ONLINE APPENDICES, NOT FOR PUBLICATION

Appendix A. Additional Tables and Figures

Appendix Table 1. Randomization timing and contact rates

Randomization date	Total	Treatment	Control	Contact rate
First cohort				
January 16, 2020	1735	871	864	56%
January 24, 2020	641	325	316	56%
February 11, 2020	739	374	365	56%
February 21, 2020	927	469	458	49%
Second cohort				
November 2, 2020	1025	519	506	53%
November 6, 2020	915	462	453	42%
November 13, 2020	213	111	102	40%
November 20, 2020	519	264	255	48%
December 6, 2020	176	91	85	43%
December 15, 2020	492	250	242	39%
January 13, 2021	168	88	80	39%
January 21, 2021	264	137	127	39%
February 5, 2021	226	115	111	36%

Notes. Each row identifies the number of students who opted in to the experiment and the date of their randomization by the authors. Contact rates identify the percent of treated students who were recorded by InsideTrack software as contacting the coach via text, email, or phone.

Appendix Table 2. Differences in background characteristics of experimental students compared to students who received outreach but did not respond

	Mean value of non-participants	Experimental sample
Current age	24.281	-0.623**
		(0.017)
Female	0.571	0.042**
		(0.006)
College-educated parent	0.290	-0.069**
		(0.005)
GPA	2.854	-0.047**
		(0.006)
High school free and reduced price lunch	0.632	0.047**
		(0.003)
High school location		
Urban	0.370	0.070**
		(0.006)
Suburban	0.355	0.012*
		(0.005)
Town/rural	0.101	-0.018**
		(0.003)
High school ethnicity		
African-American	0.058	0.006*
		(0.003)
Asian	0.076	-0.011**
		(0.003)
Hispanic	0.554	0.124**
		(0.006)
White	0.129	-0.057**
		(0.004)
College enrollment outcomes		
Any enrollment	0.345	0.009+
		(0.005)
Community college	0.253	0.042**
		(0.005)
CSU/UC	0.123	-0.053**
		(0.004)

Notes: Point estimates and robust standard errors in parentheses from a regression of the listed characteristic on an indicator for being part of the experimental sample; ** $p<0.01$, * $p<0.05$, + $p<0.1$. Regression sample is limited to participants who were sent an email or text with an offer to participate in the experiment in either of the two cohorts (n = 161,526). Unless otherwise stated all values are from the student's first FAFSA submission. High school free and reduced-price lunch (FRPL) percentage and urbanicity are from the Common Core of Data linked to each student's high school identifier from their original Cal Grant application. Ethnicity is derived from a match between CSAC and the California Department of Education but excludes students whose first Cal Grant year was 2014.

Appendix Table 3: Intent-to-treat estimates of the offer of coaching on communications with coaches

	(1)	(2)	(3)	(4)
	Any format	Text	Email	Phone
Total communications	2.889**	2.344**	0.171**	0.336**
	(0.170)	(0.157)	(0.015)	(0.021)
Control group mean	0.025	0.015	0.002	0.007
Had at least one communication	0.486**	0.389**	0.086**	0.176**
	(0.009)	(0.009)	(0.006)	(0.007)
Control group mean	0.009	0.004	0.001	0.006
Had at least two communications	0.298**	0.234**	0.036**	0.068**
	(0.009)	(0.008)	(0.003)	(0.005)
Control group mean	0.002	0.001	0.000	0.001

Notes: Point estimates and robust standard errors clustered by randomization strata (de Chaisemartin & Ramirez-Cuellar, 2020) in parentheses from a regression of communications between participants and coaches on assignment to treatment; ** $p < 0.01$. InsideTrack records all incoming communications between coach and students according to one of three formats: text/SMS; emails; and phone calls. Coaches may agree to student requests for alternate formats, such as video calls, but coaches do not offer this format and must wait for students to initiate any alternative. A small number of control group students were mistakenly re-assigned to the treatment sample (as described in the text), which accounts for why control group means are not completely null. Regressions also include randomization block fixed effects (mutually exclusive groups defined by cohort, round of randomization, and year of first and last Cal Grant receipt) and pre-registered covariates (indicators for female and having a college-educated parent, zip code level median household income, high school percent free & reduced-price lunch, high school urbanicity dummies (urban, suburban, town, rural), and dummies for students with missing values).

