NBER WORKING PAPER SERIES

DIGITAL MESSAGING TO IMPROVE COLLEGE ENROLLMENT AND SUCCESS

Christopher Avery
Benjamin L. Castleman
Michael Hurwitz
Bridget T. Long
Lindsay C. Page

Working Paper 27897
http://www.nber.org/papers/w27897

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2020

We gratefully acknowledge the contributions of many organizations and their staff members in their contributions to this work, including: The College Board, uAspire, the Austin Chamber of Commerce, the participating Texas school districts, the Ray Marshall Center at the University of Texas at Austin, Signal Vine and One Logos Solutions. We gratefully acknowledge funding support for this work from the Institute for Education Sciences (IES), U.S. Department of Education, through Grant #R305A140121 to Harvard University. We thank Aaron Anthony, Alberto Guzman-Alvarez, and Danielle Lowry for excellent research assistance. The opinions expressed are those of the authors and do not represent views of IES or the U.S. Department of Education. All errors are our own. This work is dedicated to the memory of our friend and colleague, Alexandra Chewning, whose care for supporting students in achieving their higher education goals was infectious and without whom this project would not have been possible. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Digital Messaging to Improve College Enrollment and Success
Christopher Avery, Benjamin L. Castleman, Michael Hurwitz, Bridget T. Long, and Lindsay C. Page
NBER Working Paper No. 27897
October 2020
JEL No. I21,I22,I23

ABSTRACT

We investigate the efficacy of text messaging campaigns to remind students about and support them with key steps in the college search, application, selection and transition process. First, in collaboration with the College Board and uAspire, both national non-profit organizations, we implemented text-message based outreach and advising to students in over 700 US high schools that primarily serve large shares of low-income students. Second, we collaborated with several school districts in the state of Texas to implement a school-based version of the intervention. In the national sample, treatment students received outreach approximately once per month from uAspire counselors, whereas in the Texas sample, treatment students received outreach once every one to two weeks from their high school counselors. In both samples, outreach began in Spring 2015 and continued through September 2016. We tested these interventions with concurrent cluster randomized control trials with randomization at the school level. In contrast to the national version of the intervention, which tended to produce null effects, the school-based intervention yielded positive and significant impacts on several college-going steps and on college enrollment for certain subgroups. We discuss key differences between the two versions of the intervention that may have contributed to these divergent results.

Christopher Avery
Harvard Kennedy School of Government
79 JFK Street
Cambridge, MA 02138
and NBER
christopher_avery@hks.harvard.edu

Benjamin L. Castleman
University of Virginia
Curry School of Education
405 Emmett Street South
P.O. Box 400277
Charlottesville, VA 22902
castleman@virginia.edu

Michael Hurwitz
College Board
1919 M Street NW
Suite 300
Washington, DC 2003
mhurwitz@collegeboard.org

Bridget T. Long
Harvard Grad School of Education
Longfellow Hall 101
13 Appian Way
Cambridge, MA 02138
and NBER
bridget_long@gse.harvard.edu

Lindsay C. Page
School of Education
University of Pittsburgh
230 South Bouquet Street
Pittsburgh, PA 15260
and NBER
lpage@pitt.edu

A data appendix is available at http://www.nber.org/data-appendix/w27897
A randomized controlled trials registry entry is available at https://www.socialscienceregistry.org/trials/6525

Digital Messaging to Improve College Enrollment and Success

I. INTRODUCTION

Educators and policy makers have invested considerable resources in recent years to increase college enrollment among low-income students. Nonetheless, substantial barriers remain, and gaps in both college entry and completion by family income appear to be widening over time (Bailey & Dynarski, 2012). Many community-based non-profit organizations offer programming and individualized advising to support students through college and financial aid application processes. Several recent randomized controlled trials demonstrate that these programs can significantly increase Free Application for Federal Student Aid (FAFSA) filing, college enrollment and in some cases college persistence (Avery, 2013; Barr and Castleman, 2018; Bettinger et al., 2012). The availability of personalized counseling in the summer between the end of high school and the start of college also significantly increases the probability of following through on plans for college enrollment for low-income students (Castleman, Page and Schooley, 2014). Similarly, near-peer advising and virtual advising interventions have produced significant increases in enrollment (Bos et al., 2012; Carrell and Sacerdote, 2017; Gurantz et al, 2020). Yet these programs are often expensive—sometimes thousands of dollars per student served—and dependent on one-on-one, in-person interactions between students and program staff. As a result, they can be challenging to scale.

Several other randomized controlled trials document the potential of lower cost, technology-focused interventions delivered by schools, colleges or community organizations to provide targeted information and advising during the college and financial aid application processes. For example, students who receive customized text message outreach regarding

FAFSA filing and their own status with the process are more likely to complete timely FAFSA filing (Page, Castleman and Meyer, 2020). Carefully timed and crafted text messages have also been demonstrated to be effective in combatting the phenomenon of "summer melt," whereby students who seem set to attend a particular college at the time of high school graduation fail to enroll (Castleman and Page 2015; 2016; 2017; Page & Gehlbach, 2017).

Motivated by these results, we conducted two complementary, large-scale experimental studies of text-message-based college guidance to high school juniors and seniors with the goal of testing whether technology-based strategies that worked at small scale could have similar results at larger scale. These studies are based on a checklist approach (Gawande 2010), providing nudges supported by formal advising services to help students navigate the college admissions process one step at a time.

We report on the results of each of the two studies in this paper. The "national study" includes a sample of more than 70,000 students in 745 US high schools distributed across 15 states, with the treatment group receiving monthly assistance from virtual advisors during the course of the college search, application, selection and transition process. The "Texas study" includes a sample of more than 20,000 students from 72 public high schools in eight school districts in the Austin and Houston areas, with the treatment group receiving weekly text messages from their high school counselors with reminders about required college-going tasks and related deadlines along with the suggestion to follow up with the school counselor (either via text or in person).[2] We designed the national study with the goal of isolating the effect of

[2] The Texas schools sample originally included 74 schools within nine districts. After the completion of the intervention, however, one of the districts chose not to extent our data sharing agreement (DSA). Therefore, we are unable to report results that include this ninth district. Because this district was among the smallest of those

text outreach coupled with virtual, text-based advising,[3] and we designed the Texas study with the goal of isolating the effect of text outreach and a combination of text-based and in-person advising. Though we observe null and, occasionally, negative impacts in the national version of the intervention, the school-based intervention yielded positive and significant impacts on several college-going steps and on college enrollment for certain subgroups. Our results also suggest that, for interactive advising models to be successful at a state or national scale, it may be necessary to (1) collaborate with local entities such as districts, schools, community-based organizations, and higher education institutions with whom students have a direct relationship; (2) limit advisor caseloads; and (3) leverage technologies that support more in-depth interactions. It may also be the case that these strategies are differentially effective for students based on where they are in the distribution of academic preparation, college readiness and socioeconomic status.

II. LITERATURE REVIEW

Traditional economic models, such as Becker (1964), assume that students are aware of both the benefits and the costs of higher education and posit that students will pursue a college education if the present discounted value (PDV) of the benefits of higher education exceeds the PDV of the costs of going to college. Though the economic and non-pecuniary benefits of higher education are well-documented (Goldin & Katz, 2008; Oreopoulos & Salvanes, 2011) and are

participating, this resulted in the loss of just over 300 student observations, with over 21,000 remaining in our sample. Based on analyses that we conducted prior to the expiration of our DSA with this districts, our results are not changed appreciably by its exclusion from our analytic sample.

[3] In the national experiment, the focal treatment is compared to more modest text-based outreach which we describe in further detail below.

particularly pronounced for students from low- and moderate-income backgrounds (Long, 2008; Dale & Krueger, 2014), students from lower-income backgrounds are less likely to enter college and less likely to earn a degree by their mid-twenties (Bailey & Dynarski, 2012). As many as half of students from lower socioeconomic backgrounds do not apply to academically rigorous institutions to which, based on their credentials, they would have a good chance of being admitted (Bowen, Chingos, and McPherson, 2009; Hoxby and Avery, 2013; Smith, Pender, & Howell, 2013).

Behavioral challenges may particularly deter disadvantaged students from attending college and may help to explain why many low-income high school seniors indicate that they want to enroll in college but do not complete the steps required to follow through on these intentions (Avery & Kane, 2004; Roderick et al., 2008). First, students who cannot afford to attend college on their own have to navigate the financial aid process as well as the college application process, which gives them both more tasks to complete and additional opportunities to fall prey to behavioral traps. Second, students from disadvantaged backgrounds often have to devote their time and energy to addressing immediate stressors like supporting their families financially or dealing with neighborhood violence (Casey, Jones, & Somerville, 2011; Keating, 2004; Steinberg, 2008). Third, students are less likely to have access to college-educated family members or college counselors who can help them weigh short-term investments against long-term gains (Lareau, 2015; Schneider, 2009), especially during the summer between high school graduation and the start of college (Arnold et al, 2009; Castleman, Arnold, & Wartman, 2013; Castleman & Page, 2013, 2015; Klasik, 2012). Given these issues, the number of steps required from application to actual enrollment (e.g., financial aid award letters, proof of health insurance, orientation and placement test registrations) and the complexity of the financial aid process (Dynarski & Scott-Clayton, 2006) may discourage many would-be students. Consistent with this view, modest changes in near-term costs in terms of time (e.g., a required college essay) or money (e.g., a fee for sending a test score to a college) often have outsized effects on student choices and outcomes (Bulman, 2015; Pallais, 2015; Smith, Hurwitz & Howell, 2015; Hurwitz, Mbekeani, Nipson & Page, 2017).

