The effect of teammate personality on team production

Stacey A. Hancock, Andrew J. Hill

The Effect of Teammate Personality on Team Production[1]

Stacey A. Hancock

Andrew J. Hill

Abstract: Many goods and services are produced in teams. We explore how teammate personality traits impact productivity on joint team tasks. Studying student teams at a large university and considering the "Big Five" personality characteristics of extroversion, agreeableness, conscientiousness, emotional stability, and openness, we find that teammate conscientiousness has a small, positive impact on team performance: a one standard deviation increase in teammate conscientiousness increases performance on a team task by about three percent of a standard deviation in our preferred specification. The effect is evident holding teammate ability and gender fixed, and suggestively operates through improved team functioning and sustained increases in student effort. We also find evidence of positive spillovers from teammate openness and negative spillovers from teammate extroversion.

Keywords: personality traits, , peer effects, team production,

JEL classification: I21, I24, J24

[1] Hancock: Department of Mathematical Sciences, Montana State University (email: stacey.hancock@montana.edu); Hill; Department of Agricultural Economics & Economics, Montana State University (email: andrew.hill6@montana.edu). The authors thank the course instructors of STAT 216 at Montana State University for cooperating with our study; we are grateful to Thomas Dee, Ulf Zölitz, and seminar participants at the 2019 Jacobs Center Workshop on the Economics of Education and the 2020 Annual Meeting of the Association for Education Finance & Policy for helpful comments and discussions; and Andrew Hill appreciates funding from the Initiative for Regulation and Applied Economics Analysis.

1. Introduction

The extent to which personality traits can predict and explain behavior has been debated for decades. Increasing evidence indicates that they causally impact schooling and labor market outcomes (Borghans et al., 2008; Almlund et al., 2011; Borghans et al., 2016). But, many goods and services are produced in teams, and we know less about whether personality factors exert externalities on those around them, their teammates. Given the established evidence that peer attributes such as ability and gender affect individual decision-making and outcomes[2], the potential for people to be impacted by the personality traits of their peers at school or work is clear. This paper adds to our understanding of peer spillovers by investigating whether the productivity of small teams is affected by the personality characteristics of team members.

We study three-person student teams in a large statistics course in college. Students are observed in multiple teams through the semester. Team production is measured by performance on team tests that are jointly taken in class by team members that have experience working together. This environment is particularly well-suited for studying peer effects within teams for three reasons. First, students are randomly assigned into teams, so potential nonrandom selection into peer groups is not a concern, and we can identify causal peer composition effects. Second, the course is structured so that students work repeatedly with their teammates for a month in a flipped classroom environment before taking the tests, providing an opportunity to observe both cumulative and contemporaneous personality spillovers. And, third, the team test questions we use to measure team production are subsets of individual test questions that are taken by students a few days before the team tests, so we have highly comparable measures of individual and team output that allow us to isolate the "value-added" of working with teammates rather than as an individual.

[2] An abundance of papers show that an individual's actions are affected by the characteristics of their peers. Prominent studies showing peers matter in educational settings include Duflo, Dupas & Kremer (2011), Lavy & Schlosser (2011) and Imberman, Kugler & Sacerdote (2012), and well-known examples of peer effects in work environments are Mas & Moretti (2009), Jackson & Bruegmann (2009) and Guryan, Kroft & Notowidigdo (2009).

We find that a one standard deviation increase in teammate conscientiousness boosts question scores on team tests by about three percent of a standard deviation. This effect is small, but is similar in magnitude to other recently-estimated peer personality effects (such as Golsteyn et al., 2019). More suggestively, we also find negative effects of teammate extroversion and positive effects of teammate openness. Our richest specification includes student fixed effects and controls for the performances of team members on the recently taken individual test, so our estimated effect can be interpreted as the impact of peer personality holding the ability of the team constant. The estimates are also robust to including teammate gender and a measure of teammate baseline statistics comprehension in the model.

Conscientious teammates may improve team performance through a variety of channels. Indirectly, conscientious students are likely to work harder and exert more effort when studying on their own, and therefore may bring more knowledge and understanding into the team environment; we consider this an indirect ability effect. Directly, conscientious peers may positively influence the study efforts of teammates through the example they set, as well as improve team functioning by bringing focus and discipline to the team. In addition, external inputs into team production by instructors or other peers may be affected by the conscientiousness of team members.

We cannot explore all potential mechanisms with the data we have available, but consider a subset of prominent candidates. We speak to indirect ability effects by controlling for the recent individual achievement of teammates, but focus our attention on probing the direct mechanisms. First, we test if conscientious students induce changes in their peers' study behavior by investigating whether the positive impacts of exposure to conscientious peers persist when the student is reassigned to a new team; if individual effects are sustained, students are likely changing their own work habits. And, second, we explore whether conscientious teammates improve team functioning by assessing if effects are larger on more complex tasks requiring more team coordination. Our results provide moderate support for both channels: teammate conscientiousness positively affects *future* individual test achievement, suggesting peer personality affects study behavior, and the impacts of

teammate conscientiousness are somewhat more evident on complex questions requiring the most group coordination, evidence of better teamwork.

In our study, we measure personality factors using the "Big Five" personality traits of extroversion, agreeableness, conscientiousness, emotional stability, and openness. It is argued that this five-factor model can fully describe an individual's personality.[3] Although similar ideas existed earlier, the "Big Five" model in its current form was motivated by Goldberg (1990) who categorized a near-exhaustive list of known personality traits into these five factors. Many similar analyses have followed. We summarize the model in colloquial terms in the below table adapted from Diener & Lucas (2020):

Trait	**Example behavior for low scores**	**Example behavior for high scores**
Extroversion	Prefers a quiet evening reading to a loud party; sober; aloof; unenthusiastic	Being the life of the party; active; optimistic; fun-loving; affectionate
Agreeableness	Quickly and confidently asserts own rights; irritable; manipulative; uncooperative; rude	Agrees with others about political opinions; good-natured; forgiving; gullible; helpful
Conscientiousness	Prefers spur-of-the-moment action to planning; unreliable; hedonistic; careless; lax	Never late for a date; organized; hardworking; neat; persevering; punctual; self-disciplined
Emotional stability	Constantly worried about little things; insecure; hypochondriacal; feeling inadequate	Not getting irritated by small annoyances; calm, unemotional; hardy; secure; self-satisfied
Openness	Prefers not to be exposed to alterative moral systems; narrow interests; inartistic; not analytical; down-to-earth	Enjoys seeing people with new types of haircuts and body piercings; curious; imaginative; untraditional

Our paper is most closely related to Golsteyn, Non & Zölitz (2019) who investigate peer personality effects on individual achievement in a business school. Studying teaching sections of up to 16 students, they find that peer persistence, which is likely highly correlated with peer conscientiousness, increases individual test performance by about two percent of a standard deviation, which is similar in magnitude to our estimated effect on team test performance. Our paper has important distinguishing features. We primarily consider team output rather

[3] The measurement of personality traits is a topic of considerable debate in the psychology literature. We do not delve into this discussion here, but point readers to Widiger (2017) for a comprehensive treatment of the "Big Five" model.

than individual output, and our peer groups are teams of three that work together on joint tasks in class rather than larger teaching sections. This is an important distinction; the extent to which individuals are affected by teammates on a joint task has parallels with many employment settings, and contrasts with the "softer" influence of peers in larger groups on individual output explored by Golsteyn et al. (2019). However, the unified evidence of hard working peers exerting positive externalities on their peers emphasizes the importance of personality traits in social interactions. This conclusion is also supported by Shure (2017), who uses within-school across-cohort variation to show that peer conscientiousness is positively related to secondary school grades, and Bietenbeck (2020), who uses data from Project STAR to show that having motivated classmates improves contemporaneous reading achievement.

In a recent paper, Weidmann and Deming (2021) conduct a novel laboratory experiment to identify individual contributions to team production. "Team players", individuals who consistently cause their team to outperform their team's predicted production based on individual skills, are found to be endowed with more social intelligence, but are no different in terms of IQ, personality, education, and gender, in comparison to other team members. Our study shares two important features with theirs – measuring team performance controlling for the task-specific skills of individual team members, and observing individuals in multiple teams – but a key difference across our settings is the extent of team interactions. In our paper, teams work together for a month rather than once in a laboratory; it is not surprising that we find that teammate personality matters for team performance given team members have interacted repeatedly and over an extended period.

Two factors separate our findings from a literature in psychology that explores correlational relationships between team personality composition and team performance. The randomization of individuals into teams in our setting allows us to identify causal effects of team composition rather than associations, and the availability of highly comparable measures of individual and team performance allows us to explore effects on team "value-added", the production of the team controlling for the productive capacity of individual team members. Peeters, Van Tujil, Rutte & Reymen (2006), in a meta-analysis of this literature, find evidence of a positive association

between team conscientiousness and team output, similar to the causal effect we identify, but they also report a positive relationship between team agreeableness and team performance, which we do not find. Bell (2007) also uses meta-analytic techniques to provide an overall assessment of the psychology literature exploring relationships between team composition and team performance. She also finds that team mean conscientiousness is positively associated with team performance, but only in field settings and not in laboratory settings. This important distinction helps explain the difference between our finding from the field and that of Weidmann and Deming discussed above from a laboratory experiment in which no personality factor was associated with being a "team player".

There is an established literature investigating peer effects in educational settings. Several papers exploit random assignment to peer groups in college as is done in our paper. Much of this work considers the impacts of college roommates (Sacerdote, 2001; Zimmerman, 2003; Stinebrickner & Stinebrickner, 2006) and peers in military academies (Carrell, Fullerton & West, 2009; Lyle, 2009; Jones & Kofoed, 2020), while more recent studies consider college cohorts and classrooms (Booij, Leuven & Oosterbeek, 2017; Feld & Zölitz, 2017; Hill, 2017; Brenoe & Zölitz, 2019). These papers provide evidence that peer ability, peer gender and peer race all affect individual behavior and decision-making in college; in other words, there are spillovers associated with fixed peer attributes in college.

A smaller set of papers in the peer effects literature speaks to the possibility of behavioral-type peer externalities that are more similar to the peer personality impacts studied in this paper. In the college environment, Eisenberg, Golberstein & Whitlock (2014) find evidence of peer effects in binge drinking; Figlio (2007) uses a creative name-based approach to show that boys more likely to be disruptive in class exert negative externalities on their peers in high school; and, in an elementary school context, Carrell & Hoekstra (2010) show that students whose in-class behavior is poor due to their home environments negatively impact their classmates. To the extent that individual behavior is predicted by the "Big Five" personality characteristics we study, our paper adds to this evidence that in-class behavior affects peers.

Our paper also speaks to a large literature on the relationships between "Big Five" personality traits and individual academic achievement. The findings we report on the causal effects of *peer* personality traits are similar to some of the associations between *individual* personality characteristics and achievement reported in the psychology literature. For example, Noftle & Robins (2007) find that individual conscientiousness is a strong predictor of both high school and college GPA; they report unconditional correlations of 0.20 to 0.24 between measures of conscientiousness and high school GPA in some models, which are an order of magnitude larger than the causal effects we document for peer conscientiousness. Saying that, magnitudes can be difficult to compare across studies and environments, and the associations between own conscientiousness and test performance that we subsequently report in this paper in a directly comparable setting are similar in magnitude to our estimated effects of teammate conscientiousness. And a meta-analysis conducted by O'Connor & Paunonen (2007) not only finds this conscientiousness-achievement relationship, but also reports suggestive evidence that school performance is negatively correlated with extroversion and positively associated with openness; the latter relationships are in the same direction as the secondary effects we find for peer extroversion and peer openness. Taken as a whole, the personality traits associated with individual achievement appear to causally affect the performance of peers in similar ways.

