Supporting Information (SI)

Contents

A Summer bridge programs at highly-selective private universities . . . S2
B Past interventions to reduce FGLI disparities . . . S2
C Additional details on randomization procedure . . . S5
C.1 Overview of randomization approach . . . S5
C.2 Post-randomization balance on observed characteristics . . . S8
D Secondary research questions . . . S8
E Additional details on estimation from pre-analysis plan . . . S9
E.1 Adjusting for unequal probability of assignment across priority tiers . . . S9
E.2 Additional details: estimation approach for ITT analysis . . . S9
E.3 Additional details: estimation approach for complier average causal effect analysis . . . S9
F Definition of compliers . . . S9
G Full regression tables/additional results: main outcomes and secondary outcomes . . . S10
G.1 Full regression tables across specifications: program difficulty measures . . . S10
G.2 Full regression tables across specifications: overall GPA, withdrawal, and too-few credits . . . S13
G.3 Full regression tables across specifications: STEM and writing GPA . . . S15
H GPA adjusted for difficulty analysis . . . S17
I Heterogeneous treatment effects analysis . . . S17
J Causal effect on visits to academic help centers . . . S19

A. Summer bridge programs at highly-selective private universities. Below is an alphabetical and non-exhaustive list of summer bridge programs at highly-selective private universities:

1. **Columbia's Academic Success Program (ASP) Summer Bridge**: five week summer session; limited details available publicly.

2. **Cornell's Prefreshman Summer program**: seven week summer session described as: "PSP offers wide-ranging opportunities to prepare students for Cornell by challenging and supporting them as they develop new ways of thinking and approaching academic work. Students become familiar with university resources while interacting with other students, faculty and staff." https://cals.cornell.edu/undergraduate-students/student-services/new-students/prefreshman-summer-program-psp

3. **Dartmouth's First-Year Student Enrichment Program (FYSEP):** four week summer session. Described as: "FYSEP provides a rigorous, dynamic and transformative experience that puts participants in a position to thrive at Dartmouth both academically and socially. The program offers sample classes with Dartmouth faculty, workshops, activities, and seminars designed to simulate life at Dartmouth and to prepare participants to handle some of the challenges they may face during the course of their first year." https://students.dartmouth.edu/fgo/programs/first-year-student-enrichment-program

4. **MIT's Interphase EDGE (previously Interphase)**: eight week summer session. Described as: "The Interphase EDGE/x curriculum is uniquely designed to impart pivotal concepts that will increase long-term academic success. In other words, the program will not only give students an "edge" on their MIT experience, it will catalyze their success beyond MIT." https://ome.mit.edu/programs/interphase-edge-empowering-discovery-gateway-excellence

5. **Princeton's Freshman Scholars Institute**: eight week summer session described as: "FSI is an 8-week program that offers scholars the opportunity to take ownership over their transition to Princeton by giving them the resources they need to shape their educational journey while preparing to become future campus leaders and peer mentors. During the program, our scholars immerse themselves in the intellectually vibrant culture at Princeton through seminar-style courses and laboratory research experiences, to engage with their fellow scholars in a variety of co-curricular, community-building activities, and to work closely with faculty members from a range of academic disciplines and fields." https://fsi.princeton.edu/faq

6. **Stanford's Summer Engineering Academy (SSEA)**: four week summer session focused on STEM and described as: "We seek to illuminate the brilliance of students who have been systemically marginalized in engineering. We welcome applications from all students, including womxn, first generation and/or low income, first in their family to study/pursue an engineering degree, or those from environments with limited exposure to engineering curriculum." https://engineering.stanford.edu/students-academics/equity-and-inclusion-initiatives/undergraduate-programs/stanford-summer

7. **University of Pennsylvania's Pre-Freshman Program (PFP)**: four week summer session described as: "Program participants receive comprehensive support services that begin with PFP and continue throughout the students' undergraduate experience at Penn. PFP is a chance for participating students to get an academic and social edge, while quickly becoming familiar with campus resources and the Penn community. The program, while academic in nature, encourages students to form lasting bonds of friendships through regular social and cultural activities. Students who have participated in the program report that it has made the difference in their life at Penn." https://pennfirstplus.upenn.edu/pre-freshman-program/

8. **Yale's First-Year Scholars**: six week summer session described as: "First-Year Scholars will have the opportunity to develop close relationships. The tight-knit community that emerges from the FSY program will help support First-Year Scholars as they enter their first year at Yale. The FSY experience, including the support systems and mentors in the program, will help position students to take advantage of the many choices and opportunities that await them as students at Yale." https://fsy.yale.edu/about

B. Past interventions to reduce FGLI disparities. Tables S1 and S2 summarize the past interventions on reducing FGLI disparities we were able to find. Our study contributes to this literature by being a relatively unusual combination of two features relevant to our design. First, our study features a long, intensive, residential program rather than a brief psychological intervention. Most existing studies focus on brief, psychological interventions. (30) is a notable exception where they study a remedial academic training program over the course of several weeks. However, unlike this study which uses volunteers, we randomize eligible students into the treatment and control conditions. Thus, the target population of our RCT more closely aligns with that of FGLI programs.

Table S1. Other interventions to promote success of FGLI students..

Study	Modality	Length	Content	University context
(16) (Stephens, Hamedani, Destin)	In-person	1 hour	Difference-education intervention	Four-year private university
(21) (Yeager, et al) Experiment 1	Online	25-35 minutes	Mindset intervention	N/A (high school students)
(21) (Yeager, et al) Experiment 2	Online	Same as above	Same as above	High-quality four-year public university
(21) (Yeager, et al) Experiment 3	Online	Same as above	Same as above	Highly selective four-year private university
(26) (Townsend, Stephens, Smallets, Hamedani)	Online	Unknown	Similar content as (16)	"Large, private research university on the West coast of the United States"
(27) (Murphy, et al)	In-person	1 hour	Reading and writing exercise with content focused on belonging	Large, broad-access, Hispanic-serving Institution in the Midwest with a racially and economically diverse body
(30) (Wathington, Pretlow, Barnett)	In-person	3-6 hours / day x 4-5 days / week x 4-5 weeks = 48 - 150 hours	Remedial academic training	Six community colleges and two nonselective four-year institutions located in Texas
(28) (Walton, et al)	Online	Under 30 minutes	Social-belonging intervention	22 universities and colleges

Notes: The table shows that, with one exception, the majority of interventions are short in duration and focused on psychological elements of belonging. In contrast, the intervention we study is much longer in duration and focuses not only on psychological elements of the adjustment to college but also academic training.