Many recent studies consider the effects of interventions designed to address particular aspects of the admissions process, hoping to replicate the positive effects of a small reduction in cost on student choices. A number have found positive effects of personalized, interactive text message interventions related to FAFSA submission or other required pre-matriculation tasks (Castleman & Page, 2015; 2016; 2017; Page & Gehlbach, 2017; Page, Castleman & Meyer, 2020) when conducted at the district or single institution level. However, larger-scale versions of these interventions (cf. Bird et al. (2019) which encouraged FAFSA completion for 450,000 low-income or first-generation students who had registered with the Common Application and, in a separate experiment, encouraged several hundred thousand prospective and current students in a large state to apply for financial aid, yielded precisely measured null effects on college enrollment, persistence, and financial aid receipt (see also Bergman, Day & Manoli, 2019, Bettinger et al., 2012). Two studies that yielded substantial positive results stand out because they were targeted at high-achieving students who had strong prospects of admission to an institution where they would qualify for generous aid packages (Hoxby & Turner, 2019; Dynarski, Libassi, Michelmore & Owen, 2018).

Despite the positive effects detected by Hoxby and Turner (2013), efforts by the College Board to scale a version of the "Expanding College Opportunities" intervention to students across a broader range of prior achievement failed to generate meaningful shifts in college enrollment (Gurantz et al, 2019). Moreover, two other recent studies that evaluated virtual one-on-one advising interventions for high-achieving, low- and moderate-income students found at most modest positive impacts on whether students attended selective colleges and universities (Gurantz et al., 2019b; Sullivan, Castleman, and Bettinger, 2019). The HAIL study from Dynarski and colleagues (2018) provided highly qualified, low-income students in the state of Michigan with personalized, mail-based outreach indicating that they could attend the University of Michigan tuition free, if accepted. This clear messaging from a trusted institution led to very sizable impacts on application to and attendance at the University of Michigan (Dynarski, Libassi, Michelmore & Owen, 2018).

Taken together, the findings from these large-scale studies suggest that information-only interventions may be an ineffective strategy to encourage completion of important college application and financial aid tasks that lead to higher rates of enrollment and persistence. An open empirical question is whether these interventions might have been effective had they offered students the opportunity to connect with additional advising and assistance, particularly from a source of advice that is already known and trusted, such as a student's own school counselor.[4] Given (1) the sizeable impacts that intensive advising interventions have had on students' postsecondary outcomes and (2) the fact that the smaller-scale texting campaigns

[4] The Bird et al., (2019) Common Application experiment included a small advising treatment arm. While the authors do not find significant impacts of this treatment, the confidence intervals of the authors' estimate include impacts that would be substantively meaningful and consistent with prior interactive texting interventions.

described above typically provided remote advising, it is possible that behaviorally informed messaging campaigns could be effective at scale if they provided opportunities for personalized counseling.

III. INTERVENTION DESIGN

We implemented two complementary randomized controlled trials to investigate whether behaviorally informed messaging campaigns could be effective at scale if they provided opportunities for personalized counseling. We describe the first study as "national" and the second study as "Texas". We implemented the national study in partnership with three core partners: the College Board; uAspire, a non-profit organization focused on college affordability; and Signal Vine, a text-messaging platform provider.

In the national study, which included a sample of over 70,000 students in 745 US high schools distributed across 15 states, treatment group students received text based outreach related to steps in the college-going process from the spring of junior year of high school through the summer after high school graduation. Messages encouraged students to respond via text to receive additional support and guidance (In Appendix A, we present the topic, timing and specific content of the intervention outreach messages) from remotely located advisors.

We used school-level information from the Common Core of Data and the Private School Survey to identify an initial list of 2,136 high schools (85% public schools) in 15 states to invite to participate in the study. We selected these schools because they had a substantial representation of students qualifying for free or reduced price lunch and a relatively low historical rate of college enrollment. The Educational Testing Service, in partnership with College Board, recruited schools from this list to participate in the study. A total of 935 schools agreed to have high school juniors complete a supplementary form with cell phone contact information as part of the October 2014 PSAT/NMSQT. A total of 745 schools returned forms and a total of 70,285 students (60,742 students of whom provided a valid cell phone number) enrolled in the study. We observed little to no difference in the characteristics of schools from the initial list that did and did not agree to participate.[5]

To ensure balance on geographic distribution as well as school sector, we conducted the randomization at the school level within state-secondary school sector groupings. Grouping by sector is helpful, given that on average, students in private schools outperform their public school peers on baseline measures. Within each state and sector (e.g., public schools in Arizona), we ordered schools on a lagged school-level measure of four-year college going,[6] and created groups of schools within which to randomize. Our target group size is four schools, however, because the number of schools within each state-sector group was not always divisible by four, in practice, we allowed for groups of schools to range in size from four to six schools.[7]

[5] Across nine measures (including average PSAT scores and college enrollment rates), we observed only one significant difference between schools that did and did not participate in the study. Specifically, 26.2% of students at participating schools vs. 25.2% of students at non-participating schools enrolled at two-year colleges after high school graduation based on a lagged measure of two-year college going.

[6] This decision was, in part, motivated by our finding that PSAT scores for the students in our sample are highly correlated with current and prior school level measures of PSAT performance but only modestly correlated with prior school-level measures of college going. Because we are most interested in college-going outcomes, our judgment was that stratifying on the prior measure of college going would maximize our ability to improve power and precision through stratification. Stratifying on a lagged measure of college going did require us to impute this missing measure for 13 schools that were too new to have had a lagged measure of college going observed.

[7] Within groups of four schools, the probability of assignment to treatment was 0.50. Because this probability varied somewhat across groups, in our analyses, we apply a weight, calculated at the school level, equal to the inverse of probability of assignment to each school's assigned experimental condition within groups.

Scheduled text messages were sent to treatment group students approximately monthly. Treatment group students additionally were each matched with advisors who were recruited and trained by uAspire specifically to work with students in this study (Please see Appendix B for detailed information on the uAspire advisor staffing model as well as procedures for project implementation, including advisor hiring, training and supervision). Although the initial text message sent to treatment group students for each topic was scripted in advance, the advisors provided individualized follow-up text communication for each response they received from a student.[8] Control group students received messages approximately once every two months. Rather than individualized feedback, control group students who responded to seek additional information received a pre-specified automatic sequence of texts on the topic of the original text (please see Appendix C for the details of this outreach).

In the Texas schools study, we collaborated directly with 72 schools within eight public school districts in the Austin and Houston areas of Texas. To ensure balance on key baseline information and to improve statistical power, we first matched sample schools into groups ("group") of approximately five schools, grouping on 2013 school-level college enrollment data publicly available through the Texas Education Agency. We then randomly selected

[8] In keeping with the uAspire philosophy of "student centeredness", advisors were trained to be responsive to whatever each student might present with in their reply to each program message, rather than reflexively redirect them to the message topic. Further, uAspire notes in its definition of "student centeredness" that "we respect each student's unique situation and do our best to make their college aspirations affordable"; this philosophy emphasizes long-run success rather than immediate college enrollment for students. As such, the uAspire advisors aimed to ensure that students were aware of the financial implications of their postsecondary choices as well as other options for matriculation in cases where a student's choice was not viable (for financial or other reasons). In this particular effort, advisors were trained to help students understand the long-term benefits of a college education and to identify options that would work for them.

approximately two of five schools for the treatment group.[9,10] For large districts, we prioritized matching schools within districts, but for smaller, often single-high school districts, we grouped schools across districts. In sum, we randomly selected 29 schools to participate in the treatment and 43 schools to serve as control.

Texas treatment group students received messages from the spring of their junior year in high school through the summer after graduation, just as in the national study, with the distinction that the ostensible sender of messages was each participating student's school counselor rather than the College Board (please see Appendix D for details of this outreach).[11] When students responded to this outreach via text, school counselors received these responses via the online messaging platform. In addition to responding via text, however, students also had the opportunity to interact with their counselors in person during the regular school day. In fact, some participating counselors tailored the outreach messages sent to their own students specifically to encourage face-to-face rather than text-based communication.[12]

[9] The uneven distribution of schools to treatment and control conditions relates to the cost of the intervention. Because we paid school counselors to staff the messaging in the summer of 2016, we faced budget constraints in terms of the number of schools to which we could assign to the treatment condition. In the participating districts, counselors are off contract in the summer months, and so we compensated them for engaging with students over the summer as part of this project.

[10] Note that the exact number of schools and treatment assignment probabilities varied modestly across the groups. Therefore, as in the national sample, we weight observations by the inverse of probability of assignment to experimental condition to handle variation in assignment probabilities across groups.

