Leadership, Communication, and Gender Peer Effects at the National Outdoor Leadership School

Matthew Pearson* Benjamin Blemings† Scott Cunningham‡

This Version: March 13, 2023; Most Recent: Web Link

Abstract

Non-cognitive skills play important roles in education and careers, but less is known about peers influences on these skills. The National Outdoor Leadership School conditionally randomly assigns students to wilderness classes which allows us to estimate causal effects of peers on self-reported non-cognitive skills. Being assigned to peers with higher self-rated non-cognitive ability reduces self-reported ratings in communication and leadership, particularly for females. This work suggests relying on strong peers may not be an effective way to bolster non-cognitive ability for college-aged individuals and may in fact contribute to fewer women in leadership positions.[1]

Keywords: Noncognitive ability, peer effects, human capital, leadership gender gap

JEL Classifications: I20, J01

*Ph.D. Moon Fabrications, California. E-mail: Matthew@moonfab.com

†Ph.D. Corresponding Author. Postdoctoral Researcher, Dyson School of Applied Economics and Management, Cornell University, Ithaca, New York 14850. E-mail: btb77@cornell.edu

‡Ph.D. Professor, Deparment of Economics, Baylor University, Waco, Texas. Email: Scott_Cunningham@Baylor.edu

[1]We are very grateful to John Gookin for help obtaining the data and for providing background information, and to the National Outdoor Leadership School for providing data and administrative resources that made this study possible. We also thank Ann Stevens, Scott Carrell, Doug Miller, Marianne Page, Hillary Hoynes, Chris Dawes, and the participants at the EJS, UC Davis, and West Virginia University brownbag seminars for many helpful comments and suggestions. Views expressed are our own and do not necessarily reflect those of the National Outdoor Leadership School.

1 Introduction

The importance of non-cognitive skills to employers has grown substantially (Deming, 2017; Edin, Fredriksson, Nybom, & Öckert, 2022), and even moreso when they complement cognitive skills (Weinberger, 2014).[2] One survey found that employers ranked a candidate's attitude and communication higher than their academic performance and test scores (Bowles, Gintis, & Osborne, 2001; Bureau of the Census, 1998). In the Future of Jobs Reports from the World Economic Forum, which projects future skills needed from LinkedIn and Coursera data, at least 4 of the top 10 projected skills employers look for in 2025 will be non-cognitive (Whiting, 2020).[3] Beyond the labor market returns to non-cognitive ability, social skills affect criminality, education, health, social group membership, marriage, macroeconomic performance, and reduces social transfers (Alan, Boneva, & Ertac, 2019; Algan et al., 2022; Balart, Oosterveen, & Webbink, 2018; Chiteji, 2010).

In addition to the growing relevance of non-cognitive skills, there are large gender disparities in business leadership positions (Bertrand & Hallock, 2001; Keller, Molina, & Olney, 2020). For example, in 2019 less than 3% of venture capital funding went to women-led companies and only one-fifth of funding went to startups with at least one women on the founder team (Aleman, 2020). The picture is equally stark in company board positions with women holding 5.3% of positions and 4.4% of CEO roles globally (Deloitte, 2018). It is important to understand whether this gap can be partially explained by leadership skill and self-perceptions. One reason to understand this gap is that gaps in executive leadership contribute to firm performance and have implications for the gender pay gap (Flabbi, Macis, Moro, & Schivardi, 2019).

One question is whether peers affect the development of non-cognitive ability, and if so, in what direction? One could imagine it going either way, but since peer selection is endogenous, identifying that effect is challenging in the face of unobserved confounders (Manski, 1993). Random assignment of the peer effects of noncognitive ability on subsequent noncognitive ability is a challenging problem since random assignment of peers is unnatural and rare.

We overcome this endogeneity problem by using conditionally random assignment of peers in an outdoor leadership program to estimate the causal effect of peers on non-cognitive skill

[2]Non-cognitive is an umbrella term that has encompassed a lot of measurements in the literature, see Duckworth and Yeager (2015) for a nice summary.

[3]These skills include learning strategies, initiative, leadership, and resilience.

accumulation. We provide first ever evidence that peer non-cognitive ability effects matter as inputs in one's own accumulation of noncognitive ability. We quantify the impact of baseline skills in communication, leadership, and attitude on ex-post non-cognitive skill outcomes using self-reported survey measures. The setting is the National Outdoor Leadership School (NOLS), where applicants are assigned to a course section from their rankings over course types. Conditional randomization tests, which control for selection into course types, finds that 4/7 individual baseline characteristics are uncorrelated with peers in the same section and 2/3 abilities are uncorrelated. Furthermore, conversations with NOLS officials confirm that non-cognitive ability is unobservable to those assigning students to sections. The correlated demographic characteristics are due to age restrictions on course sign-up and pre-requisites for some courses.

We document the evolution of gender gaps in baseline skills by age, on which we have data in exact days, through the formative years of high-school-new college graduates. We show that there is no gender gap in communication, that high school and advanced undergraduate college student men rate their leadership higher than women (i.e. the gender gap in leadership is u-shaped by age), and that women rate their attitudes as better until the later periods of undergraduate education. We believe that this is an important descriptive result which can inform the ongoing research on the education gender gaps (Delaney & Devereux, 2021). For example, Croson and Gneezy (2009) summarizes the evidence that women are more pro-social and we add nuance to this finding by documenting that women only believe they are more pro-social until around age 21.

Peers with higher average non-cognitive ability reduce one's own self reported non-cognitive ability after the completion of an outdoor course, controlling for student's baseline self-rating and course type fixed effects. These effects vary by gender, with men and women being affected in communication, but only women reducing their leadership due to higher average peer leadership. Furthermore, we consider alternative empirical specifications that are consistent with other mechanisms through which peers could exert influence. The linear-in-means model seems appropriate for male communication and female leadership as the inclusion of the maximum and minimum values have less explanatory power than the average term. In contrast, when the maximum and minimum communication values are included for women, the linear-in-means term becomes indistinguishable from 0, but the minimum and maximum coefficients are both negative. When the

lead communicator is male, women rate their own communication higher than when a women is the highest ranked communicator.

These results are an important contribution to the literature on the effects of peer non-cognitive ability and on the development of non-cognitive ability. While the link between peer non-cognitive characteristics and own academic achievement has been researched in both the secondary (Neidell & Waldfogel, 2010; Shure, 2021) and post-secondary schooling contexts (Golsteyn, Non, & Zölitz, 2021), this is the first research to document a link between peer non-cogntive abilities and one's own non-cognitive abilities in any context. In addition to the personality traits, persistence, and Big 5 characteristics used previously, we add the non-cognitive abilities of communication, leadership, and attitude. This also adds to an increasingly large literature on the determinants of non-cognitive abilities such as schools (Jackson, Porter, Easton, Blanchard, & Kiguel, 2020), parents (Bertrand & Pan, 2013), siblings (Hayduk & Toussaint-Comeau, n.d.), teachers (Kraft, 2019), and relative obesity (Huang, Liu, & Zuppann, 2022). Closely related to peer non-cognitive behavior, peers exposed to domestic violence or that exhibit disruptive behavior or exposure to peers with special needs reduces ones' own test scores and later in life income (Balestra, Eugster, & Liebert, 2022; Carrell, Hoekstra, & Kuka, 2018; Carrell & Hoekstra, 2010; Figlio, 2007).

We also document that women's beliefs in their own leadership abilities are particularly negatively impacted by higher quality peers. While men's leadership is unaffected by peers, women's own leadership is reduced by 0.2 standard deviations in response to a peer group with 1 unit higher leadership scores. In addition to apprehension leading male teams (Born, Ranehill, & Sandberg, 2022) and lower rewards than otherwise equal men (Grossman, Eckel, Komai, & Zhan, 2019), an additional explanation for the gender gap in leadership positions could be that exposure to peers with higher leadership ability reduces women's, and not men's, self-perception of their ability to lead. Other findings have shown that women are less likely to speak publicly (De Paola, Lombardo, Pupo, & Scoppa, 2021), but our results suggest that this is not due to higher communication ability peers, since men and women are approximately equally affected in communication- though through potentially different peer mechanisms.

While our use of survey-collected outcomes is less than ideal, since it may only weakly correlate with actual ability which is of ultimate interest, we believe that effects on self-perception are important as there is a vast literature on how beliefs affect behavior across many fields of eco-

nomics (Dupas, 2011; McCluskey & Rausser, 2001; Sloan, McCutchan, & Eldred, 2017; Viscusi, 1991; Wang, 2014). With respect to peers and gender, believing that one is lower in the ability distribution compared to peers reduces willingness to compete (Bedard & Fischer, 2019). Also, there is a gender application (for promotion at the European Central Bank) bias that is partly explained by preferences for competition (Hospido, Laeven, & Lamo, 2022). We document how peers cause one to change their self-perceptions which prior research suggests affects willingness to compete which in turn affects willingness to apply for competitive promotions to positions of leadership. Thus, it is reasonable to conclude that leadership peer effects likely play a role in the gender gap in positions of leadership.

Including peer gender composition does not alter any of these conclusions. Instead, a higher proportion of men has some negative effects over and above those of peer non-cognitive abilities on women's self perceptions of their leadership. This suggests that previous studies using peer gender composition are not capturing aspects of non cognitive ability. Additionally, we do not find peer effects of non-cognitive abilities on grades, but we do find that higher male proportions reduce grades specifically, a higher proportion of men reduce the likelihood of men getting at least an A- by 14.7-17.6 percentage points. Other studies find that more females improve test scores, social acclimation, general satisfaction, belonging, and self-worth (Gong, Lu, & Song, 2021; Orr, 2021). Though our grades measure does not purely reflect cognitive course mastery like a schooling grade would, we largely find results that are consistent with Gong et al. (2021) in that additional males harm grades in a unique, heretofore unexamined setting that is favorable for estimating peer effects.

There is a large literature on gender wage gaps and the findings related to leadership gender gaps are germane to this issue given that executives are at the top of the pay scale. Some explanations of the wage gap, which likely apply to the leadership gap as well, include career disruptions due to motherhood and differences in preferences for competition/difficult majors (Bertrand, Goldin, & Katz, 2010; Buser, Niederle, & Oosterbeek, 2014; Croson & Gneezy, 2009; Landaud, Ly, & Maurin, 2020) and workplace attributes (Flory, Leibbrandt, & List, 2014; Wiswall & Zafar, 2017). The literature has become increasingly concerned with understanding factors contributing to jobs that women have such as the role of mentors (Kofoed et al., 2019), perceptions (Hospido et al., 2022), relative underconfidence (Adamecz-Völgyi & Shure, 2022), and influences of peers (Jones & Kofoed, 2020). Frequently, the work on peer effects on gender gaps focuses on the effects

of gender composition (Fischer, 2017; Gong et al., 2021); however, gender composition could be colinear with peers' perceptions of their own noncognitive abilities. We use our unique data in which we see observe leadership ability prior to peer interaction and after and find that gender composition and peer self-perceived leadership ability are not colinear in explaining post course self-perceptions of leadership. To our knowledge, no other study is able to simultaneously examine the effect of peer gender composition and peer non-cognitive ability on both non-cognitive ability and an objective measure of an instructor-assigned grade.

The rest of the paper is organized as follows. Section 2 provides details on NOLS. Section 3 describes the origin of non-cognitive measures. Section 4 outlines the estimation procedure and presents evidence supporting the conditional random assignment of students to courses. Section 5 discusses the results and Section 6 offers a brief conclusion.

2 The National Outdoor Leadership School

We analyze a sample of students from the National Outdoor Leadership School (NOLS). NOLS was founded in 1965 to train skilled outdoor leaders, but soon broadened its focus to include non-professionals and novices. Since that time, it has grown into the largest backcountry permit holder in the U.S. with over 100,000 graduates.[4]

The courses range from 14 days to a full year, though the most common are 30-day and semester-length courses. The courses focus on one or more of a diverse array of outdoor skills, including backpacking, mountaineering, river canoeing/kayaking, sea kayaking, rock climbing, and horsepacking.[5] In addition to outdoor skills that vary by course, all courses target instruction to a common set of noncognitive skills including interpersonal and group communication, leadership, teamwork, and environmental ethics. All courses are available for college credit that can be transferred to almost any college or university in the United States.[6]

[4] https://www.nols.edu/en/about/history/.

[5] Courses focusing on multiple skill areas are broken into sections in which one skill is emphasized. Most courses that are 30 days or fewer in length have one section and longer courses may have several sections.

[6] The typical transfer student is awarded 16 semester credit hours for a NOLS Semester course. See https://www.nols.edu/en/about/resources/college-credit/.

Most NOLS students are between the ages of 16 and 22 and receive college credit for a NOLS course. NOLS enrolls students from age 14 (where they are eligible for courses that are only offered to 14–15-year-olds) to well above 40 (some courses are restricted to 40+).[7] NOLS attempts to ensure diversity in enrollment. Outreach programs on college campuses and an extensive alumni representative network are used to reach potential students.

The total cost, including tuition, travel, room, board, college transfer credit expenses, gear, and personal expenses, of a popular semester-length course, the Semester in the Rockies, is about $12,400 in 2008. NOLS awards more than $800,000 in scholarships each year and attracts students from a wide range of educational and socioeconomic backgrounds.[8] All courses are eligible for Federal Student Aid, AmeriCorps benefits, Veterans Affairs benefits, and in some cases employer tuition benefits.

This setting is new and offers several advantages for investigating peer effects. Since the classroom is the wilderness, individuals only interact with those in their classroom which means we observe the entire peer group. Furthermore, these courses don't last more than 30 days on average, which could reduce the impact of endogenous network formation (Carrell, Sacerdote, & West, 2013).

3 Data

The research question is how peer non-cognitive ability affects non-cognitive ability formation. At a minimum this requires data on non-cognitive ability at the individual level with information about ones' peers. Additionally, to estimate the causal effects of peers, these peer groups should be assigned exogenously. The NOLS data in this paper fulfills these requirements. In addition to rare measures of non-cognitive ability, these are measured for both before and after the individual interacts with the peers, addressing the reflection problem in estimating causal peer effects (Manski, 1993).

[7] NOLS also instructs custom and professional training courses for schools, organizations, and corporations including: NASA, the Kellogg School of Business, the Wharton School, the United States Naval Academy, Google, and Fidelity Bank and Trust. These groups are concerned with human capital formation and productivity, which suggests the skills taught at NOLS are determinants of educational and labor market success.

[8] See https://www.nols.edu/en/expeditions/planning/financial-aid/.

The administrative data comes from NOLS and the non-cognitive ability measures are from a survey developed by a research team affiliated with NOLS and the University of Utah called the Student Outcome Assessment Project (SOAP).[9] The survey was designed to assess the instructional effectiveness of course characteristics. It was administered to students in various courses over a two year period in 2005 and 2006; the sample in this paper includes 3,773 students, 87 course types, and 358 course sections.

At the individual level, demographics, prior experience, and self-rated ability are observed. The demographic characteristics are age (in days), race (6 categories), and gender. The individual-level data contains instructor assigned grades for the course (from F-A+) and which course and which section of that course the student participated in.

3.1 Non-Cognitive Skills

The outcomes of interest are the noncognitive skills of communication, leadership, and attitude. The actual questions, measuring various aspects of these abilities and chosen to measure targeted outcomes in the NOLS curriculm, are shown in Table A.1.[10] To each question, students rated themselves on a scale of 1 (worst) to 8 (best).

The survey underwent several pilot trials prior to being implemented to ensure that the survey questions adequately and consistently measure the associated abilities.[11] Surveys were administered within two hours of students returning from the field, typically at the end of the course. For semester-length courses, surveys were conducted during a convenient break between parts close to the 30-day time frame of the typical NOLS course.

At the same time these responses were collected, students also *retrospectively* reported their skills as they were just before the course began. This retrospective structure was chosen to avoid response-shift bias, a serious problem observed in pilot trials and in many surveys of pre-treatment

[9] The student survey is shown in Figures A.1 and A.2.

[10] NOLS's definition of these skills are available at: https://blog.nols.edu/topic/leadership.

[11] See Sibthorp, Paisley, and Gookin (2007) for details.

levels of instruments. This bias occurs when the treatment itself changes the respondent's metric of evaluating the instrument (Howard, 1980; Sibthorp et al., 2007).[12]

Combining Survey Questions Into Single Measures of Ability There are 4 questions that serve as the basis for each non-cognitive ability, but the outcomes and treatments of interest are the single underlying abilities. To simplify the analysis, the answers for each of the 4 survey questions, for each skill, are combined into a single measure of that skill. To combine the questions, an average is taken across each of the 4 questions, for each skill, for each individual. Throughout the paper, each average of 4 questions is referred to as composite communication, leadership, and attitude respectively.[13]

3.2 Descriptive Statistics

Table 1 presents summary statistics of student demographics, student pre and post cognitive ability, course type attributes, and course section compositions. Panel A shows the sample is composed of 64% males, aged 21.14 years old on average, and 52% have taken a prior course at NOLS. A majority, 91%, are white; however, race information is missing for over 300 individuals which leads to them being dropped in regressions that include race.

Panel B presents individual's composite communication, leadership, and attitude from before the course.[14] On average, students rate themselves around a 5.5 for each composite ability.[15] Panel C presents the same information for after the course and the instructor assigned grades. On average, the students rate themselves around 6.5 for each composite ability, an improvement from the baseline average of 5.5. Students receive a B+ on average in the course and the standard deviation is about 1.5 grades (B- to A).

[12] Also, the uniform time frame of surveying students approximately 30 days into the course mitigates a problem observed in pilot trials where pre-treatment scores would creep up as time passed. Similar to "teacher's bias" students begin to forget what it was like to not know the things that they now know very well.

[13] Appendix B uses principal components analysis to show the same questions load on the same factors at the group and individual levels.

[14] Table A.5 shows summary statistics which suggest the disaggregated questions are similar.

[15] While the questions are on an integer scale, the process of averaging 4 questions leads to decimal values.

Panel D shows course type characteristics. Most of the courses are in the summer and in the United States. On average courses last for 37.99 days, but that is slightly skewed upwards with the median length being 30 days. Panel E shows course section level composition. On average, the proportion of students in courses that are males is 64% and that are white is 82%. The average number of students in a section is 10.54, with a median of 11.

3.2.1 Distributions by Gender, Time, and Age

We are interested in leveraging the rich variation in our data to characterize soft skills. Figure A.4 shows the distributions of the composite abilities by gender and before/after the course. There is a shift to the right for both genders in all skills after the course. Broadly speaking, the central tendency of the before and after distributions do not seem to vary much across gender.

Furthermore, we have granular variation in age due to observing exact birthdate and course start date and students answering the same validated non-cognitive questions in the same setting. Figure 1 plots the gender gap in skills by age in quarter years with an overlapped local polynomial regression. Figure 1A shows no discernible relationship between age and communication gender gap since it always fluctuates around 0. As shown in Figure 1B, there seems to be a u-shaped relationship in the gender gap of leadership. Women who are 18-20 years old rate their leadership higher or the same, but at the other ages women rate themselves lower. In contrast, Figure 1C shows women rate their attitudes better all the way until age 21 when it returns to approximate parity.[16]

4 Method

To estimate effects of baseline peer ability, ordinary least squares (OLS) is used. The linear-in-means regressions are specified,

$$(1) \qquad y_{i,t+1} = \beta \overline{Ability}_{-i,t} + \gamma Ability_{i,t} + X_i\psi + \mu_C + \phi_Y + \theta_M + \delta \overline{Ability}_{-i,t,C} + e_i.$$

[16]Of course, these findings may not generalize to populations outside of NOLS.

The *i* stands for individual, *-i* is for everyone in the course section besides person *i*, *t* is prior to the course (i.e., baseline), *t*+1 is after the course, *C* is for course type, *Y* is for course section year, and *M* is for course section month. Standard errors are clustered by course section to account for correlated errors within course sections, potential model mis-specification, and because treatment is at the course section level.

Our outcomes, $y_{i,t+1}$, are the non-cognitive abilities of communication, leadership, and attitude and instructor-assigned grades.[17] The $\overline{Ability}_{-i,t,C}$ accounts for the fact that one cannot be one's own peer which mechanistically creates a negative correlation between peer average and one's own scores (Guryan, Kroft, & Notowidigdo, 2009). Finally, we include a control for the individuals own baseline ability. We prefer this specification, because using a difference as the outcome imposes additional, unnecessary functional form assumptions. The reflection problem in estimating peer effects, described by Manski (1993), is that it is hard to know whether one affects peers or vice-versa. Typically, it is considered adequate to control for baseline measures of performance, because this adjusts comparisons to be within baseline abilities. Our measure of baseline ability achieves this in theory, but is measured at the same time as the post course ability and may be more prone to error than objective measures such as a previous years' grades.

4.1 Peer Effect Specification

The coefficient of interest, representing the effect of peers, is β. It represents the estimate for the leave-out average baseline peer ability on individual *i*'s post course ability. The leave-out average peer effect is calculated as the average ability of all of individual *i*'s coursemates, excluding individual *i*. The leave out baseline composite peer ability is calculated,

$$\overline{Ability}_{-i,t} = \frac{\sum_{j \neq i}^{n_g} A_{j,t}}{n_g - 1},$$

in which subscript g stands for group and A for ability.

[17] We recognize that our skill outcomes are survey measures and so it may be more appropriate to call them self-perceptions of beliefs, we went with a less verbose name for the outcomes in the equation for appearance.

An important control variable is individual i's baseline ability. Students with different initial ability may be differently affected by their peers. By including this control, estimates compare those of the same baseline ability that are assigned to higher or lower skilled peer groups.