Table S2. Other interventions to promote success of FGLI students (continued)

Study	Definition of first-generation low income/disadvantage	Measures of hardship collected beyond first-generation status	Opt in or automatically randomized (auto)?	Sample size	Outcomes
(16) (Stephens, Hamedani, Destin)	College students who do not have parents with 4-year degrees	N/A	Opt-in (answer a recruitment email)	N = 168; 147 with outcomes	Self-reported willingness to seek academic help; self-reported psychosocial outcomes; first-year GPA
(21) (Yeager, et al) Experiment 1	N/A	N/A	Auto	N = 584	Full-time enrollment during first year of college
(21) (Yeager, et al) Experiment 2	"Continuing-generation students, regardless of race / ethnicity, and African-American and Hispanic / Latino students, regardless of social class constituted the "disadvantaged" group."	N/A	Auto	N = 7,335	Full-time enrollment enrollment during first-year of college; first-year GPA
(21) (Yeager, et al) Experiment 3	"All Asian students and all continuing-generation European American students as not facing group-based disadvantages (i.e. "advantaged") and all African American, Latino, Native, Pacific Islander, and first-generation European American students as facing group-based disadvantages ("disadvantaged")."	N/A	Auto	N = 1,592	First-year GPA
(26) (Townsend, Stephens, Smallets, Hamedani)	College students who do not have parents with 4-year degrees (first-generation students) (also examine subgroup differences classifying White and Asian students advantaged; Black and Hispanic students as disadvantaged)	Pell receipt (56% of first-generation students)	Opt-in (respond to recruitment efforts)	N = 133	Second-year GPA; self-reported psychosocial outcomes
(30) (Wathington, Pretlow, Barnett)	Low score on placement exam for college matriculation	Self-reported socioeconomic variables (received public assistance i.e. food stamps, welfare, Section 8 housing, qualified for FRL, students held job for pay)	Opt in based on who responds to invitation after low score on placement exam	N = 1318	Persistence; credit accumulation; completion of math and writing courses
(27) (Murphy, et al)	African American, Latino, and Native American students as well as first-generation college students of any racial-ethnic background.	n/a	Auto	N=1063	Full-time enrollment for two years following intervention
(28) (Walton, et al)	First generation students and non-white or Asian students.	n/a	Randomized after opt-in to recruitment text	N=26,911	First year completion rate as full-time student.

Notes: This table shows two additional contributions of the present study relative to past research: (1) automatic randomization of eligible students, rather than asking students to opt in and (2) the collection of additional measures of hardship beyond generation status and race/ethnicity, such as family death and food insecurity, that help us understand the degree of disadvantage in the analytic sample.

C. Additional details on randomization procedure. The randomization procedure was designed to accommodate two interests: (1) construct a valid randomized experiment for use in estimating causal effects, while (2) trying to give a higher probability of receiving an invitation to students that program administrators designated as high-priority for the program. To implement this, we first obtained the following two measures of priority assigned by administrators to each student:

1. **Ordered categorical measure of the student's priority tier for *SB Program***: administrators placed each eligible student into one of four or five tiers, depending on the cohort.
   (a) *No invitation (lowest priority)*: these are students whom the administrators designate as lowest-priority. This tier only existed as a tier for the Summer 2017 cohort
   (b) *Invited to online SB Program*: these students are given slightly higher priority, but are not designated for an on-campus version of the program.
   (c) *Low priority for on-campus SB Program*
   (d) *Medium priority for on-campus SB Program*
   (e) *High priority for on-campus SB Program*
2. **Binary measure of whether the student was designated for an invitation**. The binary measure collapses the tiers into two categories:
   (a) *Not invited:* students who were either designated for no invitation or designated for SB Program online (categories (a) and (b) above).
   (b) *Invited:* students invited to on-campus program who fall into any of the three priority tiers.

Given the above priority measures, we developed a randomization procedure that satisfies the following two goals: (1) achieving balance along pre-treatment attributes between those invited to *SB Program* and those not invited; (2) giving higher-priority students a better shot at an invitation.

C.1. Overview of randomization approach. Figure S1 provides a step-by-step guide to the randomization approach. While our pre-analysis plan shows summer-specific details, the Figure shows the more general steps and high-level, summer-specific differences. In different years, we used the pre-treatment attributes outlined in Table S3 in different ways to satisfy the aforementioned two goals of randomization. The pre-analysis plan posted at https://osf.io/qh75m reports (1) precise details of the exact process for each of the summers, (2) the distribution of propensity scores before and after randomization, (3) the specific treatment assignment probabilities for different priority tiers, and (4) balance on pre-treatment attributes before and after randomization.

Table S4 shows the results of the randomization process and the relationship between: (1) the administrator-designated priority tiers and (2) the count of each student in a tier selected for an invitation, with Figure S2 showing the same information. Two points are worth noting. First, the analytic sample was comprised only of students flagged as "high priority" but not the highest priority to attend; the highest-priority students were automatically issued invitations rather than randomized with some probability. Second, due to other constraints such as the total number of slots available each summer, the randomization gave higher priority tiers a higher odds of receiving an invitation but the probabilities did not end up being strictly monotonically increasing.

Fig. S1. Randomization procedure

Step 1: Begin with pool of administrator-selected students identified as potentially eligible for invitation. We observe:
- Administrator-designated information (Y/N invite flag; priority tier)
- Pre-treatment student attributes (e.g., test scores; hardship flags)

↓

Step 2: Fit two models to estimate propensity scores for use in randomization, with student pre-treatment attributes as the predictors:
1. Logistic regression predicting Y/N invite by administrator
2. Ordered logistic regression predicting priority tier

↓

Step 3: Randomize within the administrator-defined priority tiers (i.e., low, medium, and high priority). Students in higher-priority tiers have higher odds of receiving the an invitation to the residential *SB Program*. Specifics vary across the summers and are described in the pre-analysis plan. Broadly:
Summers 2017 and 2019: within each priority tier, block on linear predictor from ordered logit and the students' highest score on a pre-matriculation standardized test (we recoded ACT scores to the SAT scale, and if students submitted both tests, took the higher of the two)
Summer 2018: within each priority tier, no blocking (did not substantially improve balance)

↓

Output: a list of students to invite to the residential version of *SB Program* + a waitlist with the order randomized. Administrators invite students who can respond by:
1. Attending in-person *SB Program*
2. Attending the online *SB Program*
3. Attending neither

Notes: The figure highlights the general steps we used to randomize students to either an invitation to the in-person version of SB Program or not, along with summer-specific variations.

Table S3. Covariates used to compare balance between those randomized to treatment versus control: data sources

Covariate	Details
Did not attend pipeline	Binary variable reflecting attendance at programs like QuestBridge; reported in application.
Neither parent attended college	Contrasts with (1) one parent attended some college; (2) 1+ parents completed college; reported in application.
Family death	Binary variable; coded by administrators using essays
Family hardship	Binary variable; coded by administrators using essays
Food insecurity	Binary variable; coded by administrators using essays
Highest test score (SAT scale)	Continuous variable; students apply with SAT score, ACT score, or both. We recoded ACT scores to the SAT scale for that year and then took the highest test score out of those reported.