In addition, evidence of personality spillovers is important for policy evaluation. Given the increasing empirical support that educational interventions improve socio-emotional skills (Dee & West, 2011; West, Kraft & Finn, 2016), the potential for socio-emotional skills to be related to the personality factors we study, and that these impacts may ultimately be more persistent than the effects on test scores (Heckman, Pinto & Savelyev, 2013), we underestimate the returns to policies if we do not incorporate positive externalities such as those found in our study into program evaluations.

2. Data

Our data come from an introductory statistics course at a large state school. Over 800 students who register for the course every semester are divided into about 20 sections, each of which is taught by a different instructor. The course is structured so that students within each section work in teams of three in a flipped classroom environment. Students belong to three different teams during the semester; they work with their first team up until the first test, a second team up until the second test, and a third team up until the final examination. This provides an ideal natural setting for investigating team production.

We made three small changes to the course for the purpose of this study.[4] First, we ensured that students were randomly assigned into teams. This was done by the course administrator, who provided instructors with team assignments through the semester. Although it is not the intention for instructors or students to choose teams, this had happened to a limited extent in the past in some sections. Second, we introduced team tests (called "exam reflections") to be taken a few days after the regular individual tests, but before the students receive any feedback on their individual tests. These team tests happen in class, each team submits a single set of answers, and the team tests consist wholly of a subset of questions from the individual tests, allowing us to observe perfectly comparable measures of individual and team outputs.[5] The last point has been a challenge in previous team production research. For example, Devereaux (2019) provides an interesting analysis of team production in tennis, but is partly limited by the extent to which singles and doubles in tennis require different skillsets. And, third, we augmented a baseline test of statistics comprehension – the Comprehensive Assessment of Outcomes in a First Statistics Course (CAOS) test (delMas, Garfield, Ooms & Chance, 2007) – that is taken online by all students in the first week of class with a 50-question personality test from the International Personality Item Pool (IPIP).[6] The IPIP test scores individuals on a scale of one to five on each of the "Big Five" personality traits: extroversion, agreeableness, conscientiousness, emotional stability, and openness. Higher scores reflect more of the associated trait.

[4] We received IRB exemption for this study; the groupwork we study falls under normal education practices.
[5] A sample exam reflection is provided in Appendix A2.
[6] The personality test is available in Appendix A1. More information about IPIP can be found at <https://ipip.ori.org/>.

Table 1 reports that 802 unique students took at least one of the personality, individual or team tests and therefore appear in our data. Out of these students, 86 percent took the personality test. Turning to the actual coursework, 94 percent took the first individual test; and this declines for each subsequent test, consistent with students dropping out or losing focus as the semester progresses. We define the main sample as students who took the personality test and at least one other test, the minimum requirement for an individual to contribute to our understanding of the relationship between personality traits and performance, and the balanced sample as students we observe taking all tests. Out of the 802 students who appear in our data, 75 percent are in the main sample and 62 percent are in the balanced sample. Teammate characteristics are computed using available data, so are based on a subset of teammates when teammate data is missing. About a third of students miss at least one test during the semester. Students are randomly assigned into teams, so, conditional on observing teammates, selection into the sample cannot affect peer characteristics, our explanatory variables of interest; nonetheless, we will explore the extent to which student characteristics systematically predict missing tests in a subsequent table.

The primary variables in our study are described in Table 2. Columns 1 and 2 report overall means, while Columns 3 and 4 split the main sample by gender. Note that there are a small number of students in the main sample without information on gender, so the sample sizes in Columns 3 and 4 do not add to the sample size in Column 1.

As mentioned above, personality traits are measured on a scale of one to five with higher scores indicating that the individual exhibits more of the respective personality trait. We report raw means in Table 2, although use normalized z-scores in the regression analysis for ease of interpretation. The first row indicates that males and females are similarly extroverted. Females are more agreeable and conscientiousness than males, and less emotionally stable and open. These gender differences are consistent with observations in the psychology literature (Weisberg, DeYoung & Hirsh, 2011; Vecchione, Alessandri, Barbaranelli & Caprara, 2012). Appendix

Table 1 reports descriptive statistics separately for the samples of students taking each test; this provides an initial check of potential selection and attrition in test-taking.

The distributions of both individual personality measures and teammate means of personality measures are plotted in Figure 1. The gray bars indicate that all the individual personality characteristics are symmetrically distributed other than agreeableness, which is skewed to the left; in other words, people tend to self-report that they are highly agreeable. We also observe that extroversion and emotional stability are supported by the full range of scores from 1 to 5, while there are almost no scores below 2 for the other three traits. We see similar distributions for teammate means of the personality characteristics in the outlined bars. The considerable amount of variation in the peer personality characteristics is not surprising since teams consist of only three members; other papers in the peer effects literature encounter more compressed variation given peer groups are often defined to be larger classes or grades.

The correlations between personality traits are shown in Table 3. These should be low given personality trait factor models such as the "Big Five" are intended to reduce the dimensionality of the personality space. This is confirmed.

Our sample is representative of the general college population given the course we study is a first year course required for majors across many disciplines. Returning to Table 2, the equal gender split we observe is broadly consistent with this (although the college we study is about 52% male). The consequent generalizability of our findings is a potential advantage over other college peer effects papers set in military academies or professional schools that may be less representative of the college population. There is no statistically discernible gender difference in baseline statistics comprehension, as well as on the first individual test. By the second test, however, females outperform males. This evidence of a widening of the gender achievement gap over time within a course is difficult to attribute to anything other than course-specific inputs such as student effort. Students score about 12 to 13 percentage points more on the team tests than the individual tests, which is

consistent with students in the team tests both collectively improving their answers and solving problems none of them could do on their own. The improvement between the individual and team tests also indicates that students apply themselves on the team tests, which is important to show because the team tests only enter the calculation of the final course grade through a small completion credit, so students may perceive less incentive to exert effort.

Table 4 describes key variables at the student-question level, which corresponds to the unit of observation in the regression analysis. The means of the teammate personality traits are expectedly similar to the student-level means reported in Table 2. Our primary outcome variable is the team question score. These are consistently higher than individual question scores, which is in line with the above observation that, on average, team test scores are greater than individual test scores.

Interestingly, the mean maximum score within a team from the individual test (reflecting the performance of the best team member on each question on the individual test) is higher than the mean team question score, reflecting that teams often underperform relative to the best team member on a question. This is consistent with existing evidence documenting a lack of synergy in almost all small-group tasks (Larson, 2013). This finding is confirmed by the three indicators that show whether a team gains, stays the same, or loses points relative to the best team member on a given question: in the main sample, 16 percent of student-question observations are associated with gains, 57 percent with staying the same, and 27 percent with losses. This asymmetry is partly driven by about two-thirds of observations coming from student-questions where the best team member scored the maximum possible score, making it impossible for the team to gain. Note that teams underperforming relative to the best team member on a question is not inconsistent with mean team question scores being greater than mean individual question scores; teams typically consist of three members, allowing for two team members to gain points on the team test relative to the individual test even if the best team member on the question loses points.

Observed team sizes are typically three in both the first and second teams, although there are some teams of two or four due to section enrollment not being a multiple of three, absences on the day of team tests, and a small number of teammates with missing data.

We explore selection into the main and balanced samples in Table 5. The table reports results from regressing a binary indicator for being in the respective sample on individual personality traits, gender, and performance on the baseline test of statistics comprehension. Recall that selection into the sample cannot bias estimates of teammate effects given random team assignment, but can affect the pool of potential teammates. There are 675 students in the sample for this analysis; these are students who both took the personality test and for whom we observe gender and the measure of baseline statistics comprehension. There appears to be a positive relationship between emotional stability and test-taking behavior, and a more suggestive positive association between baseline test performance and sample inclusion. Overall, our interpretation of Table 5 is that there is minimal systematic selection into the sample, especially on personality traits such as conscientiousness for which we subsequently find impacts.

Finally, Table 6 is addressed at supporting the claim that students are randomly assigned into teams. We regress the teammate mean of each "Big Five" personality trait on a vector of individual and team characteristics. The first seven rows of this table reveal that teammate personality traits and individual characteristics are generally unrelated; out of 35 estimated parameters across the five models, three are statistically significant at conventional levels (and one would expect two of 35 tests to be statistically significant just by chance). We do, however, observe that teammate female share is correlated with teammate personality. This is expected; if females, on average, are more agreeable and conscientious, and less emotionally stable and open (as shown in Table 1), teammate personality and teammate gender will be systematically related. We observe very small but sometimes precisely estimated relationships between teammate baseline test performance and teammate characteristics.

3. Empirical Strategy

Before explaining the empirical approach behind our main investigation of the impact of teammate personality on team performance, we introduce our analysis of associations between students' personality measures and their own test scores. We estimate the simple model:

$$IndQScore_{iq} = \alpha_0 + \alpha_1 IndPers_i + \alpha_2 Q_q + \alpha_3 X_i + \varepsilon_{iq}, \qquad (1)$$

where the question score of student i on question q on the individual test $IndQScore_{iq}$ is regressed on a vector of student i's personality characteristics $IndPers_i$, question fixed effects Q_q, and a vector of other individual controls X_i. The vector of parameters α_1 reports the partial correlations between students' individual test scores and normalized measures of their extroversion, agreeableness, conscientiousness, emotional stability, and openness, the "Big Five" personality traits. Note that this exercise does not provide causal estimates; we are not isolating exogenous variation in individual personality traits, and there are many factors correlated with the personality measures in the error term. The data are at the student-question level; question fixed effects Q_q are included to capture question-specific factors, such as question difficulty and maximum possible question score, and the errors are clustered at the student level.

Recall that students first take the test as an individual and then a few days later as a team. Although our primary focus is on the team outcome, we first report how teammate personality traits affect performance on the individual test (Equation 2). The effects we observe on the individual test can be attributed to changes in comprehension or study behavior caused by working with a specific set of teammates for a month in the flipped classroom environment. This parallels the general set-up in many educational peer effects papers that explore how exposure to a given set of peers affects a subsequent individual task (such as a test) or a decision (such as major choice). We take this a step further in our main analysis and consider peer effects in group tasks. Our

primary model (Equation 3) involves regressing a measure of team production on a vector of teammate personality traits.

$$IndQScore_{itq} = \delta_0 + \delta_1 TeamPers_{-i,t} + \delta_2 Q_q + \delta_3 IndPers_i + \delta_4 X_i + \delta_5 OtherTeamChars_{-i,t} + \varepsilon_{itq}. \quad (2)$$

$$TeamQScore_{[i]tq} = \beta_0 + \beta_1 TeamPers_{-i,t} + \beta_2 Q_q + \beta_3 IndPers_i + \beta_4 X_i + \beta_5 OtherTeamChars_{-i,t} + \beta_6 IndQScore_{itq} + \varepsilon_{itq}. (3)$$

Team production is measured by the score of student i in team t on question q on the team test, $TeamQScore_{[i]tq}$, where the [i] indicates that student i shares this score with their teammates. For ease of interpretation, question scores are normalized across all the student-question observations to have a mean of zero and a standard deviation of one. This is preferred to normalizing at the individual question level, which would force each question score to have the same amount of score variation, although using this alternative normalization does not affect the findings. The explanatory variable of interest $TeamPers_{-i,t}$ is a vector of normalized measures of the "Big Five" personality traits of extroversion, agreeableness, conscientiousness, emotional stability, and openness. Given teams typically consist of three members, mean teammate personality measures are simply averages of the characteristics of the two other team members. Students are randomly assigned into teams, so teammate characteristics are orthogonal to observed or unobserved individual factors affecting team production. Therefore, the vector of parameters β_1 provides the causal impacts of each teammate personality trait on team performance. A variety of other controls are included depending on the model, some of which have important implications for how we interpret the coefficients on the $TeamPers_{-i,t}$ vector. [7]

[7] An alternative specification is at the team level, modeling team output as a function of mean team personality traits. Doing so, however, does not allow us to capture individual-level factors that affect a student across the two teams in which they are observed. Saying that, a team level analysis revealed similar findings to what we observe in the student level analysis, although some of the estimates are expectedly less precise given the smaller sample and fewer controls. These results are reported in Appendix Table 3.