Other individual level pre-treatment covariates are included in X. The demographic variables included are continuous age (in days), 6 indicator variables for different ethnicity categories, and a binary variable for male. The final variable is a binary variable for whether or not the individual has previous experience in a NOLS course.

The last fixed effects are for course month, θ_M, and course year, ϕ_Y. These indicators control for trends in the types of individuals who select into NOLS over time. Also, these control for the potential that weather may be correlated with peer effects and/or ability formation.

4.2 Identification, Conditional Randomness, and Measurement Error

Identification of causal peer effects depends on the classmates in the same course section being as good as random. The process by which NOLS assigns students to sections is a good reason to believe that this condition is met, because students do not self-select their peers. The only choice that NOLS applicants have control over is their preference ordering over course types in the course catalogue.[18,19] After students submit course preferences, a NOLS official assigns applicants to a course section. At the time the student is assigned, non-cognitive ability measures are not available to NOLS officials, because the survey is administered at the end of the course.

Similar to the heterogeneity across university courses, there are distinctive differences across NOLS course types (as shown in Table 1, Panel D). It is probable that different types of students

[18] They rank three courses in order of preference, and they are not always assigned their first choice. According to NOLS admissions, there are rare exceptions, such as friends taking a course together who wish to be assigned to the same group, or siblings taking a course at the same time who may wish to be assigned to different groups. A NOLS admissions officer stated that "very few, less than five percent" of students know their potential peers prior to enrolling in the course.

[19] Another situation where assignment may be non-random is when NOLS admissions makes an effort to ensure that any one female (or in very rare cases, male) is not the only female in a group. This means that if a course has only one female enrolled, the admissions officers will either attempt to find another or offer her the option to switch to a different course. Thus the characteristics of lone females in a group may differ systematically from the rest of the sample because they have chosen not to select out of the group. For most courses the gender ratio is determined by enrollment, but for popular courses with multiple sections, the admissions staff forms some groups that are roughly half male and the remaining males are sorted into all-male groups (note the bump on the right tail of Figure A.3).

choose to take different course types.[20] To account for this selection into courses, the models include μ_C which is a vector of course type fixed effects. By controlling for selection into course type, it makes it plausible that the peers in a students specific course section are as good as random, allowing for the identification of causal peer effects.

4.2.1 Conditional Randomization

An empirical fact that would be true if the students are randomly assigned, enabling identification of causal effects, is baseline characteristics of the students in course sections would be uncorrelated with the baseline characteristics of the other students in the same section. To investigate whether peers are as good as randomly assigned, Figure 2 presents regression coefficients from regressions of own baseline characteristics on the average baseline group characteristic. The blue, open dots are from regressions without any additional controls and show a statistically significant correlation for every measured characteristic. This is not unexpected, since different types of students are likely to prefer different course types; however, this correlation is not conducive for estimating causal effects without a plan to account for selection into course type.

Course type is observable, so selection into course type can be accounted for by including course type indicators in the regression of individual baseline characteristics on group average baseline characteristics. These indicators enable the relationship between individual and group characteristics to be estimated within the same course type. The estimated relationship between individual and peer characteristics are displayed in the orange, closed dots in Figure 2. After controlling for selection into course type, 6 out of 7 of the individual characteristics are uncorrelated with group characteristics.[21] The one coefficient that remains statistically significant is group age. This is likely due to age restrictions on courses that are decided by NOLS which leads to discontinuities in the ages that are eligible, making global linear regression mis-specified.[22,23]

[20] For instance, those with stronger leadership skills may be more likely to take a course requiring technical skill; however, these course indicator variables allows β to be estimated across individuals who are in the same course type, but different course section with different peer group ability.

[21] The way conditioning on course types removes the relationship between peer baseline ability and own baseline ability can be seen in the binned scatter plots, one set bivariate and one set conditional on course type indicators, in Figure A.5.

[22] Figures A.6 and A.7 show these discontinuities around age 16 and 22.

[23] Table A.2 shows that while age does have a global linear effect, there is no relationship of peer age on individual age within the age groups of 14-15 and 16-22, compared to 22 and above.

These conditional random assignment tests do not adjust for the mechanistic negative correlation that arises from individuals being unable to be their own peer (Guryan et al., 2009). Table A.3 employs the fix suggested by Guryan et al. (2009) by including the course type (urn) leave out average. As shown in Table A.3, this causes some of the peer characteristics to be associated, in a statistically significant way, with ones' own characteristics. As shown in Panel A, attitude and communication now have associations of 0.11-0.12 which is still smaller than the unconditional association of 0.5 (shown in Figure 2). As shown by Jochmans (2023), this test from Guryan et al. (2009) is size correct, but has low power and courses with different numbers of students causes additional problems.[24]

We address these issues using a test developed by Jochmans (2023). In this test, baseline communication is no longer correlated with peer communication. Previous experience is associated with peer previous experience, due to the presence of prerequisites for some courses. Of the other 5 characteristics (race, gender, and skills), only attitude is significantly correlated with peer attitude which could just be a result of multiple testing. We know that the officials that assign students do not observe attitude (or other skills), so this is not the result of selection. We provide robustness checks that include attitude, age, and previous experience and none of the results are affected. Peer attitude is also the only skill that doesn't affect ones' own attitude. The absence of correlations between individual and group characteristics, at the beginning of the course, supports the conditional random assignment of peers which addresses the selection problem in estimating causal peer effects Manski (1993).

Measurement Error in Ability One might be concerned about the connections that Angrist (2014) makes between peer effects estimations and instrumental variable (IV) designs. One of the main conclusions of Angrist (2014) is that estimated peer effects from a leave one out OLS approach is equivalent to an IV approach, in which group indicators are (weak) instruments for peer group ability. These coefficients may differ due to peer effects or any other reason. For example, IV estimates and OLS estimates may differ due to measurement error in the explanatory variable which attenuates IV estimates less than OLS estimates (Feld & Zölitz, 2017, pg. 391). We have shown evidence consistent with the conditional random assignment of students, which means

[24]There is substantial variation in the number of students in courses at NOLS, ranging from 427 to only a single section of 4 students.

that measurement error in the peer ability explanatory variable leads only to attenuation bias (Feld & Zölitz, 2017, pg. 393).

4.3 Binned Scatterplots

Next, binned scatterplots are presented in Figure 3, providing a non-parametric visualization of the relationship between baseline peer ability and post course own ability. All Figures control for own baseline ability, to ensure comparisons are made between those of similar baseline ability, and course-type fixed effects to address selection into course type.[25] Across all skill categories and both genders, a linear line of best fit is a reasonable specification.

As shown in Figure 3A by the nearly equal slopes, an increase in baseline, peer communication ability has approximately equal negative effects on both men and women's post-course communication. In contrast, Figure 3B shows that an increase in peer leadership has a very small, if any, effect on men's post-course leadership. However, an increase in peer leadership has a large, relative to communication or men's leadership, negative effect on women's post-course leadership. The slope of the men's line is only .027 and the slope of the women's line is nearly 5 times larger at 0.133. For attitude, Figure 3C shows that there is barely any effect of baseline peer attitude on post-course peer attitude for either gender.[26]

5 Results

To understand the magnitude and preciseness of the relationships presented in the bineed scatterplots, Table 2 presents regression results. As shown by Panel A, an increase in baseline peer communication ability has a negative and statistically significant relationship (reject above 99% confidence) with own, post-course communication regardless of which controls are included. Column 3 of Panel A, in which individual covariates and course type fixed effects are included so that the estimates may have a causal interpretation, reports that a 1 unit increase in peer, baseline

[25] They are included through a semi-parametric adjustment described in Cattaneo, Crump, Farrell, and Feng (2019), but as shown in Figure A.10 the results are similar without the course type fixed effects.

[26] These binned scatterplots all use 10 equally sized bins, but the bias-variance optimal number of bins is between 3 and 6 (Cattaneo et al., 2019). The optimal bins are shown in Figure A.9 and the conclusions are the same.

communication (approximately 1.25 standard deviations) causes a 0.08 decrease (approximately one-tenth of a standard deviation) in own, post-course communication.[27] The reduction in magnitude when controls are added could be classic omitted variables bias that is addressed or it could be that the course type fixed effects addresses potential measurement error from self-reported surveys which becomes only attenuation bias.

As shown in Panel B of Table 2, nearly the exact same pattern is true of leadership. In column 1, a 1 unit increase in peer, baseline leadership is associated with a 0.13 reduction in own, post-course leadership. As shown in column 3 of Panel B, adding controls and fixed effects reduces the level of statistical significance to 95% confidence and the magnitude of the estimate from 0.13 to 0.07.[28] In contrast to leadership and communication, Panel C shows that there are no significant effects of peer attitude once fixed effects and/or covariates are included.

The negative estimates suggest that stronger peers have negative influences on one's own self-perception of their own non-cognitive ability. As shown by Table A.4, which is a table summarizing peer effects models that is reproduced from Sacerdote (2011) that is based on Hoxby and Weingarth (2006) and Lazear (2001), one plausible mechanism generating this result is called an "invidiuous comparison." When invidiuous comparisons exist, the outcomes of students are so-called "harmed" by the presence of better achieving peers. This could be from students comparing their own ability to others and adjusting their beliefs about their own ability.

When interpreting the results, one should be careful not to conclude that actual ability decreased. It could be the case, although it seems somewhat unlikely, that self-perceived ability is lowered but actual ability is unchanged or increased given the large literature on how beliefs influence actions (Dupas, 2011; McCluskey & Rausser, 2001; Sloan et al., 2017; Viscusi, 1991; Wang, 2014). Focusing on perceptions of one's own ability, perceiving that you are less able than you believed prior to interacting with a peer group could affect levels of self-investment in human capital and future aspirations.

Next, we rule out several alternative explanations for the results. One might be concerned that the way that the composite skills are averaged over 4 different questions relating to a particular

[27] Because it is an average of 4 questions, a reduction of 1 would imply they scored 1 lower on each question or 4 lower on one question.

[28] The standard deviations are nearly exactly the same as communication or attitude leading to a similar magnitude interpretation as offered above for communication.

skill are generating spurious results. For example, this could happen if there are some positive peer effects for some questions but negative peer effects for others which would alter the conclusions. In order to address this concern, the averaging is abandoned in order to look at the effect of peer's answers to each individual question on one's own answer and these results are presented in Table A.6. As evidenced by every coefficient being negative, the estimates are not being driven by the averaging into composite skills. The only communication question in which there are no reported effects is on reading others' body language, while leading a discussion, giving constructive feedback, and expressing ideas clearly are all negatively affected by peers. For leadership, there are no effects on initiative for group tasks or making decisions in a timely manner, but there are effects for taking responsibility without being asked and being good at making decisions. It seems reasonable that body language, decision timeliness, and initiative in group tasks are less likely to be affected by peer comparisons than the questions that are affected by peers.

The primary specifications do not account for the variation in peer ability which could be concerning if the variation in peer ability also explains outcomes. In order to account for this possibility, regressions are estimated which account for variance by including the leave-out standard deviation of peer ability ability as a separate control as in Hanushek, Kain, Markman, and Rivkin (2003). As shown in column 1 of Table A.7, standard deviation is not statistically significant and it's inclusion has no effects on the estimates for leave-out mean.

Finally, one may be concerned about specification error due to the bounded nature of the dependent variable which ranges from 1 to 8. As shown in Figure A.4, it is possible that this upper bound is relevant for the after course dependent variables. To address the bounded dependent variable, non-parametric kernel regression is used. These regressions are estimated,

$$Ability_{i,t+1} = f(\overline{Ability}_{-i,t}, Ability_{i,t}, X_i, \mu_C, \phi_Y, \theta_M) + e_i, \tag{2}$$

in which all subscripts and variables are the same as Equation 1.[29] As shown in Table A.8, this more flexible specification produces very similar estimates. One way the estimates are different

[29]The method does not assume that $f(...)$ is linear or linear in parameters and is estimated using local linear regression. Continuous covariates use an Epanechnikov kernel and discrete covariates use the Li-Racine Kernel (Li, Lu, & Ullah, 2003). As recommended by Cattaneo and Jansson (2018), standard errors are calculated by bootstrapping.

is that, even once covariates are included, there is a small negative effect on attitude which is statistically significant at 90% confidence.

5.1 Treatment Effects by Gender

Next, research shows that group dynamics are influenced by gender differences, so it's possible that there are heterogeneous treatment effects across gender. To understand how gender influences the results, Table 3 disaggregates the sample into men and women. The sample size becomes 2,124 men and 1,154 women which is not alarmingly small for either group.

As shown in Panel A of Table 3, there is a negative peer effect for men (column 1) and women (column 4) in communication. In contrast, Panel B shows no effect of peers on men's leadership. Meanwhile, a 1 unit increase in peer leadership reduces female leadership by 0.16 and this is statistically significant above 99% confidence. While other studies have shown that women are reluctant to lead male groups (Born et al., 2022) or that women are less rewarded for leading (Grossman et al., 2019), this is the first study to identify that peers with better leadership ability have uniquely harmful effects on women's perception of their own leadership ability. Disaggregation does not change any of the peer effects on attitude.

5.1.1 Robustness

Next, there are other specification and sampling choices that could be appropriate so we plot the linear in means peer effect along with 95% confidence intervals for these other choices in Figure 4. First, one could wish to see the outcome be the difference between baseline and post skills, instead of using baseline as a control. Row 1 of each subfigure in Figure 4 shows that this does not alter the conclusions of the linear-in-means estimates. Second, one could be concerned about the age restrictions on courses making including age as a continuous control a mis-specification. To address this possibility, the second rows of each subfigure in Figure 4 re-parameterize age as 5 dummies (13-15, 15-17, 17-21, 22-40, 40+) which roughly correspond to NOLS age restrictions for some courses and the results remain similar.

Finally, the before and after measures are both elicited at the end of the course to avoid response-shift bias which could also cause some skepticism that the before controls account for

reflection. As noted above, the courses range from 8-94 days, but the data is elicited after about 30 days for nearly all courses. This makes it somewhat odd to consider the after measures as truly after for courses where the measures were elicited up to 64 days before the end of the course. In row 3, we drop all courses that were longer than 30 days and the relative magnitude of the communication effect on men and women switches with the effect on women become insignificant at 95% confidence. This could be also due to wider standard errors that results from having less women than men and even less when some courses are excluded. The other skills remain largely unchanged which suggests the retrospective elicitation is of minimal concern. Finally, there could be concern about the imbalanced variables which Figure A.8 suggest are previous experience and baseline attitude, so the fourth row includes these as additional controls and shows the results are unaffected.

5.1.2 Gender Composition

There is also a large literature on the effects of peer's gender composition, so it could be that self-rated non-cognitive ability is just proxying for gender composition and therefore has no additional explanatory power for our outcomes. To address this, we include the leave-out proportion of males in a group as an additional explanatory variable. Prior to including both peer ability and gender composition in the same regression, we first separately estimate the impact of gender composition. There are no statistically significant effects of male proportion on communication (Panel A) or attitude (Panel C) on either men (column 2) or women (column 5). For men, there is no effect of male proportion on leadership, but an increase in male proportion does reduce women's self-reported leadership by 0.35 and this is statistically significant at 99% confidence.

Separately estimated, male proportion and non-cognitive ability have effects on various skills; however, one could worry about colinearity or that they capture similar aspects. If either of these were the case, then including both variables in the same regression could affect the point estimates and/or standard errors to the point of insignificance or even cause sign-flipping. Columns 3 and 6 include both peer baseline ability (linearly) and proportion male as separate regressors and the results are essentially the same across all panels and both genders. For communication, attitude, and male leadership we cannot reject that the effects of gender composition are 0 or that peer effects and gender composition exhibit equally sized effects. We are able to reject, at 10% confidence,

that proportion male and leadership peer effects are equal with proportion male having a larger coefficient. While it appears that the effect of male proportion is much larger, it is an explanatory variable that varies from 0 to 1 so it is difficult to compare.

Heterogeneous Treatment Effects Other than Gender One might expect heterogeneous treatment effects by age. For instance, non-cognitive skills are believed to be malleable longer than cognitive ones due to the way the brain develops (Cunha, Heckman, Lochner, & Masterov, 2006, pg. 702). As such, columns 1 and 2 of Table A.9 splits the sample into above and below the median age (19.58 years). As shown by Panel A, columns 1 and 2, greater peer communication ability harms both older and younger groups of people. In contrast, a 1 unit increase in baseline peer leadership has negative effects on older individuals, but not younger ones.[30,31]

Course Characteristics Characteristics of the course, such as the number of days a course lasts or the number of course-mates, could matter for the effects of peers. In accordance with examining this potential heterogeneity, Table A.10 presents results which are split at the median number of individuals in the class (11) and at the median length of the course in (30 days). The most consistent result is that courses with more members have stronger negative effects. A 1 unit increase in peer ability leads to reductions of 0.15 (p-value = 0.005) and 0.12 (p-value = 0.008) for communication and leadership respectively. This is consistent with the invidious comparison that is likely behind the negative effects, since more members means more people to compare ones self to.

5.2 Alternative Peer Effects Models

Next, as we believe it to be reasonable that our negative effects of strong peers is consistent with the invidious comparison model, it is reasonable to understand how well other peer models explain one's post-course non-cognitive ability survey answers. Two such models focus on the maximum and minimum scores of one's peers which we model empirically by taking the (leave-out) first and

[30] Figure A.11 disaggregates age by quintiles. As this reduces power, naturally confidence intervals are much wider. The main pattern that emerges is that the oldest students are most harmed. Going off point estimates, effects appear weaker in college-aged individuals compared to high school or older individuals.

[31] Those with previous experience have lower communication scores from better quality peers. Those without previous experience have lower leadership scores from better quality peers.

last ranked individual scores in the course section. We begin by including these without the average peer score. As shown by Panel A of Table 4, the maximum peer reduces male communication (column 1) and the maximum and minimum values of peers reduces women's communication (column 4).

In column 2 and 5, we also include the average leave out peer score effects. The maximum score becomes statistically insignificant for men and renders the average effect statistically insignificant by increasing the standard errors more than the magnitude. When the leave out average is added for women, it becomes statistically insignificant and the max and min coefficients remain statistically significant. This suggests that for females the peer effects of communication operate through the maximum and minimum communication ability members of the group and that the largest effect is through the highest ranked peer.

So far, the models ignore the potential that the gender of the top or bottom ranked individual may matter. To account for this we code dummy variables for whether the leaveout maximum or minimum ability individual is male.[32] As shown in column 3 of Table 4, including these additional interactions does not matter much for men, making us hesitant to conclude that communication operates through the minimum and maximum channels for men.

On the other hand, when these interactions are included for women, the average effect of peer leadership stays similarly sized and statistically insignificant. The minimum and maximum coefficients have slightly larger magnitudes (i.e. more negative). The interaction of a dummy that the top communicator is male and the top communication score yields a statistically significant coefficient of 0.012. This suggests that when men are the top communicator in the group that it is less harmful to women's perceptions of their own communication than when the top communicator is a women.

We repeat this same exercise for leadership in Panel B. For men, when all the variables are included (column 3) none of them are statistically significant which is similar to regressions without the other measures for men. For the women, there are initially effects of the maximum leader in the group, these go to zero when the average peer leadership is included in columns 2 and 3. For attitude, there are some statistically significant and/or large coefficients in places, but no obvious systemic patterns emerge.

[32]We break ties by using the fraction of individuals male. For even gender splits, observations are dropped. Everyone has the same leave out maximum and minimum individual except for the top individual who receives the values of those second from the bottom or top.

5.3 Instructor-Assigned Grades

Finally, it may be that the negative impact is entirely driven by the self-reported aspect of the dependent variable and it would be interesting to be able to compare previous work on grades. To address this possibility, there's a non self-reported post course dependent variable which is the instructor-assigned grade. While this addresses the self-reporting it comes at the cost of not measuring only non-cognitive skills. It's also worth noting that the grades have both similarities and differences with typical schooling. A similiarity is that these grades are actual competencies (though they are outdoor) and can be used for college credit; while on the other hand, the grades also include leadership, teamwork, and other environmental studies which typical grades do not.[33]

The results with instructor-assigned grades are presented in Table 5. As shown in Panel A, there is a negative and stable, though statistically insignificant effect of proportion male regardless of which non-cognitive ability variables are included. One's baseline non-cognitive score coefficients are statistically significant in all columns, but peer average characteristics are all statistically insignificant due to smaller magnitudes and larger standard errors than own characteristics. As shown by column 4, when own communication and attitude are included in the same regression there coefficients change, with communication becoming negative, but the effect of leadership remains unchanged.

Work on grades often uses an aggregated GPA from several courses as the dependent variable which can take on many values, but our grades data is a discrete, ordered score from 1 course. As shown in Figure A.13, there are only about 10 values for grades in this data and there is not much variation with most of the data centered around B+ and A-. This may be due to students being able to opt into NOLS while schooling is more universal with less alternative options. An alternative way to specify the outcome, which is arguably needed due to this limited variation, is to sensibly dichotomize grades into a dummy variable. We create a dummy variable which equals 1 for students who receive at least an A- and 0 otherwise and use this as the dependent variable. This can be interpreted as the outcome that an instructor felt a student did excellent in the class.

