lang, and Oreopoulos (2009), which was also conducted at the Mississauga campus of U of T but resulted in negligible effects. In that study, one treatment arm had student coaches email once a week and offer to meet at a student service office. In contrast, this study had student coaches aggressively initiate contact and build trust with students over time, in person, and through text. Our coaches were able to clearly understand the problems students were facing through a series of open-ended gentle questions. Upon understanding the problem, the coaches could provide clear advice, resulting in most conversations ending with students knowing at least one specific action to take to help them solve their current problems.

Our text messaging campaign sent weekly messages of academic advice, resource information, and motivation, but we did not initiate contact with individual students to specifically ask how they were doing or whether they wanted to talk about something specific. The text messaging team often did invite students to reply to our texts and share their concerns, but we were unable to do this from the perspective of a coach—that is, we did not present ourselves as a real person (with a name), and we did not try to establish a rapport with the students. Our inability to reach out to all students and softly guide the conversation likely prevented us from learning the key details of their specific problems. Although we would provide answers and advice to the questions we received, we did not have as much information on the students' backgrounds as our coaches did, and thus could not tailor our responses to each student's specific circumstances.

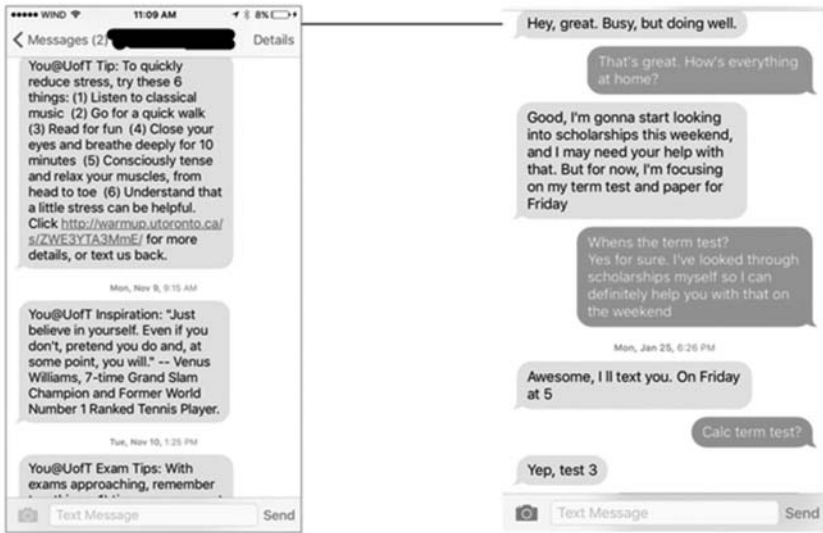
Our coaches were also able to build trust with their students by often fulfilling a support role for them. Figure 4 provides an example of how the coaching service was far more effective than the text messaging campaign in this respect. The text messages attempted to “nudge” students in the right direction, rather than provide tailored support. The left panel of Figure 4 shows three consecutive text messages, in which we provide a tip on stress management, an inspirational quote, and a time-management tip around the exam period. As shown in this example, it was very often the case that students would not respond to such messages. In contrast, the student–coach interaction in the right panel of Figure 4 shows our coaches offering more of a supportive role to students rather than trying to simply nudge them toward a certain path. The coach starts by asking an open-ended question, to which the student responds and guides the conversation forward. In this particular example, the coach assures the student that they will be available to help with a pending deadline and shows a genuine interest in the events in the student's life.

Coaches also kept a record of their evolving conversations with each student and could check in with students to ask how previously discussed issues were being resolved. Although we kept a record of all text message conversations we had with students, a lack of resources did not allow us to regularly check in with students to see how previous events had unfolded. A lack of these regular checkups likely prevented

---

Universities can help make this type of volunteer work attractive by creating a system that official recognizes students volunteer investments. The U of T Co-Curricular Record, for example, is designed to give students explicit credit for their experiences outside of the classroom (<https://ccr.utoronto.ca/home.htm>).

## ‘Nudge’ vs. ‘Support’



**Figure 4**  
*Contrasting Text Messaging and Coaching*

us from helping students effectively with the problems they told us about and from establishing the trust required for students to share additional problems.

In sum, the two key features that distinguish the coaching service from the texting campaign are that coaches proactively initiated discussion with students about their problems and could establish relationships based on trust in which students felt comfortable to discuss their issues openly. The ability of our coaches to slowly guide the conversations and inquire about previously discussed events likely contributed in large part to establishing the required trust between students and coaches.

Future work that attempts to improve academic outcomes in higher education with interventions that use technology to maintain constant contact with students should keep in mind that simply nudging students in the right direction may not be enough. Prior work on text message interventions with college students—most notably, Castleman and Page (2014a, 2014b)—finds that a relatively small number of text messages can lead to increased college matriculation and renewal of financial aid. However, these interventions target significantly different outcomes. Filling out the forms for college enrollment or FAFSA renewal requires students to take the necessary action once, whereas improving academic performance requires that students fundamentally alter their study habits and sustain their efforts over a period of several months.<sup>22</sup>

22. While they analyze a low-income population in West Virginia, Castleman and Meyer (2016) also find no effects of a text messaging campaign on student GPA, which supports our conclusion that simple nudges are ineffective at improving college performance.

