# Non-cognitive Peer Effects and the Origins of the Gender Gap in Leadership

Matthew Pearson,* Benjamin Blemings,† Scott Cunningham‡

This Version: July 31, 2022; 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, & Ockert, 2017), and even moreso where they are complementary to 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 et al., 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]

In addition to the growing relevance of non-cognitive skills, there are large gender disparities in business leadership positions. 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 formation.

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 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

[2]Non-cognitive is an umbrella term that has encompassed a lot of measurements in the literature. For example, the early literature used whether an individual had an internal or external locus of control (Bowles, Gintis, & Osborne, 2001), whether an individual had behavioral incidents as a child (Heckman & Rubinstein, 2001), and concentration endurance on tests (Balart, Oosterveen, & Webbink, 2018).

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

applicants are assigned to a course section from their rankings over course types. After controlling for students selecting into different course types, there is no relationship between an individual's baseline non-cognitive abilities and average baseline abilities of others in the same section. This baseline balance in observable characteristics also extends to basic demographic characteristics which suggests that cohort composition is as good as randomly assigned conditional on course types (like e.g. Feld and Zölitz (2017).)

Peers with higher 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. Specifically, if the baseline peer communication or leadership increases by 1 (a 1.25 standard deviation increase), then the individuals rank themselves lower in communication or leadership respectively by about 0.1 standard deviations. There are no effects of peer attitude on attitude.

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.), and the gender composition of peers (Gong, Lu, & Song, 2021; Orr, 2021). Closely related to peer non-cognitive behavior, peers exposed to domestic violence or that exhibit disruptive behavior reduces ones' own test scores and later in life income (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, 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.

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]

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

[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/.

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).

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

[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/.

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 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]

[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.

[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.

**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 Summary 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).

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.

[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.4 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.

Table 1: Summary 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 |
| **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 |
| **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 | 10.54 | 11.00 | 2.63 | 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 inititative 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.

# 4 Method

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

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

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.

## 4.1 Peer Effect Specification

The coefficient of interest, representing the effect of peers, is $\beta$. 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.

Equation 1 specifies peer ability as linear-in-means which may lead to mis-specification error. The potential mis-specification is due to many students near the upper bound of the composite abilities (8), as shown by Figure A.9. Furthermore, there may be non-linear effects of peer ability. Robustness checks address potential mis-specification using non-parametric kernel regression.

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.

**Demographic Controls** 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.

**Course Section Timing Fixed Effects** The last fixed effects are for course month, $\theta_M$, and course year, $\phi_Y$. 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.[16,17] 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

[16]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.

[17]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.[18] To account for this selection into courses, the models include $\mu_C$ 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 1 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.

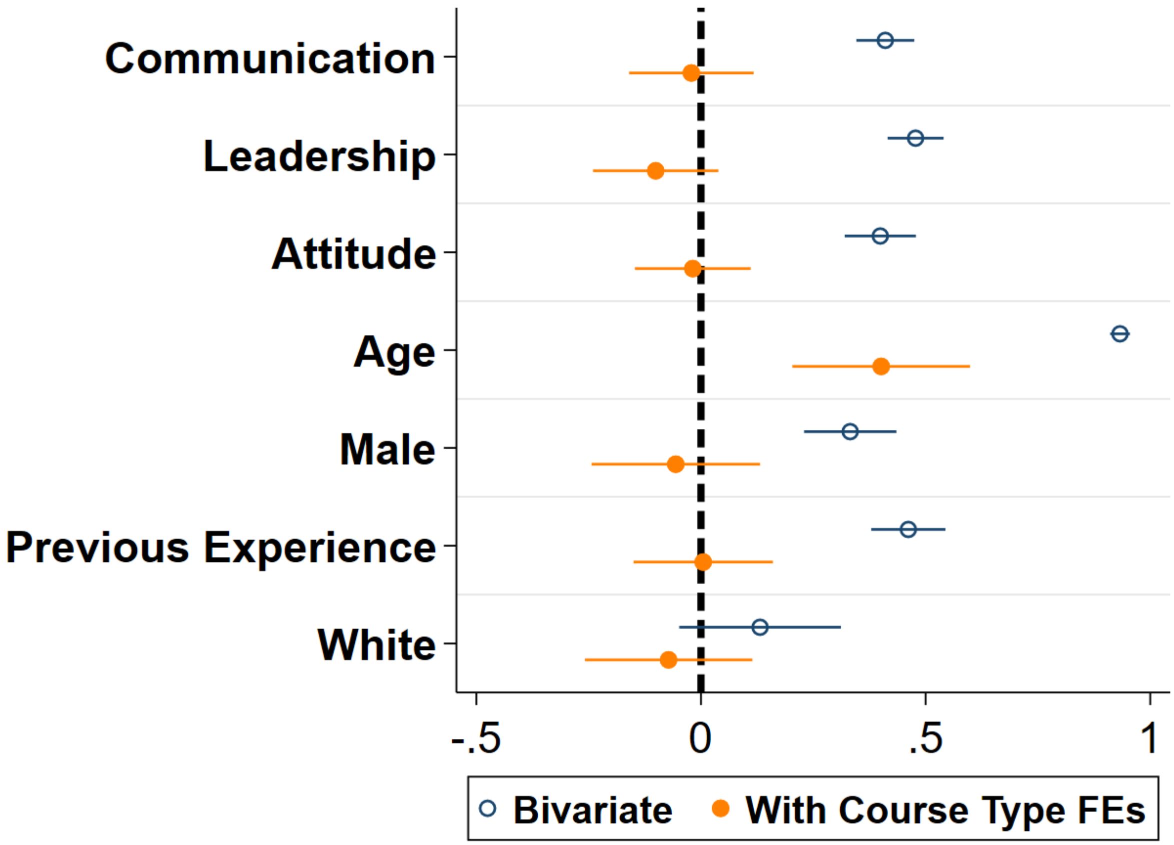
Course type is observable, so selection into course type can be accounted for by including course type indicators in the regression of individual baseline characteristics on group average baseline characteristics. Intuitively, these indicators enable the relationship between individual and group characteristics 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 1. After controlling for selection into course type, 6 out of 7 of the individual characteristics are uncorrelated with group characteristics.[19] 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

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

[19]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.4.

to discontinuities in the ages that are eligible, making global linear regression mis-specified.[20,21] 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 the selection problem in estimating causal peer effects Manski (1993).

Figure 1: Baseline Characteristics Regressed on Same Group Characteristics




**Note:** N = 3,773. 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.

[20]Figures A.5 and A.6 show these discontinuities around age 16 and 22.

[21]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. Overall, it is reasonable to conclude that peers are randomly assigned, after conditioning on course types, so that $\beta$ can be considered a causal effect of peers.

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

## 4.3 Binned Scatterplots

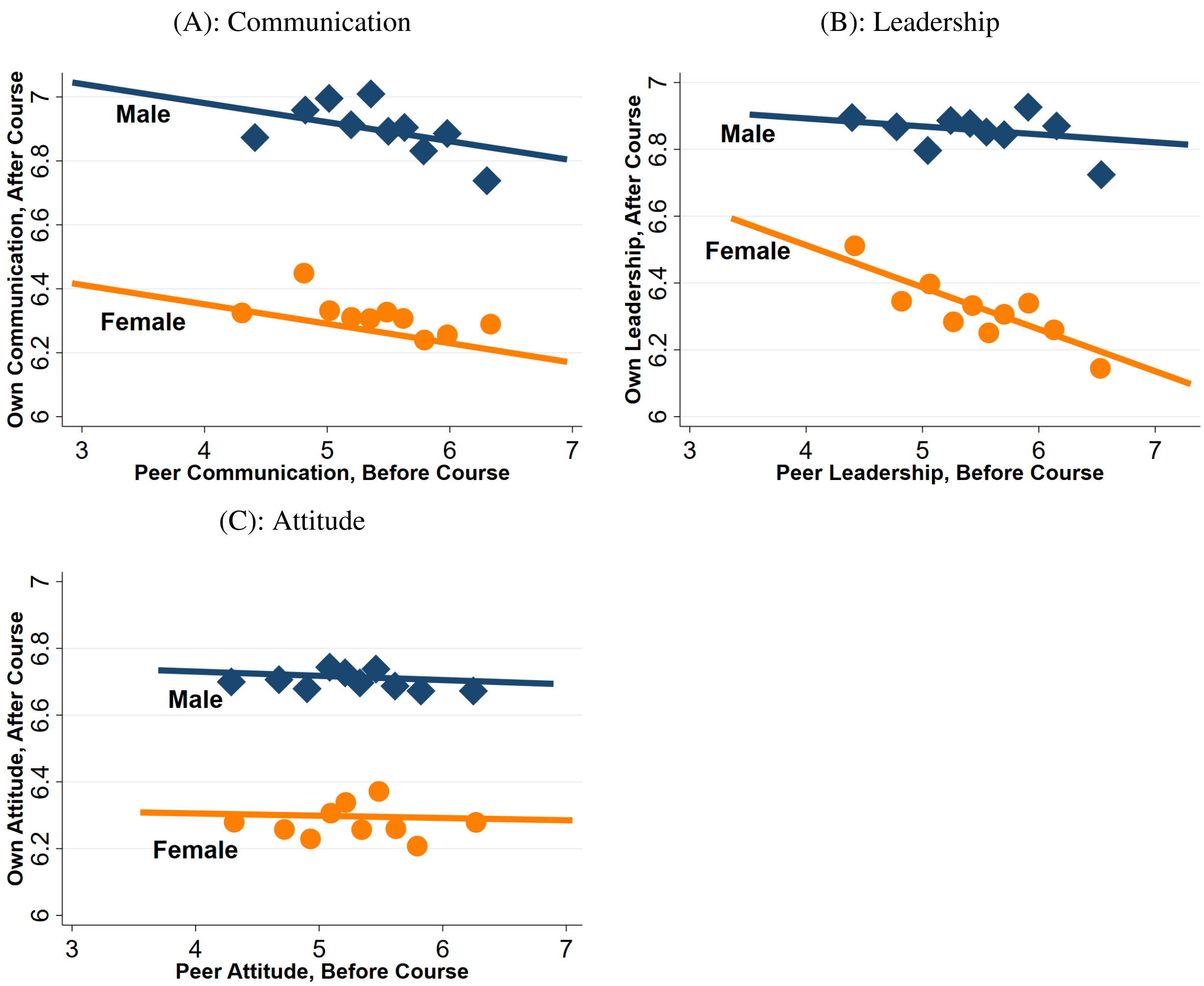
Next, binned scatterplots are presented in Figure 2, providing a non-parametric visualization of the relationship between baseline peer ability and post course own ability. All Figures control for own baseline ability, to ensure comparisons are made between those of similar baseline ability, and course-type fixed effects to address selection into course type.[22] Across all skill categories and both genders, a linear line of best fit is a reasonable specification.

