Supplementary Information for

When and Why Working Together Benefits First-Generation College Students:

A Registered Report

Pilot Study: Additional Methods

Participants

Participants were 209 students enrolled in a large undergraduate course at a West Coast public university in the United States. Of the 209 students, 177 took an extra credit survey (in which they reported demographics) at the end of the academic term. Thus, we only have self-reported demographic information for 177 students. However, we obtained objective first-generation college student status for 194 students from the university registrar. We used self-reported FG college student status from the self-report survey for 13 students who did not have registrar data. Thus, we have complete data for all of our primary variables of interest (reported in the main manuscript) for 207 students. The two students who did not have data for FG student status were excluded from analysis.

We asked participants about their gender identity with the item "My gender is…" and the answer options were "Female", "Male", and "Other". From the 177 students who took the extra credit survey, 35 identified as male, 141 identified as female, and 1 person identified as "other." Students reported being between the ages of 18 to 47 (M = 21.01, SD = 3.03). 154 students self-reported having entirely or mostly been raised in the United States. Participants were asked about their racial or ethnic group they belong to with the item: "Please check all racial or ethnic groups that you belong" and the answer options were "African-American, Black, African, or Caribbean", "Asian-American, Asian, or Pacific Islander", "European-American, White, Anglo, or Caucasian", "Hispanic-American, Latino(a), or Chicano(a)", "Native-American or American Indian", and "Other (please specify):___". 67 students identified as East Asian/Asian American, 5 as African American, 61 as Latino/a, 21 as Non-Hispanic White, 4 as other and 19 as having more than one ethnicity. Taken together, 44% of the sample was classified as coming from an underrepresented minority background (URM). For analyses including race/ethnicity and other

control variables, all students with missing data on the demographic questionnaire (because they had missing data on one or more variables included in the multiple regression model) were excluded from analysis. For the moderator analysis reported in the main text (i.e., meeting mode), we calculated the average meeting mode response for each team with the data from the students who took the survey.

Team Composition

All 209 students were randomly assigned through the learning management system Canvas to teams. After the drop/add course deadline, the class consisted of 43 teams: 37 teams of 5 students and 6 teams of 4 students. After excluding the students who did not have data on the FG student status variable, the final composition was 36 teams of 5 students, 6 teams of 4 students, and 1 team of 3 students. This is the final team composition used in the analyses reported in the main text. Of these teams, 4 teams had 1 FG college student, 12 teams had 2 FG college students, 16 teams had 3 FG college students, 9 teams had 4 FG college students, and 2 teams had 5 FG college students. In percentages, no teams were 0% FG college students, 2 teams were 100% FG college students, and 16 teams were 60% FG college student composition (which was the modal social class composition). In terms of the teams' gender composition, 3 teams had 1 woman, 7 teams had 2 women, 15 teams had 3 women, 11 teams had 4 women, and 7 teams had 5 women. The teams' racial/ethnic compositions were as follows: 4 teams had no URM students, 14 teams had 1 URM student, 15 teams had 2 URM students, 7 teams had 3 URM students, 3 teams had 4 URM students.

Teamwork Grades

Each team read two scientific articles from major psychological science journals and synthesized the information from the articles into a poster (one separate poster for each article). Two teaching assistants who were blind to our hypotheses graded all posters (i.e., they each

graded half of the submitted posters). They used the same grading rubric (see Figure S1) to grade each poster.

Figure S1

Grading Rubric Used by Teaching Assistants to Determine Grades on Teamwork Projects

Poster Grading Rubric

Team:_____

Introduction	**4 total pts**
Clarity & Comprehension: Completeness:	____/2 ____/2 ____/4
Hypotheses/Research Question	**2 total pts**
Clarity & Comprehension: Completeness:	____/1 ____/1 ____/2
Research Methods	**4 total pts**
Clarity & Comprehension: Completeness:	____/2 ____/2 ____/4
Results	**4 total pts**
Clarity & Comprehension: Completeness:	____/2 ____/2 ____/4
Discussion	**4 total pts**
Clarity & Comprehension: Completeness:	____/2 ____/2 ____/4
Overall communication and design	2 pts
Total:	**____/20**

Plagiarism

If we find evidence of plagiarism in any of the 5 sections listed above, you will receive 0 points for that section. Plagiarism is defined as "the practice of taking someone else's work or ideas and passing them off as one's own." For this assignment, this means that you should not simply copy and paste whole sentences from the research article. It's ok to use the same language and words (broadly defined) but if we find a *sentence* from the research article on your poster, we will consider this plagiarism. Please try to use your own sentence structure and describe the research in your own words as much as possible.