When the goal is to improve students' academic performance, a more personalized approach may be required, in which coaches or mentors initially guide students through a series of gentle conversations and subsequently show a proactive interest in students' lives. These conversations need not necessarily occur during face-to-face meetings, but the available evidence suggests that they should occur frequently and be initiated by the coaches. While such an intervention is likely to be costlier than the text messaging campaign in this study, it is also likely to be more effective than the current messaging campaign but still less costly than the personalized coaching treatment analyzed here.

## VI. Conclusion

Building on recent insights from social psychology and behavioral economics, we estimated the effects of the following three treatments on students' academic outcomes: (i) an online exercise designed to affirm students' values and goals in university, (ii) a two-way text and email messaging campaign that provided students with academic advice and motivation, and (iii) a one-on-one coaching program in which students were matched with upper-year undergraduate students acting as coaches. We found no effect from either the online exercise or the messaging campaign, across the general student population and many student subgroups. In contrast, we found large and significant effects from the coaching program, which increased average course grades by 0.3 standard deviations and GPA by 0.35 standard deviations. Coached students also failed fewer credits and earned more credits, on average.

Contrasting the designs of the text messaging campaign treatment and the coaching treatment, we argued that coaches being proactive in contacting students was a critical feature of the program's success. Coaches were also better able to keep an account of previously discussed issues, subsequently inquire about how those issues were being resolved, and build the required trust that made students feel comfortable enough to keep returning for help.

While our results are robust to additional controls, alternative parameterizations of treatment status, missing observations, and alternative standard error calculations, we continue to interpret the estimated effects of the coaching treatment with caution, given a coaching sample size of only 24 students. However, we do interpret the results as being in line with previous studies that find benefits from personal coaching, such as Bettinger and Baker (2014).

As a followup to explore our conclusions that personalization is critical for coaching effectiveness, we are conducting a modified text messaging treatment that uses "virtual" coaches, each assigned to about 100 students, to interact separately with each assigned student. We are also expanding the sample size of the personal coaching treatment to include ten personal coaches, who are each assigned six students, in an attempt to replicate the effects of one-on-one coaching with a larger sample.

Technology-based interventions are attractive because they are relatively inexpensive and scalable, and can be implemented across a wide range of settings. With continued research, there exists real potential for developing a consensus around the types of student services that are most effective for improving student progress and wellbeing.

## References

- Ander, Roseanna, Jonathan Guryan, and Jens Ludwig. 2016. "Improving Academic Outcomes for Disadvantaged Students: Scaling Up Individualized Tutorials." Report prepared for the Brookings Institute. Washington, DC: Brookings Institute.
- 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):136–63.
- Angrist, Joshua David, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- 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. 2016. "Thrivers and Divers: Using Non-Academic Measures to Predict College Success and Failure." NBER Working Paper 22629. Cambridge, MA: National Bureau of Economic Research.
- Belley, Philippe, and Lance Lochner. 2007. "The Changing Role of Family Income and Ability in Determining Educational Achievement." *Journal of Human Capital* 1(1):37–89.
- Bergman, Peter. 2016. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment." Columbia University Working Paper. New York, Columbia University.
- 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):1205–42.
- Bloom, Benjamin. 1984. "The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring." *Educational Researcher* 13(6):4–16.
- Bound, John, Michael Lovenheim, and Sarah E. Turner. 2012. "Increasing Time to Baccalaureate Degree in the United States." *Education Finance and Policy* 7(4):375–424.
- Bound, John, and Sarah E. Turner. 2007. "Cohort Crowding: How Resources Affect Collegiate Attainment." *Journal of Public Economics* 91(5–6):877–99.
- 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.
- Castleman, Benjamin, and Lindsay Page. 2014a. "Summer Nudging: Can Personalized Text Messages and Peer Mentor Outreach Increase College Going Among Low-Income High School Graduates?" Center for Education Policy and Workforce Competitiveness Working Paper 9. Charlottesville, VA: University of Virginia.
- . 2014b. "Year Financial Aid Nudges: An Experiment to Increase FAFSA Renewal and College Persistence." Center for Education Policy and Workforce Competitiveness Working Paper 28. Charlottesville, VA: University of Virginia.
- Cohen, Geoffrey, and Julio Garcia. 2014. "Educational Theory, Practice, and Policy and the Wisdom of Social Psychology." *Policy Insights from the Behavioral and Brain Sciences* 1(1):13–20.
- Cohen, Geoffrey, and David Sherman. 2014. "The Psychology of Change: Self-Affirmation and Social Psychological Intervention." *Annual Reviews of Psychology* 65:333–71.
- Cook, Philip J., Kenneth Dodge, George Farkas, Roland G. Fryer, Jr, Jonathan Guryan, Jens Ludwig, Susan Mayer, Harold Pollack, and Laurence Steinberg. 2014. "The (Surprising) Efficacy of Academic and Behavioral Intervention with Disadvantaged Youth: Results from a