As shown in Figure 2A 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 2B 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 2C shows that there is barely any effect of baseline peer attitude on post-course peer attitude for either gender.[23]

[22]They are included through a semi-parametric adjustment described in Cattaneo, Crump, Farrell, and Feng (2019), but as shown in Figure A.8 the results are similar without the course type fixed effects.

[23]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.7 and the conclusions are the same.

Figure 2: How Peer Baseline Ability Affects Own Post-Course Ability




**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.7.

# 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 communication (approximately 1.25 standard deviations) causes a 0.08 decrease (approximately one-tenth of a standard deviation) in own, post-course communication.[24] 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.[25] 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.3, 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.

[24] 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.

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

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

| Panel A: Communication | | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Baseline Peer Average Composite Communication | -0.14*** | -0.11*** | -0.08*** |
| | (0.02) | (0.02) | (0.03) |
| Observations | 3661 | 3265 | 3265 |
| $R^2$ | 0.478 | 0.483 | 0.521 |
| Covariates | - | X | X |
| Course FEs | - | - | X |

| Panel B: Leadership | | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Baseline Peer Average Composite Leadership | -0.13*** | -0.11*** | -0.07** |
| | (0.02) | (0.02) | (0.03) |
| Observations | 3682 | 3286 | 3286 |
| $R^2$ | 0.454 | 0.457 | 0.488 |
| Covariates | - | X | X |
| Course FEs | - | - | X |

| Panel C: Attitude | | | |
|---|---|---|---|
| | (1) | (2) | (3) |
| Baseline Peer Average Composite Attitude | -0.08*** | -0.05 | -0.00 |
| | (0.03) | (0.03) | (0.03) |
| Observations | 3669 | 3278 | 3278 |
| $R^2$ | 0.420 | 0.422 | 0.454 |
| Covariates | - | X | X |
| Course FEs | - | - | 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.

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. 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.

**Disaggregated Skills** 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 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.5. 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.

**Variance Controls** 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.6, standard deviation is not statistically significant and it's inclusion has no effects on the estimates for leave-out mean.

**Non-Parametric Kernel Regression** 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.9, 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.[26] As shown in Table A.7, this more flexible specification produces very similar estimates. One way the estimates are different is that, even once covariates are included, there is a small negative effect on attitude which is statistically significant at 90% confidence.

**Instructor-Assigned Grades** Finally, it may be that the negative impact is entirely driven by the self-reported aspect of the dependent variable. 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. As shown in column 2 of Table A.9, better peer ability has positive impacts on grades which are of nearly equal magnitude; however, they are two to three times less precise which makes them statistically insignificant.

## 5.1 Heterogeneous Treatment Effects

The estimates reported so far are average effects, but peers may influence various subgroups differently. For example, the peer effects literature documents several ways in which the influence of peers depends on one's gender (e.g. Gong et al. (2021)). In the following results, we consider the effect of the entire (leave-one out) course section's baseline peer ability on different subsets of individuals.

The results, which are estimated separately by gender and median age, are presented in Table 3.[27] As shown in Panel A, peer communication has negative effects on both men and women.

[26]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.

[27]There are too few non-white students to have meaningful estimates of peer effects for white and non-white individuals.

A 1 unit increase in peer communication reduces post course communication for men by 0.07 (p-value = 0.061) and for women by 0.1 (p-value = 0.013).

The nearly equally sized effects for men and women on communication stands in contrast to the effects on leadership. A 1 unit increase in peer leadership reduces male leadership by only 0.03, although this is not statistically significant with a standard error of 0.04. 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.

#### 5.1.1 Age

Next, 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 3 and 4 of Table 3 splits the sample into above and below the median age (19.58 years).

As shown by Panel A, columns 3 and 4, greater peer communication ability harms both older and younger groups of people. A 1 unit increase in baseline, peer communication leads to 0.09 decrease (p-value = 0.028) in post-course, own communication for those above 19.58 and a 0.07 decrease (p-value = 0.09) for those younger than 19.58. In contrast, a 1 unit increase in baseline peer leadership has negative effects on older individuals, but not younger ones. A 1 unit increase in peer, baseline leadership reduces leadership for those above 19.58 years old by 0.1 which is statistically significant at 99% confidence. There are no effects for attitude when split by gender or age.[28,29]

[28] 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, college-aged individuals are generally less harmed than high school or older individuals.

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

Table 3: Heterogeneous Effects of Peer Non-Cognitive Ability on Own Non-Cognitive Ability by Demographic Characteristics

| Panel A: Communication | Gender | | Age | | Previous Experience | |
|---|---|---|---|---|---|---|
| | (1) Male | (2) Female | (3) > Median | (4) ≤ Median | (5) Yes | (6) No |
| Baseline Peer Average Comm. | -0.07* | -0.10** | -0.09** | -0.07* | -0.11*** | -0.05 |
| | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) | (0.04) |
| Observations | 2111 | 1154 | 1633 | 1632 | 1725 | 1540 |
| $R^2$ | 0.497 | 0.613 | 0.585 | 0.481 | 0.544 | 0.530 |

| Panel B: Leadership | Gender | | Age | | Previous Experience | |
|---|---|---|---|---|---|---|
| | (1) Male | (2) Female | (3) > Median | (4) ≤ Median | (5) Yes | (6) No |
| Baseline Peer Average Lead | -0.03 | -0.16*** | -0.10*** | -0.04 | -0.04 | -0.10** |
| | (0.04) | (0.04) | (0.03) | (0.04) | (0.04) | (0.04) |
| Observations | 2124 | 1162 | 1647 | 1639 | 1737 | 1549 |
| $R^2$ | 0.480 | 0.549 | 0.537 | 0.460 | 0.509 | 0.496 |

| Panel C: Attitude | Gender | | Age | | Previous Experience | |
|---|---|---|---|---|---|---|
| | (1) Male | (2) Female | (3) > Median | (4) ≤ Median | (5) Yes | (6) No |
| Baseline Peer Average Att. | 0.01 | -0.03 | -0.01 | -0.00 | 0.01 | -0.00 |
| | (0.04) | (0.05) | (0.04) | (0.05) | (0.04) | (0.05) |
| Observations | 2124 | 1154 | 1640 | 1638 | 1733 | 1545 |
| $R^2$ | 0.451 | 0.504 | 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.

**Course Characteristics** Finally, 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.8 presents results which are split by 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.