Question fixed effects Q_q control for any differences across questions in difficulty and maximum possible question score. We control for individual personality traits $IndPers_i$, as well as a vector X_i of other individual characteristics, such as gender. Some specifications include a control for the student's score on the corresponding question on the individual test. This is intended to control for the cumulative effects of teammate exposure on individual comprehension acquired during the month of working with teammates in the flipped classroom environment. The model with this control provides an estimate of the "value-added" impact of working on a specific task with teammates of certain personalities, holding the prior knowledge that the student brings into the team environment fixed. With the above set of individual controls, our experiment is essentially to investigate the effects of varying teammate personality holding an individual's own characteristics, notably their personality and ability, fixed.

Personality characteristics are correlated with other individual attributes, such as ability and gender. To ensure that our teammate personality effects are not simply reflecting the impacts of teammate ability or teammate gender, we include a vector of other teammate characteristics $OtherTeamChars_{-i,t}$ in some specifications. The effect of teammate ability is captured in the regression model by including both the maximum and mean scores obtained by teammates of student i on the corresponding question q on the individual test. Our intention is to control for teammate potential measured at the start of the team test, the maximum ability or understanding of other team members on each question, as well as average teammate ability. In some sense, this measures the production possibilities frontier of teammates before working on the task together as a team.

Controlling for teammate ability affects our interpretation of the estimates. As discussed in the introduction, teammate personality can both directly and indirectly affect team performance. For example, first, a conscientious teammate may affect her teammates directly though her conscientiousness; she may work harder and with more diligence in team meetings and during flipped classroom exercises, and this may directly affect her team's production. Second, her conscientiousness may affect team performance indirectly through other

personal attributes that are functions of her conscientiousness, the most likely being her ability or comprehension; a conscientiousness teammate has worked harder in the past, has developed more knowledge and understanding, and is therefore more productive in a team environment.

When we do not control for teammate ability, our estimates reflect the combined direct and indirect effects of teammate personality. These are our parameters of interest if we are concerned with the overall, cumulative impacts of peer personality on team production. If, on the other hand, we fully control for teammate ability, we isolate the direct effect of peer personality on a joint task holding all else equal. We can think of this as the effect of a teammate experiencing a transitory shock to one of her personality characteristics during a group project. Given we do not anticipate fully capturing the many potential indirect effects of peer personality in our model, we primarily control for teammate ability to ensure that we are identifying peer personality effects, not peer ability effects, and only secondarily as a suggestive step towards separating the direct and indirect impacts of teammate personality.

We include controls for teammate gender, as well as teammate performance on the baseline test of statistics comprehension, in subsequent robustness checks. The latter provides a measure of underlying teammate ability that cannot be affected by team composition. The former is included because personality traits are not evenly distributed across genders and it has been shown in many contexts that greater shares of female peers are associated with increases in student performance (Lavy & Schlosser, 2011, among others). By controlling for teammate gender, we ensure that any productivity effects attributed to teammate personality are not teammate gender effects. These controls are especially important given the correlations we observe in Table 6.

The inclusion of teammate gender in the model also has the potential to provide insights into the gender peer effects literature. Specifically, it allows us to explore the extent to which gender peer effects may be explained by gender differences in personality traits. An outstanding question in the peer effects literature is why females tend to exert positive spillovers on their peers in educational environments; it may be the case that this is due

to differences in personality traits, a hypothesis that could be supported by our study if we find that gender peer effects are no longer evident when we control for peer personality.

4. Results

4.1. Main estimates

We begin the results section by reporting the correlations between student performance on the individual tests and their own personality traits in Table 7 (Equation 1). Each column corresponds to a regression model, all specifications include question fixed effects, individual controls are changed as we move from left to right, and the unit of observation is a student-question. As discussed above, these estimates are not causal.

Only one personality trait is statistically associated with achievement; a one standard deviation increase in conscientiousness is associated with a 0.021 to 0.027 standard deviation increase in question scores. Golsteyn et al. (2019) finds a similar correlation, although they consider the related personality measure of persistence rather than conscientiousness. We also observe a gender gap, with females outperforming males, as well as a positive association between baseline statistics comprehension and question scores, although this is relatively small (a correlation coefficient of 0.25), indicating that the baseline comprehension measure is not particularly predictive of subsequent test performance.[8] These associations between own personality factors and test scores are reported graphically in Appendix Figure 1.

Table 8 reports the effects of teammate personality on individual performance (Equation 2), which we show before the main results on team performance. Column 1 includes only the five mean teammate personality

[8] For comparison, results from a student-level regression of baseline statistics comprehension scores on the five personality traits are reported in Appendix Table 2. They are different to the estimates in Table 7, providing further evidence of the generally weak relationship between the measure of baseline statistics comprehension and test scores.

measures, Column 2 adds controls for the student's own personality traits, Column 3 includes both a gender indicator and a control for performance on the baseline test, and Column 4 replaces all the time-constant individual controls with student fixed effects. Models were not pre-specified, so checking sensitivity to controls throughout the analysis is helpful for ensuring that findings do not rely on particular specifications.

Conscientiousness is the only peer personality trait that statistically affects individual question scores. Columns 1 to 3 indicate that exposure to conscientious teammates during the month of flipped classroom groupwork increases question scores on the subsequent individual test; a one standard deviation increase in teammate conscientiousness improves question performance by a little under 0.03 standard deviations. However, this effect is not evident when student fixed effects are included in Column 4. Here, rather than estimating the impacts of peer personality using both across and within student comparisons, the identifying variation is only within students. In other words, the estimate is generated by comparing average performance after working with the first set of randomly assigned teammates to average performance after working with the second set of randomly assigned teammates. This model controls for all (fixed) unobserved and observed differences between students, but consequently leaves less identifying variation. We interpret the mixed results in Table 8 as suggestive evidence of positive spillovers from teammate conscientiousness on individual achievement.

The approach taken in Table 8 – considering impacts on an individual outcome after exposure to a set of peers – is standard in the peer effects literature. The overarching idea is that interactions with peers in classes, grades, or college affect a variety of attitudes and behaviors, and these subsequently affect a student's individual achievement or decision-making. As discussed above, one of the primary contributions of this paper is taking this a step further and assessing how peers affect each other when working together on a joint task, which we turn to now.

Our main results are reported in Table 9: the effect of teammate personality on team production (Equation 3). Recall that the team test is taken together by team members a few days after the individual test, but before

students receive any feedback on their individual tests. The controls are varied as we move across columns from left to right, question fixed effects are included in all specifications (which is largely responsible for the high R^2s), and student fixed effects replace fixed individual controls in Column 6. Errors are clustered at both the student level to account for correlated student-specific shocks across questions (and tests) and at the team level given scores on a given question are necessarily the same for all team members.

Columns 1 to 6 of Table 9 all reveal that teammate conscientiousness positively affects team performance; a one standard deviation increase in teammate conscientiousness increases team question scores by between 0.026 and 0.039 standard deviations. This is a small effect that only explains a small share of the variation in team performance, although the magnitude of our estimated teammate conscientiousness effect is similar to the peer persistence effect in Golsteyn et al. (2019). We also observe suggestive evidence that the effect of teammate extroversion is negative, although the precision of the estimated coefficient depends on the model. And, in our fullest specification (Column 6), we find that teammate openness positively affects team production. Taking the estimates in this table together, the best teammates are certainly hard-working, and, more suggestively, introverts open to new ideas. The positive peer conscientiousness and openness impacts are not surprising. Less expectedly, extroverts appear to negatively affect joint tasks on average, although perhaps their benefits in the team environment come in the form of conversation and entertainment.

The extent to which estimates respond to varying the controls adds to our understanding of the effect. Column 1 reports the effects of teammate personality characteristics controlling only for fixed question characteristics; we observe a modest, precise impact of teammate conscientiousness on team question scores. Moving from Columns 1 to 2, the addition of individual personality traits does not affect this estimate, which is expected given the generally weak correlation between individual personality and test scores reported in Table 7. Controlling for teammate ability in Column 3 explains about a quarter of the teammate conscientiousness effect (with the caveat that we do not have the power to statistically differentiate the estimates in Columns 2 and 3). This indicates that although some of the peer conscientiousness effect may be operating through peer ability, a

considerable share of the effect extends beyond that. Columns 4 and 5 add controls for the student's question score on the individual test, first without and then with the control for teammate ability. The effect of teammate conscientiousness is attenuated, but only slightly, providing evidence that teammate conscientiousness directly affects team production even after controlling for any cumulative effects on individual understanding from working with peers in the flipped class environment. The teammate conscientiousness impact remains relatively precise, although is smaller in magnitude, when individual controls (Column 5) or student fixed effects (Column 6) are added. The latter model is particularly demanding of the data given we only observe each student in two teams.

The regression results in Columns 1, 5 and 6 are presented graphically in the three panels of Figure 2. The graphical representation provides a reminder that the teammate conscientiousness effect is not just statistically significant, but is larger in magnitude than the effects of the other teammate personality traits.. The figure also highlights the persistence of the negative effect of teammate extroversion across specifications.

We report in Table 10 whether teammate personality effects are related to peer gender or teammate underlying ability. Column 1 reports the benchmark effects of teammate female share and teammate baseline performance without controlling for own or teammate personality. Column 2 reports results from the specification with personality controls (corresponding to Table 9, Column 2), but adding controls for teammate female share and teammate baseline test performance. The effect of a one standard deviation increase in peer conscientiousness is 0.036 standard deviations of a question score, similar to the estimates in Table 9. When we add further controls and student fixed effects in Columns 3 and 4, the magnitude of the teammate conscientiousness effect drops to slightly below the corresponding estimates in Table 9, Columns 5 and 6. There is less evidence that peer extroversion affects team performance in the models with teammate gender controls, which is somewhat surprising given we do not observe mean differences in extroversion by gender in Table 2. Figure 3 plots the comparisons between the models without (Table 9) and with (Table 10) teammate female share and teammate baseline test performance; the difference between the estimates in the left and right columns are generally small,

showing that peer personality effects and gender peer effects are generally orthogonal inputs into team production.