As shown by Panel B of Table 5, this produces mostly similar results. There are no significant peer effects and there are significant effects of own baseline ability whose magnitude and sign

[33] See https://web.archive.org/web/20080208094643/http://www.nols.edu/courses/find/byskill/.

depend on which baseline skills are included. The main difference is that the proportion of male students is now statistically significant across all combinations of explanatory variables. A greater number of men in the group reduces the probability that students are assigned grades of at least an A-. More specifically, going from no men to all men in the group reduces the probability of getting at least an A- by 15-18 percentage points (pp). This result is driven by an increasing proportion of men being more harmful to men than women (Table A.11) and is also statistically significant when estimated by a logistic regression (Table A.12).

6 Conclusion

Non-cognitive abilities are increasingly important and there are large disparities in female representation in business leadership positions. To assess the importance of peer non-cognitive ability on non-cognitive ability, we leverage the conditional random assignment of students to courses in the National Outdoor Leadership School. We find that higher non-cognitive ability peers reduces self-perceptions in leadership and communication, but not necessarily attitude. There are heterogeneous treatment effects with women's self perception of their leadership being particularly harmed by higher leadership ability peers.

There are several limitations to these results. First, the reported effects are for self-perception which may not accurately reflect actual skill formation. Ideally we have more objective measures of non-cognitive ability but those types of data are not collected. But we nonetheless think this self-reported measure has value because it suggests that peers with higher non-cognitive ability either reduce one's own non-cognitive abilities when exposed for a length of time, or they change one's reference points such that they are less confident. Either of these are important for thinking about the accumulation of non-cognitive skill. Self-perceptions of owns ability could become a self-fulfilling prophecy and is therefore important to understand. Next, NOLS is focused on outdoor education which means that their students may not be representative of the general population which limits the external validity of the findings.

The results tell a cautionary tale for making policy which aims to address the lack of women in business leadership positions. Namely, assigning women to groups in which there are other strong leaders may have the opposite of the intended impacts of increasing their perceptions of their own

leadership ability. Instead of expecting individuals to improve their own non-cognitive abilities from peers, other interventions which increase self-perception of non-cognitive abilities should be considered.

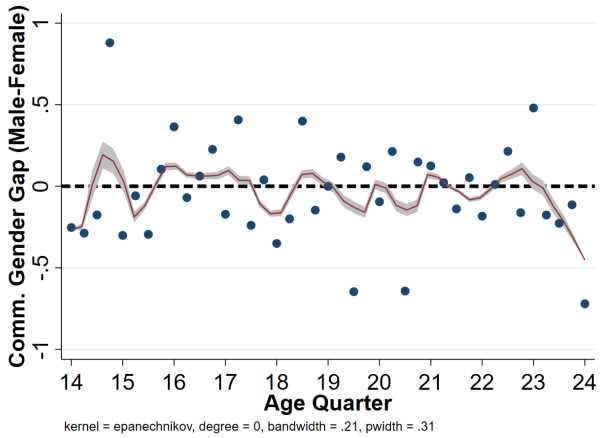
Table 1: Descriptive Statistics

Panel A: Demographics						
	Mean	Median	SD	Min	Max	Count
Male	0.64	1.00	0.48	0.0	1.0	3773
Age	21.14	19.58	6.49	13.6	70.4	3706
White	0.91	1.00	0.29	0.0	1.0	3403
Previous Experience	0.52	1.00	0.50	0.0	1.0	3773
Leave-out Proportion Male	0.58	0.58	0.17	0.0	0.9	3773
Panel B: Pre-Course						
	Mean	Median	SD	Min	Max	Count
Composite Communication[a]	5.48	5.50	1.17	1.0	8.0	3709
Composite Leadership[b]	5.57	5.75	1.28	1.0	8.0	3708
Composite Attitude[c]	5.37	5.50	1.21	1.0	8.0	3702
Panel C: After Course						
	Mean	Median	SD	Min	Max	Count
Composite Communication[a]	6.49	6.50	0.83	1.0	8.0	3693
Composite Leadership[b]	6.68	6.75	0.84	1.0	8.0	3713
Composite Attitude[c]	6.48	6.50	0.89	1.0	8.3	3698
Instructor Grade	9.05	9.00	1.53	2.0	12.0	3091
Panel D: Course Level						
	Mean	Median	SD	Min	Max	Count
Course in US	0.59	1.00	0.50	0.0	1.0	87
Technical Course	0.25	0.00	0.44	0.0	1.0	60
Cultural Course	0.07	0.00	0.25	0.0	1.0	60
Summer Course	0.80	1.00	0.40	0.0	1.0	60
Course Days	37.99	30.00	22.89	9.0	94.0	77
Panel E: Section Level						
	Mean	Median	SD	Min	Max	Count
Proportion Male	0.64	0.64	0.19	0.0	1.0	358
Proportion White	0.82	0.82	0.14	0.0	1.0	358
Number of Members	10.55	11.00	2.64	2.0	22.0	358

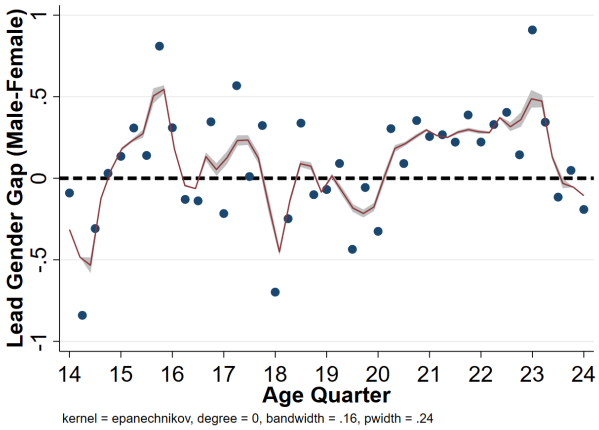
Note: Data is on National Outdoor Leadership School (NOLS) students from 2005-2006. In Panel A, previous experience means that the student has participated in a NOLS course in the past. Composite abilities are averages of the 4 questions that make up each ability as displayed in Table A.1. Panel B is from before the course begins. Panel C is from the end of the course. Instructor grade is a conversion from letter grades to an integer scale from 2-12. Panel D shows course level characteristics. Panel E shows course section level composition information. [a] Composite communication is made up of 4 questions regarding communication. 1) I can lead others in a discussion. 2) I give constructive feedback. 3) I am good at reading other people's body language. 4) I express my ideas clearly. [b] Composite leadership comes from 4 questions. 1) I take initiative in completing group tasks. 2) I often take responsibility without being asked. 3) I am good at making decisions. 4) I make decisions in a timely manner. [c] Composite attitude comes from 4 questions regarding attitude. 1) I am patient with others. 2) I place emphasis on group goals above personal goals. 3) I maintain a positive attitude in adverse conditions. 4) I can manage conflict that occurs between group members.

Figure 1: Gender Gap in Baseline Abilities by Age Quarter at NOLS

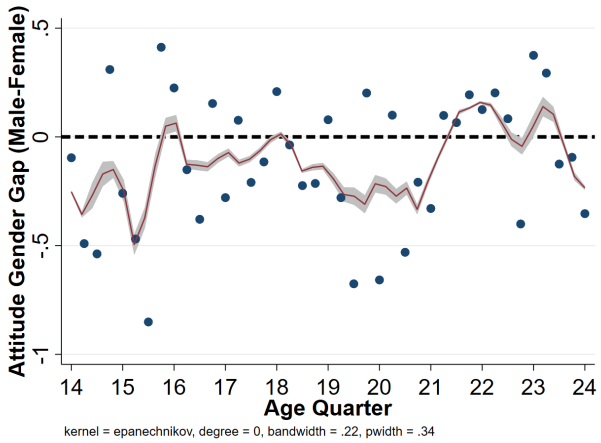
(A): Communication



(B): Leadership

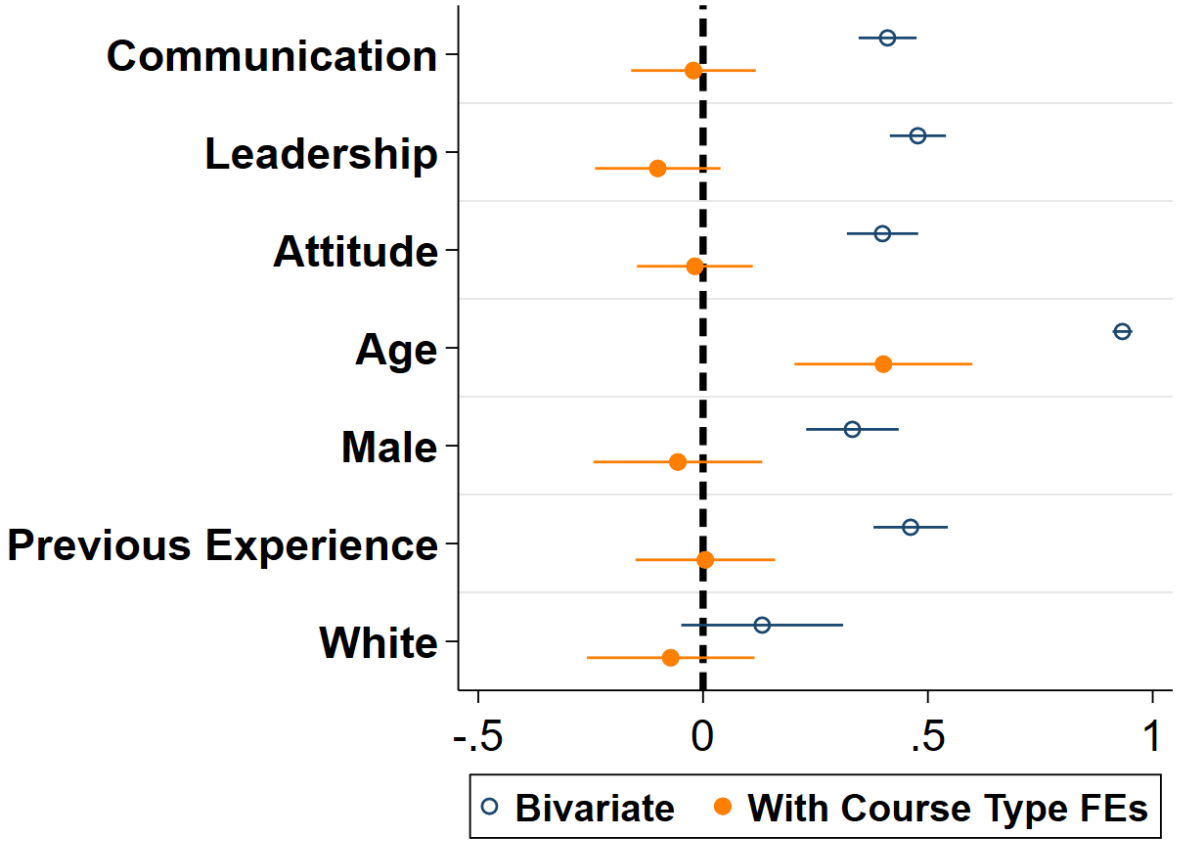


(C): Attitude



Note: Figures show average composite male baseline skills minus average composite female baseline skills conditional on age in quarters (14, 14.25, 14.5) etc. Positive numbers means higher male scores. Local polynomial regression. Frequency weights. 95% confidence intervals. (a) shows the gender gap in composite communication. (b) shows the gender gap in composite leadership. (c) shows the gender gap in composite attitude.

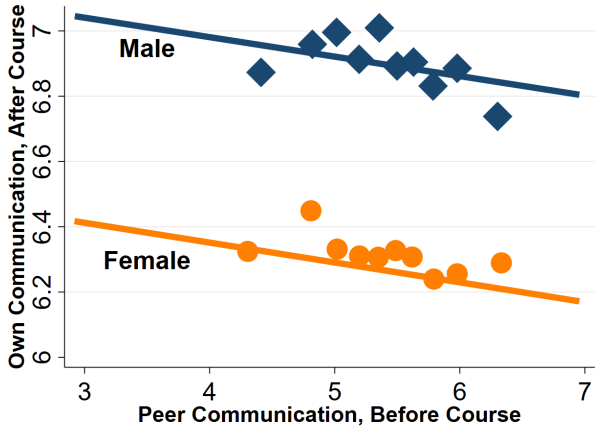
Figure 2: Baseline Characteristics Regressed on Same Group Characteristics



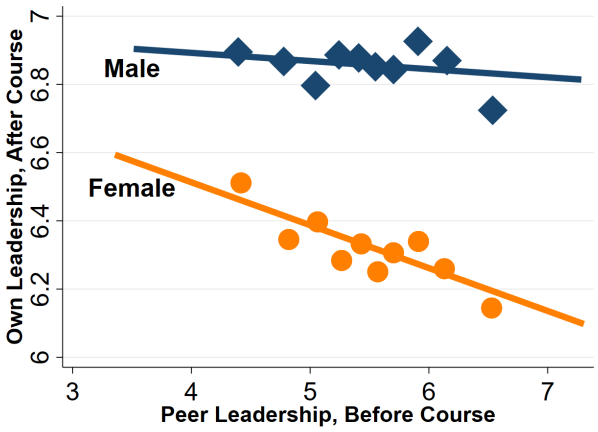
Note: The coefficients are from a regression of baseline characteristics on average baseline group characteristics. The whiskers represent 95% confidence intervals, in which the standard errors are clustered at the course section level. The blue, open dots are from bivariate regressions. The orange, closed dots are from regressions that include dummy variables for each course type.

Figure 3: How Peer Baseline Ability Affects Own Post-Course Ability

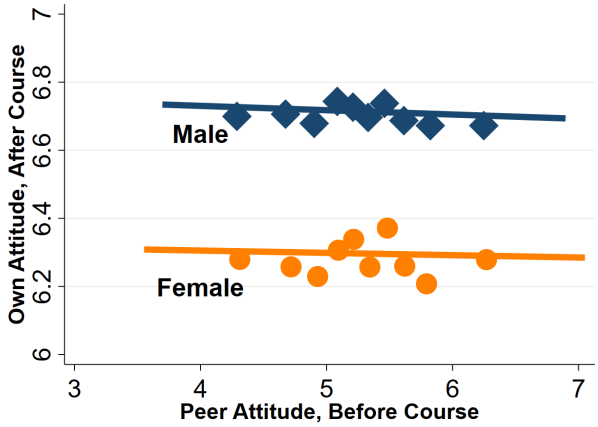
(A): Communication



(B): Leadership



(C): Attitude



Note: All figures show the correlation between baseline leave-one out peer ability and own post-course ability. The correlations control for course type indicators and baseline own ability using methods from Cattaneo et al. (2019). The x-axis is peer leave one-out baseline ability. The y-axis is own post-course ability. Panel (A) shows communication, Panel (B) shows leadership, and Panel (C) shows attitude. The ability of each individual in each category is an average of 4 questions that are answered on a scale of 1-8. The number of bins in each figure, for each gender, are set to 10. The bias-variance optimal number of bins is between 3 and 6 and these figures, which are qualitatively similar, are shown in Figure A.9.

Table 2: The Effect of Baseline Peer Average Composite Ability on Own Composite Ability

Panel A: Communication				
	(1)	(2)	(3)	(4)
Baseline Peer Average Communication	-0.136***	-0.111***	-0.078**	-0.073**
	(0.023)	(0.024)	(0.030)	(0.030)
Observations	3661	3265	3265	3265
R^2	0.478	0.483	0.521	0.522
Covariates	-	X	X	X
Course FEs	-	-	X	X
Course Comm. LO Avg.	-	-	-	X
Panel B: Leadership				
	(1)	(2)	(3)	(4)
Baseline Peer Average Leadership	-0.125***	-0.106***	-0.066**	-0.061**
	(0.019)	(0.021)	(0.026)	(0.026)
Observations	3682	3286	3286	3286
R^2	0.454	0.457	0.488	0.489
Covariates	-	X	X	X
Course FEs	-	-	X	X
Course Lead LO Avg.	-	-	-	X
Panel C: Attitude				
	(1)	(2)	(3)	(4)
Baseline Peer Average Attitude	-0.077***	-0.046	-0.000	-0.001
	(0.026)	(0.030)	(0.032)	(0.032)
Observations	3669	3278	3278	3278
R^2	0.419	0.422	0.454	0.454
Covariates	-	X	X	X
Course Type FEs	-	-	X	X
Course Att. LO Avg.	-	-	-	X

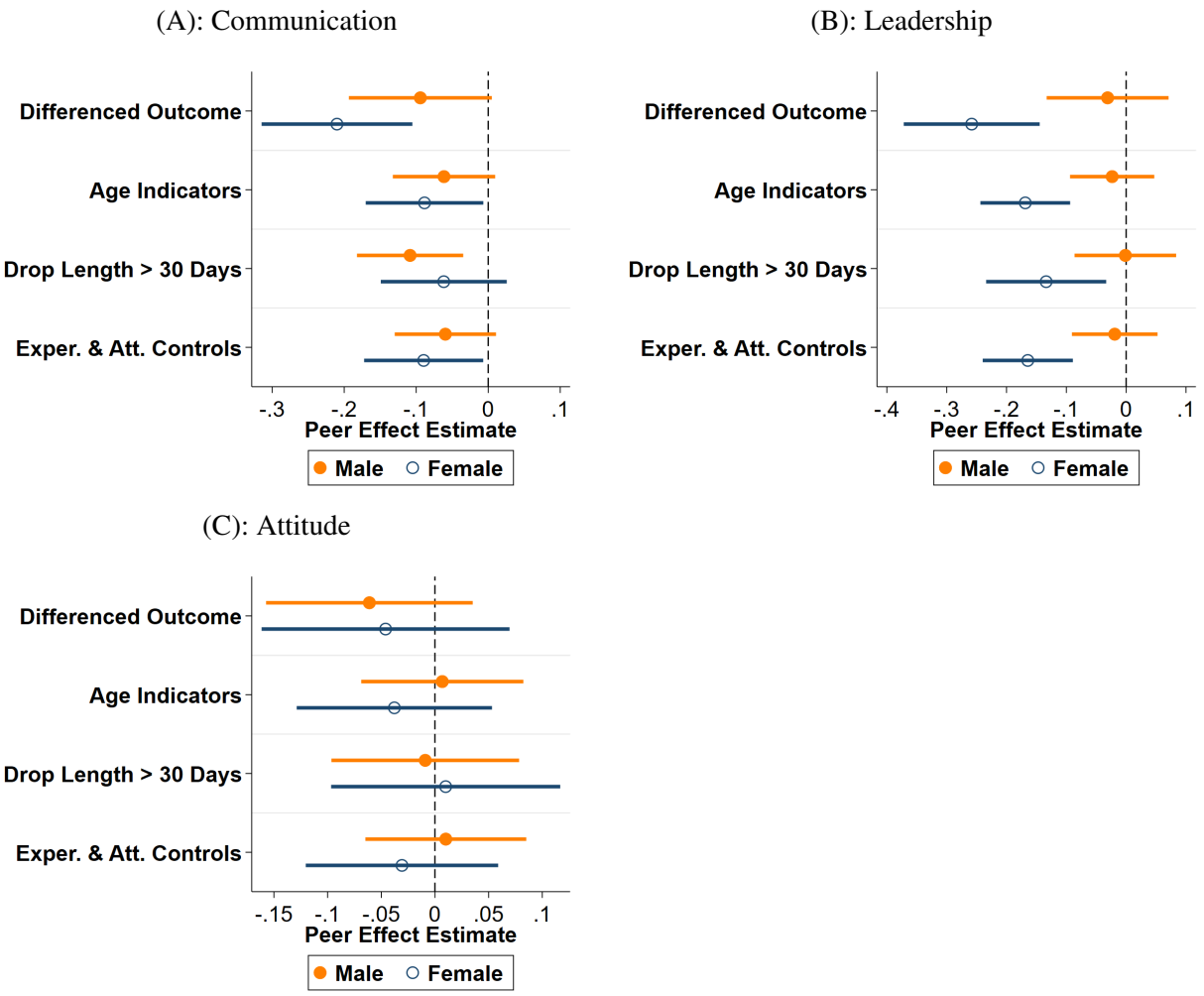
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual composite communication and the independent variable is pre-course peer average composite communication. In Panel B, the dependent variable is post-course individual composite leadership and the independent variable is pre-course peer average composite leadership. In Panel C, the dependent variable is post-course individual composite attitude and the independent variable is pre-course peer average composite attitude. All columns control for the individual's pre-course ability for the outcome of interest in that panel. Column 1 includes no additional covariates. Column 2 adds controls for male, age, and race. Column 3 adds indicators for course type and fixed effects for course section month and year. Column 4 adds the course leave out (LO) average to account for one being unable to be one's own peer (Guryan et al., 2009).

Table 3: OLS of Peer Group on Noncognitive, Gender Composition Controls

Panel A: Communication						
	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Peer Average Comm.	-0.07*		-0.07*	-0.10**		-0.10**
	(0.04)		(0.04)	(0.04)		(0.04)
Leave-out Proportion Male		0.07	0.08		-0.11	-0.11
		(0.12)	(0.12)		(0.13)	(0.13)
Observations	2111	2111	2111	1154	1154	1154
R^2	0.497	0.496	0.497	0.613	0.611	0.613
P-value (Comp = Peer)			0.25			0.94
Panel B: Leadership						
	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Peer Average Lead	-0.03		-0.03	-0.16***		-0.17***
	(0.04)		(0.04)	(0.04)		(0.04)
Leave-out Proportion Male		-0.01	-0.02		-0.35***	-0.38***
		(0.10)	(0.10)		(0.12)	(0.12)
Observations	2124	2124	2124	1162	1162	1162
R^2	0.480	0.480	0.480	0.549	0.545	0.551
P-value (Comp = Peer)			0.89			0.09
Panel C: Attitude						
	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Peer Average Att.	0.01		0.01	-0.03		-0.04
	(0.04)		(0.04)	(0.05)		(0.05)
Leave-out Proportion Male		0.08	0.08		-0.21	-0.23
		(0.13)	(0.13)		(0.15)	(0.16)
Observations	2124	2124	2124	1154	1154	1154
R^2	0.451	0.451	0.451	0.504	0.505	0.505
P-value (Comp = Peer)			0.60			0.22

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual composite communication and the independent variable is pre-course peer average composite communication. In Panel B, the dependent variable is post-course individual composite leadership and the independent variable is pre-course peer average composite leadership. In Panel C, the dependent variable is post-course individual composite attitude and the independent variable is pre-course peer average composite attitude. All columns control for the individual's pre-course ability for the outcome of interest in that panel. Columns 1-3 use the men-only sample and columns 4-6 use the women-only sample.