- Randomized Experiment in Chicago." NBER Working Paper 19862. Cambridge, MA: National Bureau of Economic Research.
- Dee, Thomas. 2015. "Social Identity and Achievement Gaps: Evidence from an Affirmation Intervention." *Journal of Research on Educational Effectiveness* 8(2):149–68.
- Dobbie, Will, and Roland G. Fryer. 2013. "Getting beneath the Veil of Effective Schools: Evidence from New York City." *American Economic Journal: Applied Economics* 5(4):28–60.
- Fryer, Roland G. 2014. "Injecting Charter School Best Practices into Traditional Public Schools: Evidence from Field Experiments." *The Quarterly Journal of Economics* 129(3):1355–1407.
- Harackiewicz, Judith M., Elizabeth A. Canning, Yoi Tibbetts, Stacy J. Priniski, and Janet S. Hyde. 2016. "Closing Achievement Gaps with a Utility-Value Intervention: Disentangling Race and Social Class." *Journal of Personality and Social Psychology* 111(5):745–65.
- Imbens, Guido W., and Michal Kolesar. 2016. "Robust Standard Errors in Small Samples: Some Practical Advice." *The Review of Economics and Statistics*. 98(4):701–12
- Kosse, Fabian, Thomas Deckers, Hannah Schildberg-Hörisch, and Armin Falk. 2016. "The Formation of Prosociality: Causal Evidence on the Role of Social Environment." IZA Working Paper 9861. Bonn, Germany: Institute for the Study of Labor.
- Kost-Smith, Lauren E., Steven J. Pollock, Noah D. Finkelstein, Geoffrey L. Cohen, Tiffany A. Ito, and Akira Miyake. 2012. "Replicating a Self-Affirmation Intervention to Address Gender Differences: Successes and Challenges." *Physics Education Research Conference, AIP Conference Proceedings* 1413(1):231–4.
- Kraft, Matthew A. 2015. "How to Make Additional Time Matter: Extending the School Day for Individualized Tutorials." *Education Finance and Policy* 10(1):81–116.
- Kraft, Matthew A., and Shaun M. Dougherty. 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 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, MA: Harvard University.
- 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, MA: National Bureau of Economic Research.
- Morisano, Dominique, Jacob Hirsh, Jordan Peterson, Robert Pihl, and Bruce Shore. 2010. "Setting, Elaborating, and Reflecting on Personal Goals Improves Academic Performance." *Journal of Applied Psychology* 95(2):255–64.
- Oreopoulos, Philip, Adam Lavecchia, and Robert S. Brown. "Pathways to Education: An Integrated Approach to Helping At-Risk High School Students." *Journal of Political Economy*. Forthcoming.
- Oreopoulos, Philip, and Uros Petronijevic. 2016. "Student Coaching: How Far Can Technology Go?" NBER Working Paper 22630. Cambridge, MA: National Bureau of Economic Research.
- Scott-Clayton, Judith. 2012. "What Explains Trends in Labor Supply among U.S. Undergraduates, 1970–2009?" NBER Working Paper 17744. Cambridge, MA: National Bureau of Economic Research.
- 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. New York: MDRC.
- Symonds, William C., Robert Schwartz, and Ronald F. Ferguson. 2011. "Pathways to Prosperity: Meeting the Challenge of Preparing Young Americans for the 21st Century." Pathways to Prosperity Project, Harvard University Graduate School of Education. Cambridge, MA: Harvard University.

- Turner, Sarah. 2004. "Going to College and Finishing College. Explaining Different Educational Outcomes." In *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It* ed. Caroline M. Hoxby, 13–62. Chicago, IL: University of Chicago Press.
- Walton, Gregory. 2014. "The New Science of Wise Psychological Interventions." *Current Directions in Psychological Science* 23(1):73–82.
- 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):468–85.
- Yeager, David, Marlone D. Henderson, David Paunesku, Gregory M. Walton, Sidney D'Mello, Brian J. Spitzer, and Angela Lee Duckworth. 2014a. "Boring but Important: A Self-Transcendent Purpose for Learning Fosters Academic Self-Regulation." *Journal of Personality and Social Psychology* 107(4):559–80.
- Yeager, David, Valerie Purdie-Vaughns, Julio Garcia, Nancy Apfel, Patti Brzustoski, Allison Master, William T. Hessert, and Matthew E. Williams. 2014b. "Breaking the Cycle of Mistrust: Wise Interventions to Provide Critical Feedback Across the Racial Divide." *Journal of Experimental Psychology* 143(2):804–24.
- 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):374–91.
- Yeager, David, and Gregory Walton. 2011. "Social-Psychological Interventions in Education: They're Not Magic." *Review of Educational Research* 81(2):267–301.