Appendix Table 4. Intent-to-treat estimates of the offer of coaching on postsecondary enrollment, using alternate enrollment data

Appendix Table 3. Intent-to-treat estimates of the offer of coaching on postsecondary enrollment, alternate enrollment data

		(1)	(2)	(3)	(4)	(5)	(6)
		CSAC enrollment data (Fall)			CSAC or NSC enrollment data (Fall)		
	N	Any	CC	CSU/UC	Any	CC	Other
All students	8040	-0.001	-0.003	0.008+	0.003	-0.002	0.012+
		(0.009)	(0.010)	(0.005)	(0.010)	(0.011)	(0.006)
Control group mean		0.357	0.297	0.067	0.407	0.313	0.107
First cohort	4042	0.020	0.015	0.014+	0.014	0.011	0.016
		(0.013)	(0.016)	(0.008)	(0.014)	(0.017)	(0.011)
Control group mean		0.439	0.350	0.102	0.509	0.372	0.157
Second cohort	3998	-0.022+	-0.020+	0.001	-0.008	-0.016	0.007
		(0.011)	(0.011)	(0.005)	(0.014)	(0.012)	(0.007)
Control group mean		0.273	0.242	0.032	0.304	0.252	0.057
Test of equality (p-value)		0.018	0.067	0.184	0.259	0.192	0.470

Notes: Columns (1) - (3) show treatment effects on postsecondary enrollment using in-state, public college enrollment data provided to CSAC, as described in the text. Columns (4) - (6) show treatment effects on postsecondary enrollment by defining a student as enrolled if they appear enrolled in either the CSAC data or NSC data. Point estimates and robust standard errors clustered by randomization strata (de Chaisemartin & Ramirez-Cuellar, 2020) in parentheses from a regression of enrollment in the specified sector on assignment to treatment; + p <0.1. Regressions also include randomization block fixed effects (mutually exclusive groups defined by cohort, round of randomization, and year of first and last Cal Grant receipt) and pre-registered covariates (indicators for female and having a college-educated parent, zip code level median household income, high school percent free & reduced-price lunch, high school urbanicity dummies (urban, suburban, town, rural), and dummies for students with missing values). The bottom row of the table shows p-values from a test of hypothesis of equal treatment effects for cohorts 1 and 2.

Appendix Table 5. Intent-to-treat estimates of the offer of coaching on postsecondary enrollment (NSC data) by randomization cohort and prior contact

	N	At least one communication (treatment group)	(1) NSC enrollment data (Fall) Any	(2) NSC enrollment data (Fall) CC	(3) NSC enrollment data (Fall) Other
All students	8040	0.494	0.000	-0.006	0.008
			(0.009)	(0.009)	(0.006)
Control group mean			0.331	0.234	0.101
First cohort	4042	0.545	0.013	0.007	0.011
			(0.013)	(0.014)	(0.011)
Control group mean			0.418	0.278	0.147
Second cohort	3998	0.444	-0.013	-0.019+	0.005
			(0.012)	(0.010)	(0.007)
Control group mean			0.241	0.19	0.053
Second cohort	2945	0.428	-0.018	-0.026*	0.007
Experienced outreach in first year			(0.015)	(0.013)	(0.008)
Control group mean			0.238	0.183	0.057
Second cohort	1053	0.490	0.002	0.003	0.001
Did not experience outreach in first year			(0.023)	(0.016)	(0.013)
Control group mean			0.25	0.208	0.042
Combination of first cohort and students	5095	0.533	0.010	0.006	0.009
who did not experience outreach in first year			(0.011)	(0.012)	(0.009)
			0.384	0.263	0.126

Notes: Point estimates and robust standard errors clustered by randomization strata (de Chaisemartin & Ramirez-Cuellar, 2020) in parentheses from a regression of enrollment in the specified sector on assignment to treatment; ** $p<0.01$, * $p<0.05$, + $p<0.1$. Regressions also include randomization block fixed effects (mutually exclusive groups defined by cohort, round of randomization, and year of first and last Cal Grant receipt) and pre-registered covariates (indicators for female and having a college-educated parent, zip code level median household income, high school percent free & reduced-price lunch, high school urbanicity dummies (urban, suburban, town, rural), and dummies for students with missing values).