Figure 4: Robustness Checks



Note: Scales differ. Figures show peer linear-in-means point estimates with 95% confidence intervals by skill and gender. The dependent variable is composite, post course composite survey answers. The standard errors are clustered at the course section level. All regressions include controls for course type, year, and month fixed effects, male gender composition, age, gender, race, and baseline own ability in that given skill. Rows 1 changes the parameterization of the outcome by using the after course - baseline ability. Rows 2 changes parameterization of the age covariate to 5 dummies: 13-15, 15-17, 17-22, 22-40, 40+. Row 3 drops courses that are longer than 30 days. Row 4 adds a previous experience indicator and baseline attitude.

Table 4: Other Peer Effects Models

Panel A: Communication	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Comm. LO Min.	-0.036	-0.013	-0.019	-0.065*	-0.056	-0.058*
	(0.027)	(0.031)	(0.033)	(0.033)	(0.035)	(0.034)
Baseline Comm. LO Max.	-0.085***	-0.054	-0.062	-0.121***	-0.107**	-0.115**
	(0.033)	(0.039)	(0.039)	(0.041)	(0.045)	(0.045)
Baseline Peer Average Comm.		-0.085	-0.081		-0.037	-0.024
		(0.054)	(0.053)		(0.056)	(0.057)
Base Comm. LO Min. * Male			0.006			-0.008
			(0.011)			(0.012)
Base Comm. LO Max. * Male			0.006			0.012**
			(0.005)			(0.006)
Observations	1735	1735	1735	922	922	922
R^2	0.514	0.515	0.516	0.614	0.614	0.616
Panel B: Leadership	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Lead LO Min.	-0.031	-0.017	-0.029	-0.034	0.008	0.022
	(0.023)	(0.024)	(0.026)	(0.026)	(0.029)	(0.029)
Baseline Lead LO Max.	-0.044	-0.024	-0.020	-0.074*	-0.022	-0.020
	(0.034)	(0.039)	(0.040)	(0.044)	(0.048)	(0.048)
Baseline Peer Average Lead		-0.057	-0.059		-0.150***	-0.148***
		(0.047)	(0.046)		(0.054)	(0.054)
Base Lead LO Min. * Male			0.017			-0.020
			(0.011)			(0.013)
Base Lead LO Max. * Male			-0.003			0.000
			(0.005)			(0.006)
Observations	1802	1802	1802	993	993	993
R^2	0.472	0.472	0.473	0.533	0.537	0.538
Panel C: Attitude	Male			Female		
	(1)	(2)	(3)	(4)	(5)	(6)
Baseline Att. LO Min.	-0.009	-0.028	-0.034	-0.052	-0.064	-0.043
	(0.023)	(0.028)	(0.030)	(0.037)	(0.044)	(0.045)
Baseline Att. LO Max.	-0.042	-0.067	-0.069	-0.050	-0.066	-0.065
	(0.040)	(0.045)	(0.046)	(0.047)	(0.050)	(0.049)
Baseline Peer Average Att.		0.070	0.073		0.047	0.040
		(0.056)	(0.056)		(0.075)	(0.076)
Base Att. LO Min. * Male			0.006			-0.031**
			(0.012)			(0.013)
Base Att. LO Max. * Male			-0.001			0.003
			(0.006)			(0.008)
Observations	1746	1746	1746	930	930	930
R^2	0.475	0.475	0.476	0.498	0.498	0.501

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual composite communication and the independent variable is pre-course peer average composite communication. In Panel B, the dependent variable is post-course individual composite leadership and the independent variable is pre-course peer average composite leadership. In Panel C, the dependent variable is post-course individual composite attitude and the independent variable is pre-course peer average composite attitude. LO stands for leave-out. The interaction terms are male = 1 if the top or bottom individual in the group is a male. Ties are broken by using the fraction of men tied for top or bottom. If these ties are exactly 50/50, then these observations/groups are dropped. For students who are top or bottom ranked, the max and min are from the second from the bottom or top rank and the corresponding gender are used.

Table 5: OLS Effect of Peer Non-Cognitive Ability on Grades

Panel A: Discrete Grade				
	(1)	(2)	(3)	(4)
Leave-out Proportion Male	-0.382	-0.330	-0.372	-0.352
	(0.281)	(0.272)	(0.280)	(0.278)
Baseline Own Communication	0.135***			-0.098**
	(0.036)			(0.043)
Baseline Peer Average Communication	0.073			0.046
	(0.091)			(0.114)
Baseline Own Leadership		0.315***		0.308***
		(0.034)		(0.038)
Baseline Peer Average Leadership		0.018		-0.009
		(0.077)		(0.102)
Baseline Own Attitude			0.233***	0.135***
			(0.035)	(0.036)
Baseline Peer Average Attitude			0.045	0.010
			(0.081)	(0.100)
Observations	2755	2754	2759	2690
R^2	0.105	0.154	0.122	0.162
Panel B: Receive $\geq$ A-				
	(1)	(2)	(3)	(4)
Leave-out Proportion Male	-0.170**	-0.147*	-0.176**	-0.176**
	(0.081)	(0.078)	(0.081)	(0.079)
Baseline Own Communication	0.035***			-0.038***
	(0.012)			(0.014)
Baseline Peer Average Communication	0.014			0.001
	(0.027)			(0.033)
Baseline Own Leadership		0.099***		0.111***
		(0.010)		(0.012)
Baseline Peer Average Leadership		0.017		0.020
		(0.024)		(0.029)
Baseline Own Attitude			0.058***	0.022*
			(0.010)	(0.012)
Baseline Peer Average Attitude			0.012	-0.002
			(0.024)	(0.028)
Observations	2755	2754	2759	2690
R^2	0.0893	0.134	0.103	0.142

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is discrete grade in the course which can be A+ (12), A (11), etc. In Panel B, the dependent variable is a dummy variable for the student getting at least an A-.

References

Adamecz-Völgyi, A., & Shure, N. (2022). The gender gap in top jobs – the role of overconfidence. *Labour Economics*, *79*, 102283. doi: https://doi.org/10.1016/j.labeco.2022.102283 [4]

Alan, S., Boneva, T., & Ertac, S. (2019, 02). Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit*. *The Quarterly Journal of Economics*, *134*(3), 1121-1162. doi: 10.1093/qje/qjz006 [1]

Aleman, X. (2020). *Startup fundraising is the most tangible gender gap. How can we overcome it?* https://techcrunch.com/2020/11/09/startup-fundraising-is-the-most-tangible-gender-gap-how-can-we-overcome-it/#:~:text=Gender%20equality%20in%20VC%20makes%20more%20business%20sense&text=In%20fact%2C%20a%20study%20by,%25%20from%20men%2Dfounded%20companies. (Accessed: March 2021) [1]

Algan, Y., Beasley, E., Côté, S., Park, J., Tremblay, R. E., & Vitaro, F. (2022, August). The impact of childhood social skills and self-control training on economic and noneconomic outcomes: Evidence from a randomized experiment using administrative data. *American Economic Review*, *112*(8), 2553-79. doi: 10.1257/aer.20200224 [1]

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108. [13]

Balart, P., Oosterveen, M., & Webbink, D. (2018). Test scores, noncognitive skills and economic growth. *Economics of Education Review*, *63*, 134–153. [1]

Balestra, S., Eugster, B., & Liebert, H. (2022, 05). Peers with Special Needs: Effects and Policies. *The Review of Economics and Statistics*, *104*(3), 602-618. doi: 10.1162/rest_a_00960 [3]

Bedard, K., & Fischer, S. (2019). Does the response to competition depend on perceived ability? Evidence from a classroom experiment. *Journal of Economic Behavior & Organization*, *159*, 146-166. doi: https://doi.org/10.1016/j.jebo.2019.01.014 [4]

Bertrand, M., Goldin, C., & Katz, L. F. (2010, July). Dynamics of the gender gap for young professionals in the financial and corporate sectors. *American Economic Journal: Applied Economics*, *2*(3), 228-55. doi: 10.1257/app.2.3.228 [4]

Bertrand, M., & Hallock, K. F. (2001). The gender gap in top corporate jobs. *ILR Review*, *55*(1), 3-21. doi: 10.1177/001979390105500101 [1]

Bertrand, M., & Pan, J. (2013). The trouble with boys: Social influences and the gender gap in disruptive behavior. *American Economic Journal: Applied Economics*, *5*(1), 32–64. [3]

Born, A., Ranehill, E., & Sandberg, A. (2022, 03). Gender and Willingness to Lead: Does the Gender Composition of Teams Matter? *The Review of Economics and Statistics*, *104*(2), 259-275. doi: 10.1162/rest_a_00955 [3, 17]

Bowles, S., Gintis, H., & Osborne, M. (2001, December). The Determinants of Earnings: A Behavioral Approach. *Journal of Economic Literature*, *39*(4), 1137–1176. Retrieved from http://www.jstor.org/stable/pdfplus/2698522.pdf [1]

Bureau of the Census. (1998). *First findings from the EQW National Employer Survey.* EQW Catalog RE01. (Accessed: July 2020) [1]

Buser, T., Niederle, M., & Oosterbeek, H. (2014, 05). Gender, Competitiveness, and Career Choices *. *The Quarterly Journal of Economics*, *129*(3), 1409-1447. doi: 10.1093/qje/qju009 [4]

Carrell, S. E., Hoekstra, M., & Kuka, E. (2018, November). The long-run effects of disruptive peers. *American Economic Review*, *108*(11), 3377-3415. doi: 10.1257/aer.20160763 [3]

Carrell, S. E., & Hoekstra, M. L. (2010, January). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics*, *2*(1), 211-28. doi: 10.1257/app.2.1.211 [3]

Carrell, S. E., Sacerdote, B. I., & West, J. E. (2013). From natural variation to optimal policy? The importance of endogenous peer group formation. *Econometrica*, *81*(3), 855–882. [6]

Cattaneo, M. D., Crump, R. K., Farrell, M. H., & Feng, Y. (2019). On binscatter. *arXiv preprint arXiv:1902.09608*. [14, 27, 44, 45, 49, 50, 59]

Cattaneo, M. D., & Jansson, M. (2018). Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency. *Econometrica*, *86*(3), 955–995. [16]

Chiteji, N. (2010). Time preference, noncognitive skills and well being across the life course: Do noncognitive skills encourage healthy behavior? *American Economic Review*, *100*(2), 200–204. [1]

Croson, R., & Gneezy, U. (2009, June). Gender differences in preferences. *Journal of Economic Literature*, *47*(2), 448-74. doi: 10.1257/jel.47.2.448 [2, 4]

Cunha, F., Heckman, J. J., Lochner, L., & Masterov, D. V. (2006). Interpreting the evidence on

life cycle skill formation. *Handbook of the Economics of Education*, *1*, 697–812. [19]

De Paola, M., Lombardo, R., Pupo, V., & Scoppa, V. (2021). Do women shy away from public speaking? A field experiment. *Labour Economics*, *70*, 102001. doi: https://doi.org/10.1016/j.labeco.2021.102001 [3]

Delaney, J., & Devereux, P. J. (2021). Gender and educational achievement: Stylized facts and causal evidence. [2]

Deloitte. (2018). *Women in the boardroom: A global perspective.* https://www2.deloitte.com/global/en/pages/risk/articles/women-in-the-boardroom-global-perspective.html. (Accessed: March 2021) [1]

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, *132*(4), 1593–1640. [1]

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237-251. doi: 10.3102/0013189X15584327 [1]

Dupas, P. (2011). Health behavior in developing countries. *Annual Review of Economics*, *3*(1), 425–449. [4, 15]

Edin, P.-A., Fredriksson, P., Nybom, M., & Öckert, B. (2022, April). The rising return to noncognitive skill. *American Economic Journal: Applied Economics*, *14*(2), 78-100. doi: 10.1257/app.20190199 [1]

Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, *35*(2), 387–428. [13, 14]

Figlio, D. N. (2007, 10). Boys Named Sue: Disruptive Children and Their Peers. *Education Finance and Policy*, *2*(4), 376-394. doi: 10.1162/edfp.2007.2.4.376 [3]

Fischer, S. (2017). The downside of good peers: How classroom composition differentially affects men's and women's STEM persistence. *Labour Economics*, *46*, 211–226. [5]

Flabbi, L., Macis, M., Moro, A., & Schivardi, F. (2019, 08). Do Female Executives Make a Difference? The Impact of Female Leadership on Gender Gaps and Firm Performance. *The Economic Journal*, *129*(622), 2390-2423. doi: 10.1093/ej/uez012 [1]

Flory, J. A., Leibbrandt, A., & List, J. A. (2014, 10). Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions. *The Review of*

Economic Studies, *82*(1), 122-155. doi: 10.1093/restud/rdu030 [4]

Golsteyn, B. H. H., Non, A., & Zölitz, U. (2021). The impact of peer personality on academic achievement. *Journal of Political Economy*, *129*(4), 1052-1099. doi: 10.1086/712638 [3]

Gong, J., Lu, Y., & Song, H. (2021). Gender peer effects on students' academic and noncognitive outcomes: Evidence and mechanisms. *Journal of Human Resources*, *56*(3), 686–710. [4, 5]

Grossman, P. J., Eckel, C., Komai, M., & Zhan, W. (2019). It pays to be a man: Rewards for leaders in a coordination game. *Journal of Economic Behavior & Organization*, *161*, 197-215. doi: https://doi.org/10.1016/j.jebo.2019.04.002 [3, 17]

Guryan, J., Kroft, K., & Notowidigdo, M. J. (2009, October). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, *1*(4), 34-68. doi: 10.1257/app.1.4.34 [10, 13, 28]

Hanushek, E. A., Kain, J. F., Markman, J. M., & Rivkin, S. G. (2003, September). Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics*, *18*(5), 527–544. [16]

Hayduk, I., & Toussaint-Comeau, M. (n.d.). Determinants of noncognitive skills: Mediating effects of siblings' interaction and parenting quality. *Contemporary Economic Policy*, *n/a*(n/a). doi: https://doi.org/10.1111/coep.12572 [3]

Hospido, L., Laeven, L., & Lamo, A. (2022, 09). The Gender Promotion Gap: Evidence from Central Banking. *The Review of Economics and Statistics*, *104*(5), 981-996. doi: 10.1162/rest_a_00988 [4]

Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, *4*(1), 93–106. [8]

Hoxby, C. M., & Weingarth, G. (2006). *Taking race out of the equation: School reassignment and the structure of peer effects.* (Working Paper) [15, 51]

Huang, W., Liu, E. M., & Zuppann, C. A. (2022). Relative obesity and the formation of noncognitive abilities during adolescence. *Journal of Human Resources*, *57*(6), 2086–2112. [3]

Jackson, C. K., Porter, S. C., Easton, J. Q., Blanchard, A., & Kiguel, S. (2020, December). School effects on socioemotional development, school-based arrests, and educational attainment. *American Economic Review: Insights*, *2*(4), 491-508. doi: 10.1257/aeri.20200029 [3]

Jochmans, K. (2023). Testing random assignment to peer groups. *Journal of Applied Econometrics*, *n/a*(n/a). doi: https://doi.org/10.1002/jae.2953 [13, 48]

Jones, T. R., & Kofoed, M. S. (2020). Do peers influence occupational preferences? Evidence from randomly-assigned peer groups at West Point. *Journal of Public Economics*, *184*, 104154. [4]

Keller, W., Molina, T., & Olney, W. W. (2020, December). *The gender gap among top business executives* (Working Paper No. 28216). National Bureau of Economic Research. doi: 10 .3386/w28216 [1]

Kofoed, M. S., et al. (2019). The effect of same-gender or same-race role models on occupation choice evidence from randomly assigned mentors at West Point. *Journal of Human Resources*, *54*(2), 430–467. [4]

Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources*, *54*(1), 1–36. [3]

Landaud, F., Ly, S. T., & Maurin, É. (2020). Competitive schools and the gender gap in the choice of field of study. *Journal of Human Resources*, *55*(1), 278–308. [4]

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics*, *116*(3), 777–803. [15, 51]

Li, Q., Lu, X., & Ullah, A. (2003). Multivariate local polynomial regression for estimating average derivatives. *Journal of Nonparametric Statistics*, *15*(4-5), 607–624. [16]

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542. [1, 6, 10, 13]

McCluskey, J. J., & Rausser, G. C. (2001). Estimation of perceived risk and its effect on property values. *Land Economics*, *77*(1), 42–55. doi: 10.2307/3146979 [4, 15]

Neidell, M., & Waldfogel, J. (2010). Cognitive and noncognitive peer effects in early education. *The Review of Economics and Statistics*, *92*(3), 562–576. [3]

Orr, C. (2021). Peer gender composition and non-cognitive factors. *Applied Economics Letters*, *0*(0), 1-4. doi: 10.1080/13504851.2021.1991560 [4]

Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far? In *Handbook of the Economics of Education* (Vol. 3, pp. 249–277). Elsevier. [15, 51]

Shure, N. (2021). Non-cognitive peer effects in secondary education. *Labour Economics*, *73*, 102074. doi: https://doi.org/10.1016/j.labeco.2021.102074 [3]

Sibthorp, J., Paisley, K., & Gookin, J. (2007, January). Exploring participant development through adventure-based programming: A model from the National Outdoor Leadership School. *Leisure Sciences*, *29*(1), 1–18. [7, 8]

Sloan, F. A., McCutchan, S. A., & Eldred, L. M. (2017). Alcohol-impaired driving and perceived risks of legal consequences. *Alcoholism: Clinical and Experimental Research*, *41*(2), 432-442. doi: https://doi.org/10.1111/acer.13298 [4, 15]

Viscusi, W. K. (1991). Age variations in risk perceptions and smoking decisions. *The Review of Economics and Statistics*, *73*(4), 577–588. Retrieved from [2022-10-27]http://www.jstor.org/stable/2109396 [4, 15]

Wang, Y. (2014, January). Dynamic implications of subjective expectations: Evidence from adult smokers. *American Economic Journal: Applied Economics*, *6*(1), 1-37. doi: 10.1257/app.6.1.1 [4, 15]

Weinberger, C. J. (2014). The increasing complementarity between cognitive and social skills. *Review of Economics and Statistics*, *96*(4), 849–861. [1]

Whiting, K. (2020). *These are the top 10 job skills of tomorrow – and how long it takes to learn them.* https://www.weforum.org/agenda/2020/10/top-10-work-skills-of-tomorrow-how-long-it-takes-to-learn-them/. (Accessed: January 2021) [1]

Wiswall, M., & Zafar, B. (2017, 08). Preference for the Workplace, Investment in Human Capital, and Gender. *The Quarterly Journal of Economics*, *133*(1), 457-507. doi: 10.1093/qje/qjx035 [4]

Online Appendix

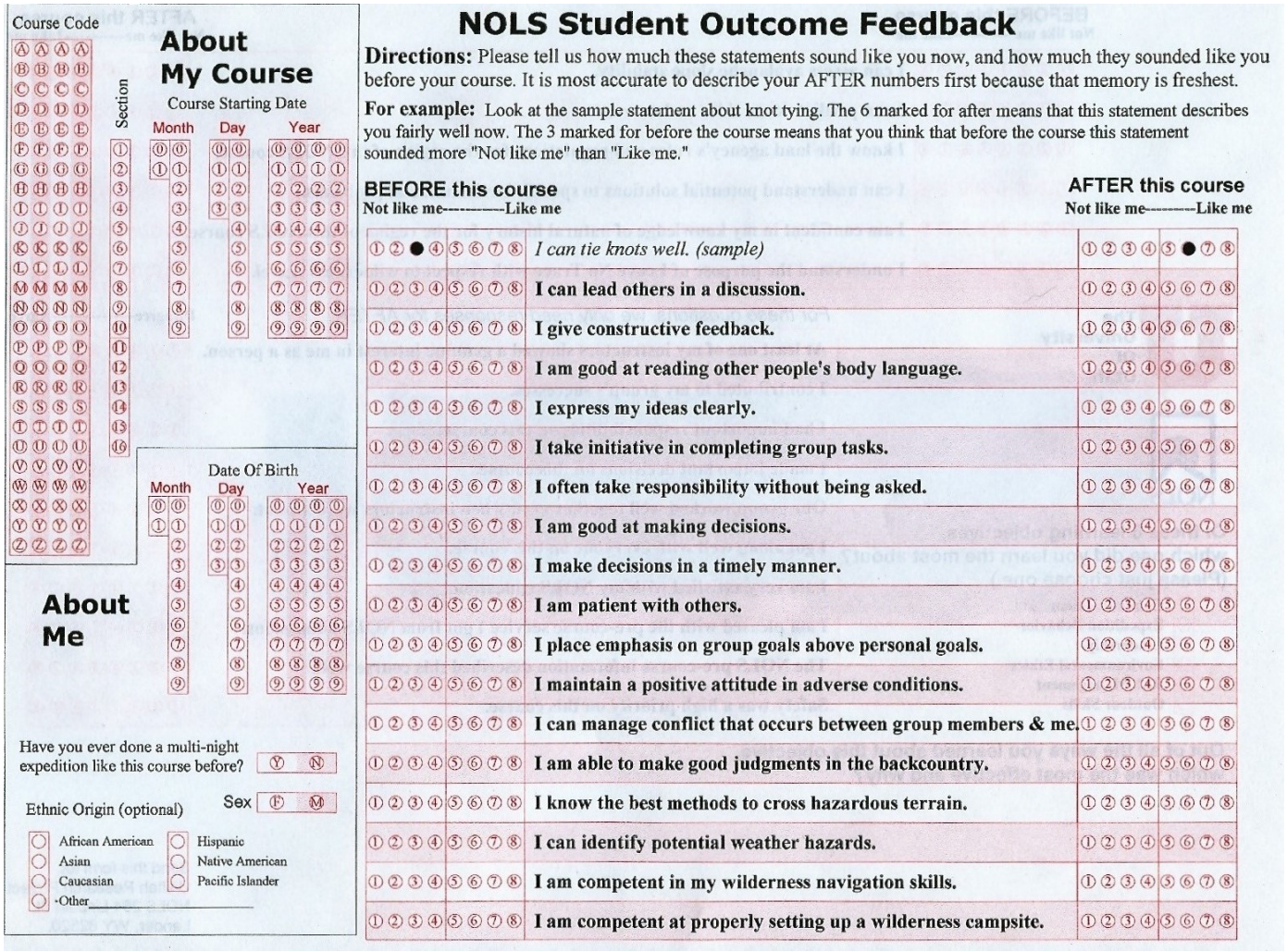
A Additional Tables and Figures

Table A.1: Composition of Noncognitive Ability Measures

Students are asked to rate how well the following statements describe them. Composite ratings for each skill are computed as the mean of the responses in each category.
COMMUNICATION
1. I can lead others in a discussion.
2. I give constructive feedback.
3. I am good at reading other people's body language.
4. I express my ideas clearly
LEADERSHIP
1. I take initiative in completing group tasks
2. I often take responsibility without being asked
3. I am good at making decisions
4. I make decisions in a timely manner
ATTITUDE
1. I am patient with others
2. I place emphasis on group goals above personal goals
3. I maintain a positive attitude in adverse conditions
4. I can manage conflict that occurs between group members

Note: These questions are chosen to measure targeted outcomes in the NOLS curriculm. To each question, students rated themselves on a scale of 1 (worst) to 8 (best). The survey underwent several pilot trials prior to being implemented to ensure that the survey questions adequately and consistently measure the associated abilities. Surveys were administered within two hours of students returning from the field, typically at the end of the course. For semester-length courses, surveys were conducted during a convenient break between parts close to the 30-day time frame of the typical NOLS course. At the same time these responses were collected, students also *retrospectively* reported their skills as they were just before the course began. This retrospective structure was chosen to avoid response-shift bias, a serious problem observed in pilot trials and in many surveys of pre-treatment levels of instruments.