We can also observe from the results in Table 10 that teammate female share has a positive effect on team production, noting that this can be given a causal interpretation since teams are randomly assigned. The relationship is not precisely estimated across all specifications, although is consistently of a relatively large magnitude. Comparing Column 1 that does not control for peer personality and the remaining columns that do control for teammate characteristics indicates that the positive effect of female peers does not appear to be explained by gender differences in the "Big Five" personality traits, at least in our linear model..

As a final robustness check of the main finding, we report the results of a placebo test in Table 13. Here, we regress the team question scores from the first test on the personality traits of the student's teammates on the second test. Given teams are randomly re-assigned after the first test, there should be no relationship between future teammate characteristics and current team production, which is what we observe in Table 13.[9]

4.2. Probing the mechanism

We have shown that there is a suggestive peer personality impact on an individual task (Table 8), an unambiguous effect of teammate personality when working on a group task (Table 9), and that this team production effect persists holding other teammate characteristics constant (Table 10). The extent to which individual and teammate controls affect the estimates partly speaks to the mechanism, but we now turn to exploring this further.

[9] We also report regressions of two predetermined characteristics on teammate personality in Appendix Table 4. There are no systematic relationships across both tests, although some coefficients statistically differ from zero, which can be expected given the large number of estimates in the table.

We consider two potential channels for teammate personality effects, recognizing that these are neither exclusive nor exhaustive. First, we investigate whether conscientious students induce their peers to persistently work harder, or, in other words, whether conscientiousness is "contagious". If so, it may be the case that a student continues to exert more effort even when subsequently working with new teammates. Second, we explore whether conscientious peers improve group functioning. This could be the case, for example, if conscientious peers are more prepared and organized, and this high level of organization and preparedness allows the team to function particularly well when working together on more complex tasks that require more team coordination.

We probe persistence in peer personality effects in Table 11. This table reports results from regressing question scores from the second individual test (which is taken after working with the second set of teammates) on the personality characteristics of the first set of teammates. The results reported in Column 1 are from a model that does not control for individual performance on the first test, a measure of understanding obtained while working with the first set of teammates, while this control is added in Column 2.

The estimates in Columns 1 and 2 indicate that exposure to conscientious peers has a persistent positive effect on test performance, although the relationship is only precisely estimated in Column 1. Given that the inclusion of first test achievement attenuates the impact, some of this effect appears to be operating through the initial comprehension acquired while working with the first set of teammates. The evidence from this table is mixed. We cannot conclude that students who experience positive teammate conscientiousness shocks are definitively still working harder in subsequent teams, but since the magnitude of the effect in Column 2 is almost identical to the contemporaneous peer personality effect estimated in Table 9, this remains a plausible hypothesis.

Table 12 is aimed at exploring the effects of teammate conscientiousness on group functioning. Questions are partitioned into those with written answers, numeric answers, and circled answers (such as multiple-choice questions), and the main regression model is estimated separately on each subsample. Our hypothesis is that

written answers require the most group coordination, and therefore offer the most opportunity to benefit from improved team functioning. Circled answers, on the other hand, require the least coordination. The results reported in Table 12 indicate that the benefits of teammate conscientiousness are most evident in the more complex questions requiring written answers (Columns 1 and 2), although the differences across question types are not statistically significant, and should therefore be interpreted cautiously. To the extent that group functioning is most tested on complex tasks, conscientious peers appear to improve a team's ability to work together. We also observe that teammate openness has a suggestive positive impact on all but the least complex questions with circled answers, although these effects are more sensitive to the specification.

Taken together, Tables 11 and 12 provide some evidence that conscientious team members induce persistent increases in effort among their teammates and contemporaneous improvements in their team's ability to coordinate on more involved problems.

Finally, we note that it is likely that the effects of peer personality interact with peer gender and ability. Unfortunately, we do not have the statistical power to explore such interactions robustly.[10]

5. Conclusion

This paper adds to an emerging literature in economics investigating how individuals are affected by the personalities of their peers. We consider team production on joint tasks rather than individual production, which distinguishes our paper from some earlier work. Our study involves analyzing students randomly assigned to three-person teams performing joint tasks in a large introductory course at a state university. We show that peer conscientiousness has a small, robustly positive impact on group performance holding the ability

[10] We experimented with a lasso approach to select interactions of teammate characteristics that were most predictive of team performance, and, although we found some patterns that were reasonably robust and credible, the relatively limited data in our study does not make it an ideal environment for applying machine learning methodologies.

and gender composition of teams fixed. Secondary analyses point to increases in study effort and improvements in team functioning as potential mechanisms. We also find suggestive evidence that teams benefit when teammates are open and introverted, while teammate agreeableness and teammate emotional stability do not have consistent effects on team production.

REFERENCES

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In *Handbook of the Economics of Education* (Vol. 4, pp. 1-181). Elsevier.

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: a meta-analysis. *Journal of Applied Psychology*, *92*(3), 595.

Bietenbeck, J. (2020). Own Motivation, Peer Motivation, and Educational Success. Working paper.

Booij, A. S., Leuven, E., & Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies*, *84*(2), 547-578.

Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*, *43*(4), 972-1059.

Borghans, L., Golsteyn, B. H., Heckman, J. J., & Humphries, J. E. (2016). What grades and achievement tests measure. *Proceedings of the National Academy of Sciences*, *113*(47), 13354-13359.

Brenoe, A. A., & Zölitz, U. (2019). Exposure to more female peers widens the gender gap in STEM participation. *Journal of Labor Economics*.

Carrell, S. E., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, *27*(3), 439-464.

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

(2010)

Dee, T. S., & West, M. R. (2011). The non-cognitive returns to class size. *Educational Evaluation and Policy Analysis*, *33*(1), 23-46.

DelMas, G., Joan, G., Ooms, A., & Chance, B. (2007). Assessing Students' Conceptual Understanding After a First Course in Statistics. *Statistics Education Research Journal*, 6(2).

Diener, E. & Lucas, R. E. (2020). Personality traits. In R. Biswas-Diener & E. Diener (Eds), Noba textbook series: Psychology. Champaign, IL: DEF publishers. Retrieved from http://noba.to/96u8ecgw

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. *American Economic Review*, *101*(5), 1739-74.

Eisenberg, D., Golberstein, E., & Whitlock, J. L. (2014). Peer effects on risky behaviors: New evidence from college roommate assignments. *Journal of Health Economics*, *33*, 126-138.

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.

Figlio, D. N. (2007). Boys named Sue: Disruptive children and their peers. *Education Finance and Policy*, *2*(4), 376-394.

Goldberg, L. R. (1990). An alternative" description of personality": the big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216.

Golsteyn, B., Non, A., & Zölitz, U. (2019). The impact of peer personality on academic achievement. *University of Zurich, Department of Economics, Working Paper*, (269).

Guryan, J., Kroft, K., & Notowidigdo, M. J. (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics*, *1*(4), 34-68.

Heckman, J., Pinto, R., & Savelyev, P. (2013). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review*, 103(6), 2052-86.

Hill, A. J. (2017). The positive influence of female college students on their male peers. *Labour Economics*, *44*, 151-160.

Imberman, S. A., Kugler, A. D., & Sacerdote, B. I. (2012). Katrina's children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, *102*(5), 2048-82.

Jackson, C. K., & Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics*, *1*(4), 85-108.

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.

Larson Jr, J. R. (2013). *In search of synergy in small group performance*. Psychology Press.

Lavy, V., & Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, *3*(2), 1-33.

Lyle, D. S. (2009). The effects of peer group heterogeneity on the production of human capital at West Point. *American Economic Journal: Applied Economics*, *1*(4), 69-84.

Mas, A., & Moretti, E. (2009). Peers at work. *American Economic Review*, *99*(1), 112-45.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, 93(1), 116–130.

O'Connor, M. C., & Paunonen, S. V. (2007). Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*, *43*(5), 971-990.

Peeters, M. A., Van Tuijl, H. F., Rutte, C. G., & Reymen, I. M. (2006). Personality and team performance: a meta-analysis. *European Journal of Personality*, *20*(5), 377-396.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, *116*(2), 681-704.

Shure, N. (2017). Non-cognitive peer effects in secondary education. Unpublished manuscript.

Stinebrickner, R., & Stinebrickner, T. R. (2006). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics*, *90*(8-9), 1435-1454.

Vecchione, M., Alessandri, G., Barbaranelli, C., & Caprara, G. (2012). Gender differences in the Big Five personality development: A longitudinal investigation from late adolescence to emerging adulthood. *Personality and Individual Differences*, *53*(6), 740-746.

Weidmann, B., & Deming, D. J. (2021). Team players: How social skills improve team performance. *Econometrica*, 89(6), 2637-2657.Weisberg, Y. J., DeYoung, C. G., & Hirsh, J. B. (2011). Gender differences in personality across the ten aspects of the Big Five. *Frontiers in Psychology*, *2*, 178.

West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., & Gabrieli, J. D. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, *38*(1), 148-170.

Widiger, T. A. (Ed.). (2017). *The Oxford handbook of the five-factor model*. Oxford University Press.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics*, *85*(1), 9-23.

Tables and Figures

Table 1. Descriptive statistics: sample inclusion

Share of students who took:	
Personality test	0.86
Individual test 1	0.94
Team test 1	0.86
Individual test 2	0.83
Team test 2	0.79
Share of students in:	
Main sample (took personality test and at least one team test)	0.75
Balanced sample (took all above tests)	0.62
Observations (unique students who took at least one of the above tests)	802

Table 2. Descriptive statistics: personality measures and test performance (means and standard deviations)

	(1) Main sample	(2) Balanced sample	(3) Males	(4) Females	(5) p-value on t-test of gender difference
Extroversion	3.04	3.06	3.04	3.05	0.265
	(0.75)	(0.75)	(0.72)	(0.78)	
Agreeableness	3.78	3.79	3.61	3.95	0.000
	(0.61)	(0.60)	(0.57)	(0.58)	
Conscientiousness	3.57	3.59	3.48	3.67	0.000
	(0.59)	(0.58)	(0.55)	(0.61)	
Emotional stability	3.13	3.16	3.28	2.98	0.000
	(0.69)	(0.68)	(0.68)	(0.66)	
Openness	3.49	3.49	3.54	3.45	0.038
	(0.50)	(0.51)	(0.48)	(0.52)	
Female	0.51	0.51	0.00	1.00	
	(0.50)	(0.50)	(0.00)	(0.00)	
Baseline test %	49.99	50.15	50.28	49.74	0.466
	(10.50)	(10.45)	(11.00)	(10.08)	
Individual test 1 %	67.00	67.50	66.35	67.48	0.593
	(17.53)	(17.16)	(17.17)	(17.96)	
Team test 1 %	79.70	79.39	78.70	80.62	0.093
	(12.78)	(12.85)	(13.56)	(12.04)	
Individual test 2 %	62.40	62.08	60.66	64.26	0.001
	(16.08)	(15.62)	(15.94)	(15.95)	
Team test 2 %	75.49	75.38	74.65	76.24	0.080
	(12.57)	(12.59)	(13.19)	(11.97)	
Observations (unique students)	603	495	292	304	

Table 3. Correlation between personality measures

	Extro.	Agree.	Consc.	Em. Stab.	Openness
Extroversion	1.00				
Agreeableness	0.32	1.00			
Conscientiousness	0.10	0.33	1.00		
Emotional stability	0.21	0.08	0.12	1.00	
Openness	0.19	0.33	0.18	0.07	1.00