# 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.

# References


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) [2]

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. [2]

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. Retrieved from https://doi.org/10.1162/rest_a_00955 doi: 10.1162/rest_a_00955 [3, 19]

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 [2]

Bureau of the Census. (1998). *First findings from the EQW National Employer Survey.* EQW Catalog RE01. (Accessed: July 2020) [2]

Carrell, S. E., Hoekstra, M., & Kuka, E. (2018, November). The long-run effects of disruptive peers. *American Economic Review*, *108*(11), 3377-3415. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aer.20160763 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. Retrieved from https://www.aeaweb.org/articles?id=10.1257/app.2.1.211 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. [5]

Cattaneo, M. D., Crump, R. K., Farrell, M. H., & Feng, Y. (2019). On binscatter. *arXiv preprint arXiv:1902.09608*. [13, 14, 31, 32, 35, 36, 44]

Cattaneo, M. D., & Jansson, M. (2018). Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency. *Econometrica*, *86*(3), 955–995. [18]

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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0927537121000361 doi: https://doi.org/10.1016/j.labeco.2021.102001 [4]

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) [2]

Deming, D. J. (2017). The growing importance of social skills in the labor market. *The Quarterly Journal of Economics*, *132*(4), 1593–1640. [2]

Edin, P.-A., Fredriksson, P., Nybom, M., & Ockert, B. (2017). The rising return to non-cognitive skill. [2]

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. [3, 13]

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

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. Retrieved from https://doi.org/10.1086/712638 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. [3, 18]

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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167268119301052 doi: https://doi.org/10.1016/j.jebo.2019.04.002 [3, 19]

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. [17]

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). Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/coep.12572 doi: https://doi.org/10.1111/coep.12572 [3]

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *American Economic Review*, *91*(2), 145–149. [2]

Howard, G. S. (1980). Response-shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review*, *4*(1), 93–106. [6]

Hoxby, C. M., & Weingarth, G. (2006). *Taking race out of the equation: School reassignment and the structure of peer effects.* (Working Paper) [15, 37]

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. Retrieved from https://www.aeaweb.org/articles?id=10.1257/aeri.20200029 doi: 10.1257/aeri.20200029 [3]

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics*, *116*(3), 777–803. [15, 37]

Li, Q., Lu, X., & Ullah, A. (2003). Multivariate local polynomial regression for estimating average derivatives. *Journal of Nonparametric Statistics*, *15*(4-5), 607–624. [18]

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542. [2, 5, 12]

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. Retrieved from https://doi.org/10.1080/13504851.2021.1991560 doi: 10.1080/13504851.2021.1991560 [3]

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, 37]

Shure, N. (2021). Non-cognitive peer effects in secondary education. *Labour Economics*, *73*, 102074. Retrieved from https://www.sciencedirect.com/science/article/pii/S0927537121001093 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. [6]

Weinberger, C. J. (2014). The increasing complementarity between cognitive and social skills. *Review of Economics and Statistics*, *96*(4), 849–861. [2]

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) [2]

# Online Appendix

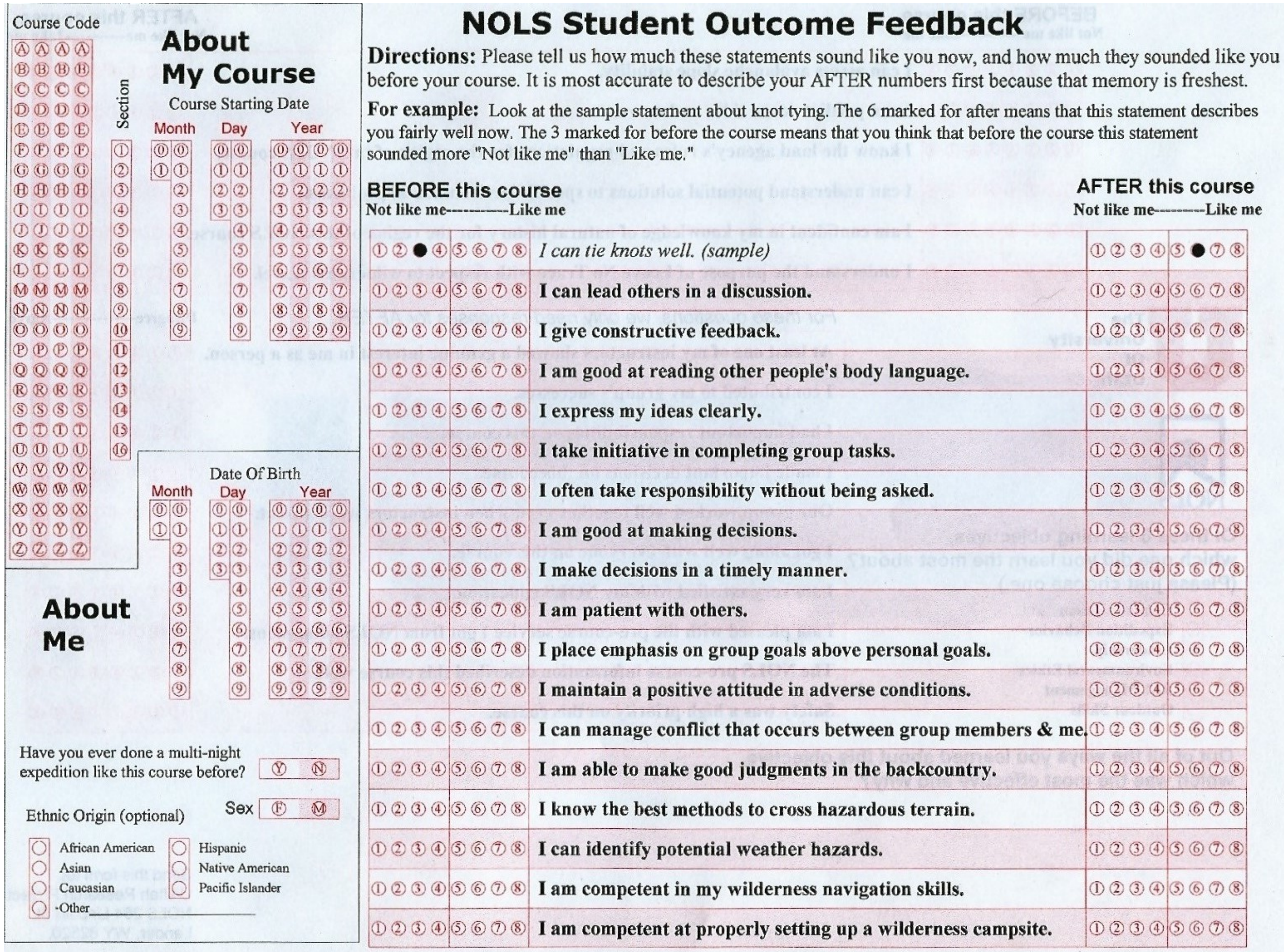
## A Additional Tables and Figures

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<br>Not like me----------Like me | | AFTER this course<br>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) |
|---|---|---|
| ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ | I can assess avalanche slope stability. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ | I can predict ocean tides and currents. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ | I know the land agency's rules and regulations for the region of this NOLS course. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ | I can understand potential solutions to specific environmental problems. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ | I am confident in my knowledge of natural history for the region of this NOLS course. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ | I understand the purpose of Leave No Trace with respect to wilderness travel. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |

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. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| I contributed to my group's successes. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| I had important responsibilities on this course. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| I made important decisions on this course. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| Our group worked well together even when instructors were absent. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| I got along well with everyone on this course. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| I am very satisfied with my NOLS education. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| I am pleased with the pre-course service I got from NOLS admissions. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| The NOLS pre-course information described this course well. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |
| Safety was a high priority on this course. | ① ② ③ ④ ⑤ ⑥ ⑦ ⑧ |

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

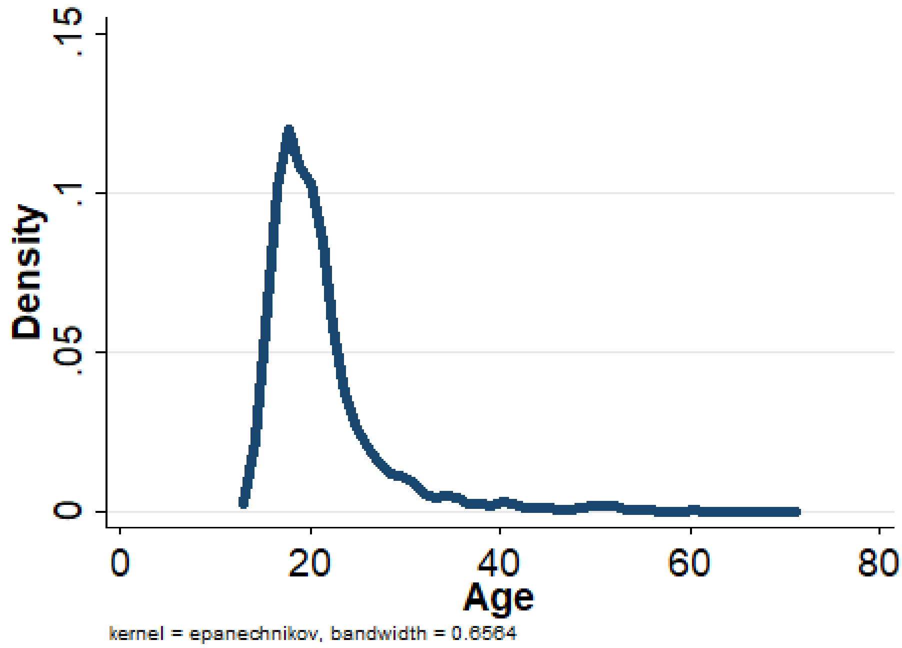
- ○ Communication
- ○ Expedition Behavior
- ○ Leadership
- ○ Environmental Ethics
- ○ Risk Management
- ○ Outdoor Skills