Fig. S2. Proportion of students by tier randomized to an invitation to in-person SB Program

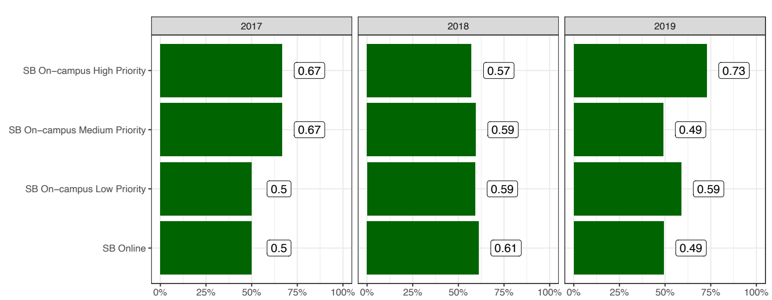


Table S4. Counts of students in treatment from each different administrator-designated priority group, by summer.

Year	Admin-designated Tier	Total Students in Tier	Tier's Proportion of Students that Summer	Count of In-Person Invites to Tier	Tier-Specific Treatment Probability
2017	SB On-campus High Priority	21	0.15	14	0.67
2017	SB On-campus Medium Priority	24	0.17	16	0.67
2017	SB On-campus Low Priority	46	0.33	23	0.50
2017	SB Online	50	0.35	25	0.50
2018	SB On-campus High Priority	14	0.11	8	0.57
2018	SB On-campus Medium Priority	37	0.28	22	0.59
2018	SB On-campus Low Priority	44	0.34	26	0.59
2018	SB Online	36	0.27	22	0.61
2019	SB On-campus High Priority	11	0.07	8	0.73
2019	SB On-campus Medium Priority	49	0.31	24	0.49
2019	SB On-campus Low Priority	34	0.22	20	0.59
2019	SB Online	63	0.40	31	0.49

C.2. Post-randomization balance on observed characteristics. There were eleven students who initially matriculated to *SB University* and were randomized but ultimately did not attend any classes at the university due to their own death, transfers, or other reasons. This brings our $N = 429$ randomization sample to an $N = 418$ analytic sample, with Table S5 showing the breakdown across summers. The treatment group attrition rate (2.9%) is similar to the control group attrition rate (2.1%), and we fail to find evidence that attrition is related to the treatment status ($\chi^2 = 0.05$, $p = 0.81$). As a result, we do not conduct a bounding exercise of imputing best and worst-case outcomes to the students who attrited from the sample.

Table S5. Count of students randomized by summer cohort. The treatment group is depicted in green (invitation to attend SB Program in person), while the control group is depicted in gray. The labels show the summer of attendance and the school year for which we measured first-year academic outcomes.

Summer cohort	Invitation	N students
Summer: 2017 SY: 17-18	Residential	78
Summer: 2017 SY: 17-18	Online or none	63
Summer: 2018 SY: 18-19	Residential	75
Summer: 2018 SY: 18-19	Online or none	53
Summer: 2019 SY: 19-20 (spring during COVID-19)	Residential	79
Summer: 2019 SY: 19-20 (spring during COVID-19)	Online or none	70

Table S6, focused on the full analytic sample rather than the cohort-specific breakdowns we show in the pre-analysis plan, shows balance (measured as the standardized mean differences) along observed characteristics between students randomized to the invitation group and students randomized to the control group.

Table S6. Attributes of students in analytic sample: the table depicts the mean and standard error around the mean of each attribute in the treatment group (invitation to residential form) and control group (invitation to online or none). The third column shows the difference in means along with a 95 percent confidence interval from a t-test of difference in means. We see that all confidence intervals cross zero, indicating a lack of estimated difference in attributes between the treatment and control groups. The pre-analysis plan shows the cohort-by-cohort standardized differences in means.

Attribute	Mean: residential	Mean: online or none	Difference
Did not attend pipeline	0.593 [0.56, 0.625]	0.518 [0.48, 0.555]	0.0753 [-0.0206, 0.171]
Family death	0.25 [0.215, 0.286]	0.245 [0.205, 0.286]	0.00528 [-0.0959, 0.106]
Family hardship	0.409 [0.376, 0.441]	0.402 [0.366, 0.439]	0.00619 [-0.088, 0.1]
Food insecurity	0.121 [0.099, 0.142]	0.126 [0.101, 0.151]	-0.00558 [-0.0703, 0.0591]
Highest test score (SAT scale)	1446.1 [1440.6, 1451.5]	1447 [1440.3, 1453.7]	-0.9741 [-17.488, 15.54]
Housing insecurity	0.105 [0.0846, 0.125]	0.101 [0.0789, 0.124]	0.00355 [-0.0559, 0.063]
Neither parent attended college	0.702 [0.672, 0.732]	0.621 [0.585, 0.657]	0.0809 [-0.0117, 0.174]

D. Secondary research questions. We pre-registered the following secondary outcomes and research questions with short summaries of our findings:

- Does *SB Program* lead to better performance in (1) STEM classes (classes in mathematics, chemistry, and physics) or (2) the required writing course during the student's first year?
    - In addition to the null effects on overall GPA, we find null effects on these specific GPAs. The full results are reported in Section G.3.
- Are the main effects moderated by pre-treatment family income?
    - Although we pre-registered this question, we were ultimately unable to obtain reliable data on the household income of student's families that was reliably and consistently collected across cohorts.
- Are the main effects moderated by pre-treatment pipeline program participation?
    - SI Section I presents these results, which find no heterogeneity.
- Are the main effects moderated by pre-treatment exposure to parents with different educational attainment/first-generation status?[‡‡]
    - SI Section I presents these results, which find no heterogeneity.
- Are the main academic outcomes mediated by student's help-seeking behavior from formal academic support resources?[§§]
    - SI Section J describes this analysis, where we find no main effects of the treatment on the mediator.

[‡‡] Originally, we pre-specified examining this controlling for family income; see above note on unreliable data on this measure.

[§§] We pre-registered using two measures: the average visits per semester and the cumulative number of visits across all enrolled semesters.

E. Additional details on estimation from pre-analysis plan. Here, we provide three forms of additional details on estimation: (1) for some of the models, reweighting to account for differential probabilities of randomization to an in-person invitation (Section E.1); (2) our estimation approaches for the ITT analysis (Section E.2); and (3) our estimation approaches for the analysis of impact on compliers (Section E.3). All these analytic decisions were pre-registered and are also found in our pre-analysis plan.