Figure A.1: Survey Taken by NOLS Students



Course Code

Section

About My Course

Course Starting Date

Month Day Year

Date Of Birth

Month Day Year

About Me

Have you ever done a multi-night expedition like this course before? Y N

Sex F M

Ethnic Origin (optional)

- African American
- Asian
- Caucasian
- Other_____
- Hispanic
- Native American
- Pacific Islander

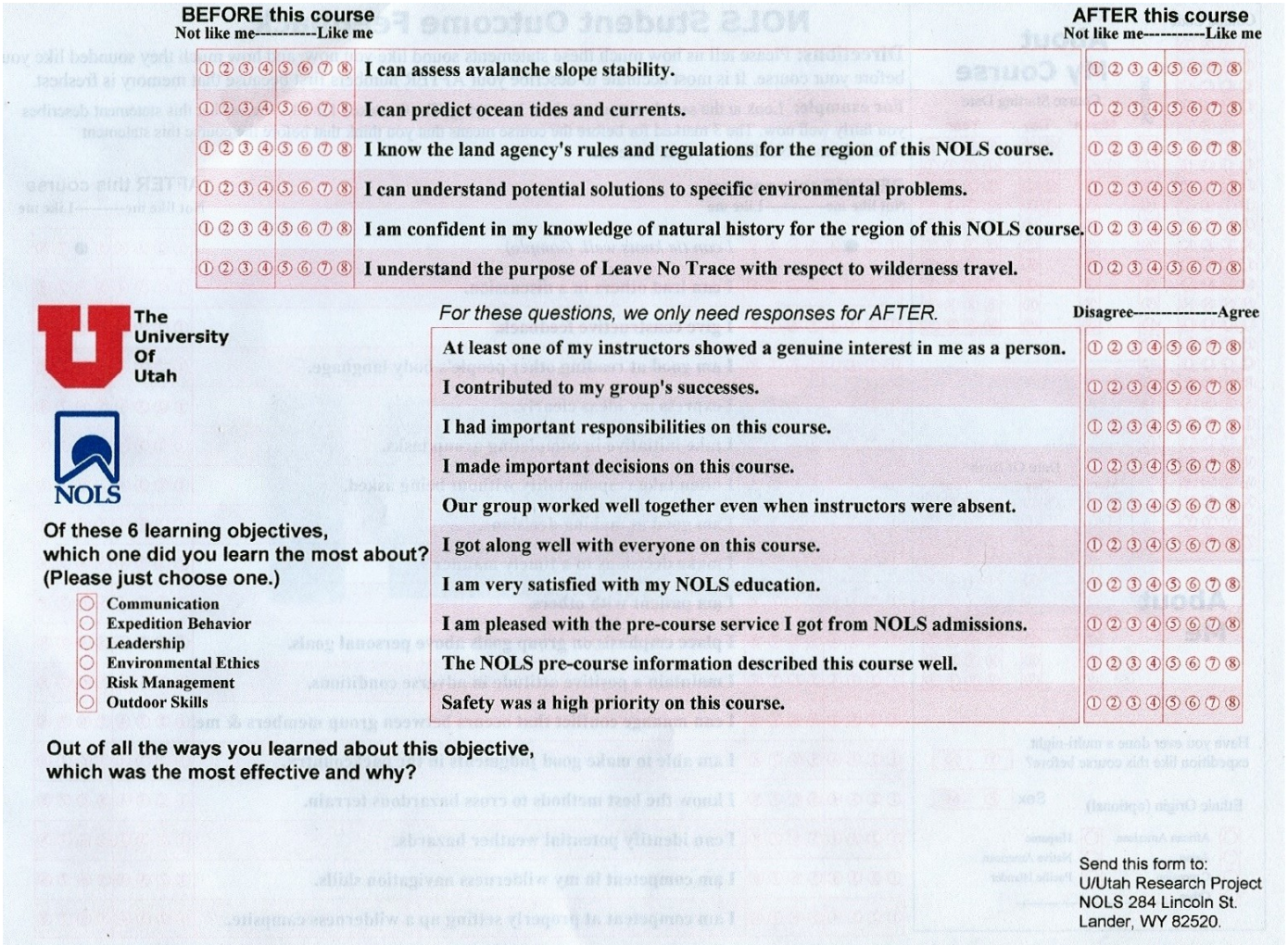
NOLS Student Outcome Feedback

Directions: Please tell us how much these statements sound like you now, and how much they sounded like you before your course. It is most accurate to describe your AFTER numbers first because that memory is freshest.

For example: Look at the sample statement about knot tying. The 6 marked for after means that this statement describes you fairly well now. The 3 marked for before the course means that you think that before the course this statement sounded more "Not like me" than "Like me."

BEFORE this course Not like me----------Like me		AFTER this course Not like me---------Like me
① ② ● ④ ⑤ ⑥ ⑦ ⑧	*I can tie knots well. (sample)*	① ② ③ ④ ⑤ ● ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I can lead others in a discussion.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I give constructive feedback.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I am good at reading other people's body language.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I express my ideas clearly.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I take initiative in completing group tasks.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I often take responsibility without being asked.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I am good at making decisions.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I make decisions in a timely manner.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I am patient with others.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I place emphasis on group goals above personal goals.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I maintain a positive attitude in adverse conditions.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I can manage conflict that occurs between group members & me.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I am able to make good judgments in the backcountry.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I know the best methods to cross hazardous terrain.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I can identify potential weather hazards.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I am competent in my wilderness navigation skills.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧
① ② ③ ④ ⑤ ⑥ ⑦ ⑧	**I am competent at properly setting up a wilderness campsite.**	① ② ③ ④ ⑤ ⑥ ⑦ ⑧

Figure A.2: Survey Taken by NOLS Students, Cont.



BEFORE this course
Not like me----------Like me

AFTER this course
Not like me----------Like me

①②③④⑤⑥⑦⑧ I can assess avalanche slope stability.	①②③④⑤⑥⑦⑧
①②③④⑤⑥⑦⑧ I can predict ocean tides and currents.	①②③④⑤⑥⑦⑧
①②③④⑤⑥⑦⑧ I know the land agency's rules and regulations for the region of this NOLS course.	①②③④⑤⑥⑦⑧
①②③④⑤⑥⑦⑧ I can understand potential solutions to specific environmental problems.	①②③④⑤⑥⑦⑧
①②③④⑤⑥⑦⑧ I am confident in my knowledge of natural history for the region of this NOLS course.	①②③④⑤⑥⑦⑧
①②③④⑤⑥⑦⑧ I understand the purpose of Leave No Trace with respect to wilderness travel.	①②③④⑤⑥⑦⑧

The University Of Utah

NOLS

For these questions, we only need responses for AFTER.

Disagree----------Agree

At least one of my instructors showed a genuine interest in me as a person.	①②③④⑤⑥⑦⑧
I contributed to my group's successes.	①②③④⑤⑥⑦⑧
I had important responsibilities on this course.	①②③④⑤⑥⑦⑧
I made important decisions on this course.	①②③④⑤⑥⑦⑧
Our group worked well together even when instructors were absent.	①②③④⑤⑥⑦⑧
I got along well with everyone on this course.	①②③④⑤⑥⑦⑧
I am very satisfied with my NOLS education.	①②③④⑤⑥⑦⑧
I am pleased with the pre-course service I got from NOLS admissions.	①②③④⑤⑥⑦⑧
The NOLS pre-course information described this course well.	①②③④⑤⑥⑦⑧
Safety was a high priority on this course.	①②③④⑤⑥⑦⑧

Of these 6 learning objectives, which one did you learn the most about? (Please just choose one.)

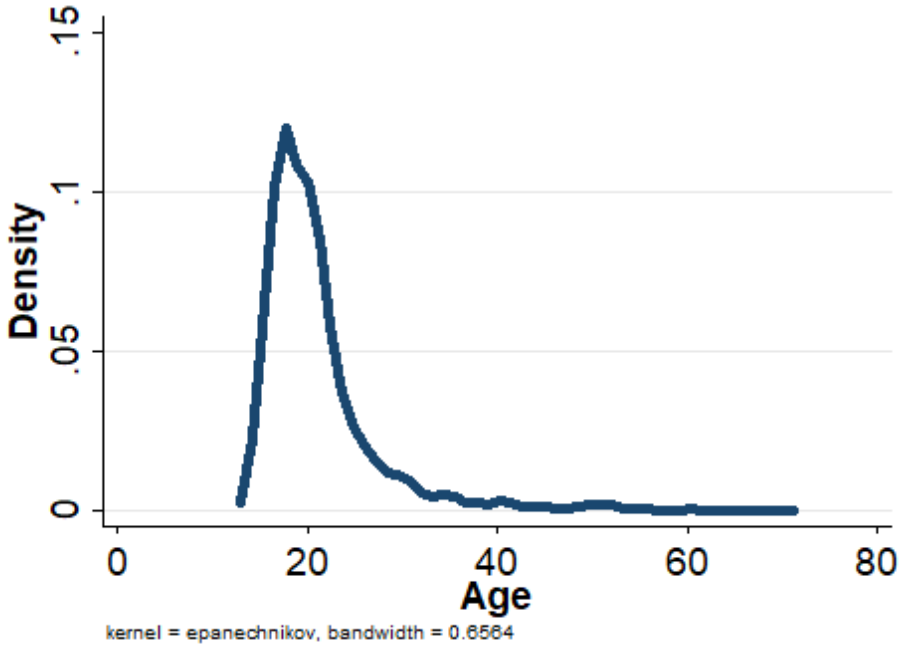
- ○ Communication
- ○ Expedition Behavior
- ○ Leadership
- ○ Environmental Ethics
- ○ Risk Management
- ○ Outdoor Skills

Out of all the ways you learned about this objective, which was the most effective and why?

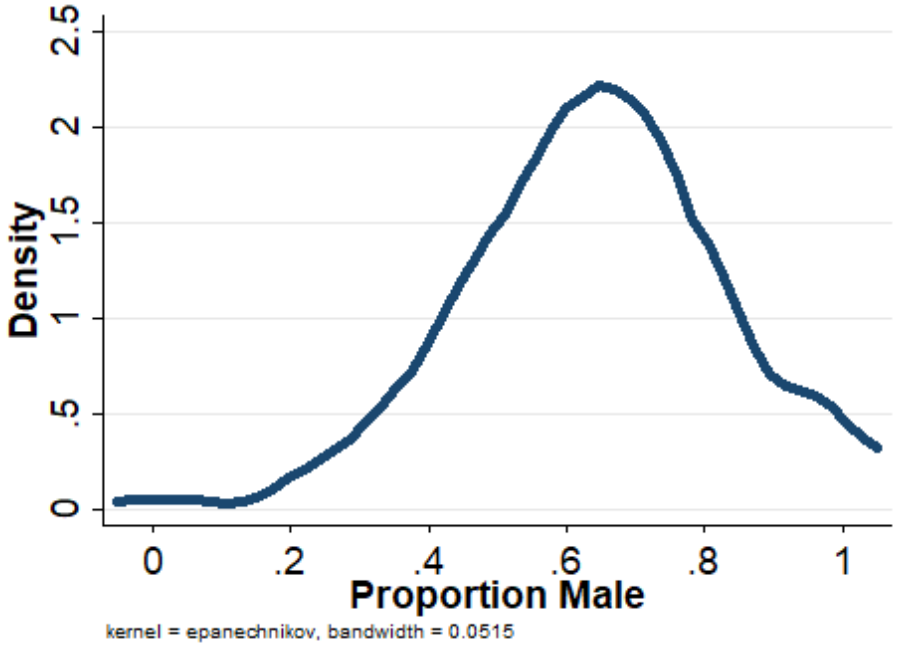
Send this form to:
U/Utah Research Project
NOLS 284 Lincoln St.
Lander, WY 82520.

Figure A.3: Distributions of Age and Proportion Male

(A): Age

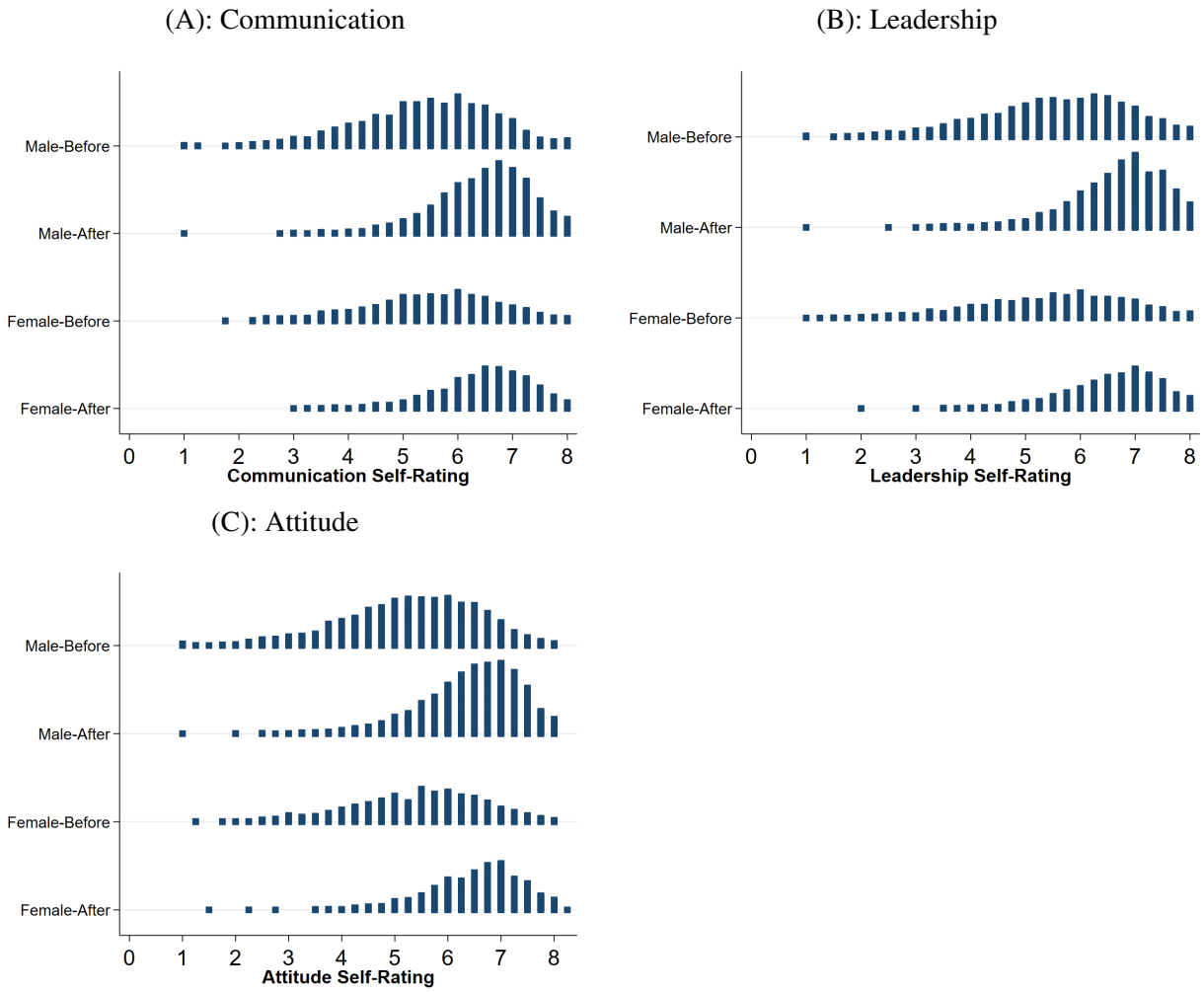


(B): Proportion Male



Note: Figures show kernel densities. Figure (A) shows kernel density of age, N = 3,773. Figure (B) shows the density of the proportion male by section, N = 358. There is bump in the right tail of (B), because NOLS makes efforts to ensure that females are not alone in sections.

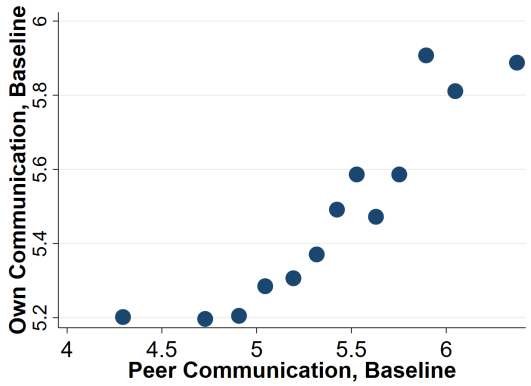
Figure A.4: Distribution of Self-Perception by Skill, Time, and Gender



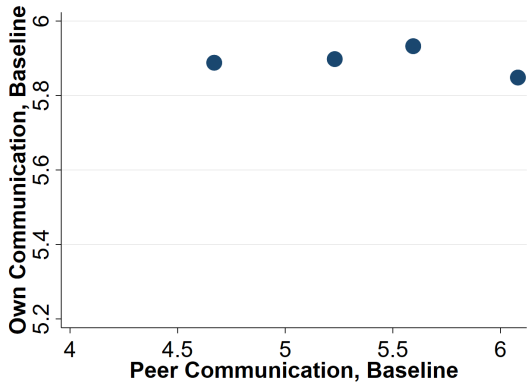
Note: Figures show the distribution of non-cognitive skills by gender and before or after the course. (a) shows composite communication. (b) shows composite leadership. (c) shows composite attitude.