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

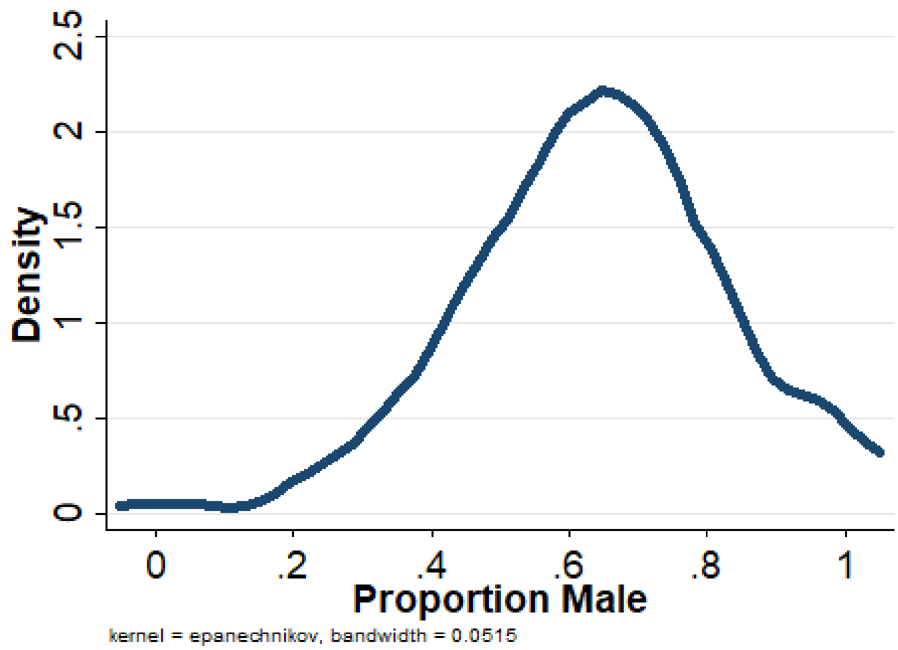
Send this form to:
U/Utah Research Project
NOLS 284 Lincoln St.
Lander, WY 82520.

Figure A.3: Distributions of Age and Proportion Male

(A): Age

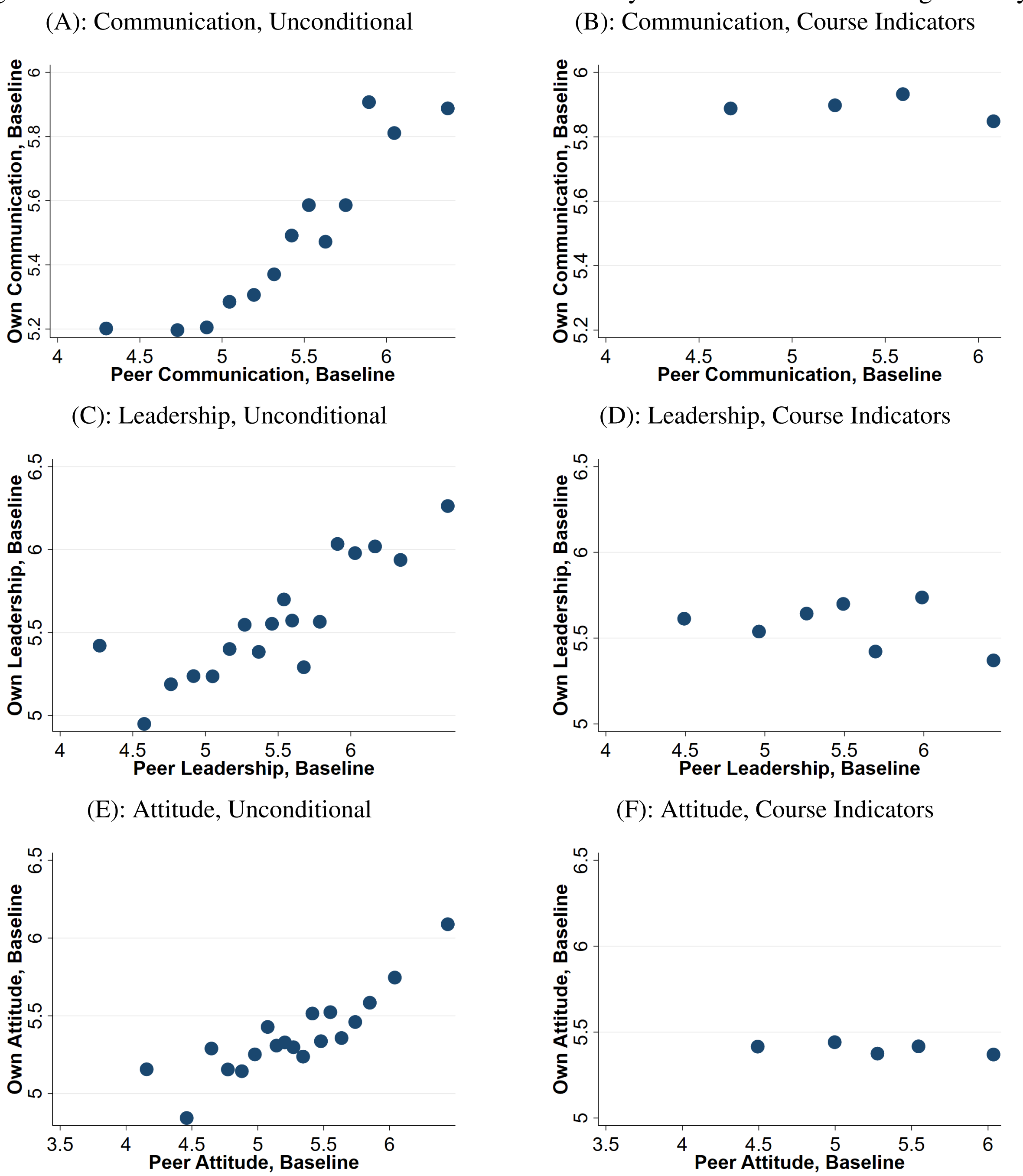



(B): Proportion Male




**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.

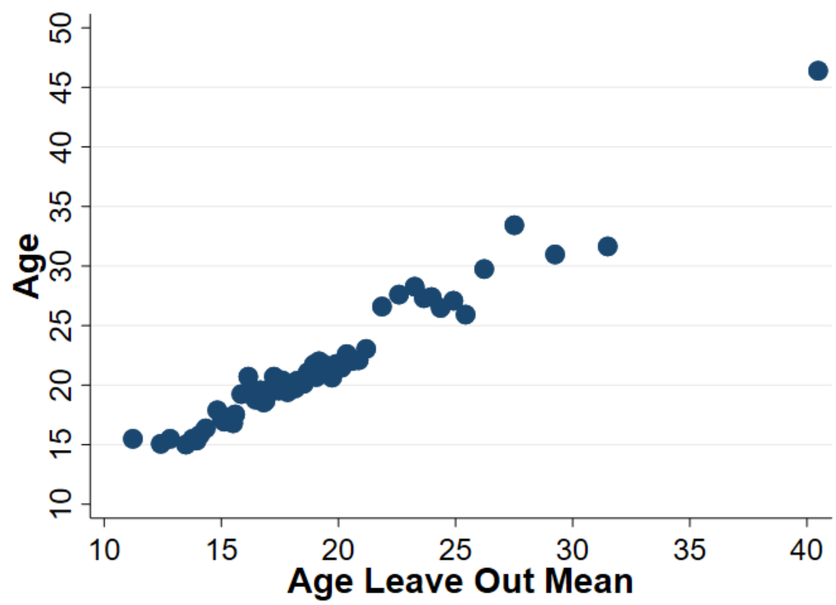
Figure A.4: Correlation Between Baseline Individual Ability and Baseline Peer Average Ability



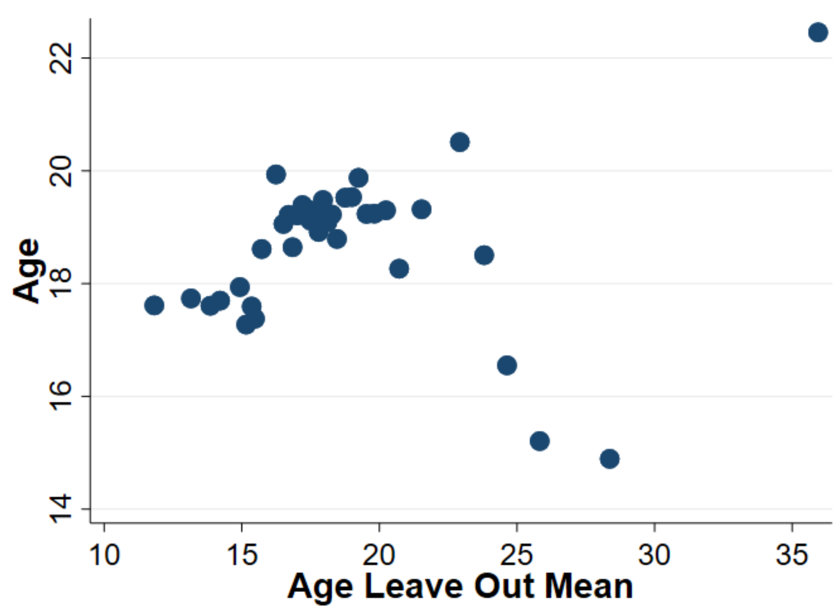

**Note:** Y-axis is baseline individual composite ability. X-axis is baseline peer average composite ability. Number of bins is optimal for variance-bias tradeoff (for integrated mean square error), as defined by Cattaneo et al. (2019). Panel A, C, and E are conditional on peer group average and own 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.5: Binned Scatter Conditional Randomization Checks for Age