Table 4. Descriptive statistics: team characteristics (means and standard deviations of raw variables)

	(1) Main sample	(2) Balanced sample
Teammate extroversion	3.05	3.05
	(0.57)	(0.60)
Teammate agreeableness	3.78	3.78
	(0.46)	(0.48)
Teammate conscientiousness	3.57	3.58
	(0.44)	(0.46)
Teammate emotional stability	3.13	3.14
	(0.52)	(0.54)
Teammate openness	3.50	3.50
	(0.38)	(0.40)
Teammate female share	0.49	0.49
	(0.41)	(0.41)
Teammate baseline test %	49.38	49.34
	(9.16)	(9.30)
Individual question score	2.38	2.38
	(1.67)	(1.66)
Team question score	2.85	2.84
	(1.55)	(1.53)
Team maximum score from individual test	3.19	3.13
	(1.42)	(1.43)
Indicator: at least one team member obtained maximum possible question score on individual test	0.70	0.66
	(0.46)	(0.47)
Indicator: team gain	0.16	0.09
	(0.37)	(0.28)
Indicator: team same	0.57	0.62
	(0.49)	(0.49)
Indicator: team loss	0.27	0.29
	(0.44)	(0.46)
Team 1 size	3.02	3.03
	(0.62)	(0.63)
Team 2 size	2.99	2.99
	(0.66)	(0.66)
Observations (student-questions)	10,404	8,469
Unique students	603	495
Unique test 1 questions	9	9
Unique test 2 questions	9	9

Table 5. Potential selection into sample

	(1) I(Main sample) *(took personality test and at least one team test)*	(2) I(Balanced sample) *(took all tests)*
Extroversion	-0.011	-0.004
	(0.017)	(0.020)
Agreeableness	-0.023	-0.001
	(0.017)	(0.022)
Conscientiousness	0.010	0.011
	(0.015)	(0.019)
Emotional stability	0.047***	0.042**
	(0.017)	(0.019)
Openness	-0.016	-0.030
	(0.017)	(0.020)
Female	0.042	-0.018
	(0.033)	(0.041)
Baseline test z-score	0.030*	0.045**
	(0.016)	(0.021)
Observations (students)	675	675
R-squared	0.019	0.019

Estimates from regressing binary indicators of sample inclusion on own personality traits, gender, and baseline statistics comprehension are reported. Personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. The sample includes all students for whom we observe the full set of explanatory variables in the table. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 6. Balance: correlation between teammate personality measures and other characteristics

	(1)	(2)	(3)	(4)	(5)
	Teammate extroversion	Teammate agreeableness	Teammate conscientiousness	Teammate emotional stability	Teammate openness
Extroversion	0.055	-0.000	-0.003	0.051*	-0.022
	(0.038)	(0.022)	(0.020)	(0.026)	(0.017)
Agreeableness	-0.062	-0.045	-0.002	0.006	-0.010
	(0.040)	(0.040)	(0.027)	(0.036)	(0.025)
Conscientiousness	0.008	0.029	0.035	0.036	0.009
	(0.036)	(0.028)	(0.033)	(0.030)	(0.024)
Emotional stability	0.027	0.010	0.025	-0.075**	-0.003
	(0.032)	(0.027)	(0.024)	(0.036)	(0.021)
Openness	-0.047	-0.017	-0.022	-0.027	-0.017
	(0.045)	(0.034)	(0.032)	(0.039)	(0.037)
Female	0.057	0.026	-0.003	-0.042	-0.002
	(0.042)	(0.032)	(0.035)	(0.033)	(0.030)
Baseline test z-score	-0.020	-0.008	-0.013	-0.044**	-0.007
	(0.024)	(0.015)	(0.018)	(0.021)	(0.014)
Teammate female share	-0.005	0.336***	0.186***	-0.309***	-0.116***
	(0.053)	(0.041)	(0.039)	(0.040)	(0.040)
Teammate baseline test z-score	-0.060*	0.071***	0.031	0.024	0.120***
	0.055	-0.000	-0.003	0.051*	-0.022
Question fixed effects	Y	Y	Y	Y	Y
Level(s) of error clustering	Student, team	Student, team	Student, team	Student, team	Student, team
Observations (student-questions)	8,469	8,469	8,469	8,469	8,469
R-squared	0.017	0.103	0.033	0.076	0.070

Estimates from regressing teammate personality traits on own personality traits, gender, baseline statistics comprehension, and various other controls are reported. Personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Teammate personality measures are calculated using the normalized individual personality measures. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 7. Correlation between individual personality measures and individual question scores

	(1)	(2)	(3)	(4)
	Individual question score			
Extroversion	-0.021	-0.021	-0.004	-0.004
	(0.014)	(0.014)	(0.014)	(0.014)
Agreeableness	0.004	-0.005	-0.005	-0.016
	(0.014)	(0.015)	(0.014)	(0.015)
Conscientiousness	0.026*	0.021	0.027**	0.022
	(0.014)	(0.015)	(0.014)	(0.014)
Emotional stability	0.013	0.021	0.003	0.011
	(0.013)	(0.014)	(0.012)	(0.013)
Openness	0.005	0.011	-0.020	-0.013
	(0.013)	(0.013)	(0.012)	(0.013)
Female		0.058**		0.060**
		(0.028)		(0.027)
Baseline test z-score			0.123***	0.122***
			(0.014)	(0.014)
Question fixed effects	Y	Y	Y	Y
Level(s) of error clustering	Student	Student	Student	Student
Observations (student-questions)	11,295	11,178	11,214	11,115
R-squared	0.503	0.503	0.514	0.515

Estimates from regressing individual question scores on own personality traits, gender and baseline statistics comprehension are reported. Both question scores and personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 8. Effect of teammate personality on individual question score

	(1)	(2)	(3)	(4)
Teammate extroversion	0.004	0.006	0.007	0.005
	(0.016)	(0.016)	(0.015)	(0.016)
Teammate agreeableness	-0.021	-0.021	-0.023	-0.005
	(0.016)	(0.016)	(0.016)	(0.018)
Teammate conscientiousness	0.027*	0.026*	0.029**	0.011
	(0.015)	(0.014)	(0.014)	(0.015)
Teammate emotional stability	-0.019	-0.019	-0.009	0.001
	(0.015)	(0.015)	(0.015)	(0.016)
Teammate openness	-0.001	-0.001	-0.006	-0.004
	(0.015)	(0.014)	(0.015)	(0.015)
Extroversion		-0.018	-0.003	
		(0.014)	(0.013)	
Agreeableness		0.007	-0.007	
		(0.014)	(0.014)	
Conscientiousness		0.027	0.025	
		(0.017)	(0.016)	
Emotional stability		0.003	-0.003	
		(0.015)	(0.015)	
Openness		0.004	-0.019	
		(0.013)	(0.013)	
Teammate max question score	-	-	-	-
Teammate mean question score	-	-	-	-
Individual question score	-	-	-	-
Female			0.024	
			(0.030)	
Baseline test z-score			0.134***	
			(0.015)	
Question fixed effects	Y	Y	Y	Y
Student fixed effects	N	N	N	Y
Level(s) of error clustering	Student, team	Student, team	Student, team	Student, team
Observations (student-questions)	8,523	8,523	8,334	8,487
R-squared	0.505	0.506	0.520	0.610

Estimates from regressing individual question scores on teammate personality traits, own personality traits, and various other controls are reported. Both question scores and personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Teammate personality measures are calculated using the normalized individual personality measures. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Table 9. Effect of teammate personality on team question score

	(1)	(2)	(3)	(4)	(5)	(6)
	Team question score					
Teammate extroversion	-0.031*	-0.030*	-0.019	-0.031**	-0.020	-0.023
	(0.016)	(0.016)	(0.014)	(0.015)	(0.014)	(0.014)
Teammate agreeableness	0.001	-0.001	-0.005	0.005	-0.003	-0.011
	(0.017)	(0.018)	(0.015)	(0.016)	(0.014)	(0.015)
Teammate conscientiousness	0.039**	0.039**	0.029*	0.034**	0.024*	0.025*
	(0.017)	(0.017)	(0.015)	(0.016)	(0.014)	(0.015)
Teammate emotional stability	-0.014	-0.015	-0.013	-0.011	-0.005	0.004
	(0.017)	(0.018)	(0.016)	(0.017)	(0.015)	(0.015)
Teammate openness	0.012	0.011	0.015	0.013	0.017	0.034**
	(0.017)	(0.018)	(0.015)	(0.016)	(0.014)	(0.017)
Extroversion		-0.014	-0.016	-0.009	-0.010	
		(0.011)	(0.010)	(0.011)	(0.009)	
Agreeableness		-0.004	0.002	-0.005	-0.005	
		(0.012)	(0.010)	(0.011)	(0.012)	
Conscientiousness		0.026**	0.017	0.021**	0.011	
		(0.012)	(0.011)	(0.010)	(0.010)	
Emotional stability		-0.005	0.000	-0.006	0.003	
		(0.013)	(0.012)	(0.012)	(0.011)	
Openness		0.004	0.010	0.005	0.008	
		(0.012)	(0.011)	(0.011)	(0.011)	
Teammate max question score			0.249***		0.258***	0.235***
			(0.026)		(0.025)	(0.025)
Teammate mean question score			0.146***		0.127***	0.114***
			(0.025)		(0.025)	(0.024)
Individual question score				0.216***	0.201***	0.176***
				(0.013)	(0.010)	(0.010)
Female					0.033	
					(0.022)	
Baseline test z-score					0.025**	
					(0.011)	
Question fixed effects	Y	Y	Y	Y	Y	Y
Student fixed effects	N	N	N	N	N	Y
Level(s) of error clustering	Student, team	Student, team	Student, team	Student, team	Student, team	Student, team
Observations (student-questions)	8,667	8,667	8,595	8,550	8,352	8,505
R-squared	0.615	0.616	0.660	0.642	0.685	0.726

Estimates from regressing team question scores on teammate personality traits, own personality traits, and various other controls are reported. Both question scores and personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Teammate personality measures are calculated using the normalized individual personality measures. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 10. Effect of teammate personality on team question score controlling for other teammate characteristics

	(1)	(2)	(3)	(4)
	Team question score			
Teammate extroversion		-0.019	-0.014	-0.017
		(0.016)	(0.014)	(0.014)
Teammate agreeableness		-0.016	-0.017	-0.017
		(0.020)	(0.015)	(0.018)
Teammate conscientiousness		0.036**	0.021	0.024
		(0.017)	(0.014)	(0.015)
Teammate emotional stability		-0.011	0.001	0.006
		(0.018)	(0.015)	(0.015)
Teammate openness		0.003	0.017	0.034*
		(0.018)	(0.015)	(0.017)
Extroversion		-0.018	-0.012	
		(0.011)	(0.009)	
Agreeableness		-0.003	-0.004	
		(0.012)	(0.012)	
Conscientiousness		0.029**	0.013	
		(0.012)	(0.010)	
Emotional stability		-0.001	0.004	
		(0.013)	(0.011)	
Openness		0.005	0.009	
		(0.012)	(0.011)	
Teammate max question score			0.265***	0.239***
			(0.025)	(0.025)
Teammate mean question score			0.114***	0.108***
			(0.026)	(0.025)
Individual question score			0.202***	0.176***
			(0.010)	(0.009)
Female			0.023	
			(0.019)	
Baseline test z-score			0.025**	
			(0.011)	
Teammate female share	0.061*	0.055	0.066**	0.039
	(0.031)	(0.038)	(0.026)	(0.030)
Teammate baseline test z-score	0.004***	0.063***	0.035**	0.028
	(0.001)	(0.016)	(0.014)	(0.017)
Question fixed effects	Y	Y	Y	Y
Student fixed effects	N	N	N	Y
Level(s) of error clustering	Student, team	Student, team	Student, team	Student, team
Observations (student-questions)	9,225	8,523	8,316	8,370
R-squared	0.614	0.619	0.686	0.726