[11] In the Texas schools intervention, we employed OneLogos Education Solutions as our technology partner for message distribution and communication. OneLogos licenses an online student information system that supports text communication between students and counselors. In addition, districts maintained discretion over the time of day for outreach message distribution. Certain districts opted for messages to be sent before the beginning of the school day to encourage students to seek same-day help on focal topics, while other districts preferred for messages to be distributed in the afternoon or early evening after the conclusion of the school day.

[12] We afforded counselors this type of modest editorial power over the message content to ensure that it aligned well with their standard counseling practices as well as with other college-going events within the school, district and/or community.

In contrast to the national study, where students received approximately one outreach message every three to four weeks, outreach was more frequent in the Texas schools study, with outgoing messages sent approximately once every one to two weeks. In addition, in some instances, students received more specialized messages based on their progress with college-going tasks as observed through the Apply Texas system. For example, we customized some messages about the financial aid process according to whether students had submitted the FAFSA, completed the FAFSA, or were selected for FAFSA verification, as in Page, Castleman and Meyer (2020). In another set of messages, we customized outreach regarding college applications according to whether students had started or submitted a college application. Although the control group schools had access to the texting platform employed in the Texas intervention, they used it much less frequently and systematically, as we illustrate below.

IV. RESEARCH DESIGN

Data Sources

We draw on data from several sources for the national study. First, as noted above, we relied on lagged measures of college going and PSAT/NMSQT performance provided by the College Board for prior cohorts of students in the process of selecting and randomizing schools for the experiment. From the National Center for Education Statistics Common Core of Data and Private School Universe Survey, we obtained data on other school-level characteristics. For example, for the public schools in our sample, we obtained a school-level measure of the share of students qualifying for free- or reduced-price school meals. For the students included in our sample, data from the College Board includes student-level demographic characteristics, a selfreported measure of high school GPA at the time of PSAT/NMSQT administration, PSAT/NMSQT and SAT scores and records of scores sent to colleges. In addition, we obtained school-level information available through Federal Student Aid (FSA) on week-by-week FAFSA submission and completion counts during the course of the intervention[13] and college enrollment data at the student-semester level from the National Student Clearinghouse. We also make use of text message level interactions between students and advisors as compiled by Signal Vine, our technology partner, during the course of the intervention.

We use similar data for the Texas schools study. As in the national study, we compile school-level information available through Federal Student Aid (FSA) on week-by-week FAFSA submission, college enrollment data from the National Student Clearinghouse, and records of text message level interactions between students and advisors as compiled by OneLogos, our technology partner for the Texas intervention.

We also observed several richer sources of administrative data for the Texas schools study than anything available for the national study. First, the participating districts provided student-level administrative records that allow us to observe information such as student race / ethnicity, gender, and an indicator of economic disadvantage corresponding to qualification for FRL. These school district records also include information on SAT taking, SAT scores and GPA. Second, the Apply TX system provides individual-level data for FAFSA completion and applications to public

[13] The school-level FSA data is available from the following site: https://studentaid.ed.gov/sa/about/data-center/student/application-volume/fafsa-completion-high-school. This site reports student counts of FAFSA submission and completion by high school. We convert these counts to rates by dividing by the number of high school seniors and analyze the completion rates at the school level. Admittedly, these data provide an imperfect indication of the effect of the intervention on FAFSA filing for students in our sample, given that the FAFSA completion rates that we calculate are measured at the school level, and not all students in sampled high schools participated in the intervention.

colleges in Texas. The FAFSA information in this system also includes the timing of FAFSA completion and whether or not each student was flagged for FAFSA verification.[14] Students selected for verification are required to complete additional steps to verify information reported on their FAFSA, therefore it represents an additional hurdle to college access for those students selected (Cochrane, 2010; Wiederspan, 2019). A limitation of both sources of financial aid application data is that they lack information on student completion of the Texas Application for State Financial Aid (TASFA), the financial aid application for undocumented students in Texas to obtain state-based financial aid.

Descriptive Statistics and Randomization Balance

In Table 1, we present descriptive statistics for the class of 2016 students included in the national study sample (i.e., the students in sampled schools who signed up to receive text-based outreach) and results for our tests of balance tests between treatment and control groups. The national study sample is 28 percent White, 19 percent Black, 36 percent Hispanic and 7 percent Asian. Consistent with national patterns of higher levels of college-going for female students, the sample is approximately 55 percent female. Participants reported a high school GPA of 3.3, on average, with average PSAT/NMSQT scores in the low 40s on each section, corresponding to approximately the 33rd percentile of the score distribution for 11th grade PSAT/NMSQT-takers. These characteristics of the national sample are generally in line with the averages for prior

[14] Whereas the student-level data provides information on the timing of final FAFSA completion only, the FSA data allows us to track both submission and completion rates over time. In addition, anecdotal evidence from participating counselors led us to question the accuracy of the student-level filing information in some instances. For example, if a student's contact information as documented in the FAFSA did not match school-level records exactly, the student-level records held by the school may not accurately reflect this student's true filing status. Therefore, the student-level data may result in downwardly biased estimates of FAFSA filing overall and of the effect of the intervention on FAFSA filing. For this reason, we report impacts on FAFSA submission and completion using both data sources.

cohorts from these schools (see Appendix Table E-1 for school-level information based on prior cohort data available through the College Board), though our sample has slightly higher GPAs, a higher representation of Hispanic students, and a lower representation of White students than the school-level averages from prior cohorts.

We assess balance both for the full sample and for the subsample of students in the national study for whom we have valid cell phone numbers.[15] Across all variables reported in Table 1, only one yields any evidence of imbalance and the associated coefficients do not suggest a difference in baseline measures of practical significance. Further, the joint tests that we run all fail to reject the null hypothesis of predictors jointly being zero. Similarly, we find no evidence of imbalance between treatment and control in terms of school level characteristics (Appendix Table E-1). In sum, we judge that our randomization procedure yielded treatment and control groups that are well balanced at baseline.

In Table 2, we present descriptive statistics for the full set of students enrolled in participating schools in the Texas schools study at the start of the intervention and results for our tests of balance tests between treatment and control groups. Approximately 56 percent of students in the Texas school study are Hispanic, 14 percent are Black, and 31 percent are White. Fifty-five percent of students are economically disadvantaged. We assess baseline equivalence at both the student and school levels. For covariates measured at both levels, we regress each baseline characteristic on an indicator for school-level random assignment using a model that includes fixed effects for group and that clusters standard errors at the school level. As shown in

[15] While 86 percent of the students in our sample provided a valid cell phone number during the initial data collection, we conducted the randomization with respect to the entire sample.

the right column of Table 2, all results indicate that our sample is well balanced according to both student- and school-level characteristics.[16]

Analysis

Given the differences across the national and Texas studies in terms of intervention structure and available data as well as constraints imposed by the data sharing agreements governing the project, we estimate impacts for the national and Texas schools samples separately.[17] Nevertheless, the structure of our outcome models is similar across the interventions. Specifically, within each sample, we use regression and linear probability models to assess intervention implementation, engagement and impact. To examine intervention participation and college enrollment outcomes, we fit models of the following general form on data at the student level:

$$Y_{ijk} = \alpha_k + \beta_1 \mathrm{Treat_{jk}} + \boldsymbol{X_{ijk}\gamma} + \boldsymbol{S_{jk}\theta} + \epsilon_{ijk}, \tag{1}$$

where for student *i* in school *j* in group *k*, Y_{ijk} is the outcome of interest, $\mathrm{Treat_{jk}}$ is an indicator for treatment assignment at the school level; and $\boldsymbol{X_{ijk}}$ and $\boldsymbol{S_{jk}}$ are vectors of baseline characteristics at the student and school levels. We include fixed effects for group (α_k) to account for the structure of the randomization and cluster standard errors at the school level.[18] Our

[16] Importantly, the schools are balanced on lagged measures of college enrollment (from the class of 2013) and FAFSA filing (from the class of 2014). A small number of participating schools were too new to have lagged college enrollment and/or FAFSA filing data available. For these schools, we imputed zero values and grouped these schools together for the sake of randomization. Therefore, within group, missingness of this school-level information is balanced.

[17] Conducting data analyses for the national and Texas interventions separately was necessitated by the structure of our data sharing agreements. For example, college enrollment data for the Texas schools sample was transferred to Harvard University for the purpose of analysis, but college enrollment data for the College Board sample was retained and analyzed by the College Board. Therefore, it was not possible to pool college enrollment data across the two experiments.

[18] Not all of the school groupings included exactly five schools. For this reason, the probability of assignment to the treatment condition varied somewhat across schools and school groupings. To handle this variation, we assign weights at the school level according to the inverse probability of assignment to the given experimental condition.

parameter of primary interest is β_1 which represents the impact of school-level random assignment to the intervention on a given student outcome.

In neither the national nor the Texas schools samples are all students within the treatment groups actually treated. This is because in both experiments some sampled students did not actually have a valid cell phone number. Therefore, β_1 represents the intent-to-treat (ITT) impact. Given potential spillover effects from participating to non-participating students within the same school, we reason that the assumptions required to use an instrumental variables strategy to derive complier average causal effects are not well met. Therefore, we focus exclusively on ITT effects.