Appendix Table 6. Intent-to-treat estimates of the offer of coaching on longer-run postsecondary enrollment (CSAC data)

		(1)	(2)	(3)	(4)	(5)	(6)
		CSAC enrollment data Fall of year 2			CSAC enrollment data Fall of year 1 and year 2		
	N	Any	CC	CSU/UC	Any	CC	CSU/UC
All students	8040	-0.001	-0.002	0.004	0.005	0.002	0.009*
		(0.009)	(0.009)	(0.006)	(0.008)	(0.007)	(0.004)
Control group mean		0.295	0.231	0.072	0.209	0.153	0.040
First cohort	4042	0.017	0.015	0.006	0.024+	0.017	0.016**
		(0.014)	(0.013)	(0.010)	(0.013)	(0.012)	(0.006)
Control group mean		0.357	0.261	0.111	0.272	0.188	0.059
Second cohort	3998	-0.020+	-0.020+	0.003	-0.013	-0.013	0.002
		(0.012)	(0.012)	(0.006)	(0.010)	(0.008)	(0.005)
Control group mean		0.230	0.201	0.032	0.145	0.118	0.020
Test of equality (p-value)		0.043	0.049	0.763	0.026	0.041	0.063

Notes: Point estimates and robust standard errors clustered by randomization strata (de Chaisemartin & Ramirez-Cuellar, 2020) in parentheses from a regression of enrollment in the specified sector on assignment to treatment; ** $p<0.01$, * $p<0.05$, + $p<0.1$. Regressions also include randomization block fixed effects (mutually exclusive groups defined by cohort, round of randomization, and year of first and last Cal Grant receipt) and pre-registered covariates (indicators for female and having a college-educated parent, zip code level median household income, high school percent free & reduced-price lunch, high school urbanicity dummies (urban, suburban, town, rural), and dummies for students with missing values). The bottom row of the table shows p-values from a test of hypothesis of equal treatment effects for cohorts 1 and 2. Enrollment outcomes are measured using CSAC data.

Appendix Table 7. Differences in background characteristics of participants assigned to the treatment group by coaching participation

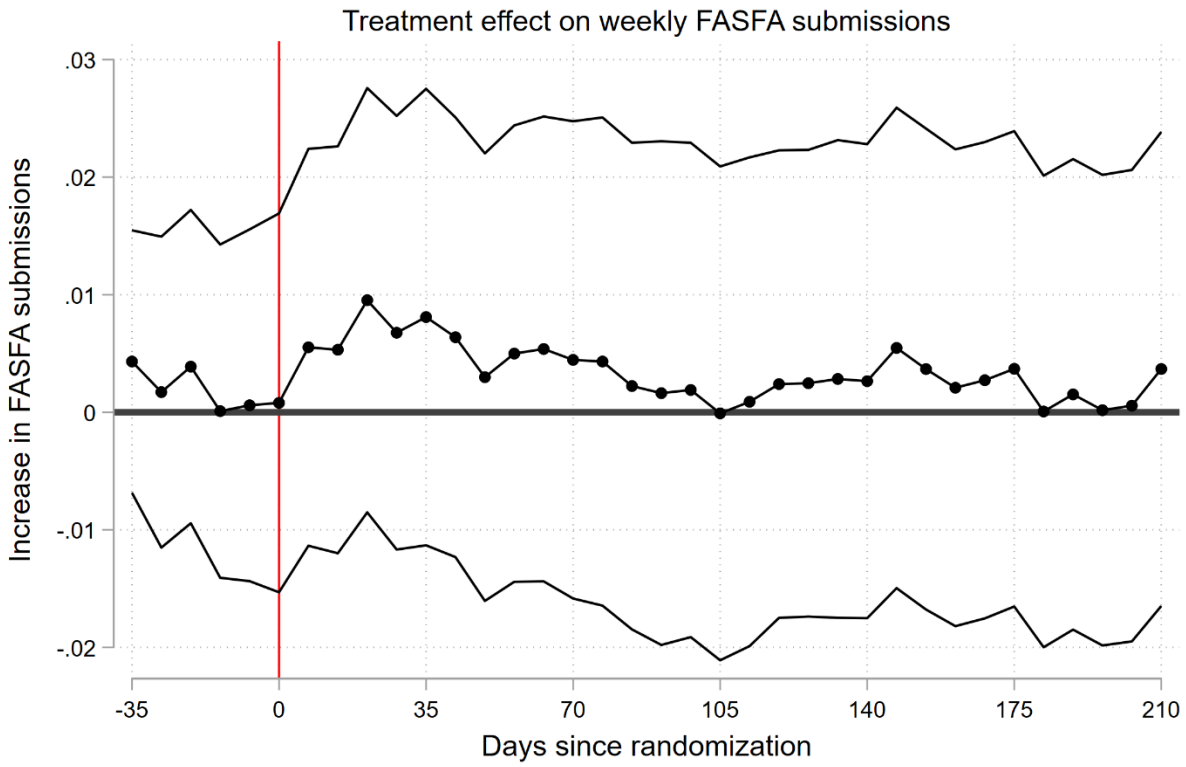
						High school urbanicity			
Pre-registered covariates	Female	College-educated parent	Zip code median HH income	Percent FRPL at high school	Missing FRPL	Urban	Suburban	Town	Rural
Had communication	-0.024	-0.015	298.779	-0.020**	0.013	-0.011	0.001	-0.008	0.005
	(0.015)	(0.013)	(666.905)	(0.007)	(0.009)	(0.016)	(0.015)	(0.006)	(0.006)
Control group mean	0.625	0.236	60,329	0.693	0.091	0.443	0.381	0.047	0.038