E.1. Adjusting for unequal probability of assignment across priority tiers. Students had unequal probabilities of receiving an invitation depending on which tier the administrators put them in and their randomization block, where relevant. As a result, we reweighted the sample using the following procedure:

1. For treatment group students, we reweight by the inverse probability of treatment assignment for each student

2. For control group students, we reweight by the inverse of one minus this treatment assignment probability

E.2. Additional details: estimation approach for ITT analysis. In the main text, we report results from the main specification: a model that adjusts for each student's randomization `block`, which is a combination of their administrator-designated priority tier and then any further strata within those tiers. Here, we also report results for three other specifications that we pre-registered:

1. **Specification with fixed effects for priority tier (shorthand: tiers):** the first alternate specification has a fixed effect for the student's priority tier, or the following, and reweights by the inverse probability of treatment weight:

$$Y_i = \alpha_{\text{tier}[i]} + \beta \text{Invite}_i + \epsilon_i \qquad [1]$$

2. **Specification with fixed effects for priority tier and control for ordered logit linear predictor (shorthand: tier and linear predictor):** the second alternative specification specification, similar to Eq. (1), contains a fixed effect for the student's priority tier. The specification also contains the student's linear predictor from the ordered logistic regression predicting tier Z_i and reweights by the inverse probability of treatment weight:

$$Y_i = \alpha_{\text{tier}[i]} + \beta \text{Invite}_i + \gamma Z_i + \epsilon_i \qquad [2]$$

3. **Specification with fixed effects for block and control for ordered logit linear predictor (shorthand: block and linear predictor):** the third alternate specification contains a fixed effect for the student's pair or block. It also contains the student's ordered logit linear predictor Z_i. It does not reweight by the inverse probability of treatment weight:

$$Y_i = \delta_{\text{block}[i]} + \beta \text{Invite}_i + \gamma Z_i + \epsilon_i \qquad [3]$$

As (33) discuss, an estimator that solely adjusts for fixed effects fails to consistently estimate the average treatment effects (in this case, the impact of an invitation to *SB Program* on student outcomes) when there is either heterogeneity in the treatment effects (e.g., the program had a stronger impact on some priority tiers than on others) or when there is variation in the proportion of treated observations across the strata. Because these conditions hold in our study, the main specification—the specification discussed in the main text that adjusts for the more granular blocks—as well as the secondary specifications outlined in Equation 1, Equation 2, and Equation 3 adjust for the inverse probability of treatment weights (see Theorem 1 of (33) for a formal justification).

E.3. Additional details: estimation approach for complier average causal effect analysis. With the ITT, we estimate the effect of *SB Program* on students who received randomized invitations to participate. However, participation was *not* required, and there were two forms of non-compliance. First, some of the treatment group students who were invited to the residential version of the program declined and attended the online version or no version. Second, some of the control group students are invited to fill spots left open due to treatment group students declining. In the main analysis, we treat control group students randomized to the waitlist as members of the control group. Here, non-compliance occurs when students on the waitlist attend *SB Program*.

We pre-registered using the following two-stage least squares to estimate the complier average causal effect that accounts for this noncompliance (see Section F for a further discussion about noncompliance).

$$\begin{aligned} Y_i &= \alpha_{block[i]} + \beta \hat{Attend}_i + \gamma Z_i + \epsilon_i & [4] \\ \text{Attend}_i &= \alpha_{block[i]} + \beta \text{Invite}_i + \gamma Z_i + \epsilon_i & [5] \end{aligned}$$

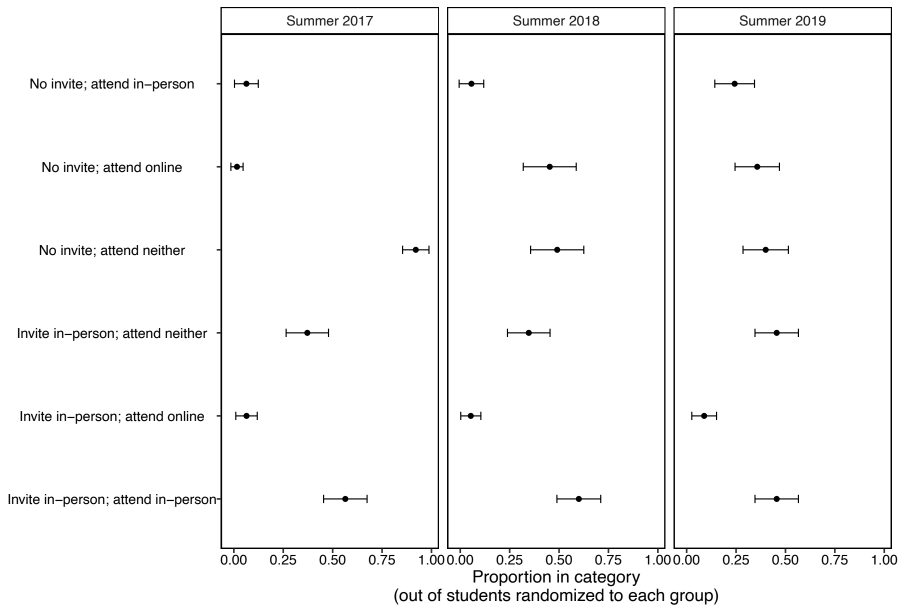
F. Definition of compliers. There are six possible combinations of the randomized treatment assignment (the basis for the ITT estimates) and treatment receipt, as shown in Table S7. Formally, let $Z_i \in \{0, 1\}$ represent a two-level treatment *assignment* where $Z_i = 1$ represents invited to in-person *SB Program* and $Z_i = 0$ represents no invitation to in-person *SB Program*. Next, we use $D_i \in \{0, 1, 2\}$ to denote a three-level *actual* treatment where $D_i = 2$ represents attends in-person *SB Program*, $D_i = 1$ represents engaging with the online content, and $D_i = 0$ represents engaging with neither.

There are three causal contrasts between *actual* treatments (D_i): (1) In-person *SB Program* versus online*SB Program*, (2) In-person *SB Program* versus none, and (3) Online *SB Program* versus none. Since we only have a two-level *randomized*

Table S7. Treatment assignment versus treatment receipt

Treatment assignment	Treatment receipt
SB in person	SB in person
SB in person	SB online
SB in person	None
Not SB in person	SB in person
Not SB in person	SB online
Not SB in person	None

Fig. S3. Invitations versus summer enrollment: count in each group as a proportion of the *each of the two randomized invitation statuses*, separated by summer cohort The figure shows point estimates and 95% confidence intervals around the sample proportions.



treatment (Z_i), we cannot identify these three causal contrasts. Instead, we identify the one causal contrast of *In-person SB Program* versus either *SB online* or none.

In order to identify this effect, we assume monotonicity: the randomized treatment of an invitation to *SB Program* never discourages students to attend *SB Program* (either in person or online), i.e., $D_i(1) \geq D_i(0)$. The data are consistent with these assumptions. First, for the empirical estimate of the probability difference $Pr(D \geq 2|Z = 1)$ - $Pr(D \geq 2|Z = 0)$ (in-person attendance rates comparing those randomized to treatment versus control), we see a positive value: $0.541 - 0.13 = 0.41$. Second, for the empirical estimate of the probability difference $Pr(D \geq 1|Z = 1)$ - $Pr(D \geq 1|Z = 0)$ (in-person or online attendance rates comparing those randomized to treatment and control), we see a positive value: $0.61 - 0.402 = 0.208$.