In both experiments, our student-level outcomes of interest include college-entrance exam taking and performance and college enrollment and persistence. In the Texas schools sample, we also consider FAFSA completion and college application submission at the student level. As noted above, for both samples we additionally examine impacts on FAFSA submission and completion based on FAFSA completion data aggregated at the school level. The associated models take a similar form with outcomes assessed at the school rather than individual level but with models that essentially weight schools according to within-school sample size.

V. RESULTS

Take up

In the national study, text messaging was the only channel of communication between students and advisors. In Appendix Table E-2, we report the responses of students to receipt of

In practice, these weights make little difference in our estimates, although the experimental results we present focus on models that incorporate these weights.

each message by topic and date. Out of the full sample of students assigned to receive the interactive text-based outreach (N = 36,521), 86% received the initial outreach message. Over the course of the intervention, near ten percent of the full sample opted out of the message outreach, with a substantial proportion (nearly four percent) opting out after the introductory outreach message. Of all students in the sample, 23,895 (65%) students engaged with the intervention by responding at least once, excluding opt out messages as engagement. Student engagement varied across message topic areas, with the highest rates of engagement on messages related to spring SAT registration and preparation; college lists and application deadlines; assistance with the college application process; and enrollment decisions and related tasks. A separate study considers the text-message conversations between advisors and students in detail (Arnold, Lewis & Owen, 2018).

In addition to this message-level analysis, we use a cluster analysis strategy to classify students who ever responded to the messaging into three groups: low, moderate and high engagers. We consider not only the number of messages that students sent within each message flow but also the length (e.g., the character count) of messages sent (please see Appendix F for more details). Examining the data in this way presents a story different from the relatively high response rates by message. Among all students who engaged in text communication, we classify most (45 percent of the full treatment group; 68 percent of those who engaged) into the low engagement group. These students sent only a few text messages, 1.8 on average, when they responded to a given outreach message, and the messages that they sent were relatively short in length. The typical low engager responded to just over two of the outreach messages. Approximately 18 percent of all students (27 percent of engaged students) engaged at a moderate level. These students sent an average of 3.6 messages when they responded to an outreach message and replied to eight to nine outreach messages across the duration of the intervention. Only a very small share of students (approximately three percent of the full treatment group and four percent of those who engaged) took up the text-based advising at a very high level. These relatively high engagers sent an average of nearly 10 messages when they engaged in message flows and responded actively to an average of seven message flows. In sum, although response rates appear high when considering the share of student who responded to each outreach message (or at all during the course of the intervention period), we observe comparatively low rates of sustained engagement with the opportunity to connect with a text-based advisor over time. A comparison of student characteristics across student engagement types reveals high and moderate engagers are disproportionately female and somewhat higher performing based on PSAT/NMSQT and GPA measures (Table F-5). In short, engagement is higher among those whom we might expect to have relatively better college going outcomes, among all students in the sample. Of course, the outreach may still be of benefit to students even if they are not responding, as it provides students with timely reminders of college-going steps to be taken.

In Appendix Table E-3, we present take up, engagement and opt out rates for the Texas schools study. In the first column of Table E-3, we observe a low level of text message usage in the control schools, with 18 percent of students in control schools receiving any text outreach at some point during the intervention period. By contrast, in the treatment schools, a large majority (86 percent) of students received text outreach through the intervention. Of those in the treatment schools who received outreach, about half of students (or 41 percent of all students) responded to the text outreach at least once, and a very small share (under 3 percent) requested to opt out of the intervention.

Compared to the national study, the rate of student opt out was lower, perhaps reflecting a greater level of trust from students in messaging sent by their own school. Also different from the national study, the typical treatment student in the Texas schools study received nearly 48 messages but sent only 1 to 2 messages over the course of the 18-month intervention period. This lower level of text-based engagement is not surprising, given that in the Texas study, students could interact with their counselors via other channels, including email and in person. Indeed, interviews with counselors across the Texas districts reveal that counselors used a variety of channels of communication to reach students including email, Naviance, social media, in-school announcements, posters / flyers, in-class presentations as well as their own personal cell phone (e.g., texting but not through the messaging platform) (Arnold, Lewis & Owen, 2018). Further, some of the Texas school counselors expressed a reticence to communicate with students via the texting platform and strongly preferred face-to-face interactions. To attend to such preferences, in these contexts we explicitly adjusted messages to encourage students with questions and/or seeking help to follow up with their counselor in person during the school day rather than to follow up via text.

Impacts on college-going outcomes

Across the national and Texas studies, we reach substantially different conclusions about the extent to which text-based outreach supports students' success in navigating college-going processes and accessing college following high school. Specifically, in the national sample, we find little evidence that student outcomes were improved by the outreach, whereas in the Texas schools study, we find a consistent pattern of positive effects. We first present results across all college-going outcomes and then consider the differences across the two studies that may have contributed to these divergent results.

- **College entrance exams**

In Tables 3 and 4, we present results associated with SAT taking, performance and score sending. Among control group students in the national sample (Table 3), approximately 68 percent took the SAT, earned an average combined math-verbal score of 906, and sent SAT scores to an average of two institutions.[19] The intervention did not appear to shift these outcomes for treatment group students, with the coefficient on SAT taking being negative although not statistically significant.

In the Texas schools sample, the baseline rate of SAT (or ACT) taking was substantially lower at 48 percent (Table 4). Therefore, there was more room for improvement in test taking rates. In addition, average SAT performance was somewhat stronger among test-takers in this context, with an average control group score of 930. In the Texas schools sample, the text-based outreach had a modest, marginally significant impact on SAT / ACT taking of 4 percentage points. In line with other efforts that increase college entrance exam taking (e.g., Hurwitz, Smith, Niu & Howell, 2015), increased test-taking occurred together with modest declines in average performance among test takers.

[19] Our college entrance exam testing data come directly from the College Board. Therefore, a limitation of our analysis here is that we are not able to observe test-taking outcomes related to the ACT.

- **FAFSA submission and completion**

We consider patterns in school-level FAFSA submission and completion rates, based on school-level FAFSA filing data made available by Federal Student Aid. In the national study, by the end of the intervention period FAFSA submission and completion rates in the control schools were 64 and 58 percent, respectively. We find no indication that the text-based outreach improved month-by-month submission and completion rates at the school level (Table 5). Of course, in some schools participating in the national study only a fraction of students in the target cohort participated in the intervention. Therefore, the school-level data may be too coarse to capture any FAFSA filing improvements induced by the outreach.

In the first two panels of Table 6, we present analogous results for the Texas schools sample. Based on the school-level data obtained by FSA, we estimate that by the end of July, FAFSA submission and completion rates were 57 percent and 52 percent respectively in the control schools, somewhat lower than in the national study control schools. In contrast to the national study, however, the text-based outreach led to substantial improvements in FAFSA filing rates in the Texas schools study, with both submission and completion being 8 to 9 percentage points higher in the treatment schools by the end of the intervention period.

Based on the student-level data (on which we will rely for student-subgroup analyses), we estimate a FAFSA completion rate that is somewhat lower at 45 percent for control schools and an impact on FAFSA completion 5 to 6 percentage points. As discussed above (see footnote 11), we reason that imperfections in data matching lead to these lower levels and impacts. Nevertheless, both data sources indicate that the outreach lead to significant improvements in FAFSA submission and completion in the Texas school context.

- **Applying to college**

As described above, we have access to data for applications to in-state public colleges for the Texas schools study and no specific data on applications for the national study. In control schools in the Texas schools study, approximately 71 percent of students applied to a public in-state college through the Apply Texas system, and the average student submitted approximately two applications (i.e., conditional on applying, the typical applicant applied to between two and three colleges) (Table 7). We estimate that the text-message outreach increased the share of students who applied to in-state public colleges by 8 percentage points, and this effect is statistically significant. The text-message outreach is estimated to increase the total number of applications by 0.11 per student, on average, although this effect was not statistically significant. Taken together, the primary effect of the outreach was to move some students from submitting no applications to submitting one.

- **College enrollment and persistence**

In Tables 8 and 9, we present impacts on college enrollment and enrollment in a two-year or four-year institution in the fall of 2016 (immediately following high school) as well as for the fall of 2017, for the national and Texas studies, respectively. Across these tables, we again observe differences in the baseline college-going rates between the two samples, such that college going is more prevalent in the national sample. In the national sample, 61 percent of students in the control schools transitioned to college immediately after high school, compared to 51 percent of students from control schools in the Texas schools sample.

In the national study (Table 8), the intervention is estimated to have a modest negative and statistically significant effect of just over one percentage point in on-time college enrollment.

We find separate negative estimated effects of the intervention on two-year and four-year enrollment, though neither estimated effect is statistically significant. We find similarly negative estimated effects of the intervention on enrollment in the fall of 2017, one year after high school graduation. These negative effects may relate to uAspire's organizational stance of helping students to understand both the long-run benefits and the financial implications of postsecondary choices that they might make. For instances, if students expressed concerns about their ability to pay for college or about the amount of debt they would have to take on, uAspire advisors might encourage students to consider waiting to enroll until they had a more financially viable plan.