Additional covariates	GPA	Missing GPA	Current age	Family size	Goal: Bachelor's degree	Hispanic	Asian	Black	White
Had communication	0.028+	-0.006	-0.121*	0.023	0.049**	-0.011	0.022*	-0.011	0.005
	(0.015)	(0.007)	(0.047)	(0.045)	(0.016)	(0.016)	(0.009)	(0.008)	(0.008)
Control group mean	2.814	0.053	23.204	3.912	0.508	0.682	0.058	0.068	0.064

	Sector of initial Cal Grant payment		Enrolled during outreach		
	CC	CSU	Any college	CC	CSU
Had communication	-0.048**	0.048**	0.019	0.010	0.013
	(0.014)	(0.014)	(0.015)	(0.015)	(0.009)
Control group mean	0.721	0.279	0.405	0.335	0.087

Notes: Point estimates and robust standard errors in parentheses from a regression of the listed characteristic on an indicator for having any contact with a coach; ** $p<0.01$, * $p<0.05$, + $p<0.1$. Regression sample is limited to participants assigned to the treatment group (N = 4,076). Unless otherwise stated all values are from the student's first FAFSA submission. High school free and reduced-price lunch (FRPL) percentage and urbanicity are from the Common Core of Data linked to each student's high school identifier from their original Cal Grant application. Ethnicity is derived from a match between CSAC and the California Department of Education but excludes students whose first Cal Grant year was 2014. "Enrolled during outreach" refers to whether NSC data shows the student was enrolled in the year in which outreach was occurring.

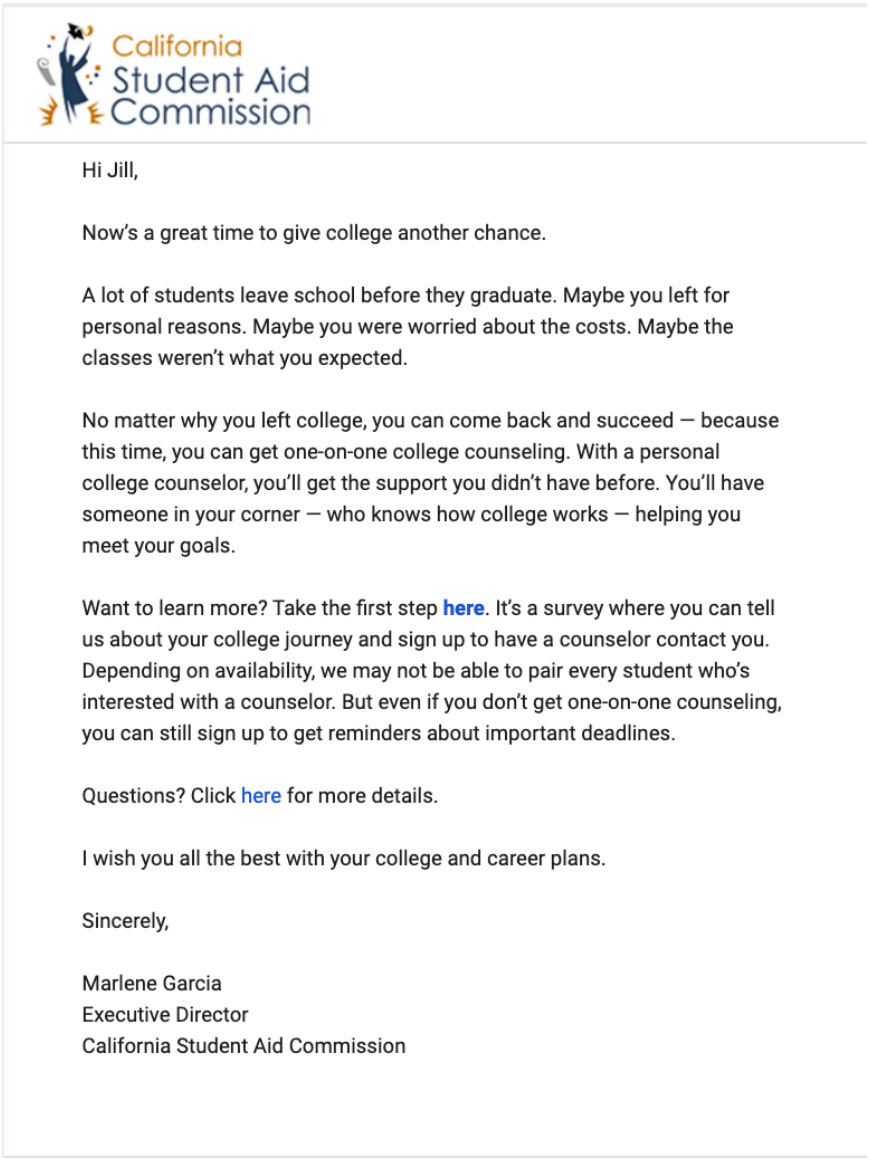
Appendix Figure 1. Regression estimates of FAFSA submission, by week