Figure S3 shows the different subgroups of treatment assignment and treatment receipt. We see that across all three summers, those invited to the residential program were most likely to either attend the residential program or attend nothing. Those *not* invited to the residential program were most likely to attend nothing in the first summer (summer 2017), but then shifted to being more likely to attend online as the online program expanded. This means that our causal contrasts are a combination of comparing the residential version to nothing and comparing the residential version to the online programming.

G. Full regression tables/additional results: main outcomes and secondary outcomes. In the main text, Table 3 shows the causal effects on measures of program difficulty, based on the primary specification, which adjusts for more granular randomization `blocks` within each tier. Similarly, Table 4 shows the results for GPA and academic withdrawal from the same specification. In this section, we present (1) the full regression tables for the main specification and the three others discussed in SI Section E.2; and (2) results for the secondary outcomes of STEM and writing GPAs, which similar to the overall GPA, show no statistically significant effects.

G.1. Full regression tables across specifications: program difficulty measures.

- Table S8 shows the full regression results for the impact on program difficulty for the main specification (adjusting for the more granular blocks), the regression that corresponds to the estimates, and the confidence intervals for the ITT presented in the main text Table 3.

- Table S9 shows the results for the specification that controls for broader priority tiers rather than more granular blocks, preserving more degrees of freedom. We see that the coefficients become somewhat smaller in magnitude. We continue to estimate a statistically significant effect on courses taken for a grade ($p = 0.002$); percent of non-introductory courses becomes marginally non-significant ($p = 0.065$).

- Table S8 presents the specification that controls for both the more granular randomization blocks and the linear predictor from the ordered logit regression of priority tier. Results are the same magnitude and significance as the main specification.
- Table S11 presents the specification that controls for the broader priority tiers rather than the more granular blocks, and also controls for the ordered logit linear predictor. Results are similar in magnitude and significance to the main specification.

Overall, the results show the robustness of the significant impact of *SB Program* on measures of first-year program difficulty across all pre-registered specifications.

Table S8. ITT estimates (main specification: blocks): effect of SB program invitation on program difficulty. The table omits the block-specific fixed effects

	Dependent variable:			
	Perc. courses > 100-level	Perc. courses graded	Total units (excludes SB)	Total units (includes SB)
Invited	0.037	0.034	−0.138	0.642
	(0.018)	(0.011)	(0.088)	(0.126)
	p = 0.040*	p = 0.002**	p = 0.117	p = 0.00000***
Constant	0.294	0.889	7.569	7.679
	(0.101)	(0.061)	(0.497)	(0.713)
	p = 0.004**	p = 0.000***	p = 0.000***	p = 0.000***
Observations	418	418	418	418
R^2	0.211	0.482	0.170	0.180
Adjusted R^2	0.049	0.376	−0.0002	0.012
Residual Std. Error (df = 346)	0.246	0.149	1.213	1.740
F Statistic (df = 71; 346)	1.304	4.543***	0.999	1.072
Note:				*p<0.05; **p<0.01; ***p<0.001

Table S9. ITT estimates (secondary specification: tiers): effect of SB program invitation on program difficulty. The table omits the tier-specific fixed effects

	Dependent variable:			
	Perc. courses > 100-level	Perc. courses graded	Total units (excludes SB)	Total units (includes SB)
Invited	0.032	0.030	−0.151	0.655
	(0.017)	(0.010)	(0.083)	(0.117)
	p = 0.065	p = 0.003**	p = 0.071	p = 0.00000***
Constant	0.377	0.752	8.374	8.452
	(0.024)	(0.014)	(0.117)	(0.163)
	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.000***
Observations	418	418	418	418
R^2	0.037	0.438	0.048	0.103
Adjusted R^2	0.009	0.421	0.020	0.077
Residual Std. Error (df = 405)	0.251	0.144	1.201	1.682
F Statistic (df = 12; 405)	1.303	26.266***	1.708	3.890***
Note:				*p<0.05; **p<0.01; ***p<0.001

Table S10. ITT estimates (secondary specification: blocks + linear predictor): effect of SB program invitation on program difficulty. The table omits the block-specific fixed effects and linear predictor coefficient.

	Dependent variable:			
	Perc. courses > 100-level	Perc. courses graded	Total units (excludes SB)	Total units (includes SB)
Invited	0.036	0.034	-0.141	0.643
	(0.018)	(0.011)	(0.088)	(0.126)
	p = 0.043*	p = 0.002**	p = 0.110	p = 0.00000***
Constant	0.322	0.897	7.732	7.652
	(0.117)	(0.071)	(0.575)	(0.825)
	p = 0.007**	p = 0.000***	p = 0.000***	p = 0.000***
Observations	418	418	418	418
R^2	0.212	0.483	0.171	0.180
Adjusted R^2	0.047	0.375	-0.002	0.009
Residual Std. Error (df = 345)	0.246	0.150	1.215	1.742
F Statistic (df = 72; 345)	1.287	4.469***	0.987	1.054
Note:				*p$<$0.05; **p$<$0.01; ***p$<$0.001

Table S11. ITT estimates (secondary specification: tiers + linear predictor): effect of SB program invitation on program difficulty. The table omits the tier-specific fixed effects and linear predictor coefficient.

	Dependent variable:			
	Perc. courses > 100-level	Perc. courses graded	Total units (excludes SB)	Total units (includes SB)
Invited	0.032	0.030	-0.153	0.657
	(0.017)	(0.010)	(0.084)	(0.117)
	p = 0.071	p = 0.003**	p = 0.069	p = 0.00000***
Constant	0.381	0.753	8.382	8.440
	(0.025)	(0.014)	(0.120)	(0.168)
	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.000***
Observations	418	418	418	418
R^2	0.038	0.438	0.048	0.104
Adjusted R^2	0.007	0.420	0.018	0.075
Residual Std. Error (df = 404)	0.251	0.144	1.202	1.684
F Statistic (df = 13; 404)	1.228	24.188***	1.580	3.590***
Note:				*p$<$0.05; **p$<$0.01; ***p$<$0.001

G.2. Full regression tables across specifications: overall GPA, withdrawal, and too-few credits.

- Table S12 shows the full regression results for the impact on GPA for the main specification (adjusting for the more granular blocks), the regression that corresponds to the estimates and confidence intervals for the ITT presented in the main text Table 4.
- Table S13 shows the results for the specification that controls for broader priority tiers rather than more granular blocks. Results are similar to the main specification in that we fail to find evidence that, amidst the higher program difficulty, there was an impact on GPA or rates of academic withdrawal.
- Table S14 presents the specification that controls for both the more granular randomization blocks and the linear predictor from the ordered logit regression of priority tier. Results are similar in magnitude and significance as the main specification.
- Table S15 presents the specification that controls for the broader priority tiers rather than the more granular blocks, and also controls for the ordered logit linear predictor. Results are similar in magnitude and significance to the main specification.