In the Texas schools study, college enrollment was approximately 2 percentage points higher in the treatment schools, although this result is not statistically significant (Table 9). This treatment effect is split essentially evenly between the two- and four-year sectors. We observe a similar pattern of results for Fall 2017 enrollment. The magnitude of the effect on college enrollment in the Texas schools study is perhaps smaller than we would expect, particularly given the strong effects of the intervention on FAFSA filing. Nevertheless, it is of a similar magnitude (but in the opposite direction) compared to the effect on college-going in the national study. Despite achieving statistical significance, neither of these two estimate magnitudes is large enough to conclude that these interventions had beneficial or deleterious effects on college-going.

Subgroup analyses

Within the national sample, we have investigated evidence of heterogeneous effects by student characteristics (including baseline measures of achievement and gender) as well as school-level measures (including school sector and locality). Overall, we find little evidence of heterogeneity, although we observe some patterns with respect to locality. For example, within the subset of schools situated in suburban areas, we observe modest positive effects of the intervention on SAT taking and score sending but negative effects on SAT performance, consistent with the idea that a broader range of students are taking the SAT. Also for students in suburban schools, the intervention increased rates of college enrollment, particularly in the two-year college sector. These positive effects are, at times, counterbalanced by negative effects for students who reside in urban areas, leaving our conclusions about the efficacy of the intervention in the national study unchanged. Because these results do not point to robust conclusions about subgroup effects, we do not report them in a table.

In the Texas schools study, we explored subgroup effects for student race, ethnicity, gender, socioeconomic status (and indicated by qualifying for free- or reduced-price meals), and academic achievement. Based on preliminary analyses and consistent with recent research and policy efforts devoted particularly to the college attainment of low-income, high achieving students (e.g., Hoxby & Turner, 2013; Hoxby & Avery 2013), we focused our attention on student subgroups defined by the joint distribution of socioeconomic status (FRL eligibility) and academic achievement (high versus low GPA).[20] We regard these findings as exploratory in nature, as they were not part of a pre-analysis plan. Nevertheless, they are helpful in exploring the connections between impacts on college-going processes and eventual college enrollment.

[20] When considering variation in effects by academic achievement, we split the sample into relatively high achievers (with GPAs greater than or equal to 3.0) and relatively low achievers (with GPAs less than 3.0). A GPA of 3.0 served essentially as a median split of the sample.

In Table 10, we report subgroup effects on all outcomes as well as an outcome indicating whether students completed all three key college-going tasks: taking the SAT, applying to at least one college, and completing the FAFSA. The intervention positively affected multiple outcomes for all subgroups except for high achieving, low-income students. The lack of impact on high achieving, low-income students may stem from a crowding out effect, as such students are highly desirable candidates and likely heavily targeted by colleges eager to enroll high achieving low-income students. We also note that we would have wanted to see positive effects of the outreach on FAFSA filing particularly for low-income students (e.g., those qualifying for FRL). It may be that for a larger share of FRL students, TASFA was the appropriate financial aid form. For such students, we are unable to observe impacts on financial aid applications.

For high achieving, non-low income students, rates of college-enrollment are already high (82 percent in control schools), and the intervention served to improve FAFSA completion moderately for this subgroup. For these students, access to financial aid may not be as critical for college attendance, although FAFSA completion may have helped them to access college loans and/or other non-need based financial aid. Relatively low-achieving, low-income students experienced increases in SAT taking and college application submission because of the outreach, but this did not translate into improved college-going outcomes for students.

In contrast, we observe large and consistent impacts for non-low income students who were relatively low achieving. For these students, we observe increases of 6 percentage points on SAT taking, 15 percentage points on college application submission, 10 percentage points on FAFSA completion, 8 percentage points on seamless college enrollment, and nearly 5 percentage points on enrollment the following fall (Fall 2017). Although not shown, these enrollment effects are split approximately evenly between two-year and four-year institutions. One possible explanation for the concentration of positive effects among this group, in particular, is that these students may receive a comparatively small share of attention either from school counselors or college advising organizations. As a result, the marginal return to the text-based outreach may have been higher for these students. The contrasting findings between lower-achieving low-income and lower-achieving non-low-income students may suggest that college sticker price shock is preventing these low-income, lower-performing students from engaging in college-going steps (Levine, Ma, and Russell, 2020).

Ideally, we would be able to examine the extent to which the intervention supports students to navigate the FAFSA verification process (when required) successfully. Although we cannot observe this directly, we can examine the relationship between selection for verification and on-time fall college enrollment among FAFSA filers. In Table 11, we present estimates of this relationship both overall and for our focal subgroups. Consider the overall results first (columns 1 and 2). Within control schools, 75 percent of FAFSA filers who were not selected for verification enroll in college on-time, whereas on-time enrollment is nearly 8 percentage points lower among those selected for verification. Within treatment schools, we observe a similar on-time enrollment rate among FAFSA filers not selected for verification, but the detrimental effect of selection for verification is somewhat lower at 6 percentage points. Of course, these verification effects are not precisely enough estimated to be distinguish statistically. Nevertheless, they are suggestive that the intervention may have helped to mitigate verification as a barrier to college access, at least partially.

The subgroup results in Table 11 are largely consistent with those in Table 10. Specifically, in Table 10, the lower performing, non-low income students were the only subgroup to experience improved rates of timely college enrollment because of the intervention. Similarly, this is the subgroup for which we observe the largest difference in the negative relationship between verification and timely enrollment. For this subgroup, FAFSA filers in the control schools were nearly 9 percentage points less likely to enroll in college if flagged for verification, whereas in treatment schools the detrimental effect is a non-significant 2.2 percentage points. For other student subgroups, we observe smaller differences in the verification effect between treatment and control settings. The ability of lower performing, non-low-income students to overcome the verification challenge may be a key reason why they were able to translate improvements in college-going processes into improvements in on-time college enrollment and persistence into the second year.

VI. DISCUSSION

We observe a consistent pattern of results that the text-message outreach did little to shape college-going outcomes in the context of the national study but led to significant improvements in several college-going processes overall as well as improvement in college enrollment for a subset of students in the context of the Texas schools study. Here, we discuss potential reasons for the differences that we observe. First, as noted, the baseline rates of success with various process outcomes (e.g., FAFSA filing, SAT taking) as well as with college enrollment were higher in the national sample control group compared to the Texas schools control group. Therefore, there was more room for improvement in outcomes in the Texas schools sample.

In addition to these contextual differences, the national and Texas interventions differed on several dimensions. In the national study, outreach was framed as coming from a representative of the College Board, whereas in the Texas schools study, students perceived the outreach as coming from their own high school counselor. Students in the Texas sample may have viewed their school counselor as a trustworthy and well-known source of information and therefore one worth paying attention to. In contrast, students may associate the College Board primarily with PSAT/NMSQT, SAT and AP exams, rather than a source of college counseling. Engaging with a familiar source of information, such as the student's own school counselor, rather than an unknown entity, may have driven differences in engagement and impact. Consistent with this view, treatment students in the national study were nearly three times as likely as treatment group students in the Texas schools study (10 percent vs. 3.5 percent) to opt out of receiving text messages- though opt-out rates were fairly small.[21] This lack of overall impact of text-based outreach when coming from a person and/or organization with whom the student has no personal connection is consistent with other recent studies of large-scale, centralized texting efforts aiming to improve educational process outcomes (Bird et al, 2019; Page et al, 2019).

Research on decision-making and advice giving also supports this notion. Advice is considered "more helpful and less intrusive" when it comes from a source that the recipient

[21] In addition, early in the national study, we found the need to send an additional message to remind students that there was a human advisor on the other end of the messaging, as described further in Appendix B.

considers to be an expert source (Goldsmith & Fitch, 1997; Bonaccio & Dalal, 2006). Students in the Texas schools may have been more likely to view their counselors as expert sources of college-going information, whereas students in the national sample may have been less confident in the outreach they were receiving.

In the Texas schools study, the messaging was also better integrated into the college-going activities and processes in students' schools. In this way, there was less possibility that the messaging could have presented information, goals or timetables that conflicted with other communication students were receiving from their schools. That this type of integrated advice would have a larger effect also has support from prior research on the quality of advice. In decision making, advice quality matters; poor quality advice leads the decision maker to quickly discount their perceptions of the advice-giver's expertise whereas the reputation as a source of good advice and information is earned slowly over time (Yaniv & Kleinberger, 2000). Therefore, even if some of the advice given in the context of the national sample was relevant, it may be that periodic experiences of irrelevant messaging could lead students to more quickly discount the advisor as a source of guidance overall. At perhaps the most extreme, in the national study, if the text outreach conflicted with advice students were receiving locally, it may have led to student confusion. Such confusion could have contributed to the modest decline in college going as a result of the outreach in the national sample.

Further, in the Texas study, we were able to exploit available student-level data to provide customized messaging to treatment group students based on their status with college applications and with the process of filing the FAFSA. For example, we relied on data in the Apply Texas system to target students for messaging according to their status with college applications and with the process of filing the FAFSA. In the case of FAFSA-related messaging, we were able to customize outreach according to whether students had submitted their FAFSA, whether it was classified as complete, and whether the student had been flagged for income verification. Such customization and targeting of outreach helps to heighten the relevance of a given message and may be ever more important the more crowded text messaging becomes as a channel of communication.