Estimates from regressing team question scores on teammate personality traits, own personality traits, and various other controls are reported. Both question scores and personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Teammate personality measures are calculated using the normalized individual personality measures. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 11. Probing the mechanism: persistence in effect of teammate personality

	(1)	(2)
	Individual question score (Test 2)	
Teammate extroversion (Test 1)	-0.025	-0.021
	(0.021)	(0.017)
Teammate agreeableness (Test 1)	-0.012	0.009
	(0.022)	(0.020)
Teammate conscientiousness (Test1)	0.046***	0.027
	(0.017)	(0.017)
Teammate emotional stability (Test 1)	0.011	0.022
	(0.025)	(0.022)
Teammate openness (Test 1)	0.022	0.017
	(0.021)	(0.019)
Extroversion	-0.003	-0.008
	(0.014)	(0.014)
Agreeableness	-0.000	0.009
	(0.017)	(0.016)
Conscientiousness	0.020	0.009
	(0.017)	(0.015)
Emotional stability	0.011	0.018
	(0.016)	(0.014)
Openness	-0.017	-0.009
	(0.016)	(0.015)
Female	0.054	0.055*
	(0.036)	(0.032)
Baseline test z-score	0.128***	0.065***
	(0.016)	(0.014)
Individual Test 1 z-score		0.171***
		(0.013)
Question fixed effects	Y	Y
Student fixed effects	N	N
Level(s) of error clustering	Student, team	Student, team
Observations (student-questions)	4,185	4,185
R-squared	0.453	0.483

Estimates from regressing individual question scores on past teammate personality traits, own personality traits, and various other controls are reported. Both question scores and personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Teammate personality measures are calculated using the normalized individual personality measures. Robust standard errors in parentheses.

Table 12. Probing the mechanism: heterogeneity by question type

	(1)	(2)	(3)	(4)	(5)	(6)
	Questions with written answers		Questions with numeric answers		Questions with circled answers	
Teammate extroversion	-0.015	-0.028	-0.025	-0.011	-0.016	-0.045*
	(0.021)	(0.023)	(0.018)	(0.026)	(0.018)	(0.023)
Teammate agreeableness	-0.001	0.008	-0.012	-0.040	0.005	-0.001
	(0.021)	(0.026)	(0.020)	(0.027)	(0.018)	(0.027)
Teammate conscientiousness	0.041**	0.043*	0.010	0.001	0.013	0.032
	(0.020)	(0.023)	(0.019)	(0.024)	(0.020)	(0.028)
Teammate emotional stability	-0.005	0.020	0.003	-0.015	-0.014	-0.002
	(0.027)	(0.026)	(0.017)	(0.024)	(0.020)	(0.028)
Teammate openness	0.025	0.041*	0.021	0.050*	0.003	0.013
	(0.021)	(0.023)	(0.018)	(0.026)	(0.023)	(0.032)
Extroversion	-0.010		-0.010		-0.008	
	(0.014)		(0.012)		(0.012)	
Agreeableness	-0.001		-0.007		-0.006	
	(0.018)		(0.014)		(0.014)	
Conscientiousness	0.017		0.005		0.009	
	(0.014)		(0.012)		(0.014)	
Emotional stability	0.007		0.004		-0.007	
	(0.020)		(0.012)		(0.014)	
Openness	0.016		0.002		0.003	
	(0.016)		(0.012)		(0.016)	
Teammate max question score	0.334***	0.288***	0.269***	0.229***	0.154***	0.144***
	(0.043)	(0.047)	(0.047)	(0.046)	(0.040)	(0.042)
Teammate mean question score	0.042	0.027	0.217***	0.221***	0.176***	0.145***
	(0.045)	(0.046)	(0.044)	(0.046)	(0.038)	(0.041)
Individual question score	0.199***	0.140***	0.248***	0.219***	0.172***	0.164***
	(0.018)	(0.020)	(0.017)	(0.019)	(0.015)	(0.015)
Female	0.072**		0.002		0.009	
	(0.035)		(0.027)		(0.031)	
Baseline test z-score	0.024		0.034***		0.013	
	(0.016)		(0.011)		(0.014)	
Question fixed effects	Y	Y	Y	Y	Y	Y
Student fixed effects	N	Y	N	Y	N	Y
Level(s) of error clustering	Student, team	Student, team	Student, team	Student, team	Student, team	Student, team
Observations (student-questions)	3,236	3,295	2,808	2,860	2,308	2,350
R-squared	0.415	0.563	0.692	0.763	0.785	0.845

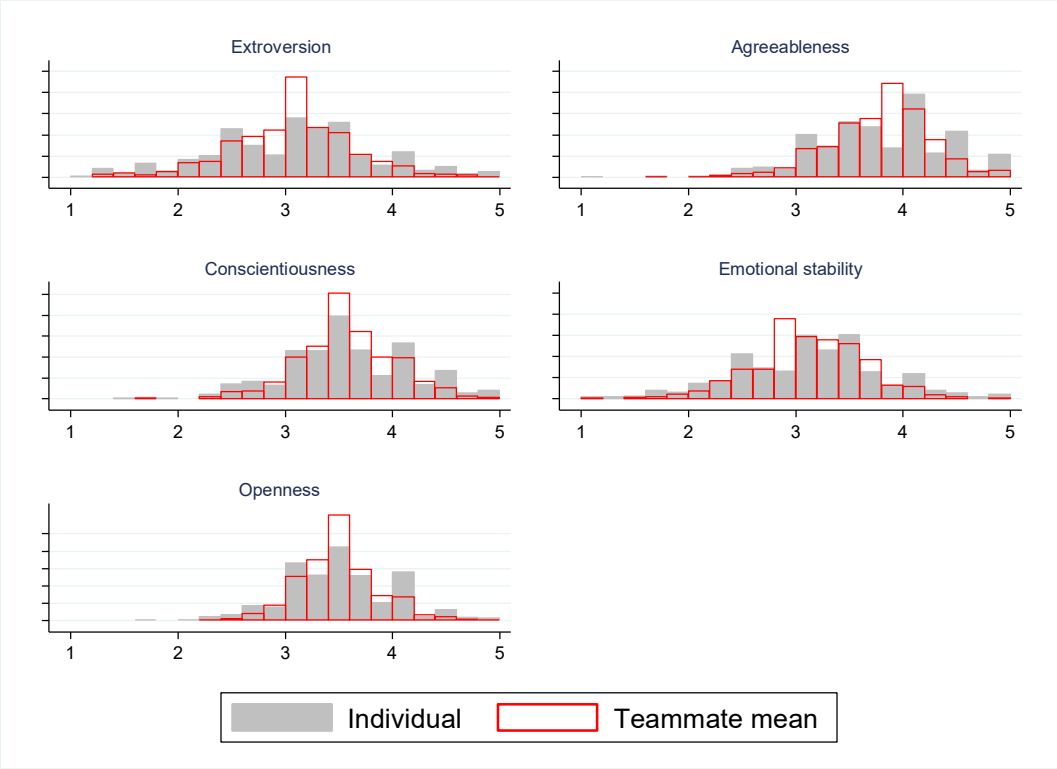
Estimates from regressing team question scores on teammate personality traits, own personality traits, and various other controls are reported. Both question scores and personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Teammate personality measures are calculated using the normalized individual personality measures. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table 13. Placebo test: effect of Test 2 teammate personality on Test 1 question score

	(1) Individual question score (Test 1)
Teammate extroversion (Test 2)	0.034
	(0.021)
Teammate agreeableness (Test 2)	-0.015
	(0.022)
Teammate conscientiousness (Test 2)	-0.013
	(0.022)
Teammate emotional stability (Test 2)	-0.031
	(0.022)
Teammate openness (Test 2)	-0.028
	(0.020)
Extroversion	-0.002
	(0.017)
Agreeableness	-0.015
	(0.019)
Conscientiousness	0.032
	(0.019)
Emotional stability	-0.020
	(0.020)
Openness	-0.017
	(0.016)
Female	0.001
	(0.037)
Baseline test z-score	0.139***
	(0.018)
Question fixed effects	Y
Student fixed effects	N
Level(s) of error clustering	Student, team
Observations (student-questions)	4,221
R-squared	0.567

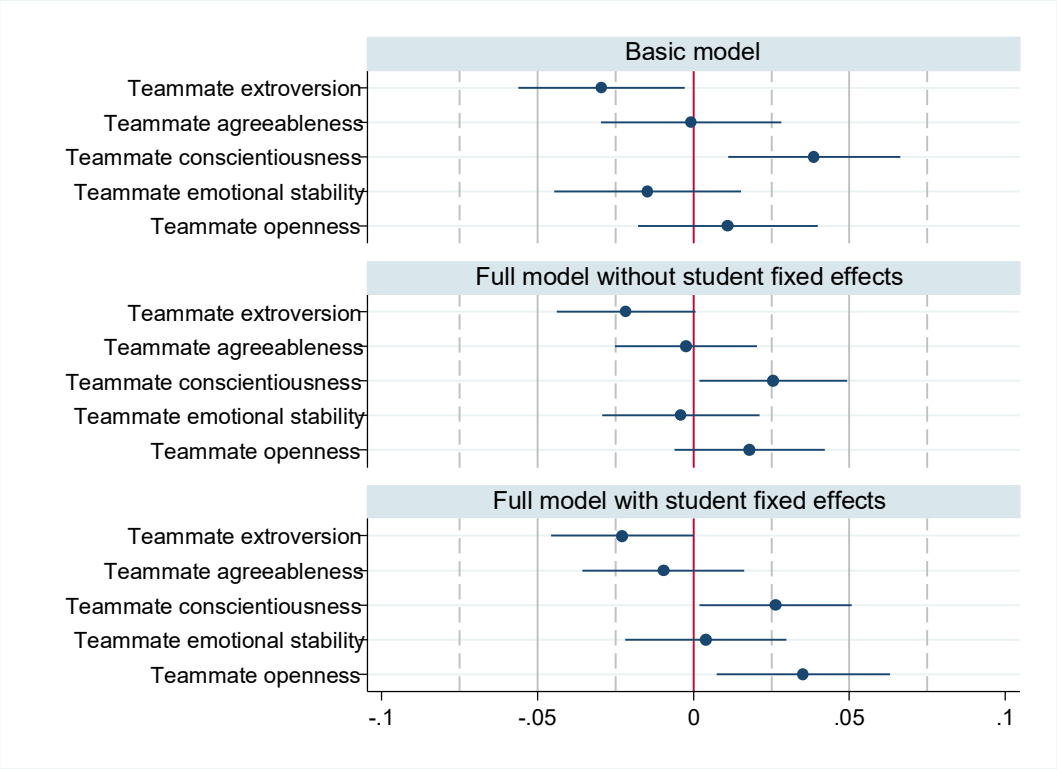
Estimates from regressing past individual question scores on future teammate personality traits, own personality traits, and various other controls are reported. Both question scores and personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Teammate personality measures are calculated using the normalized individual personality measures. Robust standard errors in parentheses.