Additional Measures

See OSF https://osf.io/zuaqj/ for complete survey.

Pilot Study: Additional Results

Raw Number Analysis

Paralleling the analysis using percentage of FG college students in a team, team performance improved linearly as the *number* of FG college students in a team increased, B = 1.45, SE B = .46, p =.003, CI [.52, 2.38], η^2 = .195.

Analyses Adjusting for Covariates

We adjusted for covariates as robustness checks. On the individual level, FG student status was negatively associated with identifying as White (r = -.23, p = .002) and Asian American (r = -.26, p < .001), as well as positively associated with identifying as Latinx (r = .47, p < .001). Thus, we adjusted for race/ethnicity. We created one dummy-coded variable to denote if students identified as an underrepresented minority or not (White or Asian American, or both = 0, any other racial/ethnic category = 1; correlation of binary URM variable with FG student status: r = .44, p < .001). In addition, students who were entirely or mostly raised in the U.S. (students who answered the question "Where were you raised?" with "In the US only" or "Mostly in the US, some in another country") were also more likely to be FG college students (r = .27, p < .001). In other words, international students are less likely to identify as FG college students. We created one dummy-coded variable to denote if students were entirely or mostly raised in the U.S (not mostly raised in the U.S. = 0, mostly or entirely raised in the U.S. = 1). No other demographic variables that we collected were significantly associated with FG student status at the individual level.

In light of the association between FG student status, race/ethnicity, and being raised in the US, we repeated our analysis on the individual level, this time adding dummy-coded race/ethnicity and U.S.-raised variables as covariates. Inclusion of these covariates did not substantially alter the relationship between individual assignment and FG student status, B = -4.08, SE B = 1.74, p = .020, 95% CI [-7.52, -.64], η^2 = .031. Similarly, an analysis with

covariates revealed no social class achievement gap emerged on team assignments, B = -.01, SE B = .58, p = .991, 95% CI [-1.15, 1.15].

On the team level, we wanted to rule out the alternative explanation that better performing teams are comprised of better performing individuals, so we adjusted for individual performance (i.e., the average of the team members' grades on individual assignment) in our analysis of team performance as a function of percentage of FG students. The results remain substantially unchanged, B = 8.29, SE B = 2.66, p = .003, CI [2.92, 13.66], η^2 = .196.

In addition, percentage of females in a team was also positively correlated with percentage of FG college students in a team (r = 0.31, p = .043). No other demographic variables that we collected were significantly associated with the percentage of FG students on the team level (p's > .06).

Thus, we repeated our analysis of team performance as a function of percentage of FG students in a team adjusting for the covariate. The results remain substantially unchanged after including the covariate, such that teams with a higher (vs. lower) percentage of FG students perform better, even when adjusting for percentage of females in the team, B = 6.52, SE B = 2.57, p = .015, CI [1.32, 11.71], η^2 = .138.

We also repeated our moderation analysis of team performance as a function of percentage of FG students in a team and meeting mode adjusting for the same covariate. Consistent with the results reported in the main text, we do not find a significant main effect of meeting mode on team performance (B = -.15, SE B = .77, p = .851, 95% CI [-1.71,1.42]), even when adjusting for percentage of females in the team. However, mirroring the results in the main text and in further support of our Extension Hypothesis 1, we find a significant interaction between meeting mode and social class composition on team performance adjusting for percentage of females in the team, (B =-10.34, SE B = 4.82, p = .038, CI [-20.09, -.60], η^2 =.108).

Specifically, among teams that work together more frequently in-person, there is a significant effect of social class composition on team performance adjusting for percentage of females in the team (B = 14.62, SE B = 4.73, p = .004, 95% CI [5.04, 24.21]). In contrast, among teams that work together more frequently asynchronously, there is no effect (B = .66, SE B = 3.56, p= .854, CI [-6.56, 7.87]).

Additional Results with Categorical Measure of Meeting Mode

We also conducted the focal moderation analysis presented in the main text instead using a categorical approach. Specifically, we constructed a variable that reflected the proportion of students in a team who selected "We primarily met in-person" to the "What was the primary way that you and your teammates communicated to complete this project?" Mirroring the results presented in the main text, we do not find a significant main effect of meeting mode on team performance (B = -.76, SE B = 1.08, p = .486, 95% CI [-2.94, 1.42]. We find a non-significant interaction between meeting mode and social class composition on team performance, (B = -11.08, SE B = -1.78, p = .083, CI [4.324, 57.702], η^2 = .123.