Table S12. ITT estimates (main specification: blocks): effect of SB program invitation on first-year GPA, rates of withdrawal, and rates of too-few crdits. The table omits the block-specific fixed effects

	Dependent variable:			
	First-year GPA (includes SB)	First-year GPA (excludes SB)	Withdrawal	Too-few credits
Invited	0.016	−0.016	−0.014	−0.042
	(0.042)	(0.043)	(0.026)	(0.031)
	p = 0.707	p = 0.714	p = 0.598	p = 0.175
Constant	2.630	2.580	1.007	0.021
	(0.236)	(0.245)	(0.148)	(0.174)
	p = 0.000***	p = 0.000***	p = 0.000***	p = 0.905
Observations	418	418	418	418
R^2	0.226	0.240	0.263	0.181
Adjusted R^2	0.068	0.084	0.112	0.014
Residual Std. Error (df = 346)	0.576	0.597	0.360	0.425
F Statistic (df = 71; 346)	1.427*	1.540**	1.741***	1.080
Note:				*p<0.05; **p<0.01; ***p<0.001

Table S13. ITT estimates (secondary specification: tiers): effect of SB program invitation on first-year GPA, rates of withdrawal, and rates of too-few credits. The table omits the tier-specific fixed effects

	Dependent variable:			
	First-year GPA (includes SB)	First-year GPA (excludes SB)	Withdrawal	Too-few credits
Invited	0.003	−0.030	−0.008	−0.029
	(0.040)	(0.042)	(0.026)	(0.029)
	p = 0.948	p = 0.470	p = 0.763	p = 0.328
Constant	3.481	3.491	0.071	0.081
	(0.056)	(0.059)	(0.036)	(0.041)
	p = 0.000***	p = 0.000***	p = 0.052	p = 0.051
Observations	418	418	418	418
R^2	0.085	0.092	0.065	0.044
Adjusted R^2	0.058	0.065	0.038	0.016
Residual Std. Error (df = 405)	0.579	0.604	0.375	0.425
F Statistic (df = 12; 405)	3.153***	3.408***	2.355**	1.569
Note:				*p<0.05; **p<0.01; ***p<0.001

Table S14. ITT estimates (secondary specification: blocks + linear predictor): effect of SB program invitation on first-year GPA, rates of withdrawal, and rates of too-few credits. The table omits the block-specific fixed effects and linear predictor coefficient.

	Dependent variable:			
	First-year GPA (includes SB)	First-year GPA (excludes SB)	Withdrawal	Too-few credits
Invited	0.019	−0.012	−0.011	−0.040
	(0.042)	(0.043)	(0.026)	(0.031)
	p = 0.650	p = 0.777	p = 0.667	p = 0.192
Constant	2.465	2.394	0.879	−0.055
	(0.273)	(0.283)	(0.170)	(0.202)
	p = 0.000***	p = 0.000***	p = 0.00000***	p = 0.787
Observations	418	418	418	418
R^2	0.230	0.244	0.268	0.183
Adjusted R^2	0.069	0.086	0.115	0.012
Residual Std. Error (df = 345)	0.575	0.597	0.359	0.426
F Statistic (df = 72; 345)	1.429*	1.545**	1.754***	1.072
Note:				*p<0.05; **p<0.01; ***p<0.001

Table S15. ITT estimates (secondary specification: tiers + linear predictor): effect of SB program invitation on first-year GPA, rates of withdrawal, and rates of too-few credits. The table omits the tier-specific fixed effects and linear predictor coefficient.

	Dependent variable:			
	First-year GPA (includes SB)	First-year GPA (excludes SB)	Withdrawal	Too-few credits
Invited	0.005	−0.027	−0.006	−0.040
	(0.040)	(0.042)	(0.026)	(0.031)
	p = 0.893	p = 0.517	p = 0.821	p = 0.192
Constant	3.465	3.473	0.060	−0.055
	(0.058)	(0.060)	(0.037)	(0.202)
	p = 0.000***	p = 0.000***	p = 0.111	p = 0.787
Observations	418	418	418	418
R^2	0.089	0.095	0.069	0.183
Adjusted R^2	0.059	0.066	0.039	0.012
Residual Std. Error	0.578 (df = 404)	0.603 (df = 404)	0.375 (df = 404)	0.426 (df = 345)
F Statistic	3.028*** (df = 13; 404)	3.280*** (df = 13; 404)	2.309** (df = 13; 404)	1.072 (df = 72; 345)
Note:				*p<0.05; **p<0.01; ***p<0.001

G.3. Full regression tables across specifications: STEM and writing GPA. We present the results for two types of subject-specific GPAs that were pre-registered as secondary outcomes: GPA in STEM coursework (either including grades earned during *SB Program* or not) and GPA in a required expository writing course. Note that the sample sizes for these outcomes are lower for two reasons. First, for the STEM-specific GPAs, we are filtering on students who took one or more STEM classes for a grade since GPA is undefined for those with zero enrollment in a graded STEM course. Second, for the writing GPAs, the last summer cohort had a semester of pass-fail only grading during the COVID-19 pandemic; we filter the sample to those with a graded writing course in which the GPA is defined.

- Table S16 shows the full regression results for the impact on subject-specific GPAs for the main specification (adjusting for the more granular blocks). We fail to reject the null of no mean difference in GPAs.
- Table S17 shows the results for the specification that controls for broader priority tiers rather than more granular blocks. Results are similar to the main specification in that we fail to find evidence of a mean difference in these subject-specific GPAs.
- Table S18 presents the specification that controls for both the more granular randomization blocks and the linear predictor from the ordered logit regression of priority tier. Results are similar in magnitude and significance as the main specification.
- Table S19 presents the specification that controls for the broader priority tiers rather than the more granular blocks, and also controls for the ordered logit linear predictor. Results are similar in magnitude and significance to the main specification.

Across specifications, we fail to reject the null of no mean difference in students' GPAs in STEM coursework or a required writing seminar. This means that the lack of statistically significant estimated difference in overall first-year GPAs extends to breakdowns of this overall GPA into particular subject areas of interest.