(A): Group Age on Age, Unconditional

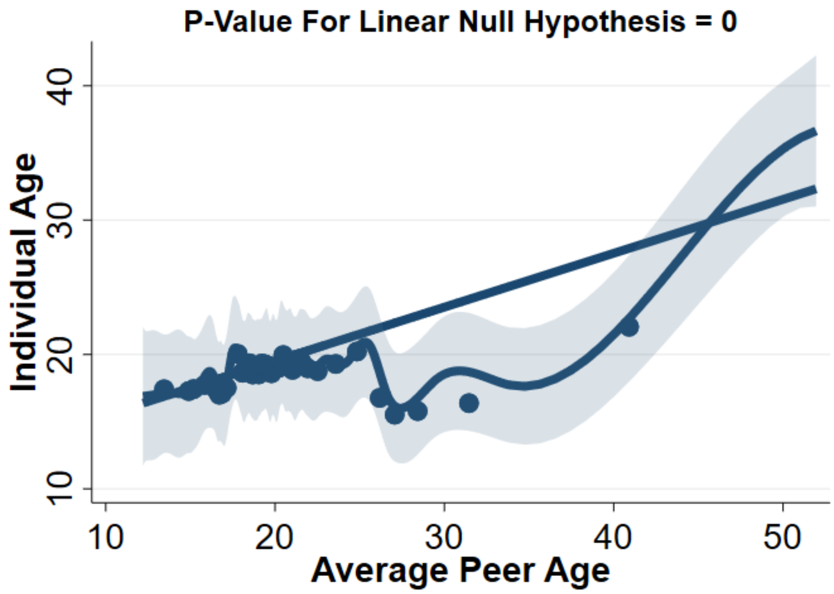



(B): Group Age on Age, Course Indicators




**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.6: Age on Meanage with Cubic B-Spline



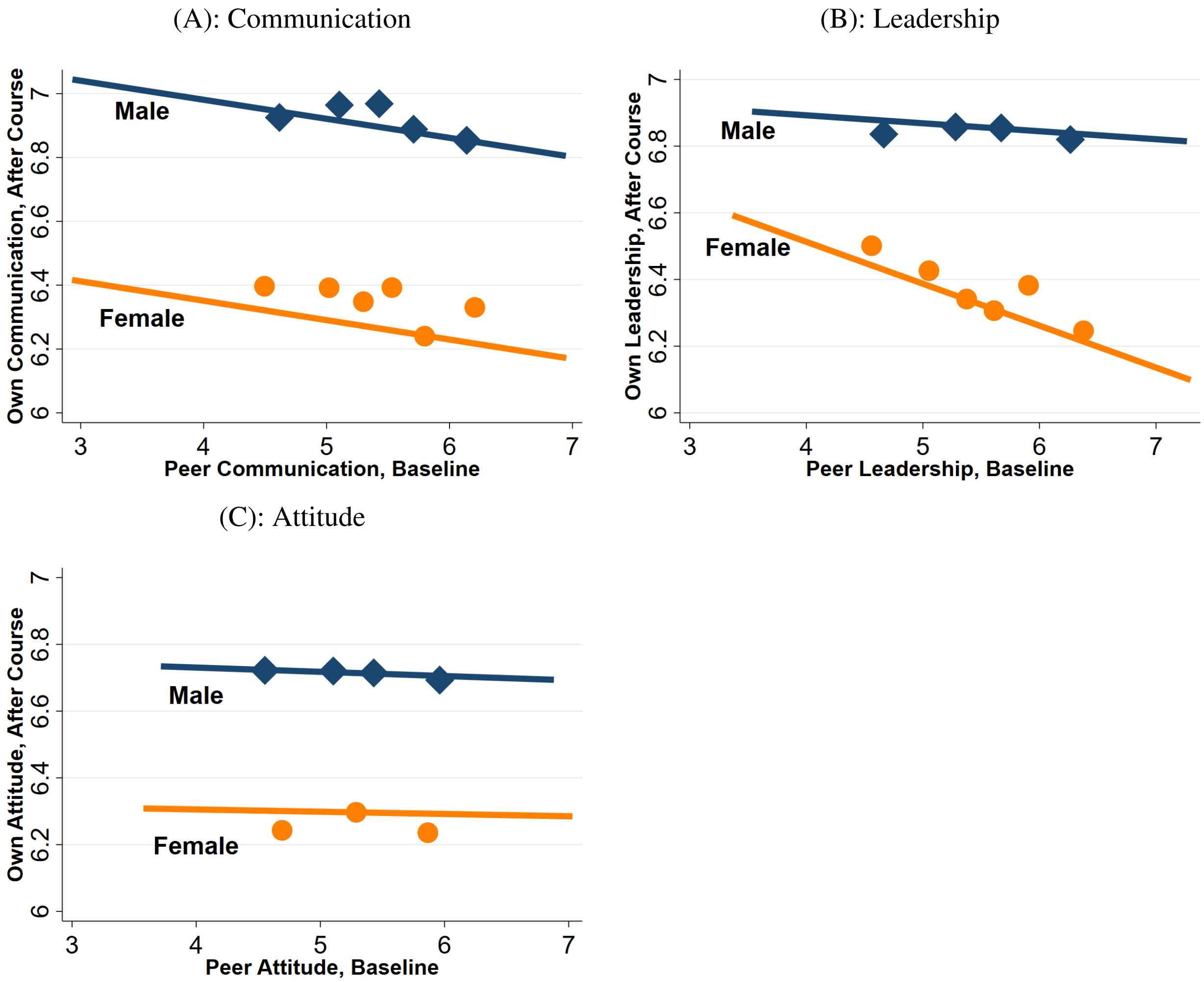

**Note:** N = 3,773. Fit and confidence bands are from a cubic B-spline. The hypothesis test is a null of global linearity. To evaluate the hypothesis, it compares the linear functional form to a cubic b-spline.

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).

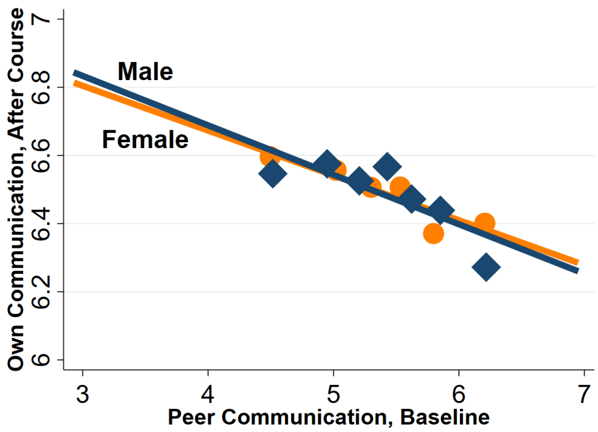
Figure A.7: How Leave One Out Peer Baseline Ability Affects Own Ability, Optimal Number of Bins



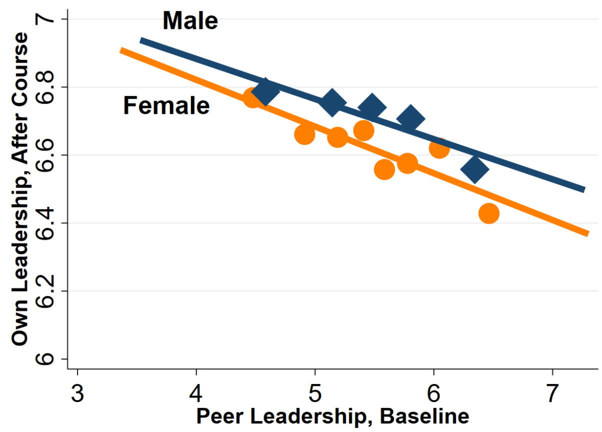

**Note:** All figures show the correlation between baseline leave-one out peer ability and own post-course ability. The correlations control for course type indicators and baseline own ability using methods from Cattaneo et al. (2019). The x-axis is peer leave one-out baseline ability. The y-axis is 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.8: How Leave One Out Peer Baseline Ability Affects Own Ability, No Course Type FE's

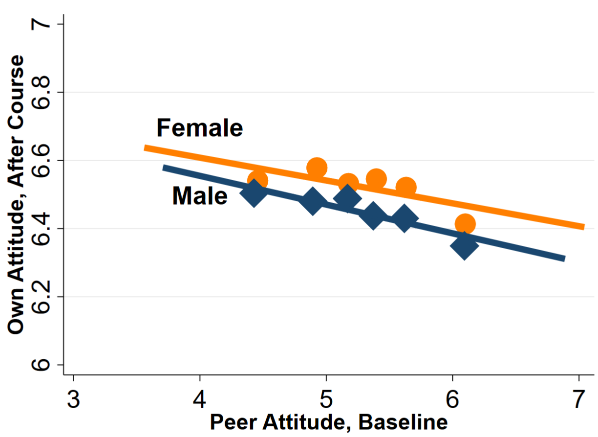
(A): Communication




(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.3: 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.4: 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.5: 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.6: Effect of Baseline Peer Ability on Own Ability, Variance Controls