A final dimension of difference between the two studies is the frequency of messaging. In the national study, students received a total of 19 automated outreach messages during the course of the intervention. In the Texas study, the typical student received more than twice as many messages over the same time span, with several messages devoted to a single topic, such as FAFSA / TASFA filing. Among other factors, the ability to process advice and information is, in part, determined by repetition (Kang & Herr, 2006). Thus, the more frequent outreach and repeated treatment of focal topics and tasks may have helped the Texas schools intervention to yield stronger impacts. This more frequent outreach was possible because in the Texas schools context, monitoring of incoming text context was distributed across counselors in all participating schools, and counselors had the ability to interact with students in a variety of ways. In the national study, text-based advisors each managed a caseload of several thousand students and were at capacity with their assigned students receiving approximately one message monthly.

We recognize that we are not able to draw a causal comparison of impacts across the two experiments. Nevertheless, our results point strongly to the relative success of this type of intervention as an integrated complement to other college-going supports within the school or local context compared to a system that attempts to offer college-going supports at large scale and in a way disconnected from the institutions, supports and local contexts within which students reside. The pattern of findings we observe lends further credence to Bird et al.'s (2019) hypothesis that nudge interventions to improve postsecondary educational outcomes are more likely to be effective when implemented locally and scaled through an expanding network approach than when they are scaled "globally," through a state- or national-level partner with only a distant or non-existent relationship to the student. Given that much college-going in the United States is a local phenomenon, it may also be important to receive support in steps like college search and choice from advisors knowledgeable about a given local context, no matter the means of communication.

WORKS CITED

American School Counselor Association. (2012). Student-to-School-Counselor Ratios. Retrieved July 26th, 2012 from: http://www.schoolcounselor.org/content.asp?contentid=658.

Arnold, K., Fleming, S., DeAnda, M., Castleman, B.L., & Wartman, K.L. (2009). The summer flood: The invisible gap among low-income students. *Thought and Action, Fall 2009*: 23-34.

Arnold, K., Lewis, J. & Owen, L. (2018). What One Million Text Messages Reveal about College Access for Low-income Students. Paper presented at the Association for the Study of Higher Education Annual Conference, Tampa, FL.

Avery, C. (2013) Evaluation of the College Possible Program. National Bureau of Economic Research working paper no. 19562.

Avery, C., & Kane, T.J. (2004). Student perceptions of college opportunities. The Boston COACH program. In C. Hoxby (ed.). *College choices: The economics of where to go, when to go, and how to pay for it.* Chicago: University of Chicago Press.

Bailey, M.J., & Dynarski, S.M. (2012). Inequality in Postsecondary Education, In G.J. Duncan and R.J. Murnane (eds.), Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances. (Russell Sage: New York, New York, September 2011).

Barr, A., and Castleman, B.L. (2018) An Engine of Economic Opportunity: Intensive Advising, College Success and Social Mobility, working paper, Texas A&M University.

Becker, G.S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education.* Chicago: University of Chicago Press.

Bergman, P., Denning, J. T., & Manoli, D. (2019). Is information enough? The effect of information about education tax benefits on student outcomes. *Journal of Policy Analysis and Management*, *38*(3), 706-731.

Bettinger, E., Long, B.T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *Quarterly Journal of Economics*, *127*(3): 1205-1242.

Bird, K. A., Castleman, B. L., Denning, J. T., Goodman, J., Lamberton, C., & Rosinger, K. O. (2019). *Nudging at Scale: Experimental Evidence from FAFSA Completion Campaigns* (No. w26158). National Bureau of Economic Research.

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human DecisionPprocesses*, *101*(2): 127-151.

Bos, J.M., Berman, J., Kane, T.J., & Tseng, F.M. (2012). The impacts of SOURCE: A program to support college enrollment through near-peer, low-cost student advising. American Institutes for Research, unpublished manuscript.

Bowen, W.G., Chingos, M.M., & McPherson, M.S. (2009). Crossing the finish line. Princeton: Princeton University Press.

Bulman, G. (2015). The effect of access to college assessments on enrollment and attainment, *American Economic Journal: Applied Economics*, 7(4): 1-36.

Carrell, S., & Sacerdote, B. (2017). Why Do College Going Interventions Work? *American Economic Journal: Applied Economics*, 9(3): 124-151.

Casey, B. J., Jones, R. M., & Somerville, L. H. (2011). Braking and accelerating of the adolescent brain. *Journal of Research on Adolescence*, *21*(1), 21-33.

Castleman, B.L., Arnold, K., & Wartman, K.L. (2012). Stemming the tide of summer melt: An experimental study of the effects of post-high school summer intervention on low-income students' college enrollment. *The Journal of Research on Educational Effectiveness 5*(1): 1 – 18.

Castleman, B.L, Page, L.C. & Schooley, K. (2014). The forgotten summer: Does the offer of college counseling after high school mitigate summer melt among college-intending, low-income high school graduates? *Journal of Policy Analysis and Management* 33(2): 320-344.

Castleman, B.L., & Page, L.C. (2014). A trickle or a torrent? Understanding the extent of summer "melt" among college-intending high school graduates. *Social Science Quarterly* 95(1): 202-220.

Castleman, B.L., & Page, L.C. (2015). Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *Journal of Economics Behavior and Organization*, 115, 114-160.

Castleman, B.L., & Page, L.C. (2016). Freshman year financial aid nudges: An experiment to increase financial aid renewal and sophomore year persistence. *Journal of Human Resources*, 51(2): 389-415.

Castleman, B.L., & Page, L.C. (2017). Parental influences on postsecondary decision making: Evidence from a text messaging experiment. *Educational Evaluation and Policy Analysis*, 39(2): 361-367.

Civic Enterprises. (2011). School counselors literature and landscape review. The College Board.

Clinedinst, M.E., & Hawkins, D.A. (2009). State of College Admission. Washington, DC: National Association for College Admission Counseling.

Cochrane, D. F. (2010). After the FAFSA: How red tape can prevent eligible students from receiving financial aid. The Institute for College Access & Success.

Dale, S., & Krueger, A. (2014). Estimating the return to college selectivity over the career using administrative earning data. *Journal of Human Resources,* 49(2): 323-358

Dynarski, S.M., & Scott-Clayton, J.E. (2006). The cost of complexity in federal student aid: Lessons from optimal tax theory and behavioral economics. *National Tax Journal 59*(2): 319-356.

Dynarski, S., Libassi, C.J., Michelmore, K., & Owen, S., (2018). Closing the gap: The effect of a targeted, tuition-free promise on college choices of high-achieving, low-income students. *National Bureau of Economic Research,* working paper no. 25349.

Feng, B., & MacGeorge, E. L. (2006). Predicting receptiveness to advice: Characteristics of the problem, the advice-giver, and the recipient. *Southern Communication Journal*, *71*(1), 67-85.

Gawande, A. (2010). *Checklist manifesto, the (HB)*. Penguin Books India.

Goldin, C.D., & Katz, L.F. (2008). *The race between education and technology.* Cambridge, MA: Harvard University Press.

Goldsmith, D. J., & Fitch, K. (1997). The normative context of advice as social support. *Human Communication Research*, 23, 454–476

Grodsky, E., & Jones, M.T. (2007). Real and imagined barriers to college entry: Perceptions of cost. *Social Science Research 36*(2): 745–766.

Gurantz, O., Howell, J., Hurwitz, M., Larson, C., Pender, M., & White, B. (2019). Realizing Your College Potential. *Impacts of College Board's RYCP Campaign on Postsecondary Enrollment (EdWorkingPaper No. 19-40)*.

Gurantz, O., Pender, M., Mabel, Z., Larson, C., & Bettinger, E. (2020). Virtual advising for high-achieving high school students. Forthcoming in *Economics of Education Review*.

Horn, L., Chen, X., & Chapman, C. (2003). *Getting ready to pay for college: What students and their parents know about the cost of college tuition and what they are doing to find out.* U. S. Department of Education, National Center for Education Statistics: Washington, DC.

Hoxby, C.M., & Avery, C. (2013). The missing `one-offs': the hidden supply of high-achieving low-income students," *Brookings Papers on Economic Activity*, Spring 2013: 1-66.

Hoxby, C.M., & Turner, S. (2013). Expanding college opportunities for high-achieving, low-income students. *Stanford Institute for Economic Policy Research*, working paper.

Hurwitz, M., & Howell, J., (2014) Estimating causal impacts of school counselors with regression discontinuity designs. *Journal of Counseling and Development* 92(3): 316-327.

Hurwitz, M., Mbekeani, P. P., Nipson, M. M., & Page, L. C. (2017). Surprising ripple effects: How changing the SAT score-sending policy for low-income students impacts college access and success. *Educational Evaluation and Policy Analysis*, 39(1), 77-103.

Hurwitz, M., Smith, J., Niu, S., & Howell, J. (2015). The Maine question: How is four-year college enrollment affected by mandatory college entrance exams? *Education Effectiveness and Policy Analysis*, 37(1): 138-159.