Notes. Point estimates and 95% confidence intervals are from separate regressions of treatment status on the likelihood of completing a FAFSA by the specific date. See text for additional control variables. Regressions measure weekly treatment effects that span the period from five weeks before treatment assignment to 30 weeks after treatment assignment. Standard errors clustered by randomization strata.

Appendix B. Sample emails and text messages

First outreach email:



California Student Aid Commission

Hi Jill,

Now's a great time to give college another chance.

A lot of students leave school before they graduate. Maybe you left for personal reasons. Maybe you were worried about the costs. Maybe the classes weren't what you expected.

No matter why you left college, you can come back and succeed — because this time, you can get one-on-one college counseling. With a personal college counselor, you'll get the support you didn't have before. You'll have someone in your corner — who knows how college works — helping you meet your goals.

Want to learn more? Take the first step **here**. It's a survey where you can tell us about your college journey and sign up to have a counselor contact you. Depending on availability, we may not be able to pair every student who's interested with a counselor. But even if you don't get one-on-one counseling, you can still sign up to get reminders about important deadlines.

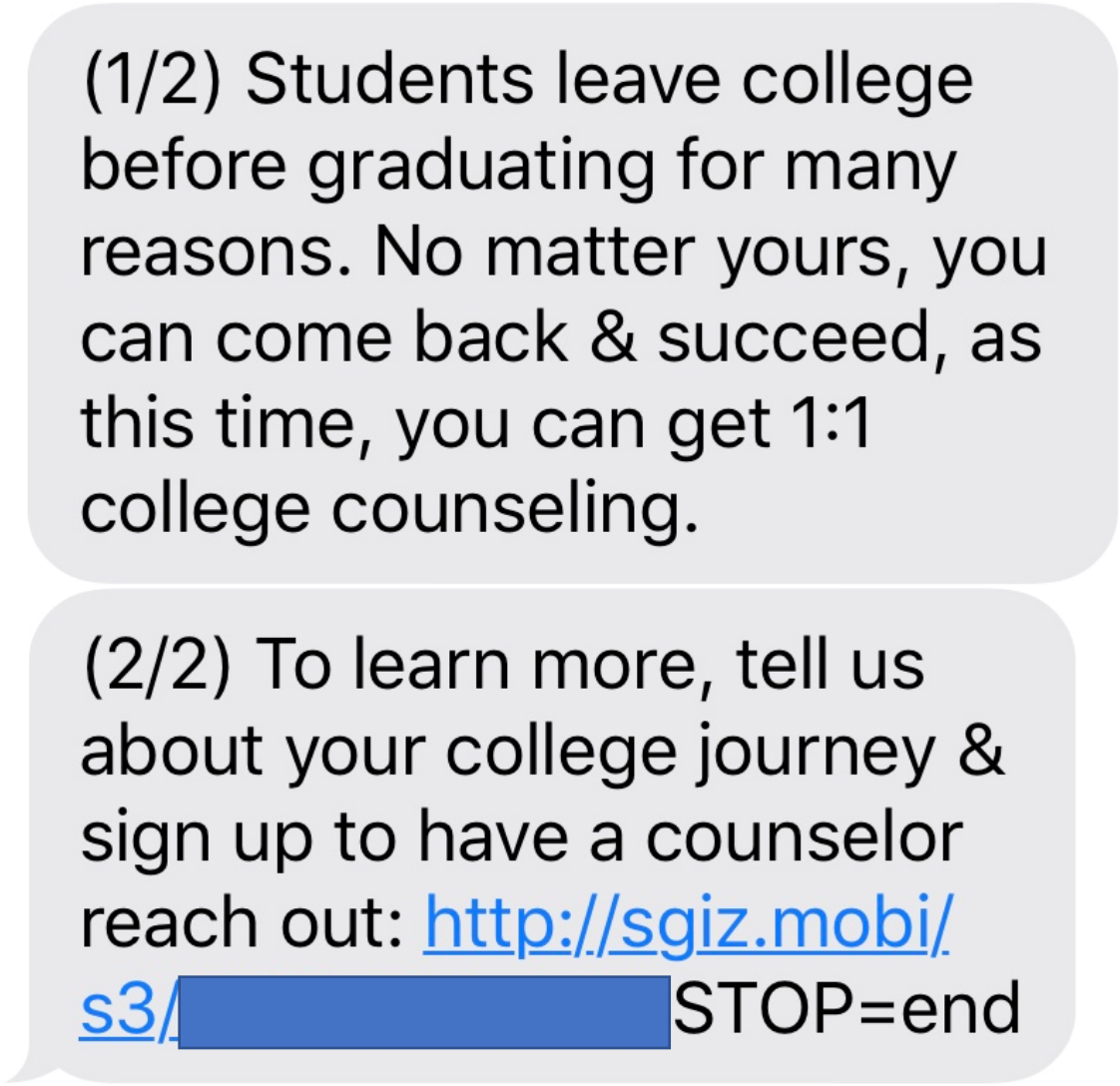
Questions? Click here for more details.

I wish you all the best with your college and career plans.