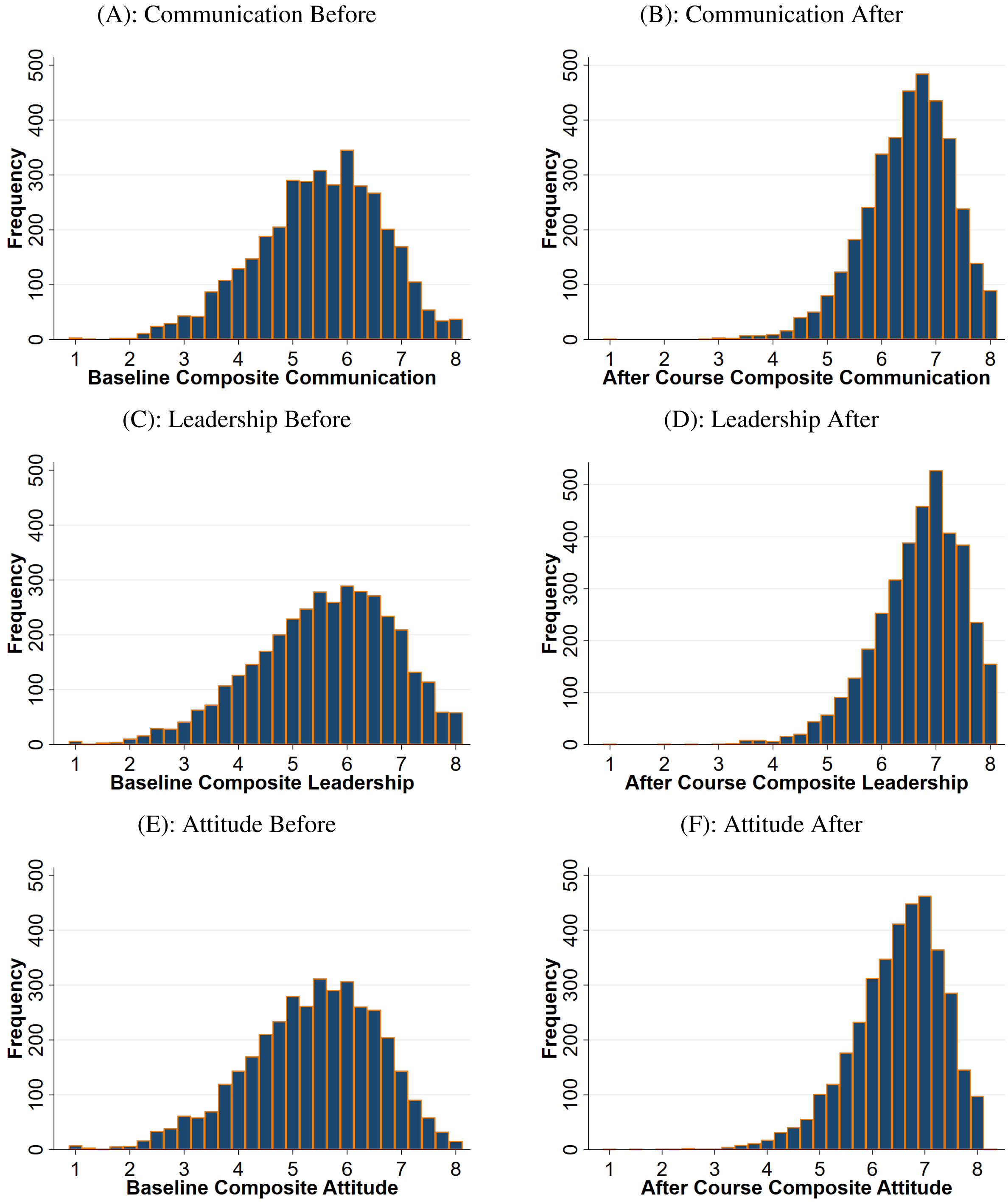
| Panel A: Communication | (1) All | (2) Male | (3) Female |
|---|---|---|---|
| Baseline Peer Average Comm. | -0.07** | -0.07* | -0.10** |
| | (0.03) | (0.04) | (0.04) |
| Baseline Peer Std. Dev. Comm. | 0.03 | 0.01 | 0.01 |
| | (0.05) | (0.06) | (0.07) |
| Observations | 3261 | 2109 | 1152 |
| $R^2$ | 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.

Figure A.9: Distributions of Individual Composite Abilities, Before and After Course




**Note:** Figures show the distributions of individual composite abilities, before and after the course ends.

Table A.7: 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.8: Heterogeneous Effects of Peer Non-Cognitive Ability on Own Non-Cognitive Ability by Course Characteristics

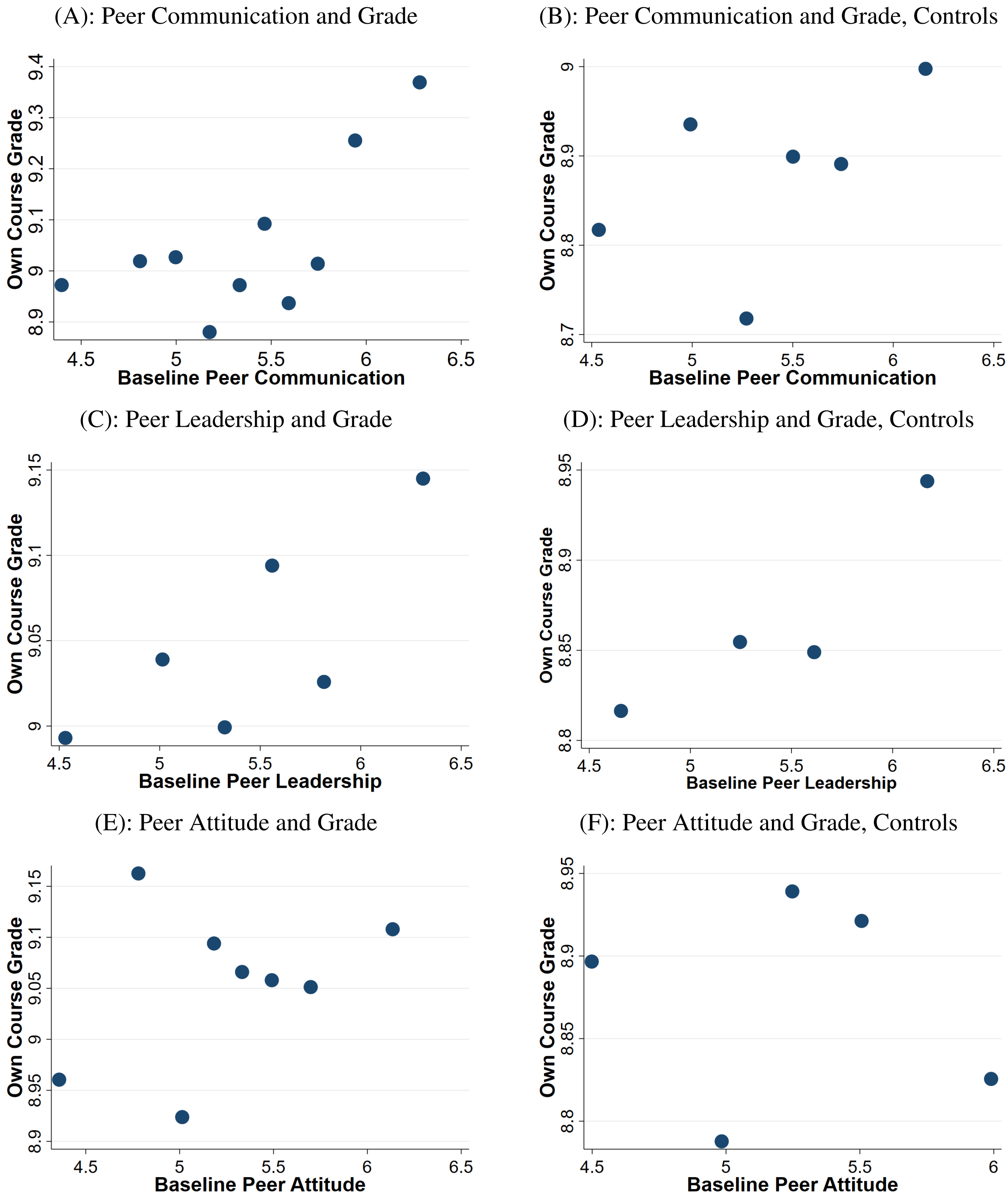
| Panel A: Communication | Course Members | | Course Length | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | $>$ Median | $\leq$ Median | $>$ Median | $\leq$ Median |
| Baseline Peer Average Comm. | -0.15*** | -0.05 | 0.02 | -0.10*** |
| | (0.05) | (0.04) | (0.07) | (0.03) |
| Observations | 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.10: Binned Scatter of Baseline Peer Ability Effects on Post Course Grade, Controlling for Own Before Ability