Kang, Y. S., & Herr, P. M. (2006). Beauty and the beholder: Toward an integrative model of communication source effects. *Journal of Consumer Research*, *33*(1): 123-130.

Keating, D.P. (2004). Cognitive and brain development. In R.M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed., pp. 45-84). New York: Wiley.

King, J.E. (2004). Missed Opportunities: Students who do not apply for financial aid. American Council on Education Issue Brief.

Klasik, D. (2012). The college application gauntlet: A systematic analysis of the steps to four-year college enrollment. *Research in Higher Education*, *53*(5), 506-549.

Kofoed, M.S., (2017). To apply or not to apply: FAFSA completion and financial aid gaps. *Research in Higher Education*. 58(1): 1-39.

Lareau, A., (2015). Cultural knowledge and social inequality. *American Sociological Review* 80(1): 1-27.

Levine, P., Ma, J., & Russell L. (2020). Do college applicants respond to changes in sticker prices even when they don't matter? *NBER working paper 26910.*

Long, M. (2008). College quality and early adult outcomes. *Economics of Education Review 27*(5): 588-602.

Mullainathan, S., & Shafir, E., (2013) Scarcity: Why having too little means so much. Times Books. New York City: NY.

National Center for Education Statistics. (2009). Common Core of Data. Washington, DC: U.S. Department of Education, Institute of Education Sciences.

Oreopoulos, P., & Salvanes, K.J. (2011). Priceless: The non-pecuniary benefits of schooling. *Journal on Economic Perspectives 25*(1): 159-184.

Page, L. C., & Gehlbach, H., (2017). How an artificially intelligent virtual assistant helps students navigate the road to college. *AERA Open*, 3(4): 1-12.

Page, L., Sacerdote, B., Goldrick-Rab, S. & Castleman, B. (2019). Financial aid nudges: A national experiment with informational interventions. The Hope Center for College, Community, and Justice, Working Paper No. 3.

Page, L. C., Castleman, B., & Meyer K. (2020). Customized nudging to improve FAFSA completion and income verification. *Educational Evaluation and Policy Analysis*, *42*(1), 3-21.

Pallais, A. (2015). Small differences that matter: Mistakes in applying to college. *Journal of Labor Economics* 33 (2) :493-520

Roderick, M., Nagaoka, J., Coca, V., Moeller, E. with Roddie, K., Gilliam, J., & Patton, D. (2008). *From high school to the future: Potholes on the road to college*. Consortium on Chicago School Research at the University of Chicago.

Schneider, B. (2009). College choice and adolescent development: Psychological and social implications of early admission. National Association of College Admission Counseling.

Seftor, N.S., Mamun, A., & Schirm, A. (2009) The impacts of regular Upward Bound on postsecondary outcomes seven to nine years after scheduled high school graduation: Final Report. *US Department of Education*.

Smith, J., Hurwitz, M., & Howell, J. (2015). Screening mechanisms and student responses in the college market. *Economics of Education Review* 44: 17-28.

Smith, J., Pender, M., & Howell, J. (2013). The full extent of student-college academic undermatch. *Economics of Education Review* 32: 247-261.

Steinberg, L., (2008). A social neuroscience perspective on adolescent risk-taking. *Developmental Review* 28:78-106.

Wiederspan, M. (2019). Impact of Verification on Iowa FAFSA Filers. Iowa College Aid Policy Brief No. 19-01.

Yaniv, I., & Kleinberger, E. (2000). Advice taking in decision making: Egocentric discounting and reputation formation. *Organizational behavior and human decision processes*, *83*(2): 260-281.

TABLES AND FIGURES

Table 1. Descriptive statistics and assessment of balance in student-level measures between treatment and control groups in the national study

	Full sample (N = 70,285)		With valid cell (N = 60,742)	
Variable	Mean	Treatment-control differential	Mean	Treatment-control differential
White	0.279	-0.015 (0.010)	0.289	-0.016 (0.010)
Black	0.186	0.003 (0.008)	0.185	0.004 (0.008)
Hispanic	0.362	0.015 (0.010)	0.359	0.016 (0.010)
Asian	0.074	-0.002 (0.004)	0.073	-0.002 (0.004)
American Indian	0.010	-0.001* (0.001)	0.010	-0.001** (0.001)
Other ethnicity	0.039	-0.001 (0.001)	0.039	-0.002 (0.001)
Male	0.452	-0.001 (0.004)	0.433	0.001 (0.004)
GPA	3.258	-0.012 (0.008) [64,551]	3.280	-0.014~ (0.008) [56,090]
PSAT/NMSQT critical reading	41.943	-0.220 (0.150) [70,238]	42.258	-0.240 (0.151) [60,707]
PSAT/NMSQT math	43.322	-0.122 (0.148) [70,230]	43.618	-0.140 (0.148) [60,696]
PSAT/NMSQT writing	40.145	-0.195 (0.150) [69,990]	40.525	-0.221 (0.151) [60,510]
F-statistic assessing joint significance (p-value)		1.16 (0.2980)		1.26 (0.2247)

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment-control differentials derived from regression models that regress each student-level covariate on a school-level indicator for treatment and fixed effects for groups within which we randomized schools. Robust standard errors (in parentheses) clustered at the school level. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups.

Table 2. Descriptive statistics and assessment of balance between treatment and control groups in Texas schools study

	Student-level		School-level	
Variable	Mean	Treatment-control differential	Mean	Treatment-control differential
White	0.305	0.002 (0.042)	0.305	0.009 (0.041)
Black	0.140	0.047 (0.034)	0.148	0.049 (0.035)
Hispanic	0.562	-0.001 (0.050)	0.549	-0.007 (0.050)
Asian	0.060	-0.025~ (0.014)	0.064	-0.027~ (0.015)
American Indian	0.030	0.002 (0.012)	0.030	0.003 (0.012)
Other race / ethnicity	0.011	-0.001 (0.002)	0.011	-0.001 (0.002)
Male	0.502	-0.001 (0.009)	0.501	-0.002 (0.009)
Economically disadvantaged	0.550	0.004 (0.043)	0.545	-0.001 (0.043)
HS GPA	3.073	0.020 (0.079) [20,694]	3.051	0.011 (0.080)
Lagged college enrollment			0.469	-0.008 (0.016)
Lagged four-year college enrollment			0.292	0.008 (0.023)
Lagged FAFSA completion			0.476	0.00 (0.032)
F-statistic assessing joint significance (p-value)		0.91 (0.5158)		1.51 (0.1707)

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment-control differentials derived from regression models that regress each student- or school-level covariate on a school-level indicator for treatment and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups. Where observations are missing, number of observations with non-missing values reported in square brackets. Otherwise, N = 21,001.

Table 3. Impact on SAT taking, performance and score sending, national study

	Took SAT		SAT Score (among test takers only)		N colleges to which student sent SAT scores	
Treatment	-0.012	-0.013	6.17	0.030	-0.100	-0.120~
	(0.015)	(0.013)	(5.93)	(1.26)	(0.075)	(0.063)
Control mean	0.675		906.23		1.999	
Covariates		X		X		X
N	70,285	70,285	47,061	47,061	70,285	70,285
R-squared	0.100	0.214	0.155	0.820	0.069	0.223

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment effects derived from regression models that regress each outcome on an indicator for treatment assignment at the school level and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. Covariates include student-level measures reported in Table 2 and school-aggregate measures of PSAT performance and percent FRL. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups.

Table 4. Impact on SAT / ACT taking and performance, Texas schools study

	Took SAT / ACT		SAT Score (zero value imputed for non test takers)		SAT Score (among test takers only)	
Treatment	0.043~	0.041~	6.169	13.886	-58.516~	-32.277~
	(0.025)	(0.024)	(32.798)	(26.692)	(32.526)	(17.634)
Control mean	0.475		443.874		930.002	
Covariates		X		X		x
N	21,001	21,001	21,001	21,001	10,516	10,516
R-squared	0.24	0.252	0.274	0.296	0.279	0.437

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment effects derived from regression models that regress each outcome on an indicator for treatment assignment at the school level and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. Student level covariates include indicators for race / ethnicity, gender and FRL status. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups.