Examining Potential Spillover Effects of Teamwork on Individual Performance

First, we tested if overall team performance and the interaction between overall team performance and FG college student status predicts individual exam performance (i.e., the average of the best two exams, which reflects the individual component of the course grade as defined on the syllabus). These results yielded non-significant results (main effect: B = -.10, SE B =.26, p = .709, 95% CI [-.608, .414]; interaction: B = -.03, SE B =.52, p = .949, 95% CI [1.06, .100]).

In addition, we tested if team performance on the first team project and the interaction between team performance on the first team project and FG college student status predict individual performance *after* the first team project (i.e., the average score on Exam 2 and Exam

3). These results yielded non-significant results (main effect: *B* = -.04, SE *B* =.13, *p* = .733, 95% CI [-.30, .21]; interaction: *B* = -.08, SE *B* =.26, *p* = .760, 95% CI [-.59, .43]).

We also tested if team performance on the first team project and the interaction between team performance on the first team project and FG college student status predicts individual performance on Exam 2 (i.e, the exam taking place right after the first team project). Again, these results yielded non-significant results (main effect: *B* = -.23, SE *B* = .48, *p* = .881, 95% CI [-1.04, .89]; interaction: *B* = -.016, SE *B* =.98, *p* = .817, 95% CI [-2.15, 1.70]).

Lastly, we tested whether team performance on the second team project and the interaction between team performance on the second team project and FG college student status predicts individual performance on Exam 3 (i.e, the exam taking place right after the second team project). These final analyses also yielded non-significant results (main effect: *B* = -.35, SE *B* =.63, *p* = .577, 95% CI [-1.59, .89]; interaction: *B* = -1.01, SE *B* =1.23, *p* = .415, 95% CI [-3.45, 1.43]).

Focal Study: Additional Methods

See Table S1-2 for intercorrelations among all variables at the individual (Table S1) and team (Table S2) level.

Participant Demographics

Students reported being between the ages of 18 to 54 (*M* = 21.15, *SD* = 3.88). 78 identified as male, 238 identified as female, and 10 people identified as “other.” 294 students self-reported having entirely or mostly been raised in the United States. 100 students identified as East Asian/Asian American, 11 as African American, 115 as Latino/a, 47 as Non-Hispanic White, 13 as other and 48 as having more than one ethnicity.

Table S1.

Means, Standard Deviations, and Correlations for Individual-Level Variables in Focal Study

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11
1. FG status	--	--											
2. URM status	--	--	.38***										
3. Age	21.15	3.88	.03	-.01									
4. Gender (1 = *man*)	0.24	0.43	-.07	-.02	.09								
5. Raised in the US (1 = *raised in the US*)	0.90	0.30	.06	.16**	-.03	.09							
6. Exam Score	80.96	13.01	-.15**	-.15**	-.11	-.13*	-.07						
7. Teamwork Score	91.42	7.34	-.00	-.09	.01	-.12*	-.06	.11*					
8. Sense of Fit	5.79	1.02		.02		.07	.01	.07	.12*	-.14*	.06		
9. Skills	5.91	0.94	.04	.17**	-.03	.03	.11	-.15**	.08	.75***			
10. Asynchronous Work	4.70	1.08	.01	-.06	.11*	-.05	.02	-.04	-.07	-.02	-.02		
11. In-Person Work	3.11	1.28	-.06	.01	-.14*	.06	-.09	-.02	.11*	.14*	.19***	-.46***	
12. Hybrid Work	2.51	1.43		.04	-.08	.01	-.03	-.15*	-.21***	-.07	.10	.13*	-.09	.17**

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Table S2.

Means, Standard Deviations, and Correlations for Team-Level Variables in Focal Study

Variable	]*M*	*SD*		1	2	3	4	5	6	7	8	9	10	11
1. % FG Students		--	--											
2. % URM Students		--	--	.20										
3. % Men Students		--	--	-.09	.24*									
4. % Students Raised in US		--	--	.04	.37**	.24*								
5. Mean Team Exam Score		81.07	5.50	-.13	.06	-.08	.15							
6. Teamwork Score		91.76	4.95	-.07	-.23	-.13	.01	.26*						
7. Mean Team Sense of Fit		5.78	0.58	.02	.12	.15	.27*	.03	.08					
8. Mean Team Skills		5.92	0.50	-.07	.20	.07	.25*	.07	.16	.74***				
9. Mean Asynchronous Work		4.69	0.71	.20	-.02	-.14	.10	-.16	-.13	-.19	-.20			
10. Mean In-Person Work		3.12	0.87	-.29*	.04	.05	-.12	.25*	.13	.24*	.29*	-.69***		
11. Mean Hybrid Work		2.53	0.66	.01	-.22	-.03	-.27*	-.40***	-.17	-.03	.09	-.03	.13	
12. In-Person and Synchronous Work		3.21	0.72	-.27*	.03	.09	-.11	.23	.16	.22	.26*	-.89***	.93***	.11

Note. * $p < .05$; ** $p < .01$; *** $p < .001$.