Figure 1. Distribution of personality measures: individual and team



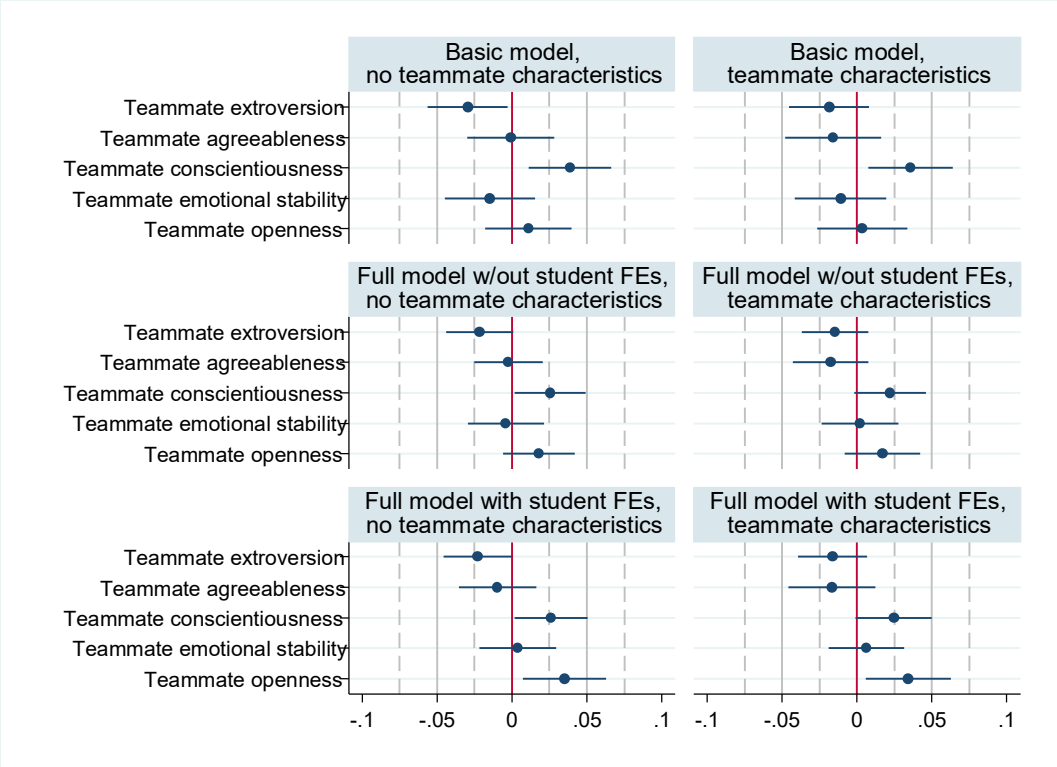
The plotted distributions are based on raw personality measures. Teammate means are calculated using the raw personality measures and the individual's own measure is excluded from the calculation.

Figure 2. Effect of teammate personality on team question score



The figure plots the estimated coefficients and 90% confidence intervals from three regressions of team question score on teammate personality measures. The units of observation in each model are student-questions.

Figure 3. Effect of teammate personality controlling for other teammate characteristics



The figure plots the estimated coefficients and 90% confidence intervals from six regressions of team question score on teammate personality measures. The units of observation in each model are student-questions. The left panel reports results from models that excludes additional teammate characteristics (teammate female share and teammate baseline test percentage score); the right panel reports results from models that include them.

ONLINE APPENDIX

Appendix Table 1. Descriptive statistics across samples

	(1)	(2)	(3)	(4)	(5)
	Students who took:				
	Personality test	Individual test 1	Team test 1	Individual test 2	Team test 2
Extroversion	3.05	3.05	3.04	3.05	3.05
	(0.76)	(0.76)	(0.75)	(0.75)	(0.75)
Agreeableness	3.79	3.77	3.77	3.77	3.77
	(0.61)	(0.62)	(0.61)	(0.61)	(0.61)
Conscientiousness	3.58	3.57	3.57	3.59	3.58
	(0.59)	(0.58)	(0.59)	(0.58)	(0.58)
Emotional stability	3.12	3.12	3.13	3.15	3.15
	(0.69)	(0.69)	(0.68)	(0.68)	(0.68)
Openness	3.50	3.49	3.49	3.48	3.48
	(0.51)	(0.50)	(0.50)	(0.50)	(0.51)
Female	0.51	0.50	0.50	0.48	0.49
	(0.50)	(0.50)	(0.50)	(0.50)	(0.50)
Baseline test %	49.83	49.26	49.22	49.57	49.36
	(10.46)	(11.58)	(11.67)	(11.50)	(11.48)
Individual test 1 %	66.84	65.97	66.19	67.49	67.25
	(17.70)	(17.60)	(17.35)	(16.68)	(16.70)
Team test 1 %	79.60	79.18	79.14	79.30	79.11
	(12.88)	(12.96)	(12.98)	(12.72)	(12.79)
Individual test 2 %	62.59	61.81	61.55	61.79	61.54
	(16.02)	(16.31)	(16.29)	(16.30)	(16.06)
Team test 2 %	76.01	75.59	75.03	75.62	75.51
	(12.45)	(12.66)	(12.82)	(12.68)	(12.80)
Team 1 size	3.02	3.01	3.01	3.00	3.01
	(0.62)	(0.62)	(0.62)	(0.63)	(0.63)
Team 2 size	2.99	2.98	2.98	2.98	2.98
	(0.66)	(0.66)	(0.67)	(0.67)	(0.67)
Observations (unique students)	686	752	690	668	630

Appendix Table 2. Correlation between individual personality measures and individual baseline test z-score

	(1)	(2)
	Baseline test z-score	
Extroversion	-0.190***	-0.191***
	(0.049)	(0.049)
Agreeableness	0.112*	0.128*
	(0.062)	(0.065)
Conscientiousness	-0.001	-0.001
	(0.063)	(0.064)
Emotional stability	0.132**	0.128**
	(0.052)	(0.054)
Openness	0.389***	0.379***
	(0.071)	(0.073)
Female		-0.040
		(0.072)
Observations (students)	681	675
R-squared	0.078	0.078

Estimates from regressing baseline test scores on own personality traits are reported. Personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Appendix Table 3. Team-level analysis: effect of team personality on team question score

	(1)	(2)
	Team question score	
Team extroversion	-0.040***	-0.036**
	(0.015)	(0.015)
Team agreeableness	0.020	-0.002
	(0.017)	(0.018)
Team conscientiousness	0.026*	0.028*
	(0.015)	(0.015)
Team emotional stability	0.012	0.026
	(0.016)	(0.016)
Team openness	0.023	0.016
	(0.015)	(0.016)
Team female share		0.106***
		(0.029)
Team baseline test z-score		0.007***
		(0.001)
Question fixed effects	Y	Y
Observations (student-questions)	4,122	4,122
R-squared	0.613	0.617

Estimates from regressing team question scores on team personality traits are reported. Both question scores and personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Team personality measures are calculated using the normalized individual personality measures. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.

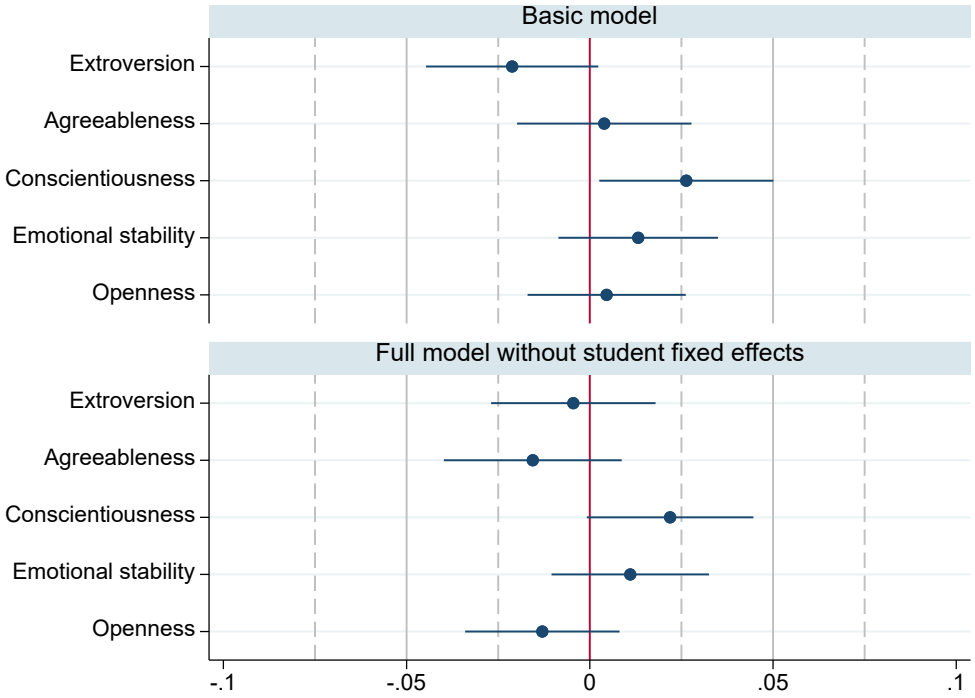
Appendix Table 4. Placebo-type test: effect of teammate characteristics on predetermined characteristics

	(1) Normalized baseline test score	(2) Female	(3) Normalized baseline test score	(4) Female
Panel A: Test 1 and Test 2				
Teammate extroversion	-0.035	0.027	-0.037	0.028
	(0.058)	(0.030)	(0.058)	(0.028)
Teammate agreeableness	-0.022	0.078**	0.006	0.005
	(0.068)	(0.040)	(0.072)	(0.036)
Teammate conscientiousness	-0.032	0.026	-0.025	-0.005
	(0.076)	(0.043)	(0.074)	(0.041)
Teammate emotional stability	-0.125**	-0.066**	-0.140**	-0.013
	(0.060)	(0.029)	(0.063)	(0.028)
Teammate openness	0.005	-0.083*	-0.015	-0.016
	(0.072)	(0.049)	(0.079)	(0.047)
Teammate female share			-0.080	0.274***
			(0.081)	(0.056)
Teammate baseline test %			-0.000	-0.000
			(0.005)	(0.002)
Panel B: Test 1				
Teammate extroversion	-0.046	0.062	-0.043	0.057
	(0.075)	(0.043)	(0.072)	(0.043)
Teammate agreeableness	-0.009	-0.025	0.024	-0.044
	(0.091)	(0.058)	(0.095)	(0.055)
Teammate conscientiousness	0.025	-0.027	0.034	-0.038
	(0.099)	(0.057)	(0.096)	(0.057)
Teammate emotional stability	-0.096	-0.028	-0.108	-0.004
	(0.082)	(0.043)	(0.078)	(0.041)
Teammate openness	0.076	-0.055	0.032	-0.023
	(0.107)	(0.072)	(0.116)	(0.073)
Teammate female share			-0.081	0.126
			(0.109)	(0.077)
Teammate baseline test %			0.005	0.000
			(0.006)	(0.003)
Panel C: Test 2				
Teammate extroversion	-0.029	-0.007	-0.039	0.004
	(0.080)	(0.042)	(0.082)	(0.035)
Teammate agreeableness	-0.037	0.177***	-0.011	0.038
	(0.101)	(0.055)	(0.112)	(0.049)
Teammate conscientiousness	-0.088	0.078	-0.089	0.023
	(0.099)	(0.053)	(0.102)	(0.042)
Teammate emotional stability	-0.142*	-0.111***	-0.158*	-0.026
	(0.077)	(0.042)	(0.087)	(0.038)
Teammate openness	-0.059	-0.098	-0.049	0.003
	(0.101)	(0.062)	(0.111)	(0.054)
Teammate female share			-0.059	0.414***
			(0.121)	(0.074)
Teammate baseline test %			-0.006	-0.001
			(0.007)	(0.002)

Estimates from regressing predetermined characteristics on teammate personality traits are reported. Personality measures are normalized to have a mean of zero and a standard deviation of one for ease of interpretation. Teammate personality measures are calculated using the normalized individual personality measures. Robust standard errors in parentheses.