Table 5. Month-by-month impact on FAFSA submission and completion, measured at school level, national study

	(1) School-level FAFSA submission			(2) School-level FAFSA completion		
By end of:	Control mean	Treatment		Control mean	Treatment	
January	0.129	0.003	0.003	0.106	0.001	0.001
		(0.004)	(0.004)		(0.003)	(0.003)
		[0.529]	[0.816]		[0.527]	[0.832]
February	0.324	0.002	0.002	0.276	-0.001	-0.001
		(0.006)	(0.006)		(0.005)	(0.005)
		[0.642]	[0.836]		[0.621]	[0.837]
March	0.459	0.001	0.001	0.403	0	0.000
		(0.006)	(0.006)		(0.005)	(0.005)
		[0.689]	[0.865]		[0.662]	[0.865]
April	0.514	-0.000	-0.000	0.457	-0.004	-0.004
		(0.006)	(0.006)		(0.005)	(0.005)
		[0.656]	[0.841]		[0.633]	[0.836]
May	0.561	0.000	-0.000	0.504	-0.003	-0.003
		(0.006)	(0.006)		(0.006)	(0.006)
		[0.634]	[0.834]		[0.619]	[0.828]
June	0.603	-0.004	-0.004	0.545	-0.006	-0.006
		(0.006)	(0.006)		(0.006)	(0.006)
		[0.616]	[0.821]		[0.605]	[0.815]
July	0.636	-0.005	-0.005	0.579	-0.006	-0.006
		(0.006)	(0.006)		(0.006)	(0.006)
		[0.604]	[0.806]		[0.598]	[0.803]
Covariates			X			X

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment effects derived from regression models that regress each outcome on an indicator for treatment assignment at the school level and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. Covariates include student-level measures reported in Table 2 and school-aggregate measures of PSAT performance and percent FRL. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups

Table 6. Month-by-month impact on FAFSA submission and completion, measured at school level, Texas schools study

By end of:	(1) School-level FAFSA submission Control mean	Treatment		(2) School-level FAFSA completion Control mean	Treatment		(3) Student-level FAFSA completion Control mean	Treatment	
January	0.114	-0.001	0.016	0.095	-0.003	0.013	0.107	0.008	0.019
		(0.022)	(0.019)		(0.019)	(0.017)		(0.017)	(0.015)
		[0.291]	[0.625]		[0.305]	[0.626]		[0.026]	[0.055]
February	0.297	0.070~	0.087**	0.260	0.062	0.083**	0.287	0.033	0.052***
		(0.040)	(0.032)		(0.037)	(0.030)		(0.021)	(0.015)
		[0.305]	[0.558]		[0.305]	[0.575]		[0.041]	[0.075]
March	0.416	0.094**	0.119**	0.372	0.083~	0.113**	0.383	0.044**	0.057***
		(0.044)	(0.035)		(0.042)	(0.033)		(0.020)	(0.015)
		[0.403]	[0.650]		[0.389]	[0.647]		[0.048]	[0.079]
April	0.488	0.072~	0.091**	0.442	0.067~	0.091**	0.413	0.044**	0.055***
		(0.037)	(0.031)		(0.037)	(0.031)		(0.019)	(0.015)
		[0.426]	[0.636]		[0.411]	[0.635]		[0.045]	[0.075]
May	0.522	0.078**	0.094**	0.474	0.075**	0.097**	0.435	0.047**	0.057***
		(0.036)	(0.030)		(0.036)	(0.030)		(0.019)	(0.014)
		[0.436]	[0.624]		[0.417]	[0.620]		[0.044]	[0.071]
June	0.526	0.106**	0.119***	0.479	0.101**	0.121***	0.447	0.047**	0.057***
		(0.040)	(0.032)		(0.040)	(0.033)		(0.019)	(0.014)
		[0.426]	[0.614]		[0.404]	[0.603]		[0.042]	[0.069]
July	0.567	0.081**	0.090**	0.520	0.079**	0.095**	0.452	0.047**	0.055***
		(0.034)	(0.029)		(0.036)	(0.031)		(0.019)	(0.014)
		[0.447]	[0.613]		[0.418]	[0.591]		[0.041]	[0.068]
Covariates			X			X			X

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment effects derived from regression models that regress each outcome on an indicator for treatment assignment at the school level and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. R-square statics for each model reported in square brackets. Student level covariates include indicators for race / ethnicity, gender and FRL status. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups. N = 21,001.

Table 7. Impacts on college application submission, Texas schools study

	Submitted any applications		Number of applications submitted	
Treatment	0.079**	0.078***	0.124	0.110
	(0.027)	(0.023)	(0.165)	(0.144)
Control mean	0.710		1.920	
Covariates		X		X
R-squared	0.055	0.079	0.212	0.247

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment effects derived from regression models that regress each outcome on an indicator for treatment assignment at the school level and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. Student level covariates include indicators for race / ethnicity, gender and FRL status. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups. N = 21,001.

Table 8. Impacts on Fall 2016 and Fall 2017 college enrollment outcomes, national study

	Control Mean	Treatment	R^2
Fall 2016 enrollment	0.608	-0.014**	0.159
		(0.006)	
Fall 2016 four-year college enrollment	0.413	-0.009	0.259
		(0.007)	
Fall 2016, two-year college enrollment	0.195	-0.005	0.067
		(0.006)	
Fall 2017 enrollment	0.556	-0.012*	0.181
		(0.006)	
Fall 2017 four-year college enrollment	0.375	-0.009	0.261
		(0.006)	
Fall 2017, two-year college enrollment	0.181	-0.003	0.056
		(0.006)	

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment effects derived from regression models that regress each outcome on an indicator for treatment assignment at the school level and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. Covariates include student-level measures reported in Table 2 and school-aggregate measures of PSAT performance and percent FRL. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups. N = 70,285.

Table 9. Impacts on Fall 2016 and Fall 2017 college enrollment outcomes, Texas schools study

	Control Mean	Treatment	R^2
Fall 2016 enrollment	0.513	0.017	0.154
		(0.015)	
Fall 2016 four-year college enrollment	0.328	0.009	0.185
		(0.016)	
Fall 2016, two-year college enrollment	0.190	0.008	0.024
		(0.009)	
Fall 2017 enrollment	0.479	0.020	0.168
		(0.014)	
Fall 2017 four-year college enrollment	0.297	0.009	0.205
		(0.015)	
Fall 2017, two-year college enrollment	0.189	0.009	0.020
		(0.009)	

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment effects derived from regression models that regress each outcome on an indicator for treatment assignment at the school level and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. Student level covariates include indicators for race / ethnicity, gender and FRL status. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups. N = 21,001.

Table 10. Sequence of outcomes for subgroups defined by GPA and FRL status, Texas schools study

	Take SAT	Submit college application	Complete FAFSA by July 2016	SAT + Apply + FAFSA	Fall 2016 enrollment	Fall 2017 enrollment
			Low GPA, non-FRL (N = 3,624)			
Treatment	0.061**	0.150***	0.096***	0.049**	0.085**	0.047~
	(0.030)	(0.037)	(0.023)	(0.016)	(0.027)	(0.024)
Control mean	0.396	0.510	0.262	0.108	0.358	0.327
R^2	0.290	0.124	0.070	0.053	0.100	0.084
			Low GPA, FRL (N = 7,001)			
Treatment	0.059**	0.083**	0.018	0.003	0.010	0.012
	(0.028)	(0.039)	(0.020)	(0.017)	(0.015)	(0.012)
Control mean	0.371	0.626	0.334	0.138	0.270	0.233
R^2	0.157	0.113	0.070	0.032	0.061	0.062
			High GPA, non-FRL (N = 5,949)			
Treatment	0.013	0.019	0.049**	0.033**	-0.028	-0.021
	(0.016)	(0.020)	(0.015)	(0.013)	(0.020)	(0.019)
Control mean	0.672	0.820	0.610	0.376	0.824	0.803
R^2	0.402	0.043	0.045	0.134	0.074	0.086
			High GPA, FRL (N = 4,427)			
Treatment	-0.005	0.029	0.016	-0.003	-0.019	-0.004
	(0.029)	(0.022)	(0.020)	(0.025)	(0.019)	(0.022)
Control mean	0.494	0.861	0.612	0.285	0.641	0.593
R^2	0.259	0.108	0.086	0.078	0.107	0.117

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001; Notes: Treatment effects derived from regression models that regress each outcome on an indicator for treatment assignment at the school level and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. Student level covariates include indicators for race / ethnicity, gender and FRL status. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups.

Table 11. Relationship between fall on-time enrollment and selection for FAFSA verification among FAFSA filers for subgroups defined by GPA and FRL status, Texas schools study

	Overall		Low GPA, non-FRL		Low GPA, FRL		High GPA, non-FRL		High GPA, FRL	
	Treatment schools	Control schools	Treatment Schools	Control Schools	Treatment Schools	Control Schools	Treatment Schools	Control Schools	Treatment Schools	Control Schools
FAFSA Verification	-0.059***	-0.076***	-0.022	-0.085**	-0.078**	-0.089***	-0.025	-0.063**	-0.071**	-0.079***
	(0.014)	(0.012)	(0.041)	(0.039)	(0.029)	(0.024)	(0.023)	(0.020)	(0.025)	(0.020)
Control mean	0.724	0.752	0.629	0.641	0.535	0.535	0.857	0.897	0.774	0.815
N	4756	6538	622	698	1228	1746	1617	2334	1289	1760
R-squared	0.087	0.118	0.113	0.155	0.072	0.089	0.063	0.056	0.074	0.067

~ p<0.10, * p<0.05, ** p<0.01, *** p<0.001

Notes: Treatment effects derived from regression models that regress each outcome on an indicator for treatment assignment at the school level and fixed effects for groups within which we randomized schools. Standard errors (in parentheses) clustered at the school level. Student level covariates include indicators for race / ethnicity, gender and FRL status. All analyses weighted by inverse of probability of assignment to experimental condition to handle variation in assignment probability across groups.