Table S16. ITT estimates (main specification: blocks): effect of SB program invitation on STEM and writing GPAs. The table omits the block-specific fixed effects

	Dependent variable:		
	STEM GPA (includes SB)	STEM GPA (excludes SB)	Writing course GPA
Invited	0.059	0.060	−0.061
	(0.065)	(0.068)	(0.048)
	p = 0.370	p = 0.381	p = 0.205
Constant	2.396	2.395	2.755
	(0.330)	(0.342)	(0.249)
	p = 0.000***	p = 0.000***	p = 0.000***
Observations	352	349	363
R^2	0.238	0.228	0.257
Adjusted R^2	0.044	0.031	0.075
Residual Std. Error	0.804 (df = 280)	0.833 (df = 277)	0.608 (df = 291)
F Statistic	1.229 (df = 71; 280)	1.155 (df = 71; 277)	1.415* (df = 71; 291)
Note:			*p<0.05; **p<0.01; ***p<0.001

Table S17. ITT estimates (secondary specification: tiers): effect of SB program invitation on STEM and writing GPAs. The table omits the tier-specific fixed effects

	Dependent variable:		
	STEM GPA (includes SB)	STEM GPA (excludes SB)	Writing course GPA
Invited	0.027	0.028	−0.050
	(0.061)	(0.063)	(0.047)
	p = 0.662	p = 0.663	p = 0.285
Constant	3.374	3.374	3.377
	(0.083)	(0.085)	(0.081)
	p = 0.000***	p = 0.000***	p = 0.000***
Observations	352	349	363
R^2	0.086	0.086	0.056
Adjusted R^2	0.053	0.053	0.024
Residual Std. Error	0.800 (df = 339)	0.824 (df = 336)	0.625 (df = 350)
F Statistic	2.647** (df = 12; 339)	2.636** (df = 12; 336)	1.738 (df = 12; 350)
Note:			*p<0.05; **p<0.01; ***p<0.001

Table S18. ITT estimates (secondary specification: blocks + linear predictor): effect of SB program invitation on STEM and writing GPAs. The table omits the block-specific fixed effects and linear predictor coefficient.

	Dependent variable:		
	STEM GPA (includes SB)	STEM GPA (excludes SB)	Writing course GPA
Invited	0.064	0.067	−0.056
	(0.065)	(0.068)	(0.048)
	p = 0.330	p = 0.331	p = 0.243
Constant	2.034	2.005	2.501
	(0.388)	(0.402)	(0.301)
	p = 0.00000***	p = 0.00001***	p = 0.000***
Observations	352	349	363
R^2	0.246	0.238	0.262
Adjusted R^2	0.051	0.039	0.079
Residual Std. Error	0.801 (df = 279)	0.830 (df = 276)	0.607 (df = 290)
F Statistic	1.264 (df = 72; 279)	1.195 (df = 72; 276)	1.432* (df = 72; 290)
Note:			*p<0.05; **p<0.01; ***p<0.001

Table S19. ITT estimates (secondary specification: tiers + linear predictor): effect of SB program invitation on STEM and writing GPAs. The table omits the tier-specific fixed effects and linear predictor coefficient.

	Dependent variable:		
	STEM GPA (includes SB)	STEM GPA (excludes SB)	Writing course GPA
Invited	0.031	0.033	−0.047
	(0.061)	(0.063)	(0.047)
	p = 0.607	p = 0.597	p = 0.321
Constant	3.344	3.340	3.362
	(0.086)	(0.088)	(0.082)
	p = 0.000***	p = 0.000***	p = 0.000***
Observations	352	349	363
R^2	0.090	0.091	0.061
Adjusted R^2	0.055	0.056	0.026
Residual Std. Error	0.799 (df = 338)	0.822 (df = 335)	0.624 (df = 349)
F Statistic	2.582** (df = 13; 338)	2.595** (df = 13; 335)	1.744 (df = 13; 349)
Note:			*p<0.05; **p<0.01; ***p<0.001

H. GPA adjusted for difficulty analysis. Based on reviewer feedback that it is important to adjust student GPAs for the difficulty of the student's course of program, we have conducted a non pre-registered analysis of the impact of an *SB Program* invitation on student GPA adjusting for course difficulty. Due to student privacy concerns, we are not able to obtain either (1) the mean GPA for each student's specific course enrollments during their first year or (2) the mean GPA for specific categories of courses (e.g., 100-level courses versus 200-level courses versus 300-level courses versus 400-level courses). Instead, we were provided less granular information that (1) focuses on the spring semester only, (2) aggregates across students from all four class years rather than subsets to Freshman, and (3) reports the following broad categories of GPA:

1. Mean GPA for spring 100 or 200-level courses across all *SB University* students
2. Mean GPA for spring 300 or 400-level courses across all *SB University* students

We use these measures to construct a "under or overperforming expected GPA" outcome variable. The construction of this variable involved the following steps:

- For the spring term within the student's first year (the term of the mean GPAs), we calculated the proportion of their total credits allocated to: (1) 100 or 200-level courses and (2) 300 or 400-level courses;
- We then constructed a "difficulty-expected GPA" based on the student-specific course composition and the mean GPA for that bucket of courses
- We then compared the student's observed GPAs to their difficulty-expected GPA; positive values indicate "outperformance" of the expected GPA; negative values indicate "underperformance"

We then estimate the same four models as we estimate for the other outcome variables. Table S20 shows the results.[¶¶] Similar to the main GPA results, across all four specifications, we fail to reject the null hypothesis of no difference in GPAs. This could be as much due to data limitations—e.g., heterogeneity across class years means that a 400-level course for a freshman is different than a 400-level course for a senior—as a true lack of difference, so we emphasize the need for more granular GPA adjustment in future research.

Table S20. ITT estimates (all specifications): effect of GPA adjusted for measures of course difficulty. Omits block and tier fixed effects

	Dependent variable:			
	Block FE	Block FE+linear predictor	Tier FE	Tier FE+linear predictor
Invited	−0.024	−0.015	−0.022	−0.013
	(0.051)	(0.053)	(0.051)	(0.053)
	p = 0.645	p = 0.781	p = 0.664	p = 0.801
Constant	0.096	−0.900	0.087	−0.984
	(0.075)	(0.294)	(0.077)	(0.341)
	p = 0.200	p = 0.003**	p = 0.259	p = 0.005**
Observations	401	401	401	401
R^2	0.095	0.245	0.095	0.246
Adjusted R^2	0.067	0.082	0.065	0.080
Residual Std. Error	0.723 (df = 388)	0.717 (df = 329)	0.724 (df = 387)	0.718 (df = 328)
F Statistic	3.381*** (df = 12; 388)	1.506** (df = 71; 329)	3.136*** (df = 13; 387)	1.485* (df = 72; 328)
Note:				*p<0.05; **p<0.01; ***p<0.001

I. Heterogeneous treatment effects analysis. We present the full regression results for the analysis of heterogeneous treatment effects across outcomes and the three moderators discussed in the main text. Across outcomes, we see no consistent heterogeneous effects of the treatment across three pre-treatment moderators: the student's standardized test scores (Table S21), whether a parent had some form of college enrollment (Table S22), or the student's participation in a preparatory pipeline program (Table S23). Overall, we fail to find evidence that the effects, where positive, are concentrated in certain groups of students. Similarly, the lack of estimated impact on GPA and withdrawal does not conceal large impacts in some groups and no impacts in others.

[¶¶] The sample size drops from $N = 418$ in the main analysis to $N = 401$ in the present analysis due to students who only took fall-term courses in their first year either due to withdrawal or a zero credit spring semester.