Team Composition

All 347 students were randomly assigned through the learning management system Canvas to teams, but $n = 2$ students were excluded from the analyses at the team level because they did not contribute to their teams (i.e., the two students received 0 points on both teamwork assignments). After dropping the two students, the class consisted of 70 teams: 66 teams of 5 students and 4 teams of 4 students. This is the final team composition used in the analyses reported in the main text. Of these teams, 3 teams had 0 FG college students, 11 teams had 1 FG college student, 20 teams had 2 FG college students, 25 teams had 3 FG college students, 8 teams had 4 FG college students, and 3 teams had 5 FG college students. In percentages, 3 teams were 0% FG college students, 3 teams were 100% FG college students, and 23 teams were 60% FG college student composition (which was the modal social class composition). In terms of the teams' gender composition, 3 teams had 1 woman, 10 teams had 2 women, 24 teams had 3 women, 22 teams had 4 women, and 11 teams had 5 women. The teams' racial/ethnic compositions were as follows: 3 teams had no URM students, 15 teams had 1 URM student, 23 teams had 2 URM students, 24 teams had 3 URM students, 4 teams had 4 URM students, and 1 team had 5 URM students.

Recommended Cutoff Ranges for Interrater Reliability

We interpreted our agreement indices based on recommendations for cutoff levels outlined by LeBreton and Senter (2008). Instead of having traditional dichotomous cutoff values, LeBreton and Senter (2008) suggest ranges of values to interpret agreement indexes (similar to how we interpret effect sizes). For the r_{wg} index, values between 0.00-0.30 represent lack of agreement, values between 0.31-0.50 represent weak agreement, values between 0.51-0.70 represent moderate agreement, values between 0.71-0.90 represent strong agreement, and values between 0.91-1.00 represent very strong agreement. Their suggested ranges for ICC(1) are: 0.01

is considered a small effect, 0.10 constitutes a medium effect, and 0.25 represents a large effect (LeBreton & Senter, 2008). Because ICC(2) values are comparable to reliability indices, LeBreton and Senter (2008) suggested cutoff values that range from 0.70 to 0.85.

Additional Measures

See OSF (https://osf.io/zuaqj/) for complete survey.

Class Session Task Focus

We also collected a measure that captured the degree to which team members stayed focused on the class project in each of two in-person class project work sessions. To assess this, we asked participants three questions: "Overall, how focused was your group on accomplishing the task, i.e. how much did members appear to be engaged and attentive to the task itself?," "Overall, how much did group members build upon each other's ideas, i.e., how much did group members integrate different members' ideas into a common solution?," and "How much positive reinforcement was given from one member to another, i.e., how much did group members accept, affirm, and complement each other's ideas?" on a scale from 1 = *None at all* to 5 = *A great deal* ($\alpha_{T1} = 0.82$, $M_{T1} = 4.01$, $SD_{T1} = 0.80$; $\alpha_{T2} = 0.85$, $M_{T2} = 4.02$, $SD_{T2} = 0.87$; $M_{avg} = 4.02$, $SD_{avg} = 0.74$).

Focal Study: Additional Results

Interrater Agreement: Working Together In-Person (vs. Remotely)

We found strong interrater agreement for working in-person (vs. remotely), ICC(1) = .28, ICC (2) = .66, and $r_{wg} = .70$.