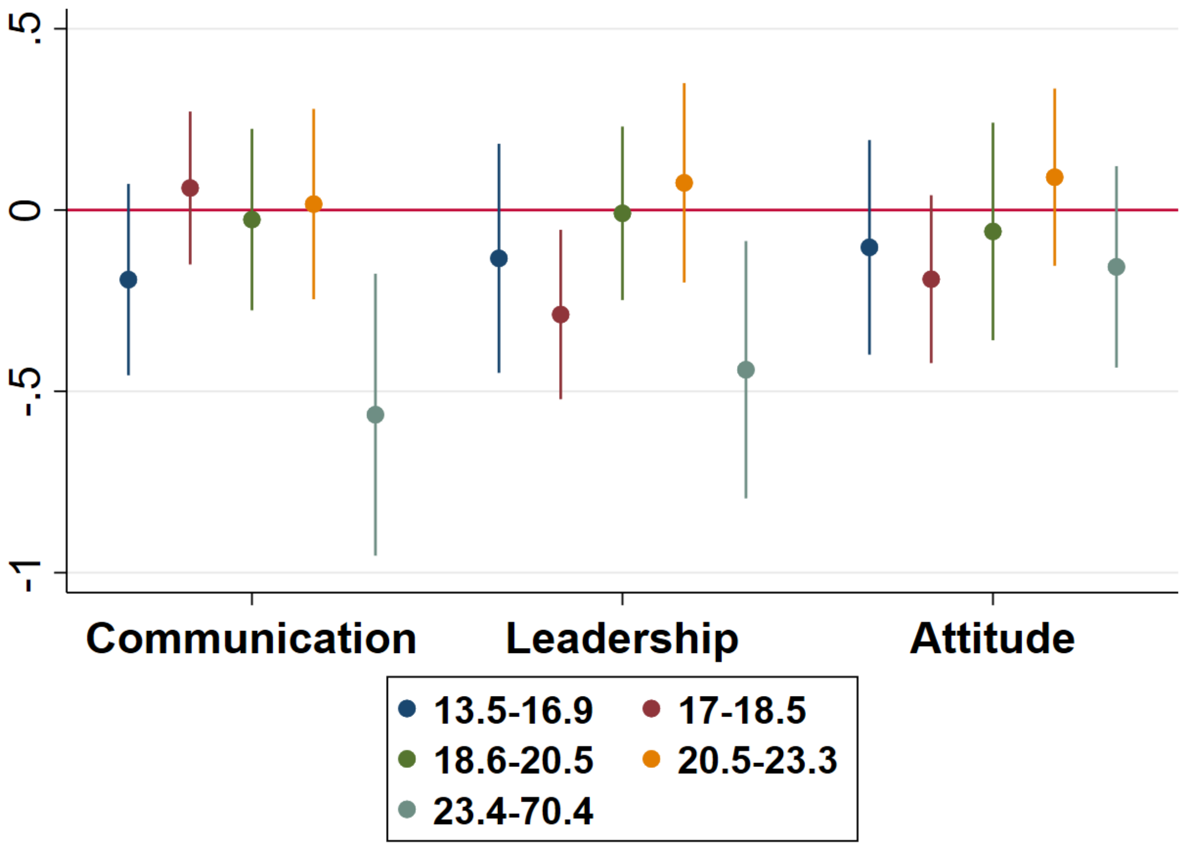

**Note:** N = 3,039 in figures A and E. N = 3,035 in Figure C. N = 2,755 in Figure B. N = 2,754 in Figure D. N= 2,759 in Figure E. Optimal bins for variance-bias tradeoff (for integrated mean square error) defined by Cattaneo et al. (2019). Panel A, C, and E are conditional on peer group average and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

Table A.9: The Effect of Baseline Peer Average Composite Ability on Own Course Grade

| Panel A: Communication | (1) | (2) |
|---|---|---|
| Baseline Peer Average Comm. | 0.15** | 0.07 |
| | (0.08) | (0.09) |
| Observations | 3039 | 2755 |
| $R^2$ | 0.0228 | 0.104 |
| Covariates | | X |
| Course Indicators | | X |
| **Panel B: Leadership** | (1) | (2) |
| Baseline Peer Average Lead | 0.06 | 0.04 |
| | (0.06) | (0.08) |
| Observations | 3035 | 2754 |
| $R^2$ | 0.0741 | 0.153 |
| Covariates | | X |
| Course Indicators | | X |
| **Panel C: Attitude** | (1) | (2) |
| Baseline Peer Average Att. | 0.06 | 0.06 |
| | (0.07) | (0.08) |
| Observations | 3039 | 2759 |
| $R^2$ | 0.0354 | 0.121 |
| Covariates | | X |
| Course Indicators | | X |
| **Panel D: All Peer Abilities** | (1) | (2) |
| Baseline Peer Average Att. | -0.10 | 0.02 |
| | (0.10) | (0.10) |
| Baseline Peer Average Comm. | 0.08 | 0.03 |
| | (0.11) | (0.11) |
| Baseline Peer Average Lead | 0.07 | 0.02 |
| | (0.09) | (0.10) |
| Observations | 2964 | 2690 |
| $R^2$ | 0.0794 | 0.160 |
| Covariates | | X |
| Course Indicators | | 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 all panels, the dependent variable is the recoded integer (1-12) grade. In Panel A, the independent variable is pre-course peer average communication for the same question. In Panel B, the independent variable is pre-course peer average leadership for that same question. In Panel C, 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, gender, and previous experience. The course indicators include course type, section year, and section month.

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.

# 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.

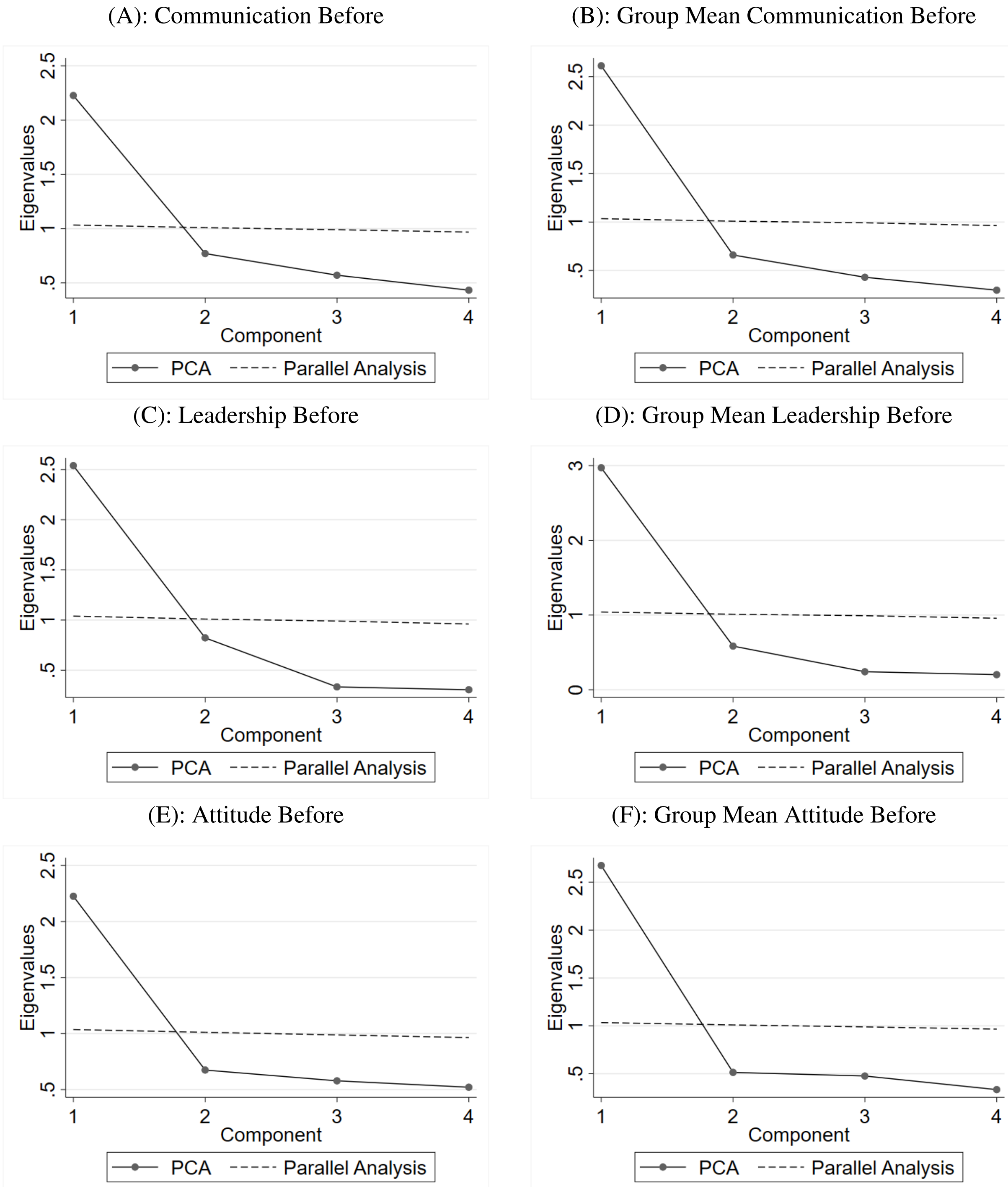
An alternative method which could be used to reduce the dimensions of each ability is principal components analysis (PCA). In PCA, the components of the correlation matrix are found, such that each component is orthogonal to the others. So an alternative way for combining the various questions into 1 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.