Figure A.5: Correlation Between Baseline Individual Ability and Baseline Peer Average Ability

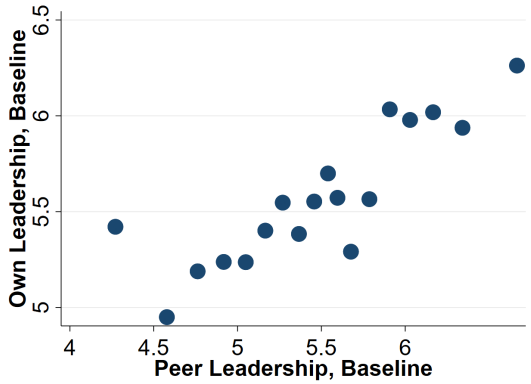
(A): Communication, Unconditional



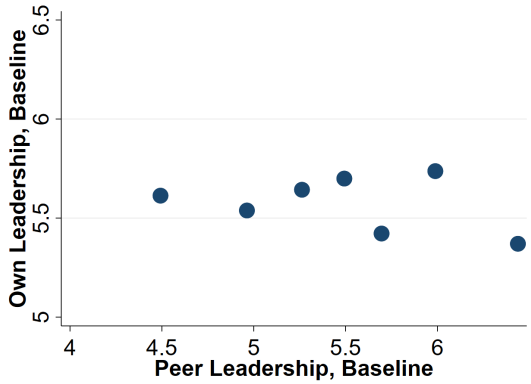
(B): Communication, Course Indicators



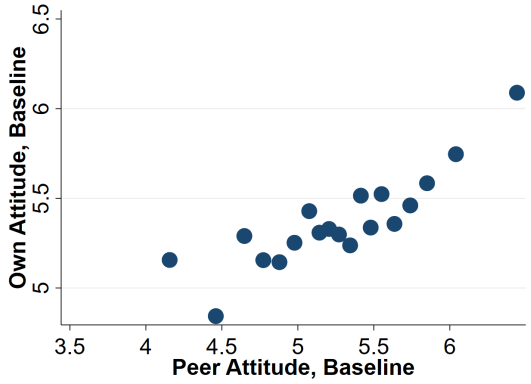
(C): Leadership, Unconditional



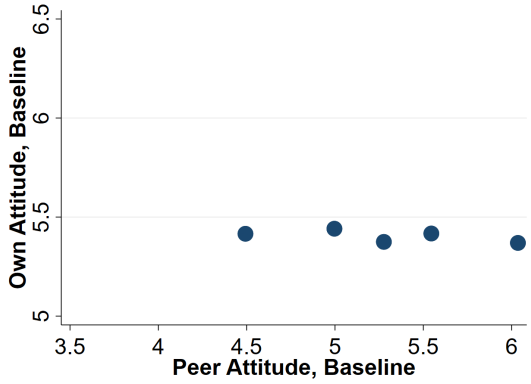
(D): Leadership, Course Indicators



(E): Attitude, Unconditional



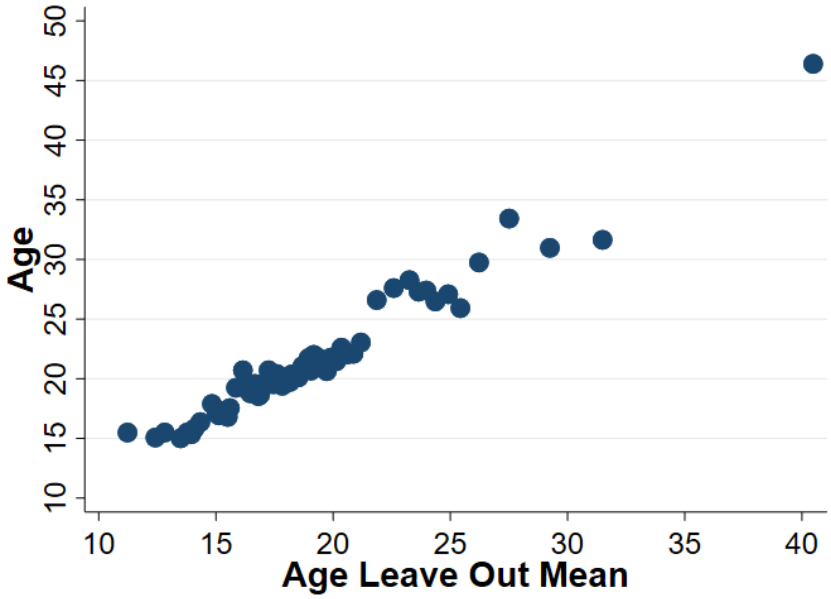
(F): Attitude, Course Indicators



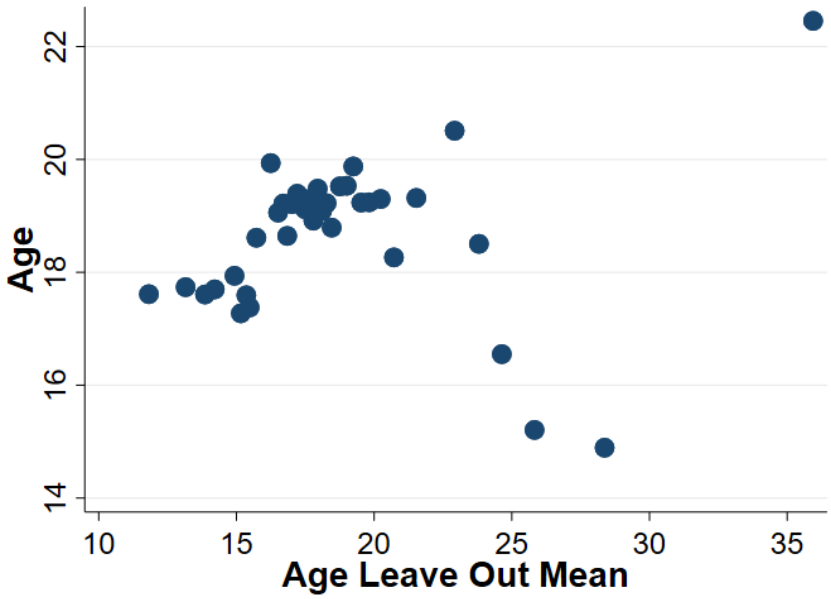
Note: Y-axis is baseline individual composite ability. X-axis is baseline peer average composite ability. Number of bins is optimal for variance-bias tradeoff (for integrated mean square error), as defined by Cattaneo et al. (2019). Panel A, C, and E are conditional on peer group average and own ability before the course. Panels B, D, and F include course indicators, gender, race, and age as covariates. All figures control for individual baseline ability. $3{,}702 \leq N \leq 3709$.

Figure A.6: Binned Scatter Conditional Randomization Checks for Age

(A): Group Age on Age, Unconditional

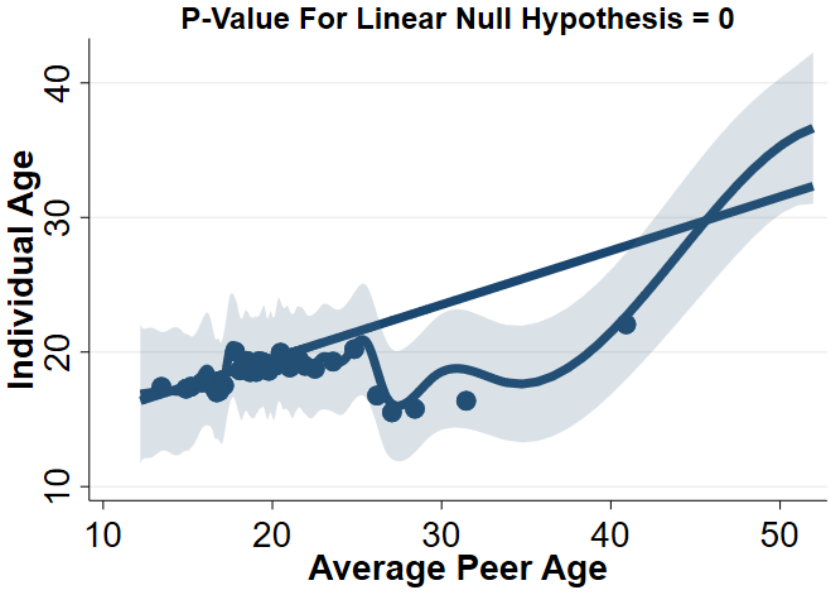


(B): Group Age on Age, Course Indicators



Note: N = 3,713 in all figures. Scales differ. Means conditional only on peer group mean. Optimal bins for variance-bias tradeoff (for integrated mean square error) defined by Cattaneo et al. (2019). Panel A is unconditional binned scatter. Panels B includes course indicators as covariates.

Figure A.7: Age on Meanage with Cubic B-Spline



Note: N = 3,773. Fit and confidence bands are from a cubic B-spline. The hypothesis test is a null of global linearity. To evaluate the hypothesis, it compares the linear functional form to a cubic b-spline.

Table A.2: Relationship Between Peer Average Age and Own Age (Within Mean Age Bins)

	(1)
Age Leave-Out Mean	0.389***
	(0.135)
meanage $<16 \times$ Age Leave-Out Mean	0.012
	(0.061)
$16 <$ meanage $< 22 \times$ Age Leave-Out Mean	0.049
	(0.045)
Observations	3706
Course Indicators	X

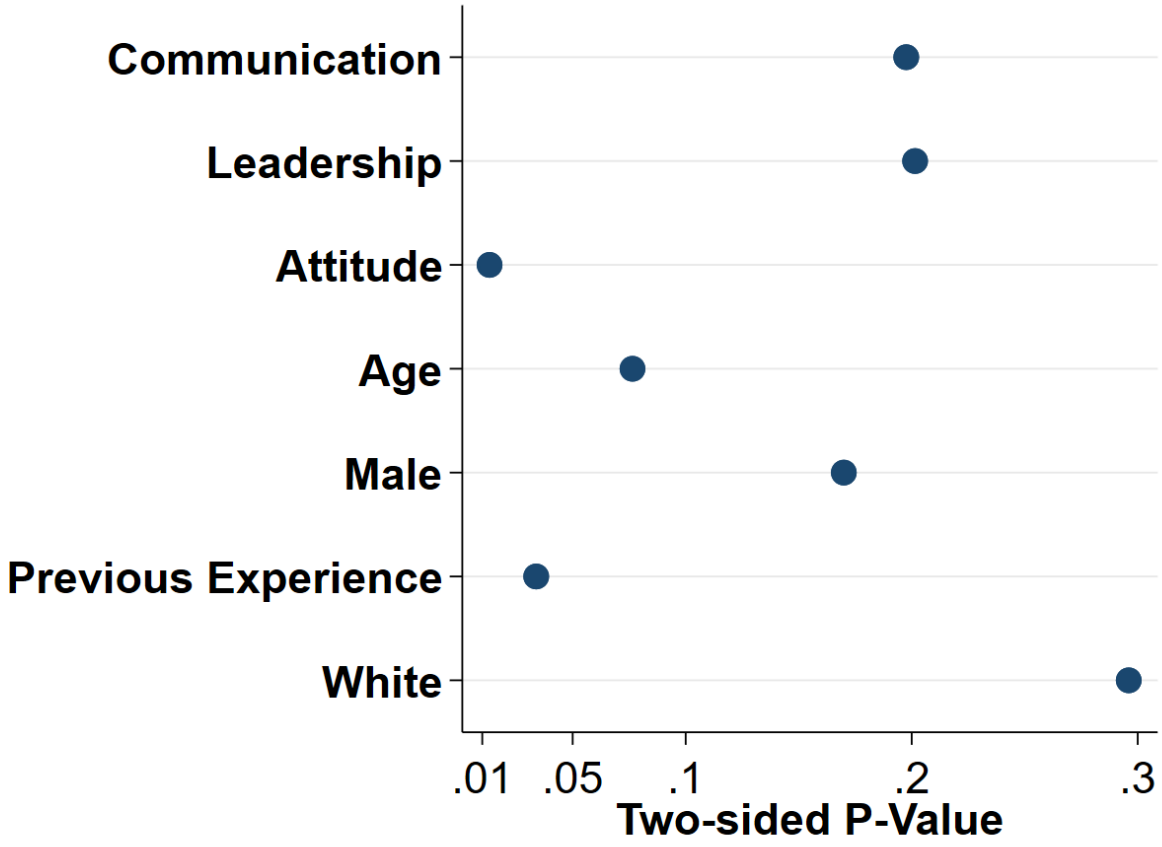
Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Cluster-robust standard errors, by courseid, in parentheses. Age Leave-Out Mean is peer average age. Meanage < 16 equals 1 if group average age is below 16, because NOLS has courses that are restricted to only 14-15 year olds. So meanage < 16 X Age Leave-Out Mean is a coefficient for effect of peer age on own age for peer groups below age 16. $16 <$ meanage < 22 is a dummy variable equalling 1 if the group average age is between 16 and 22, because these are the most common age of NOLS students (see Section 2).

Table A.3: Conditional Randomization, Urn Averages

Panel A: Abilities	(1)	(2)	(3)	
Pre-Course Communication Leave-Out Mean	0.12**			
	(0.05)			
Course Comm. LO Avg.	-17.62***			
	(1.56)			
Pre-Course Leadership Leave-Out Mean		0.05		
		(0.05)		
Course Lead LO Avg.		-17.79***		
		(1.48)		
Pre-Course Attitude Leave-Out Mean			0.11**	
			(0.04)	
Course Att. LO Avg.			-15.65***	
			(1.42)	
Observations	3709	3708	3702	
R^2	0.434	0.456	0.423	
Panel B: Other	(1)	(2)	(3)	(4)
LO Proportion White	0.07			
	(0.07)			
Course White LO Avg.	-12.82***			
	(1.95)			
LO Proportion Prev. Exp.		0.00		
		(0.00)		
Course Prev. Exp. LO Avg.		1.00***		
		(0.00)		
Age Leave-Out Mean			0.44***	
			(0.06)	
Course Age LO Avg.			-9.43***	
			(1.35)	
LO Proportion Male				0.18***
				(0.06)
Course Male LO Avg.				-15.36***
				(1.48)
Observations	3403	3773	3706	3773
R^2	0.370	1	0.782	0.397

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS.

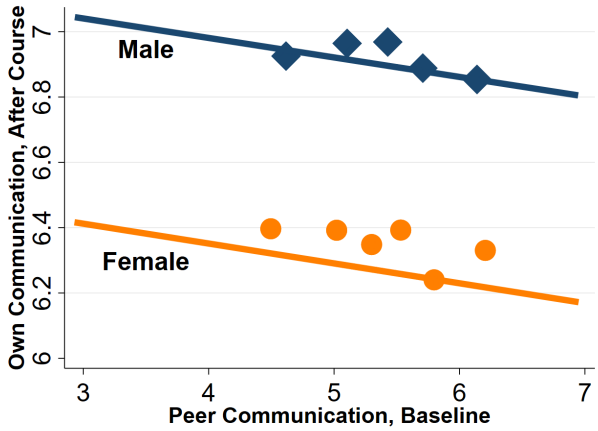
Figure A.8: Jochman Test for Conditional Random Assignment P-Values



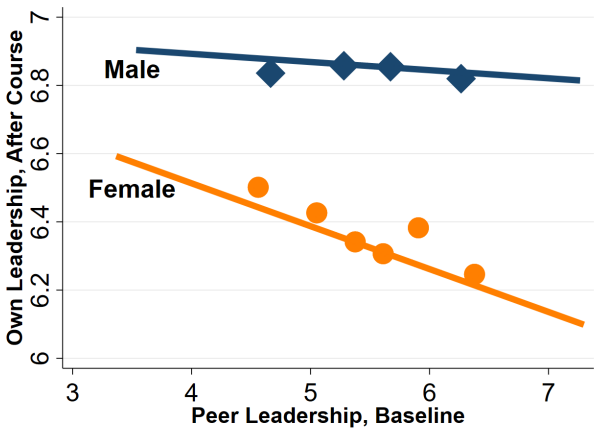
Note: Figure shows two-sided p-values from Jochmans (2023). The null is conditional random assignment.

Figure A.9: How Leave One Out Peer Baseline Ability Affects Own Ability, Optimal Number of Bins

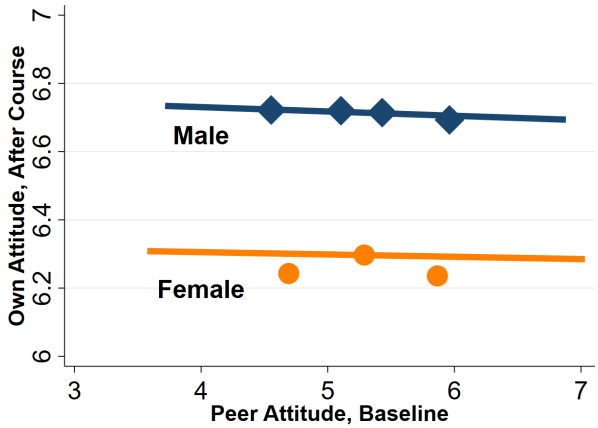
(A): Communication



(B): Leadership



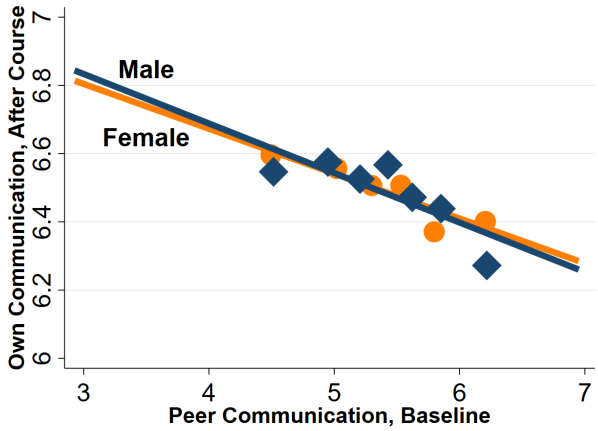
(C): Attitude



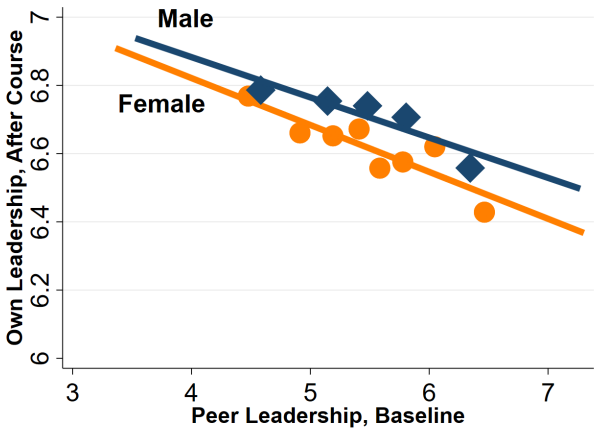
Note: All figures show the correlation between baseline leave-one out peer ability and own post-course ability. The correlations control for course type indicators and baseline own ability using methods from Cattaneo et al. (2019). The x-axis is peer leave one-out baseline ability. The y-axis is own post-course ability. Panel (A) shows communication, Panel (B) shows leadership, and Panel (C) shows attitude. The ability of each individual in each category is an average of 4 questions that are answered on a scale of 1-8. The number of bins in each optimally trades off bias and variance according to Cattaneo et al. (2019).

Figure A.10: How Leave One Out Peer Baseline Ability Affects Own Ability, No Course Type FE's

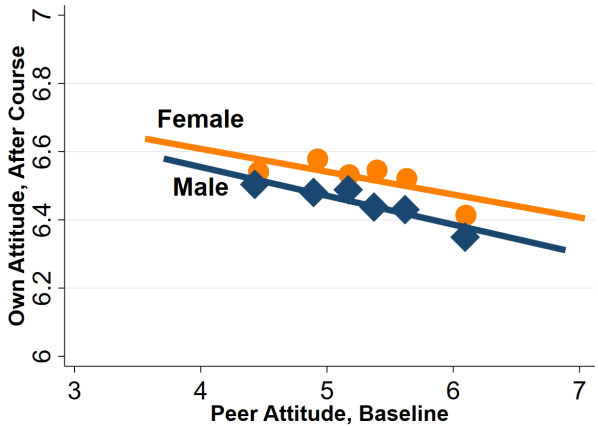
(A): Communication



(B): Leadership



(C): Attitude



Note: All figures show the correlation between baseline leave-one out peer ability and own post-course ability. The correlations control baseline own ability using methods from Cattaneo et al. (2019). The x-axis is peer leave one-out baseline ability. The y-axis is own post-course ability. Panel (A) shows communication, Panel (B) shows leadership, and Panel (C) shows attitude. The ability of each individual in each category is an average of 4 questions that are answered on a scale of 1-8. The number of bins in each optimally trades off bias and variance according to Cattaneo et al. (2019). Blue diamonds are male and orange dots are female.

Table A.4: Models of Peer Effects

Model	Homogenous Effects	Description
Linear-in-means	Yes	Only the mean of peers background matters
Bad apple	Yes	One disruptive individual harms everyone
Shining light	Yes	One excellent student provides great example
Invidious Comparison	No	Outcomes harmed by presence of better achieving peers
Boutique/ Tracking	No	Students perform best when surrounded by others like themselves
Focus	Yes	Classroom homogeneity is good, regardless of student i's ability relative to homogenous class
Rainbow	Yes	Classroom heterogeneity good for everyone
Single crossing	No	Positive effects from high ability classmate is weakly monotonically increasing in own ability

Note: Table reproduced from Sacerdote (2011). Models from Lazear (2001) and Hoxby and Weingarth (2006).