Interrater Agreement: Hybrid Meetings

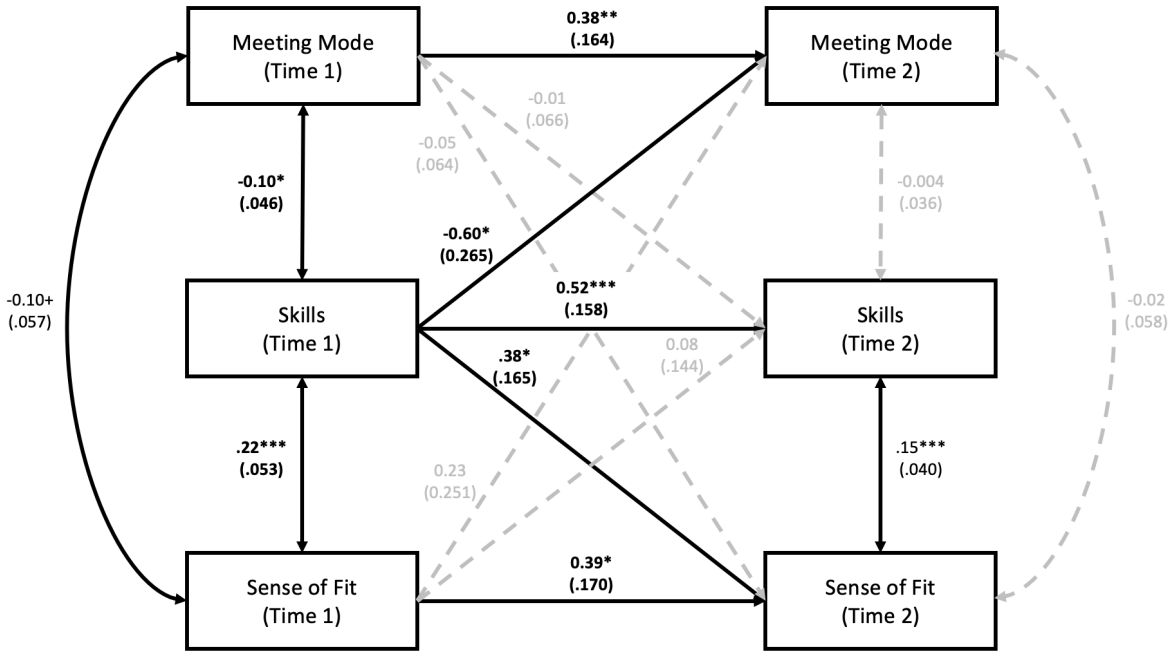
We found weak interrater agreement for hybrid work, ICC(1) = -.01, ICC (2) = -.08, and $r_{wg} = .48$.

Additional Analyses: Cross-Lagged Analyses

Given that we collected two waves of data across the semester after each team project assignment submission, we were also able to conduct cross-lagged analyses to better test our theorized causal ordering. Specifically, we entered our combined measure of synchronous and in-person (vs. asynchronous and remote) work moderator and our proposed mechanisms of skills and fit at Time 1 and Time 2 (see Figure S2). Cross-lagged analyses enabled us to test the effect of meeting mode at Time 1 on skills and fit and Time 2, and the reverse causal order. The model demonstrated good fit (CFI = 0.997, TLI = 0.984, RMSEA = .035, 90% CI of RMSEA [0.000, 0.202], SRMR = 0.042). We did not obtain evidence of our proposed indirect effects through the percentage of FG students on the team (see Figure S2 and Table S1). Of note, though exploratory, team-level reported skills at Time 1 significantly predicted both sense of fit and in-person and synchronous work at Time 2. Though skills were not connected to the percentage of FG students in the team, these results do suggest that it may be particularly important to foster a strong sense of effective group processes early in the life cycle of a team for other downstream benefits (e.g., fit, opting to work together synchronously and in-person) to emerge later on.

Figure S2

Cross-Lagged Analysis of Synchronous and In-Person Meeting Mode at Time 1 on Skill and Fit at Time 2.



Note. Parentheses represent standard errors. + $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table S3.

Path Estimates for Cross-Lagged Analysis of Meeting Mode at Time 1 on Skill and Fit at Time 2

Path	Estimate	*SE*	*Z*	*p*	95% CI Lower	95% CI Upper
In-Person + Synchronous T2 ~ In-Person + Synchronous T1	**0.367**	**0.119**	**3.074**	**.002**	**0.133**	**0.600**
Sense of Fit T2 ~ Sense of Fit T1	**0.393**	**0.17**	**2.309**	**.021**	**0.059**	**0.726**
Skills T2 ~ Skills T1	**0.517**	**0.157**	**3.283**	**.001**	**0.208**	**0.825**
Skills T2 ~ In-Person + Synchronous T1	0.015	0.066	0.225	.822	-0.114	0.144
Skills T2 ~ Sense of Fit T1	0.08	0.144	0.559	.576	-0.201	0.362
Sense of Fit T2 ~ In-Person + Synchronous T1	0.055	0.064	0.861	.389	-0.07	0.18
Sense of Fit T2 ~ Skills T1	**0.378**	**0.165**	**2.296**	**.022**	**0.055**	**0.701**
In-Person + Synchronous T1 ~ Sense of Fit T1	-0.242	0.251	-0.967	.333	-0.734	0.249
In-Person + Synchronous T2 ~ Skills T1	**0.603**	**0.265**	**2.271**	**.023**	**0.083**	**1.123**
In-Person + Synchronous T1 ~ % FG students in team	-0.83	0.51	-1.628	.104	-1.829	0.169
Sense of Fit T1 ~ % FG students in team	0.439	0.273	1.609	.108	-0.096	0.973
Skills T1 ~ % FG students in team	0.151	0.267	0.565	.572	-0.372	0.673
Teamwork Final Score ~ In-Person + Synchronous T2	0.408	0.582	0.701	.483	-0.733	1.549
Teamwork Final Score ~ Sense of Fit T2	0.602	1.242	0.485	.628	-1.833	3.037
Teamwork Final Score ~ Skills T2	-0.648	1.424	-0.455	.649	-3.439	2.142