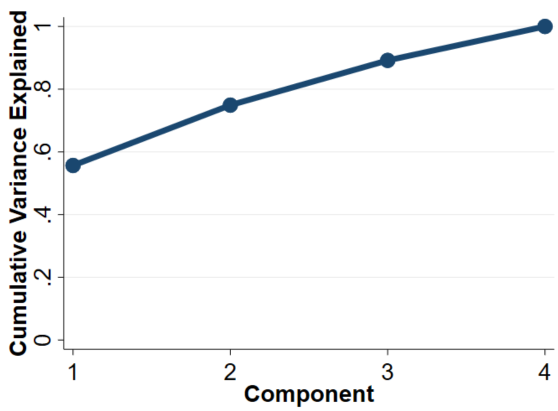
Figure B.1: Parallel Analysis



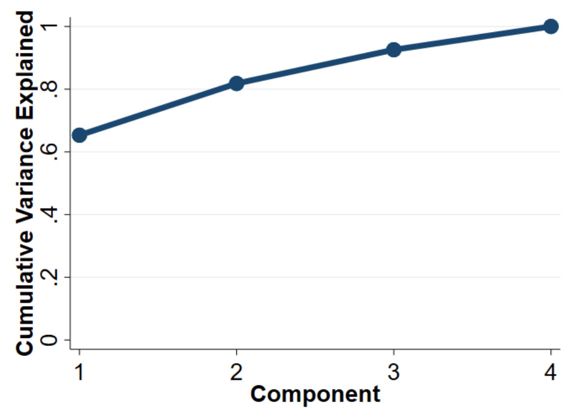

**Note:** Eigenvalues come from principal components analysis (PCA), of the correlation matrix, for the 4 questions for each ability. The parallel analysis uses 10 repetitions. Figures A, C, and E use the individual questions. Figures B, D, and F use the group means.

Figure B.2: Cumulative Variance

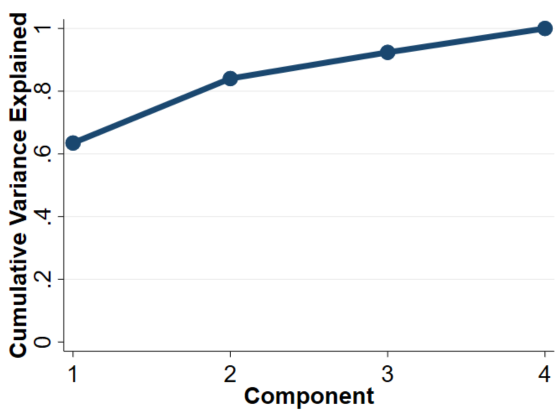
(A): Communication Before



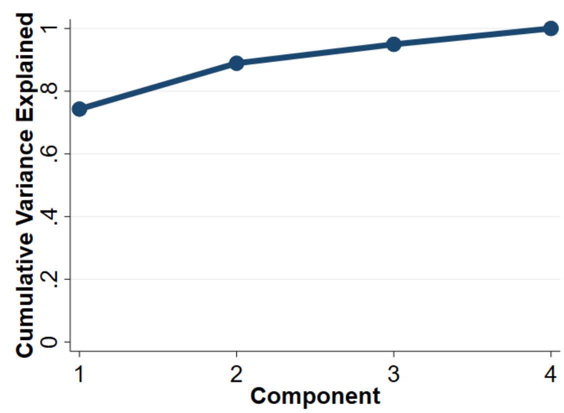

(B): Group Mean Communication Before



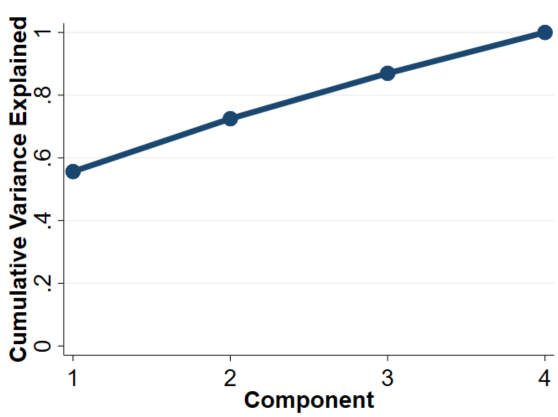

(C): Leadership Before



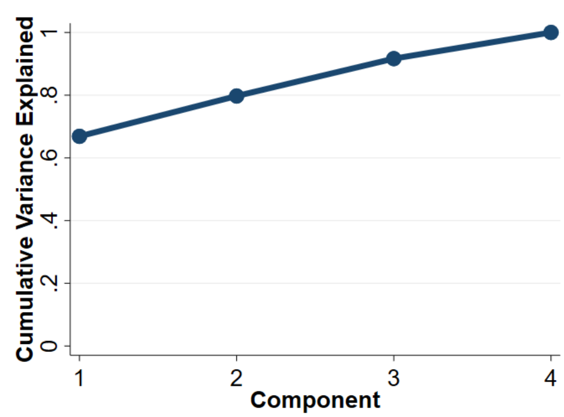

(D): Group Mean Leadership Before




(E): Attitude Before

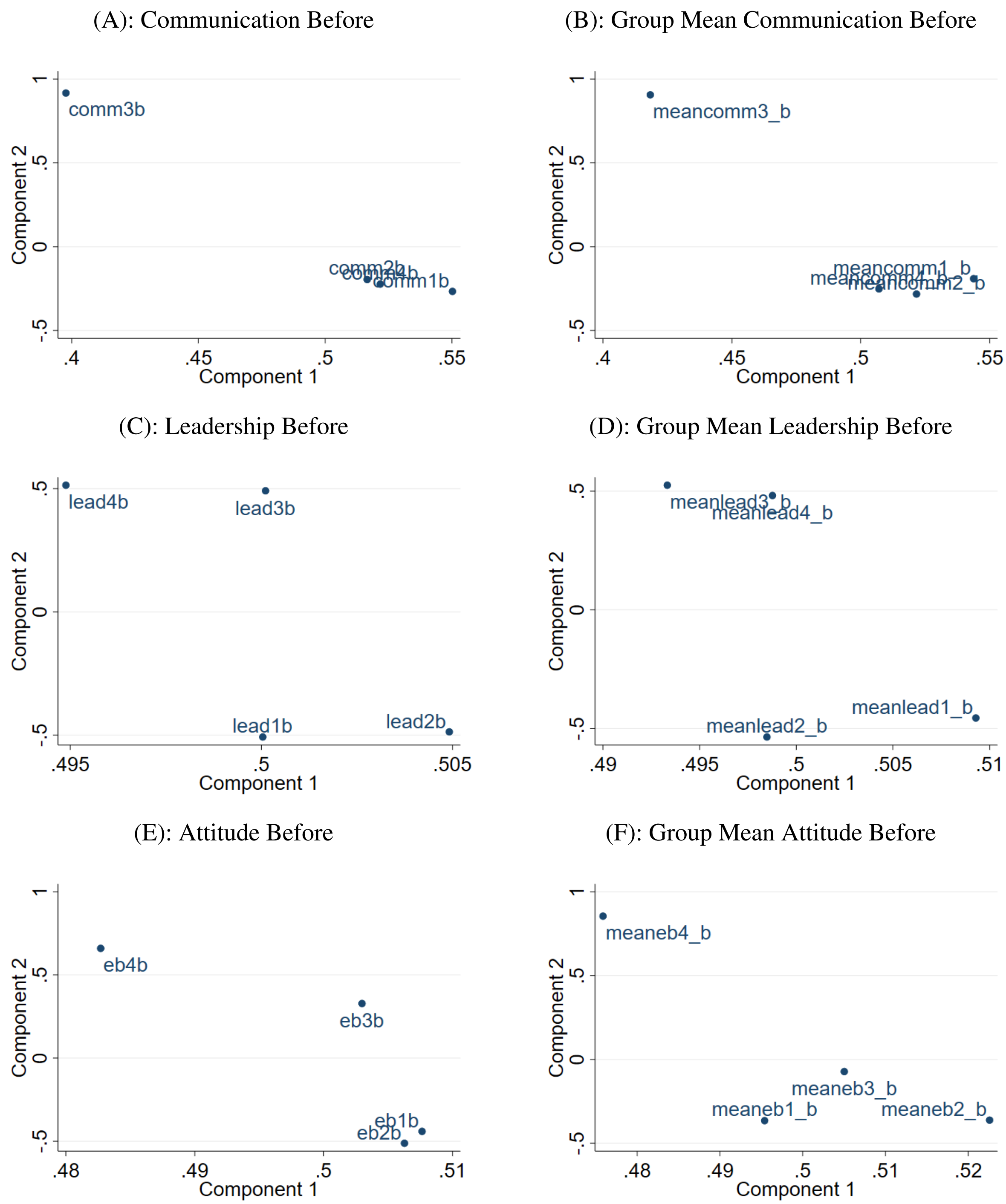



(F): Group Mean Attitude Before




**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




**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.