Table S21. Heterogeneous treatment effects: pre-treatment standardized test scores (recoded to SAT and taking highest score). The table omits the block-specific fixed effects

	Dependent variable:					
	First-year GPA (includes SB)	Withdrawal	Perc. courses > 100-level	Total units (no SB)	Total units (with SB)	Perc. courses graded
	(1)	(2)	(3)	(4)	(5)	(6)
Invited	0.200	0.708	0.011	−3.994	−2.839	−0.205
	(0.733)	(0.474)	(0.328)	(2.281)	(1.598)	(0.199)
	p = 0.785	p = 0.136	p = 0.975	p = 0.081	p = 0.077	p = 0.305
Test score	0.002	0.0003	0.00003	−0.0001	−0.001	−0.0001
	(0.0004)	(0.0003)	(0.0002)	(0.001)	(0.001)	(0.0001)
	p = 0.00004***	p = 0.179	p = 0.859	p = 0.937	p = 0.549	p = 0.500
Invited x Test score	−0.0001	−0.0005	0.00002	0.003	0.002	0.0002
	(0.001)	(0.0003)	(0.0002)	(0.002)	(0.001)	(0.0001)
	p = 0.802	p = 0.129	p = 0.932	p = 0.044*	p = 0.093	p = 0.231
Constant	0.402	0.522	0.140	7.821	8.195	0.998
	(0.596)	(0.385)	(0.266)	(1.855)	(1.299)	(0.162)
	p = 0.500	p = 0.177	p = 0.599	p = 0.00004***	p = 0.000***	p = 0.000***
Observations	415	415	415	415	415	415
R^2	0.294	0.251	0.208	0.195	0.194	0.474
Adjusted R^2	0.143	0.091	0.039	0.023	0.022	0.362
Note:						*p<0.05; **p<0.01; ***p<0.001

Table S22. Heterogeneous treatment effects: pre-treatment parent educational attainment (some college = one parent attended some college). The table omits the block-specific fixed effects

	Dependent variable:					
	First-year GPA (includes SB)	Withdrawal	Perc. courses > 100-level	Total units (no SB)	Total units (with SB)	Perc. courses graded
	(1)	(2)	(3)	(4)	(5)	(6)
Invited	0.027	−0.003	0.037	0.538	−0.231	0.041
	(0.053)	(0.033)	(0.023)	(0.161)	(0.112)	(0.014)
	p = 0.609	p = 0.935	p = 0.105	p = 0.001***	p = 0.040*	p = 0.004**
Some college	0.031	0.005	0.028	−0.189	−0.172	−0.001
	(0.071)	(0.044)	(0.030)	(0.214)	(0.148)	(0.018)
	p = 0.668	p = 0.902	p = 0.363	p = 0.379	p = 0.247	p = 0.938
Invited x Some college	−0.019	−0.030	0.011	0.342	0.282	−0.023
	(0.095)	(0.059)	(0.041)	(0.287)	(0.199)	(0.025)
	p = 0.843	p = 0.612	p = 0.788	p = 0.235	p = 0.158	p = 0.360
Constant	2.675	1.002	0.184	7.642	7.487	0.889
	(0.240)	(0.148)	(0.102)	(0.724)	(0.502)	(0.062)
	p = 0.000***	p = 0.000***	p = 0.074	p = 0.000***	p = 0.000***	p = 0.000***
Observations	418	418	418	418	418	418
R^2	0.227	0.247	0.213	0.183	0.190	0.475
Adjusted R^2	0.062	0.087	0.046	0.010	0.019	0.364
Note:						*p<0.05; **p<0.01; ***p<0.001

Table S23. Heterogeneous treatment effects: pre-treatment pipeline program participation. The table omits the block-specific fixed effects

	Dependent variable:					
	First-year GPA (includes SB)	Withdrawal	Perc. courses > 100-level	Total units (no SB)	Total units (with SB)	Perc. courses graded
	(1)	(2)	(3)	(4)	(5)	(6)
Invited	0.021	-0.002	0.044	0.573	-0.098	0.009
	(0.059)	(0.037)	(0.025)	(0.176)	(0.124)	(0.015)
	p = 0.726	p = 0.954	p = 0.081	p = 0.002**	p = 0.432	p = 0.548
Yes pipeline	0.094	-0.007	-0.010	0.235	0.106	-0.038
	(0.066)	(0.041)	(0.028)	(0.199)	(0.139)	(0.017)
	p = 0.158	p = 0.863	p = 0.725	p = 0.238	p = 0.448	p = 0.028*
Invited x Yes pipeline	0.021	-0.026	-0.019	0.262	-0.064	0.053
	(0.089)	(0.055)	(0.038)	(0.267)	(0.187)	(0.023)
	p = 0.812	p = 0.635	p = 0.626	p = 0.327	p = 0.733	p = 0.022*
Constant	2.603	1.024	0.198	7.287	7.370	0.900
	(0.240)	(0.149)	(0.103)	(0.721)	(0.506)	(0.062)
	p = 0.000***	p = 0.000***	p = 0.056	p = 0.000***	p = 0.000***	p = 0.000***
Observations	418	418	418	418	418	418
R^2	0.238	0.248	0.210	0.200	0.187	0.481
Adjusted R^2	0.077	0.088	0.043	0.030	0.015	0.371
Note:						*p<0.05; **p<0.01; ***p<0.001

J. Causal effect on visits to academic help centers. The main text discusses our pre-registered mediator, where *SB Program*'s impact on the difficulty of students' first-year coursework might have operated through students taking more advantage of academic help centeres. While the main text reports each group's mean rates of visits, Figure S4 shows the broader distribution of visit counts, contrasting the treatment and control groups. We see slightly higher rates of visiting at least once among the treatment group students, but potentially few systematic differences.

To estimate whether the student's degree of help seeking mediates the main effects we see for certain outcomes, we pre-registered relying on two models: a model that predicts the mediator and a model that predicts the outcome. First, in the mediator model, we estimated the effect of the treatment on help-seeking using the following specification, where M_i is the cumulative number of visits across semesters.[***] We report the results from the specification that controls for the more granular blocks:

$$M_i = \alpha_{\text{block}[i]} + \beta_1 \text{Invite}_i + \epsilon_i \qquad [6]$$

Since the treatment was randomized, we expect that the student's level of help seeking is randomly assigned conditional on the treatment status. These results showed no significant impact of the treatment on the mediator. In particular, we fail to reject the null of no average effect on the total count of visits ($\beta = 0.07$ [-0.41, 0.56], $p = 0.76$, control group mean $= 0.77$).[†††] This shows that our pre-registered mediator of measured academic help center visits does not explain the program's positive impacts on measures of first-year difficulty. Since the mediator model showed no impact of the treatment on the mediator, we do not estimate an outcomes model to conduct a full mediation analysis.

[***] We pre-registered this formulation of the mediator as well as a second formulation of examining the average number of visits per semester. Because the modal number of visits is zero total across both semesters, we focus on the first formulation.

[†††] We also find no estimated effects of the treatment on the binary measures of any visits to the help center.

Fig. S4. Proportion of students in each group with different numbers of first-year visits to SB University's academic help center