Teamwork Final Score ~ In-Person + Synchronous T1	0.314	0.745	0.422	.673	-1.146	1.774
Teamwork Final Score ~ Sense of Fit T1	-1.194	1.256	-0.951	.341	-3.655	1.267
Teamwork Final Score ~ Skills T1	1.941	1.546	1.255	.209	-1.09	4.971
Teamwork Final Score ~ % FG students in team	-0.675	2.443	-0.276	.782	-5.464	4.114

Note. **Bold** values indicate significant path estimates (i.e., $p < .05$).

Raw Number Analysis

Paralleling the analysis using percentage of FG college students in a team, the association between team performance and the *number* of FG college students was nonsignificant, $B = -.20$, SE $B = .52$, $p = .706$, CI [-1.24, .84].

Analyses Adjusting for Covariates

We adjusted for covariates as robustness checks. On the individual level, FG student status was positively associated with identifying as a member of an underrepresented minority group ($r = .40$, $p < .001$). Thus, we adjusted for race/ethnicity as a covariate. We created one dummy-coded variable to denote if students identified as an underrepresented minority or not (White, Asian American, or both = 0, any other racial/ethnic category = 1). In contrast to the pilot, students who were entirely or mostly raised in the U.S. (students who answered the question "Where were you raised?" with "In the US only" or "Mostly in the US, some in another country") were not more likely to be FG college students ($r = .06$, $p = .314$). No other demographic variables that we collected were significantly associated with FG student status at the individual level (all p's $\geq .542$).

In light of the association between FG student status and race/ethnicity, we repeated our analysis on the individual level, this time adding dummy-coded race/ethnicity as a covariate. Inclusion of this covariate did not substantially alter the relationship between individual

assignment and FG student status, B = -2.41, SE B = 1.40, p = .086, 95% CI [-5.16, 0.34], η^2 = .02. Similarly, an analysis with covariates revealed no social class achievement gap emerged on team assignments, B =-.36, SE B = .67, p = .597, 95% CI [-1.68, .96].

On the team level, we wanted to rule out the alternative explanation that better performing teams are comprised of better performing individuals, so we adjusted for individual performance (i.e., the average of the team members' grades on individual assignment) in our analysis of team performance as a function of percentage of FG students. The results remain substantially unchanged, B = -.79, SE B = 2.54, p = .755, CI [-5.86, 4.27].

In addition, again on the team level, the correlation between percentage of FG students on the team and percentage of URM students on the team reached significance (r = 0.29, p = .015). Thus, as a robustness check, we repeated our analysis of team performance as a function of percentage of FG students in a team adjusting for percentage of URMs in a team. The results remain substantially unchanged, B = -.14, SE B = 2.66, p = .958, CI [-5.44, 5.17].

We also repeated our moderation analysis of team performance as a function of percentage of FG students in a team and meeting mode adjusting for the same URM covariate. Consistent with the results reported in the main text, we do not find a significant main effect of meeting mode on team performance (B = -.54, SE B = .79, p = .496, 95% CI [-2.11, 1.03]), even when adjusting for percentage of URMs. Mirroring the results in the main text, we do not find a significant interaction between meeting mode and social class composition on team performance adjusting for percentage of URMs in the team, (B = -4.00, SE B = 2.57, p = .125, 95% CI [-9.15, 1.14]).

Analyses Including Non-Contributors

In addition to our focal team-level analyses excluding the 2 non-contributors, we also conducted robustness checks including the 2 non-contributors and utilizing the mode of the

teamwork final score as our outcome variable (i.e., the score that the group earned, not adjusted downward for the lower score that the 2 non-contributors earned) and adjusting for a dummy code reflecting teams with a non-contributing member (=1) vs. those without a non-contributing member (=0). These analyses mirrored the results in the main text, B = -1.86, SE B = 2.62, p = .480, 95% CI [-7.074, 3.364].