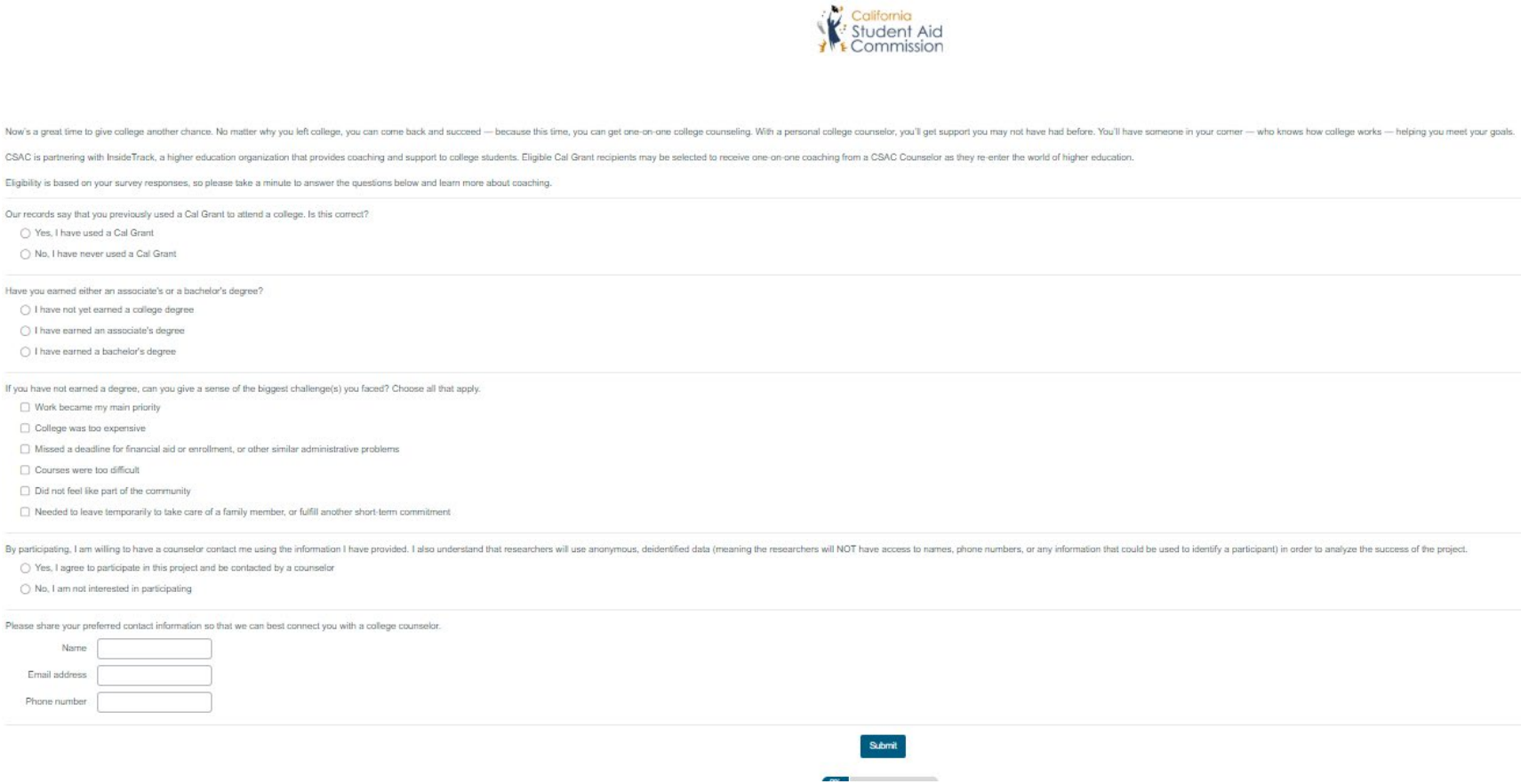
Sincerely,

Marlene Garcia
Executive Director
California Student Aid Commission

First outreach text:



Initial opt-in survey:



California Student Aid Commission

Now's a great time to give college another chance. No matter why you left college, you can come back and succeed — because this time, you can get one-on-one college counseling. With a personal college counselor, you'll get support you may not have had before. You'll have someone in your corner — who knows how college works — helping you meet your goals.

CSAC is partnering with InsideTrack, a higher education organization that provides coaching and support to college students. Eligible Cal Grant recipients may be selected to receive one-on-one coaching from a CSAC Counselor as they re-enter the world of higher education.

Eligibility is based on your survey responses, so please take a minute to answer the questions below and learn more about coaching.

Our records say that you previously used a Cal Grant to attend a college. Is this correct?

- ○ Yes, I have used a Cal Grant
- ○ No, I have never used a Cal Grant

Have you earned either an associate's or a bachelor's degree?

- ○ I have not yet earned a college degree
- ○ I have earned an associate's degree
- ○ I have earned a bachelor's degree

If you have not earned a degree, can you give a sense of the biggest challenge(s) you faced? Choose all that apply.

- ☐ Work became my main priority
- ☐ College was too expensive
- ☐ Missed a deadline for financial aid or enrollment, or other similar administrative problems
- ☐ Courses were too difficult
- ☐ Did not feel like part of the community
- ☐ Needed to leave temporarily to take care of a family member, or fulfill another short-term commitment

By participating, I am willing to have a counselor contact me using the information I have provided. I also understand that researchers will use anonymous, deidentified data (meaning the researchers will NOT have access to names, phone numbers, or any information that could be used to identify a participant) in order to analyze the success of the project.

- ○ Yes, I agree to participate in this project and be contacted by a counselor
- ○ No, I am not interested in participating

Please share your preferred contact information so that we can best connect you with a college counselor.

Name

Email address

Phone number

Submit

Appendix C. Discussion of enrollment outcomes measures

We used two distinct data sources to estimate treatment effects on postsecondary enrollment outcomes. The first data source is National Student Clearinghouse (NSC), which provides enrollment and degree receipt information at most colleges nationwide (Dynarski, Hemelt, & Hyman, 2015). NSC data has been used as a primary source of enrollment outcomes across many prior research studies.

The second data source is enrollment data from California's public colleges, which we refer to as "CSAC enrollment" data. These data include dummy variables that identify fall enrollment in the California State University (CSU) and University of California (UC) systems and term-level enrollment (Fall and Spring) in California community colleges. CSAC is provided these data each fall to verify public college enrollment, in order to properly administer its state aid programs.