Table A.5: Descriptive Statistics on Raw Questions

Panel A: Before						
	Mean	Median	SD	Min	Max	Count
Communication, Q1	5.44	6.00	1.63	1.0	8.0	3915
Communication, Q2	5.05	5.00	1.62	1.0	8.0	3917
Communication, Q3	5.91	6.00	1.50	1.0	8.0	3918
Communication, Q4	5.52	6.00	1.54	1.0	8.0	3920
Leadership, Q1	5.67	6.00	1.59	1.0	8.0	3918
Leadership, Q2	5.75	6.00	1.62	1.0	8.0	3909
Leadership, Q3	5.55	6.00	1.55	1.0	9.0	3922
Leadership, Q4	5.31	5.00	1.65	1.0	9.0	3910
Attitude, Q1	5.28	5.00	1.77	1.0	8.0	3913
Attitude, Q2	5.09	5.00	1.56	1.0	8.0	3918
Attitude, Q3	5.77	6.00	1.64	1.0	8.0	3919
Attitude, Q4	5.36	5.00	1.50	1.0	8.0	3905
Panel B: After						
	Mean	Median	SD	Min	Max	Count
Communication, Q1	6.54	7.00	1.15	1.0	8.0	3911
Communication, Q2	6.43	7.00	1.15	1.0	8.0	3920
Communication, Q3	6.51	7.00	1.23	1.0	8.0	3916
Communication, Q4	6.46	7.00	1.13	1.0	8.0	3923
Leadership, Q1	6.89	7.00	1.05	1.0	8.0	3919
Leadership, Q2	6.80	7.00	1.07	1.0	8.0	3922
Leadership, Q3	6.64	7.00	1.08	1.0	8.0	3925
Leadership, Q4	6.39	7.00	1.20	1.0	8.0	3922
Attitude, Q1	6.36	7.00	1.36	1.0	8.0	3914
Attitude, Q2	6.23	6.00	1.26	1.0	8.0	3921
Attitude, Q3	6.89	7.00	1.14	1.0	8.0	3919
Attitude, Q4	6.41	7.00	1.17	1.0	9.0	3917

Note: Table shows descriptive statistics for individual survey answers without any aggregation across skills or peer groups. Survey questions are shown in Table A.1.

Table A.6: The Effect of Baseline Peer Ability on Own Noncognitive Ability, By Individual Question

Panel A: Communication	Comm 1		Comm 2		Comm 3		Comm 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peer Comm. 1	-0.14***	-0.11***						
	(0.02)	(0.03)						
Peer Comm. 2			-0.12***	-0.09**				
			(0.03)	(0.04)				
Peer Comm. 3					-0.03	-0.02		
					(0.03)	(0.03)		
Peer Comm. 4							-0.18***	-0.11***
							(0.03)	(0.03)
Observations	3720	3316	3727	3322	3725	3318	3729	3325
R-Sq	0.496	0.524	0.376	0.412	0.615	0.640	0.471	0.501
Individual Covariates		X		X		X		X
Course Indicators		X		X		X		X
Panel B: Leadership	Lead 1		Lead 2		Lead 3		Lead 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peer Lead. 1	-0.11***	-0.04						
	(0.02)	(0.03)						
Peer Lead. 2			-0.08***	-0.07**				
			(0.02)	(0.03)				
Peer Lead. 3					-0.14***	-0.07**		
					(0.02)	(0.03)		
Peer Lead. 4							-0.08***	-0.01
							(0.02)	(0.03)
Observations	3726	3321	3728	3322	3735	3329	3723	3323
R-Sq	0.376	0.404	0.422	0.452	0.462	0.492	0.483	0.507
Individual Covariates		X		X		X		X
Course Indicators		X		X		X		X
Panel C: Attitude	Att 1		Att 2		Att 3		Att 4	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Peer Att. 1	-0.06**	-0.03						
	(0.03)	(0.03)						
Peer Att. 2			-0.06**	-0.00				
			(0.03)	(0.04)				
Peer Att. 3					-0.07***	-0.04		
					(0.03)	(0.03)		
Peer Att. 4							-0.06**	0.00
							(0.03)	(0.03)
Observations	3719	3316	3731	3327	3729	3327	3720	3321
R-Sq	0.454	0.472	0.410	0.434	0.403	0.437	0.422	0.463
Individual Covariates		X		X		X		X
Course Indicators		X		X		X		X

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual communication for a specific question and the independent variable is pre-course peer average communication for the same question. In Panel B, the dependent variable is post-course individual leadership for a single question and the independent variable is pre-course peer average leadership for that same question. In Panel C, the dependent variable is post-course individual attitude for a single question and the independent variable is pre-course peer average attitude for the same question. All columns control for the individual's pre-course ability for the outcome of interest in that panel. The individual covariates include age, race, and previous experience. The course indicators include course type, section year, and section month.

Table A.7: Effect of Baseline Peer Ability on Own Ability, Variance Controls

Panel A: Communication	(1) All	(2) Male	(3) Female
Baseline Peer Average Comm.	-0.07**	-0.07*	-0.10**
	(0.03)	(0.04)	(0.04)
Baseline Peer Std. Dev. Comm.	0.03	0.01	0.01
	(0.05)	(0.06)	(0.07)
Observations	3261	2109	1152
R^2	0.522	0.497	0.614
Panel B: Leadership	(1) All	(2) Male	(3) Female
Baseline Peer Average Lead	-0.06**	-0.03	-0.16***
	(0.03)	(0.04)	(0.04)
Baseline Peer Std. Dev. Lead	0.04	0.05	-0.01
	(0.04)	(0.05)	(0.07)
Observations	3282	2122	1160
R^2	0.488	0.480	0.549
Panel C: Attitude	(1) All	(2) Male	(3) Female
Baseline Peer Average Att.	-0.01	-0.00	-0.04
	(0.04)	(0.04)	(0.05)
Baseline Peer Std. Dev. Att	-0.05	-0.06	-0.04
	(0.06)	(0.07)	(0.09)
Observations	3274	2122	1152
R^2	0.454	0.451	0.504

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual composite communication and the independent variable is pre-course peer average composite communication. In Panel B, the dependent variable is post-course individual composite leadership and the independent variable is pre-course peer average composite leadership. In Panel C, the dependent variable is post-course individual composite attitude and the independent variable is pre-course peer average composite attitude. All columns include control for the individual's pre-course ability for the outcome of interest in that panel, course type indicators, and month and year fixed effects. Column 1 uses the full sample, Column 2 uses only men, and Column 3 use only women.

Table A.8: Non-Parametric Kernel Regression of Peer Group on Noncognitive After

Panel A: Communication		
	(1)	(2)
Pre Comm. Leave-Out Mean	-0.14***	-0.11***
	(0.02)	(0.02)
Observations	3658	3265
R-Sq	0.514	0.495
Age Covariate		X
Male & White		X
Course Indicators		X
Panel B: Leadership		
	(1)	(2)
Pre Lead Leave-Out Mean	-0.13***	-0.10***
	(0.02)	(0.02)
Observations	3682	3266
R-Sq	0.484	0.566
Age Covariate		X
Male & White		X
Course Indicators		X
Panel C: Attitude		
	(1)	(2)
Pre Att. Leave-Out Mean	-0.08***	-0.04*
	(0.02)	(0.02)
Observations	3668	3278
R-Sq	0.449	0.429
Age Covariate		X
Male & White		X
Course Indicators		X

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are bootstrapped for 100 replications. Column 4 uses slightly less replications. All columns include a control for pre-course noncognitive ability. Column 2 adds continuous age. Column 3 adds male and white binary variables. Column 4 includes course type indicators.

Table A.9: Heterogeneity in Peer Effects by Demographics

Panel A: Communication				
	Age		Previous Experience	
	(1)	(2)	(3)	(4)
	$>$ Median	$\leq$ Median	Yes	No
Baseline Peer Average Comm.	-0.09**	-0.07*	-0.11***	-0.05
	(0.04)	(0.04)	(0.04)	(0.04)
Observations	1633	1632	1725	1540
R^2	0.585	0.481	0.544	0.530
Panel B: Leadership				
	Age		Previous Experience	
	(1)	(2)	(3)	(4)
	$>$ Median	$\leq$ Median	Yes	No
Baseline Peer Average Lead	-0.10***	-0.04	-0.04	-0.10**
	(0.03)	(0.04)	(0.04)	(0.04)
Observations	1647	1639	1737	1549
R^2	0.537	0.460	0.509	0.496
Panel C: Attitude				
	Age		Previous Experience	
	(1)	(2)	(3)	(4)
	$>$ Median	$\leq$ Median	Yes	No
Baseline Peer Average Att.	-0.01	-0.00	0.01	-0.00
	(0.04)	(0.05)	(0.04)	(0.05)
Observations	1640	1638	1733	1545
R^2	0.494	0.440	0.474	0.462

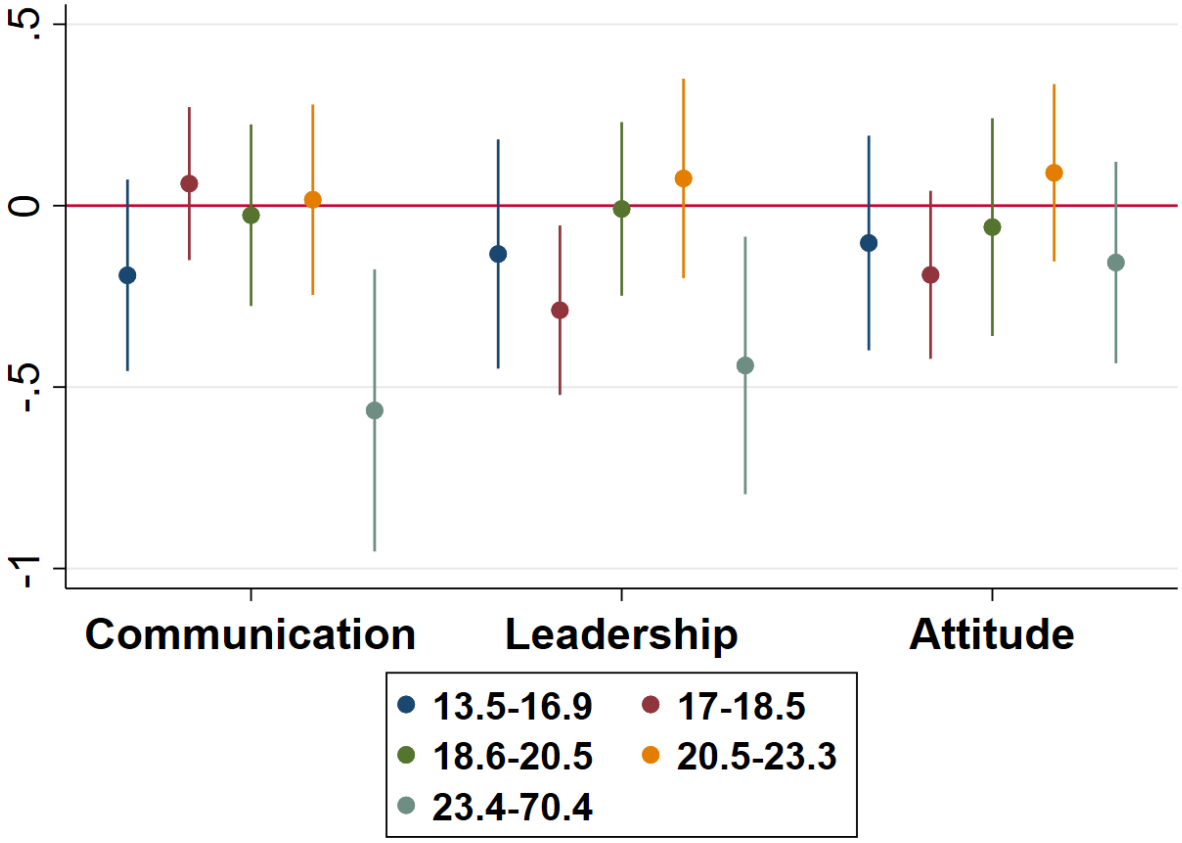
Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual composite communication and the independent variable is pre-course peer average composite communication. In Panel B, the dependent variable is post-course individual composite leadership and the independent variable is pre-course peer average composite leadership. In Panel C, the dependent variable is post-course individual composite attitude and the independent variable is pre-course peer average composite attitude. All columns control for the individual's pre-course ability for the outcome of interest in that panel. Additional controls include age (in days), race indicators, and gender. All regressions also include course indicators, section year, and section month fixed effects. The median age is 19.58 years.

Table A.10: Heterogeneous Effects of Peer Non-Cognitive Ability on Own Non-Cognitive Ability by Course Characteristics

Panel A: Communication	Course Members		Course Length	
	(1)	(2)	(3)	(4)
	$>$ Median	$\leq$ Median	$>$ Median	$\leq$ Median
Baseline Peer Average Comm.	-0.15***	-0.05	0.02	-0.10***
	(0.05)	(0.04)	(0.07)	(0.03)
Observations	1336	1929	1065	2058
R^2	0.511	0.549	0.479	0.544
Panel B: Leadership	Course Members		Course Length	
	(1)	(2)	(3)	(4)
	$>$ Median	$\leq$ Median	$>$ Median	$\leq$ Median
Baseline Peer Average Lead	-0.12***	-0.03	-0.09*	-0.05
	(0.04)	(0.04)	(0.05)	(0.03)
Observations	1334	1952	1077	2065
R^2	0.516	0.484	0.468	0.487
Panel C: Attitude	Course Members		Course Length	
	(1)	(2)	(3)	(4)
	$>$ Median	$\leq$ Median	$>$ Median	$\leq$ Median
Baseline Peer Average Att.	-0.05	0.01	-0.02	0.01
	(0.06)	(0.04)	(0.05)	(0.04)
Observations	1339	1939	1066	2067
R^2	0.469	0.456	0.431	0.462

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. In Panel A, the dependent variable is post-course individual composite communication and the independent variable is pre-course peer average composite communication. In Panel B, the dependent variable is post-course individual composite leadership and the independent variable is pre-course peer average composite leadership. In Panel C, the dependent variable is post-course individual composite attitude and the independent variable is pre-course peer average composite attitude. All columns control for the individual's pre-course ability for the outcome of interest in that panel. Additional controls in all columns include age (in days), race indicators, and gender. All regressions also include course indicators, section year, and section month fixed effects. The median number of course members is 11 and the median number of days is 30.

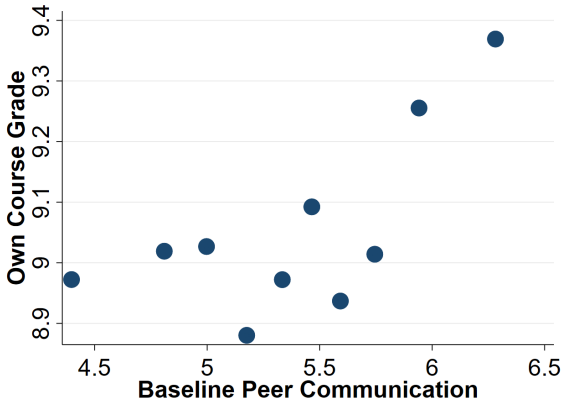
Figure A.11: Heterogeneous Treatment Effects by Age Quintile



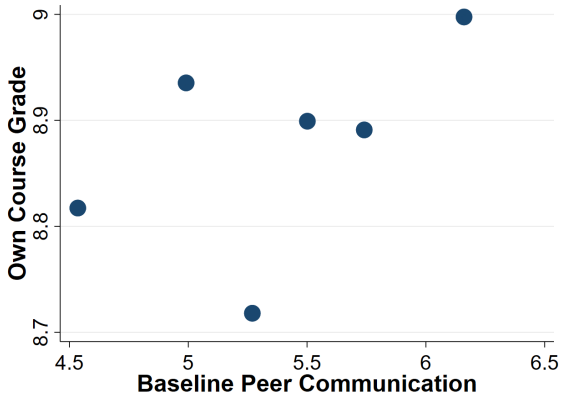
Note: The dots show point estimates from estimating Equation 1 on different age quintiles. The whiskers represent 95% confidence intervals, in which the standard errors are clustered at the course section level. The ages in each quintile in years are shown in the legend.

Figure A.12: Binned Scatter of Baseline Peer Ability Effects on Post Course Grade

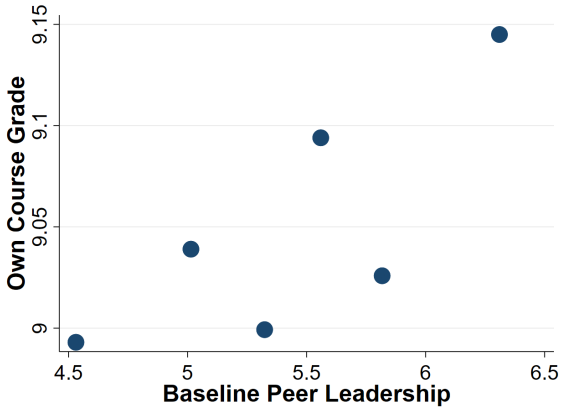
(A): Peer Communication and Grade



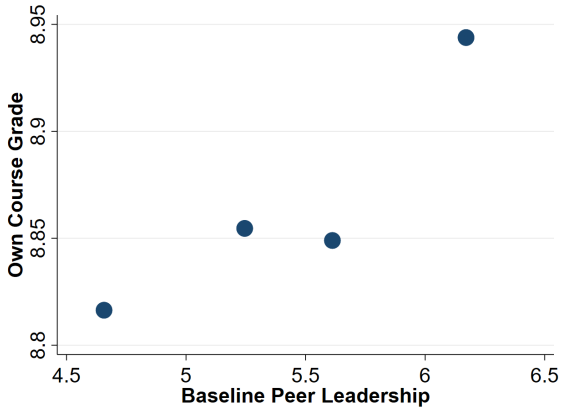
(B): Peer Communication and Grade, Controls



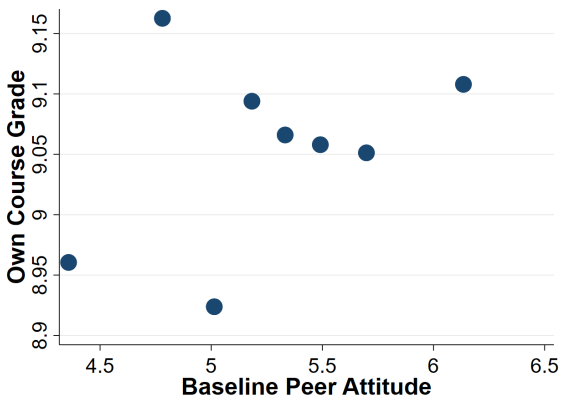
(C): Peer Leadership and Grade



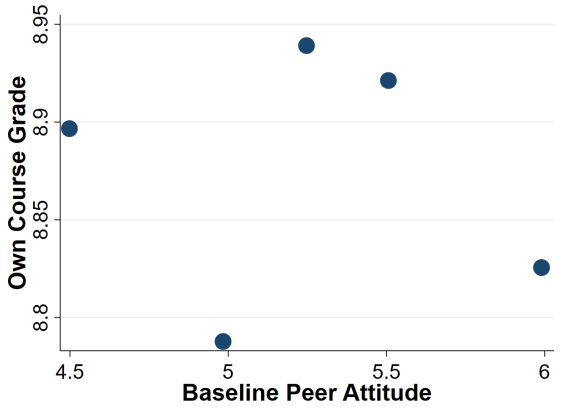
(D): Peer Leadership and Grade, Controls



(E): Peer Attitude and Grade

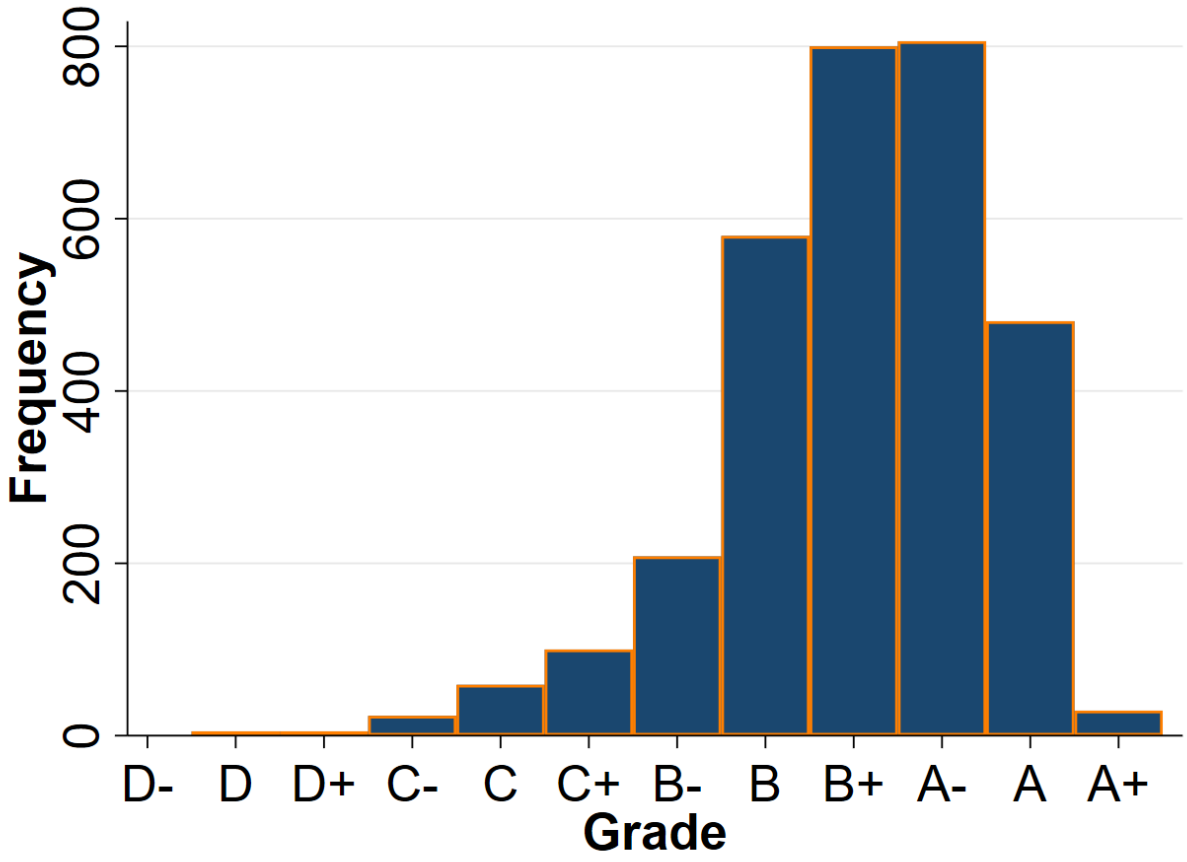


(F): Peer Attitude and Grade, Controls



Note: N = 3,039 in figures A and E. N = 3,035 in Figure C. N = 2,755 in Figure B. N = 2,754 in Figure D. N= 2,759 in Figure E. Optimal bins for variance-bias tradeoff (for integrated mean square error) defined by Cattaneo et al. (2019). Panel A, C, and E are conditional on peer group average and own ability before the course. Panels B, D, and F include course indicators as covariates.