We next conducted moderation analyses. Mirroring the results in the main text, when utilizing teamwork final score mode as our outcome and including dummy coded non-contributors, the interaction between percentage of FG students in a team and asynchronous meeting mode was not significant, B = -5.25, SE B = 3.26, p = .112, 95% CI [-11.751, 1.254].

Finally, similar to the exploratory moderation analyses in the main text combining in-person and synchronous teamwork into a single variable, the interaction between percentage of FG students in a team and in-person and synchronous teamwork adjusting for groups with non-contributors was marginally significant, B = 6.45, SE B = 3.34, p = .058, 95% CI [-0.220, 13.116]. While substantively similar, these results failed to reach conventional levels of significance.

Finally, mirroring the exploratory moderation analyses in the main text combining in-person and synchronous teamwork into a single variable, the interaction between percentage of FG students in a team and in-person and synchronous teamwork adjusting for groups with non-contributors and hybrid work was significant, B = 7.45, SE B = 3.31, p = .025, 95% CI [0.962, 14.174].

We next investigated the simple effects of percentage of FG students. In line with the results reported in the main text, a floodlight analysis to identify the Johnson-Neyman significance region revealed two statistically significant transition points. First, the effect of the percentage of FG students on teamwork performance was statistically significant for values of

working together synchronously and in-person above 5.36. Specifically, when teams frequently worked together synchronously and in-person (i.e., score > 5.36; the top 3% of our sample), the effect of percentage of FG students on teamwork performance was positive and significant ($p <$ .05). Next, the results indicated that the effect of the percentage of FG students on teamwork performance was statistically significant for values of working together synchronously and in-person below 2.37. Specifically, when teams rarely worked together synchronously and in-person (i.e., score < 2.37; the bottom 11% of our sample), the effect of percentage of FG students on teamwork performance was negative and significant ($p < .05$). For values from 2.37 to 5.36, the effect was non-significant.

Integrative Data Analysis: Additional Results

Table S4.

Integrative Data Analysis of Individual Performance as a Function of FG College Student Status and Study (At the Individual Level)

Predictor	*B*	*SE B*	*t*	*p*	95% CI Lower	95% CI Upper
Study	-.016	.065	-0.25	.802	-.145	.112
FG Student Status (FG)	**-.305**	**.087**	**-3.47**	**.001**	**-.478**	**-.132**
Study × FG Student Status	.005	.088	0.06	.949	-.167	.178

Note. **Bold** values indicate significant path estimates (i.e., $p < .05$). Study coded such that −1 = pilot study and 1 = focal study 1B. FG student status coded such that 0 = CG student status and 1 = FG student status. CI = confidence interval.

Table S5.

Integrative Data Analysis of Teamwork Performance as a Function of FG College Student Status and Study, Adjusting for Percentage of FG College Students on a Team (At the Individual Level)

Predictor	*B*	*SE B*	*t*	*p*	95% CI Lower	95% CI Upper
Study	.116	.071	-0.25	.104	-.02	.256
FG Student Status (FG)	.048	.102	-3.47	.638	-.153	.249
Study × FG Student Status	-.178	.094	-1.91	.057	-.362	.005
Percentage of FG Students on a Team	**.134**	**.050**	**2.71**	**.007**	**.037**	**.232**

Note. **Bold** values indicate significant path estimates (i.e., $p < .05$). Study coded such that −1 = pilot study and 1 = focal study 1B. FG student status coded such that 0 = CG student status and 1 = FG student status. Percentage of FG Students on a Team was *Z* -scored after the data sets were pooled. CI = confidence interval.

Table S6.

Integrative Data Analysis of Teamwork Performance as a Function of Percentage of FG College Students on a Team and Study (At the Team Level)

Predictor	*B*	*SE B*	*t*	*p*	95% CI Lower	95% CI Upper
Study	.049	.096	0.51	.610	-.142	.240
Percentage of FG Students on a Team	**.211**	**.102**	**2.06**	**.041**	**.008**	**.413**
Study × Percentage of FG Students on a Team	**-.276**	**.102**	**-2.70**	**.008**	**-.478**	**-.074**

Note. **Bold** values indicate significant path estimates (i.e., $p < .05$). Study coded such that −1 = pilot study and 1 = focal study 1B. Percentage of FG Students on a Team was *Z*-scored after the data sets were pooled. CI = confidence interval.