These two data sources differ in a number of ways. The first is who is covered by the data sources. CSAC enrollment data only covers in-state public colleges. In addition, CSAC enrollment data only requires CSU and UC colleges to share enrollment data for students who submitted the FAFSA in that year, though in practice we observe that some non-FAFSA submitting students are identified in the CSAC enrollment data. In contrast, NSC theoretically covers most colleges nationwide, including in-state private colleges and out of state colleges, though is subject to concerns around "FERPA blocking" that might limit what data is actually shared; see NSC's Impact of Directory Information Blocks on StudentTracker Results (2017) for more details. NSC covers a broader range of colleges, though in our study almost all students enroll in a community college or CSU, thus limiting potential differences along this dimension.

The second key difference is each dataset is linked to experimental sample members in a different way. CSAC enrollment data includes student SSN, which is used for matching purposes. In contrast, NSC data are matching via a proprietary process that focuses on a student's name and birthdate. Thus matching rates are likely more accurate in the CSAC enrollment data.

Finally, the timing of the data linkage also differs. CSAC enrollment data is provided to CSAC from each college during September each year, but the exact time at which these data are transferred varies by college and may reflect slightly different enrollment dates. In our study, the NSC data were matched in March 2022, compared to the CSAC enrollment data which were shared in Fall 2021. If there were updates or data errors that were corrected between Fall and Spring, NSC would be more likely to have corrected data. One area where this might theoretically occur – though we do not know how this actually occurs in practice – is if students withdraw later in the Fall term, in such a way that they

are considered to have been "not enrolled" in the NSC data by the time we have received it.

Ultimately, both data sources produced similar results and led us to the same conclusions, though we describe some potential differences below. Any statistics reported here are relevant to the sample of former college students who received state aid at a community college or California State University college, and may not be broadly applicable.

In short, we find that: (1) The NSC and CSAC enrollment data align in most cases; (2) CSAC enrollment data is better at identifying community college enrollment, and; (3) NSC enrollment data is able to identify a small amount of additional CSU enrollment and alternative, non-public enrollment. We discuss each finding in turn:

1. For 87% of individuals in our 8,040 student sample, the two data sources match on enrollment (i.e., both data sources say the student is enrolled or not enrolled). This number is even higher for just community college enrollment, for which the two data sets align 90% of the time (i.e., both data sources say the student is enrolled in a CC or not enrolled in a CC).

2. When the two data sources DO NOT match on community college enrollment, it is mostly (83%) that the CSAC data identifies community college enrollment but the NSC does not. In our control group sample, the CSAC enrollment data finds that 29.7% of students enrolled in a community college in the Fall, compared to 23.4% when using the NSC data.

   While we do not know exactly why this occurs, two potential explanations are that (i) NSC's matching by name and birthdate produces "non-matches" in their data and (ii) NSC data may be updated such that students who drop out through the semester are listed as enrolled in CSAC enrollment data but not in the (potentially updated) NSC data.

   An additional explanation may be that a few community colleges report inconsistently to the NSC. One concern with NSC data generally is that some specific colleges may choose not to participate in the StudentTracker service entirely. In our data, students who showed up as enrolled in a CC in the CSAC data but not in the NSC data attended 88 distinct community colleges, but only 6 colleges accounted for a third of this "missingness". Thus, some CCs may have large "FERPA blocking" policies that inhibit our ability to observe many students in the NSC data.

3. When the two data sources DO NOT match on enrollment other than community colleges, the student shows up as enrolled in the NSC but not in the CSAC data 83% of the time. In our control group sample, the NSC enrollment data finds that 10.1% of students enroll in something other than an in-state community college in the Fall, compared to 6.7% of students enrolling in a CSU or UC in the CSAC enrollment data.

   One-half of this "missing" enrollment (though it's worth reiterating that this is just a small part of the experimental sample) appears in a CSU; as noted above, this may be as CSAC only receives CSU enrollment data if the student also submitted the FAFSA in that year. The other half of this "missing" enrollment covers tens of different universities, many of them private colleges which would never show up in the CSAC enrollment data.

In sum, we do not find that these differences in the data sources alter any of our findings, though this may be due to our specific sample of older, "non-traditional" students who primarily enroll in community colleges and the CSU system. Alternate research projects with different samples may produce different conclusions.