Figure A.13: Grades Distribution



Note: Figure shows the instructor-assigned grades of NOLS students.

Table A.11: OLS Effect of Peer Non-Cognitive Ability on Getting at Least an A-, By Gender

Panel A: Men	(1)	(2)	(3)	(4)
Leave-out Proportion Male	-0.184*	-0.163*	-0.181*	-0.179*
	(0.103)	(0.098)	(0.100)	(0.099)
Baseline Own Communication	0.043***			-0.033**
	(0.014)			(0.017)
Baseline Peer Average Communication	0.010			0.006
	(0.031)			(0.038)
Baseline Own Leadership		0.103***		0.116***
		(0.012)		(0.015)
Baseline Peer Average Leadership		0.015		0.023
		(0.029)		(0.036)
Baseline Own Attitude			0.060***	0.018
			(0.013)	(0.014)
Baseline Peer Average Attitude			-0.015	-0.029
			(0.030)	(0.036)
Observations	1774	1773	1778	1731
R^2	0.117	0.167	0.135	0.177
Panel B: Women	(1)	(2)	(3)	(4)
Leave-out Proportion Male	-0.122	-0.114	-0.116	-0.117
	(0.133)	(0.130)	(0.140)	(0.135)
Baseline Own Communication	0.028			-0.045*
	(0.021)			(0.025)
Baseline Peer Average Communication	0.023			-0.025
	(0.049)			(0.061)
Baseline Own Leadership		0.097***		0.102***
		(0.017)		(0.022)
Baseline Peer Average Leadership		0.032		0.035
		(0.039)		(0.050)
Baseline Own Attitude			0.067***	0.040*
			(0.018)	(0.022)
Baseline Peer Average Attitude			0.067*	0.057
			(0.039)	(0.049)
Observations	981	981	981	959
R^2	0.113	0.150	0.127	0.162

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by OLS. The dependent variable is a binary variable which takes on a value of 1 if a student got at least an A- as graded by the instructor. Panel A includes only men. Panel B includes only women.

Table A.12: Logit Effect of Peers on Getting at Least an A-

Panel A: Men	(1)	(2)	(3)	(4)
Leave-out Proportion Male	-0.825*	-0.777*	-0.826*	-0.856*
	(0.444)	(0.453)	(0.437)	(0.467)
Baseline Own Communication	0.187***			-0.159*
	(0.068)			(0.084)
Baseline Peer Average Communication	0.038			0.017
	(0.132)			(0.175)
Baseline Own Leadership		0.478***		0.558***
		(0.070)		(0.088)
Baseline Peer Average Leadership		0.064		0.115
		(0.135)		(0.171)
Baseline Own Attitude			0.237***	0.065
			(0.066)	(0.074)
Baseline Peer Average Attitude			-0.079	-0.161
			(0.135)	(0.171)
Observations	1751	1750	1755	1708
Panel B: Women	(1)	(2)	(3)	(4)
Leave-out Proportion Male	-0.505	-0.495	-0.488	-0.513
	(0.545)	(0.555)	(0.579)	(0.576)
Baseline Own Communication	0.098			-0.222*
	(0.095)			(0.122)
Baseline Peer Average Communication	0.097			-0.141
	(0.201)			(0.264)
Baseline Own Leadership		0.438***		0.486***
		(0.095)		(0.115)
Baseline Peer Average Leadership		0.145		0.182
		(0.173)		(0.220)
Baseline Own Attitude			0.290***	0.190*
			(0.091)	(0.104)
Baseline Peer Average Attitude			0.294*	0.264
			(0.171)	(0.219)
Observations	952	950	952	928

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Cluster-robust standard errors, by course section, in parentheses. All regressions estimated by logistic regression. The dependent variable is a binary variable which takes on a value of 1 if a student got at least an A- as graded by the instructor. Panel A includes only men. Panel B includes only women.

B Principal Components Analysis

There are 4 questions that are used for each ability. The primary method for combining these questions into information about the underlying ability is to take the average. While an advantage is that it creates a more easily interpretable coefficient for regression analysis, it abstracts away from the potentially different aspects of each ability covered by each individual question.

An alternative method which could be used to reduce the dimensions of each ability is principal components analysis (PCA). In PCA, the components of the correlation matrix are found, such that each component is orthogonal to the others. So an alternative way for combining the various questions into 1 measure of skill by finding these components. While this may not give interpretable regression coefficients, it can at least inform us of whether aggregating the questions, then aggregating by peers is affected by choice of method for aggregating the questions.

The first step is investigating whether the questions are suitably correlated for a PCA. To investigate this, the Kaiser-Meyer-Olkin (KMO) statistics (which range from 0 to 1) are calculated. Too low of a KMO statistic (typically 0.5 or 0.6) implies too little correlation amongst the variables being considered for PCA. The overall KMO statistics for communication, leadership, and attitude are 0.7384, 0.6904, and 0.7545 respectively, while the individual questions of each are never below 0.686 (lead question 1). The conclusion is that PCA is statistically justifiable as a way to reduce the dimensions of the before individual abilities. Also, the KMO statistics are never below 0.75 once the leave out mean is calculated within groups, making PCA appropriate for the mean group score for each question also.

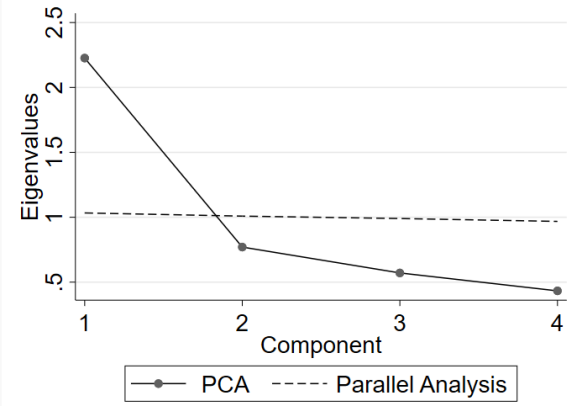
Second, we are interested in how many components would be required to be retained for regression analysis. Figure B.1 plots the eigenvalues of each component. Sometimes, the so-called "elbow rule" is used to decide how many components to retain, but this can be subjective. Instead, a parallel analysis is performed, which uses a correlation matrix from randomly generated data as a benchmark. The parallel analysis crosses the actual eigenvalues plot prior to component 2 in all figures implying that only 1 component should be retained for regression from either the individual or group means of abilities. Figure B.2 shows that the first component alone always explains about 60% of the data and explains about the same for individual or group mean questions. The

first component usually explains a little bit more variance for the group mean than the individual questions.

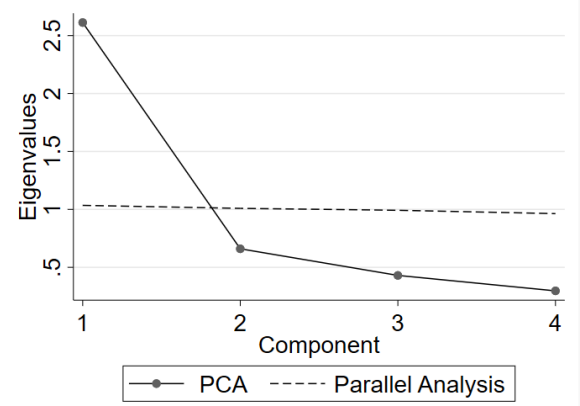
Finally, we can see which survey questions load on which components in Figure B.3. For communication, question 3, which asks about body language, loads heavily on component 2 while all the other questions load heavily on component 1 for both individuals and group menas. For leadership, question 4 loads heavily on component 2 and question 2 loads heavily on component 1 in the individual responses. Leadership question 4 loads equally on each component which is the only difference between group and individual leadership questions. Finally for attitude for both individual and group averages, questions 1 and 2 load heavily on component 1 and question 4 loads heavily on component 2. Overall, the analysis concludes that aggregating to group means does not alter how the questions load or how much variance is explained by each component.

Figure B.1: Parallel Analysis

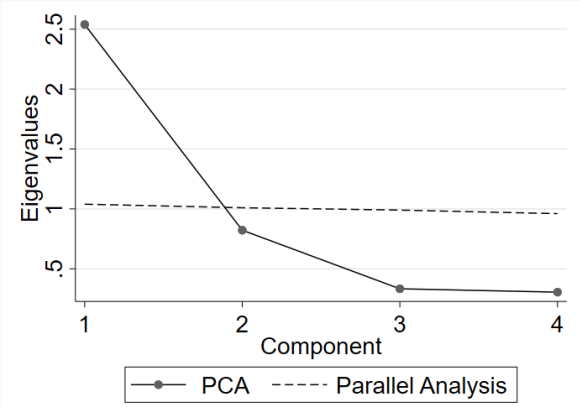
(A): Communication Before



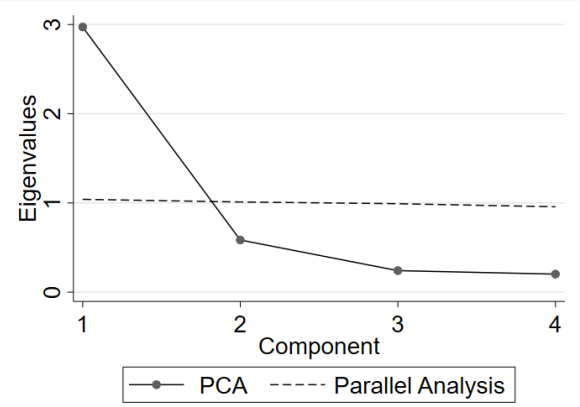
(B): Group Mean Communication Before



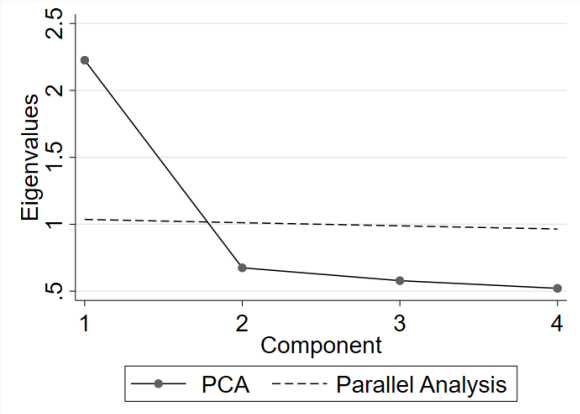
(C): Leadership Before



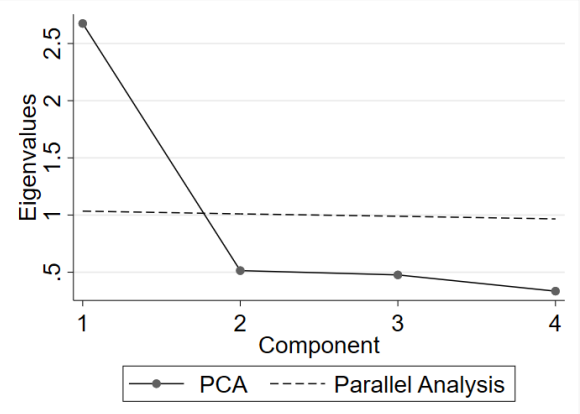
(D): Group Mean Leadership Before



(E): Attitude Before



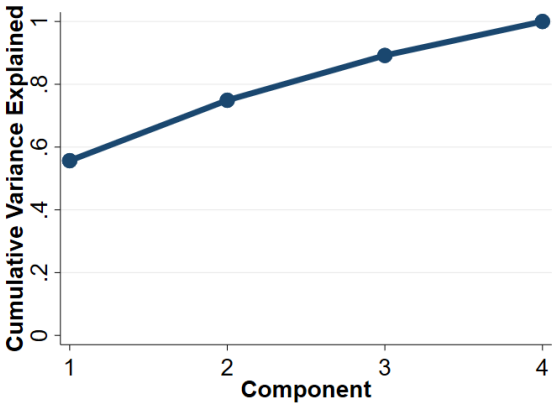
(F): Group Mean Attitude Before



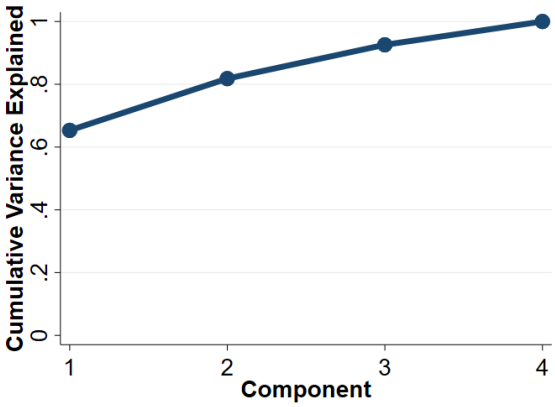
Note: Eigenvalues come from principal components analysis (PCA), of the correlation matrix, for the 4 questions for each ability. The parallel analysis uses 10 repetitions. Figures A, C, and E use the individual questions. Figures B, D, and F use the group means.

Figure B.2: Cumulative Variance

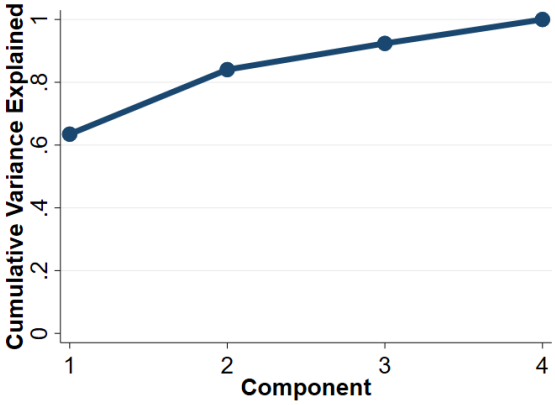
(A): Communication Before



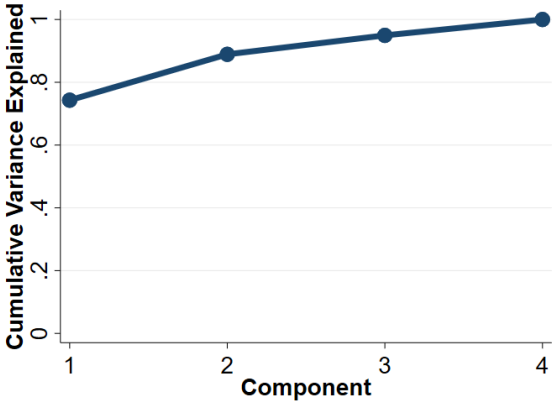
(B): Group Mean Communication Before



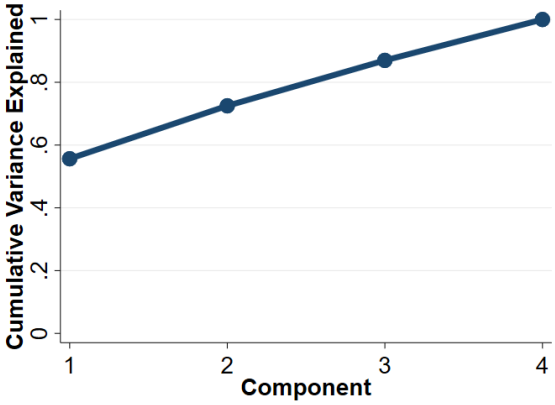
(C): Leadership Before



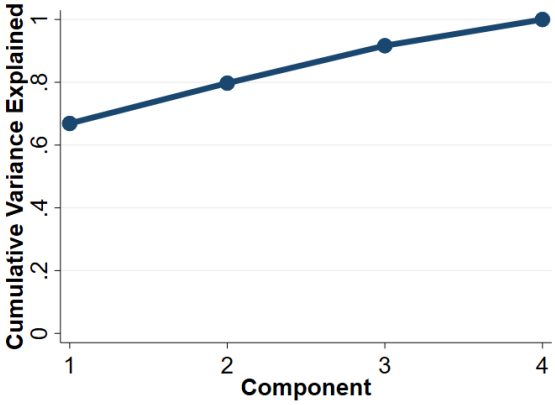
(D): Group Mean Leadership Before



(E): Attitude Before



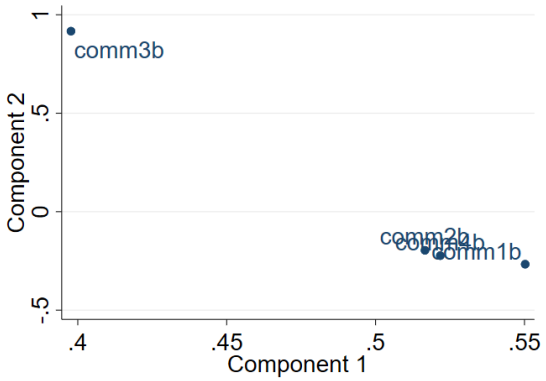
(F): Group Mean Attitude Before



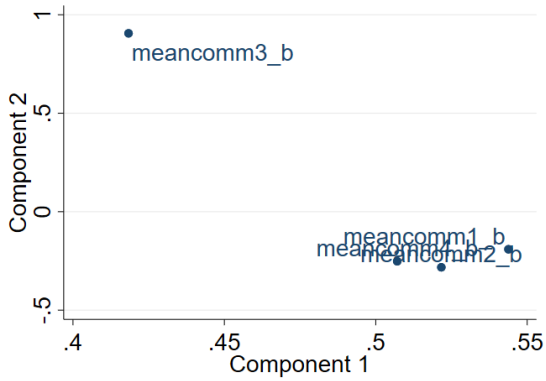
Note: Figures show the cumulative variance explained by each successive component. Panels A, C, and E use the individual measures. Panels B, D, and F use the peer measures.

Figure B.3: Loading on Components 1 and 2

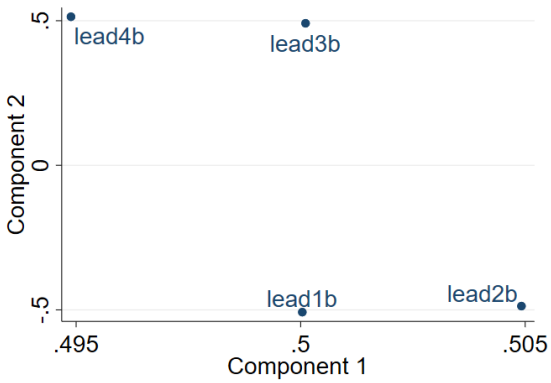
(A): Communication Before



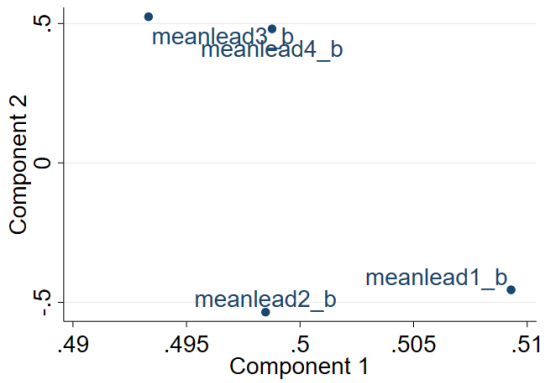
(B): Group Mean Communication Before



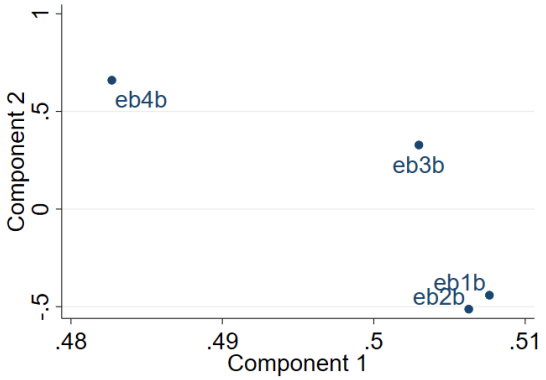
(C): Leadership Before



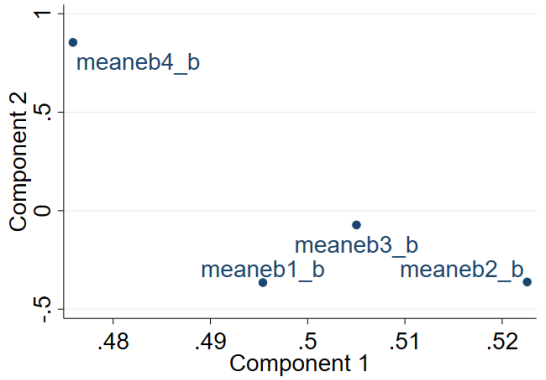
(D): Group Mean Leadership Before



(E): Attitude Before



(F): Group Mean Attitude Before



Note: Figures show how each question loads onto each of the first 2 components. Figures A, C, and E use individual measures. Figures B, D, and F use peer measures. Comm3b stands for communication question 3, before the course. Meancomm3_b stands for peer average communication on question 3, before the course.