Table S7.

Integrative Data Analysis of Teamwork Performance as a Function of Percentage of FG College Students on a Team, Meeting Mode (Composite Measure), and Study (At the Team Level)

Predictor	*B*	*SE B*	*t*	*p*	95% CI Lower	95% CI Upper
Study	.071	.095	0.74	.458	-.118	.259
Percentage of FG Students on a Team	**.269**	**.103**	**2.61**	**.010**	**.065**	**.474**
Study × Percentage of FG Students on a Team	**-.308**	**.103**	**-2.99**	**.003**	**-.513**	**-.104**
Meeting Mode (Composite)	-.085	.099	-0.86	.391	-.281	.111
Study × Meeting Mode	-.149	.099	-1.51	.135	-.345	.047
Percentage of FG Students on a Team × Meeting Mode	**-.347**	**.125**	**-2.77**	**.007**	**-.595**	**.098**
Study × Percentage of FG Students on a Team × Meeting Mode	.149	.125	1.19	.236	-.099	.398

Note. **Bold** values indicate significant path estimates (i.e., $p < .05$). Study coded such that −1 = pilot study and 1 = focal study 1B. Percentage of FG Students on a Team was *Z*-scored after the data sets were pooled. Meeting Mode was *Z*-scored within each data set (higher numbers mean more asynchronous and remote work), CI = confidence interval.

Table S8.

Integrative Data Analysis of Teamwork Performance as a Function of Percentage of FG College Students on a Team, Meeting Mode (1-Item Measure), and Study (At the Team Level)

Predictor	*B*	*SE B*	*t*	*p*	95% CI Lower	95% CI Upper
Study	.059	.095	0.63	.533	-.129	.248
Percentage of FG Students on a Team	**.264**	**.103**	**2.56**	**.012**	**.059**	**.468**
Study × Percentage of FG Students on a Team	**-.314**	**.103**	**-3.05**	**.003**	**-.518**	**-.110**
Meeting Mode (1-Item)	-.060	.097	-0.62	.537	-.252	.132
Study × Meeting Mode	-.124	.097	-1.28	.204	-.316	.068
Percentage of FG Students on a Team × Meeting Mode	**-.331**	**.125**	**-2.66**	**.009**	**-.578**	**-.0834**
Study × Percentage of FG Students on a Team × Meeting Mode	.165	.125	1.32	.188	-.082	.412

Integrative Data Analysis (IDA): Exploratory Moderator

In this IDA, we used the composite meeting mode variable that captures the highly correlated in-person and synchronous dimensions of the meeting mode because it is more practically relevant and more similar to the meeting mode item we used in the pilot. In order test the robustness of our exploratory meeting mode moderator, we reran our analyses with the pooled dataset, again allowing the study to predict team performance as a main effect and in interaction with all other predictor variables: percentage of FG college students on the team and asynchronous and remote (vs. synchronous and in-person) meeting mode. The results with the exploratory moderator mirror the results we reported with the preregistered moderator. Again, wee find that the interaction between meeting mode and percentage of FG students on a team is

robust across both studies: B = -.35, SE B = .13, p = .007, 95% CI -.59, -.10], η^2 =.068. The exploratory moderator is also homogenous across studies; we do not find a significant three-way interaction between team social class composition, team performance, and study (p = .236). Specifically, among teams that work together more frequently in-person and synchronously, there is a positive significant effect of social class composition on team performance, B = .51, SE B = .15, p = .001, 95% CI [.20, .81], η^2 = .099. In contrast, among teams that work together more frequently remotely and asynchronously, there is no effect, B = -.12, SE B = .13, p = .398, CI [-.38, .15]. Moreover, for teams with a high percentage of FG students (i.e., +1 *SD*), there was a significant negative effect of working together asynchronously and remotely, B = -.43, SE B = .15, t = -2.82, p = .006, 95% CI [-.73, -.13], η^2 = .059. In contrast, for teams with a low percentage of FG students (i.e., -1 *SD*), the effect was nonsignificant, B = .19, SE B = .14, t = 1.37, p = .172, 95% CI [-.08, .46] (see Table S7 in the SI for detailed results).

Taken together, across studies, we find that working together (vs. individually) improves the performance of FG college students. In addition, we find that teams with a higher percentage of individuals from working-class contexts perform better than teams with a lower percentage of individuals from working-class contexts, although this association was more pronounced in the pilot study than in the focal study. Lastly, we find a significant and robust moderation of this effect, such that the beneficial effects of working together for FG college students is most pronounced when teams work together primarily in-person and synchronously.