*** p<0.01, ** p<0.05, * p<0.1

Appendix Figure 1. Correlation between individual personality measures and individual question scores



1

The figure plots the estimated coefficients and 90% confidence intervals from two regressions of individual question score on own personality measures. The units of observation in each model are student-questions.

Appendix A1: Personality Test

How Accurately Can You Describe Yourself?

Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Indicate for each statement whether it is 1. Very Inaccurate, 2. Moderately Inaccurate, 3. Neither Accurate Nor Inaccurate, 4. Moderately Accurate, or 5. Very Accurate as a description of you.

1	Am the life of the party.	(1+)
2	Feel little concern for others.	(2-)
3	Am always prepared.	(3+)
4	Get stressed out easily.	(4-)
5	Have a rich vocabulary.	(5+)
6	Don't talk a lot.	(1-)
7	Am interested in people.	(2+)
8	Leave my belongings around.	(3-)
9	Am relaxed most of the time.	(4+)
10	Have difficulty understanding abstract ideas.	(5-)
11	Feel comfortable around people.	(1+)
12	Insult people.	(2-)
13	Pay attention to details.	(3+)
14	Worry about things.	(4-)
15	Have a vivid imagination.	(5+)
16	Keep in the background.	(1-)
17	Sympathize with others' feelings.	(2+)
18	Make a mess of things.	(3-)
19	Seldom feel blue.	(4+)
20	Am not interested in abstract ideas.	(5-)
21	Start conversations.	(1+)
22	Am not interested in other people's problems.	(2-)
23	Get chores done right away.	(3+)
24	Am easily disturbed.	(4-)
25	Have excellent ideas.	(5+)
26	Have little to say.	(1-)
27	Have a soft heart.	(2+)
28	Often forget to put things back in their proper place.	(3-)
29	Get upset easily.	(4-)
30	Do not have a good imagination.	(5-)
31	Talk to a lot of different people at parties.	(1+)
32	Am not really interested in others.	(2-)

33	Like order.	(3+)
34	Change my mood a lot.	(4-)
35	Am quick to understand things.	(5+)
36	Don't like to draw attention to myself.	(1-)
37	Take time out for others.	(2+)
38	Shirk my duties.	(3-)
39	Have frequent mood swings.	(4-)
40	Use difficult words.	(5+)
41	Don't mind being the center of attention.	(1+)
42	Feel others' emotions.	(2+)
43	Follow a schedule.	(3+)
44	Get irritated easily.	(4-)
45	Spend time reflecting on things.	(5+)
46	Am quiet around strangers.	(1-)
47	Make people feel at ease.	(2+)
48	Am exacting in my work.	(3+)
49	Often feel blue.	(4-)
50	Am full of ideas.	(5+)

Note. These five scales were developed to measure the Big-Five factor markers reported in the following article: Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. Psychological Assessment, 4, 26-42.

The numbers in parentheses after each item indicate the scale on which that item is scored (i.e., of the five factors: (1) Extroversion, (2) Agreeableness, (3) Conscientiousness, (4) Emotional Stability, or (5) Openness) and its direction of scoring (+ or -).

Appendix A2: Exam 1 Team Reflection

Team Member Names: Your team should be comprised of the team of 3-4 people with whom you worked with during Unit 1.

1._____ 3._____

2._____ 4._____

Instructions: Each team will turn in one Exam 1 Team Reflection worksheet. Discuss each question with your teammates, agree on a solution, then write your solution on this worksheet. You may use your book and notes, but only work with your team members; you will not receive help from your instructor. You will have 30 minutes to complete this reflection.

1. [2 pts each] For each of the following statements, circle whether it is true or false:

   a. Convenience samples often lead to non-response bias. True / False

   b. If selection bias is present in a study, we can decrease the bias by collecting a larger sample. True / False

   c. Using a sampling frame that differs from the population of interest will lead to selection bias. True / False

   d. If we take a simple random sample from the population of interest, but our sample size is too small, then we cannot generalize our sample results to the population. True / False

2. Research done in the mid 1980s indicated that 80% of grizzly bears in the greater Yellowstone ecosystem entered their den for hibernation by the last day of November. This has come into question as of late. Researchers have hypothesized that climate change has postponed bears entering hibernation until later in the season; that is, researchers believe that a smaller percentage of bears are entering their den for hibernation by the last day of November. Sixty-two Yellowstone grizzly bears are being tracked via radio monitors, which allow scientists to pinpoint when the bears enter their dens with a high degree of accuracy. Of the sixty-two Yellowstone grizzly bears being tracked, forty-two had entered their den for hibernation by the last day of November.

   a. [4 pts] Write the null and alternative hypotheses **in words** in the context of the problem.

      H_0:

      H_a:

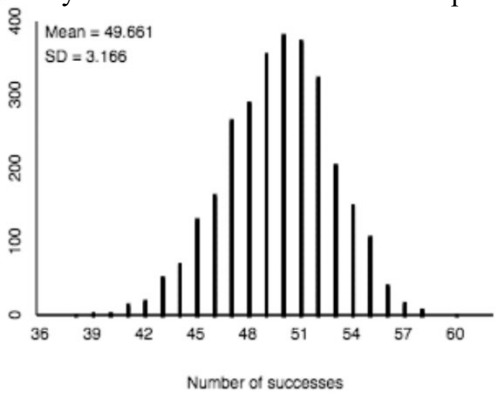
   b. [3 pts] What values would you enter for each of the following in the One Proportion applet to conduct a simulation-based hypothesis test with these data?

      Probability of success (π): ____

      Sample size (n): ____

      Number of samples: ____

c. [3 pts] A distribution of simulated "Number of successes," assuming the null hypothesis is true, is shown below. Circle the portion of the plot below that you would count to calculate the p-value.



d. [3 pts] Where is the distribution of simulated "Number of successes" centered? Explain, in context, why this makes sense.

e. [4 pts] How could we use coins, a spinner, or cards to simulate one of the samples in the above simulation? Circle **all that apply**.

(A) Flip a coin 62 times, where heads represents "hibernate by November 30." Count the number of heads in 62 flips.

(B) Create a spinner with 80% shaded to represent "hibernate by November 30." Spin the spinner 62 times and count the number of times the spinner lands in the shaded area.

(C) Have 8 black cards and 2 red cards where black represents "hibernate by November 30". Draw with replacement from the cards 62 times and count the number of times a black card is drawn.

(D) Create a spinner with 68% shaded to represent "hibernate by November 30." Spin the spinner 62 times and count the number of times the spinner lands in the shaded area.

f. [4 pts] The simulation-based p-value for this test is 0.0153. **Interpret** this value in the context of the problem. (Note: This should *not* be a strength of evidence statement, but rather an interpretation of what the value 0.0153 represents in context of the problem.)

g. [4 pts] Recall that in this sample, 42 of 62 bears had entered their den for hibernation by the last day of November. For each of the following sample results, circle whether the p-value would increase, decrease, or remain the same if we were to track another sample of Yellowstone grizzly bears.

(i) 126 of 186 bears had entered their den for hibernation by the last day of November:

Increase Decrease Remain the same

(ii) 45 of 62 bears had entered their den for hibernation by the last day of November:

Increase Decrease Remain the same

h. [2 pts] Discuss as a group how you studied for Exam 1, including what seemed to work and what didn't. Write down suggestions for how to study better for Exam 2.

Appendix A3: Exam 2 Team Reflection

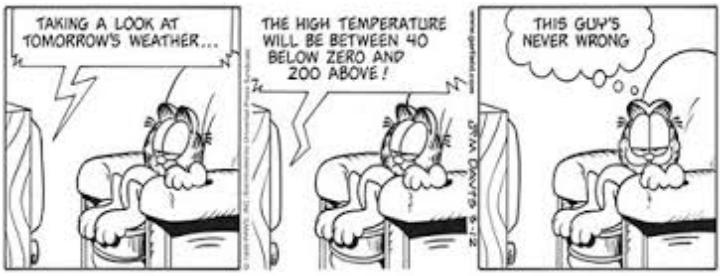
Team Member Names: Your team should be comprised of the team of 3-4 people with whom you worked with during Unit 2.

1._____ 3._____

2._____ 4._____

Instructions: Each team will turn in one Exam 2 Team Reflection worksheet. Discuss each question with your teammates, agree on a solution, then write your solution on this worksheet. You may use your book and notes, but only work with your team members; you will not receive help from your instructor. You will have 30 minutes to complete this reflection.

1. Use this Garfield cartoon to answer the following questions.



   a. [3 pts] What is the margin of error for the interval reported in the cartoon?

   b. [4 pts] Suppose the interval in the cartoon used a 99% confidence level and was calculated using the high temperatures in a sample of 150 days with similar conditions to tomorrow's prediction. Which of the following statements are true? Select **all** that apply.
      (A) There is a 99% chance that the true mean high temperature among days with similar conditions will be between 40 below and 200 above.
      (B) The sample mean high temperature is equal to 80 degrees.
      (C) The long-run proportion of intervals computed over many samples of 150 days that will contain the true mean high temperature among days with similar conditions is 0.99.
      (D) If the confidence level was lowered to 95%, the interval would be narrower.

2. The paper "The Effect of Multitasking on the Grade Performance of Business Students" (*Research in Higher Education Journal,* 2010) describes a study in which 62 undergraduate business students were randomly assigned to one of two lecture groups. Students in the first group were asked to listen to a lecture but were told that they were permitted to use cell phones to send text messages during the lecture. Students in the second group listened to the same lecture but were not permitted to send text messages. Afterwards, students in both groups took a quiz on material covered in the lecture. The researchers want to know if these data provide evidence to support the researcher's claim that the mean quiz score for the texting group is significantly lower than the mean quiz score for the no-texting group.
   a. [3 pts] Is this an observational study or a randomized experiment? Circle one.
      (A) Observational study, because the researchers observed which group had a higher mean quiz score.
      (B) Randomized experiment, because the students took the same quiz.
      (C) Randomized experiment, because students were randomly assigned to lecture groups.
      (D) Observational study, because the researchers did not take a random sample of students.

b. [3 pts] If the evidence against the null were strong, would you be justified in concluding that you have found strong evidence of a cause-and-effect relationship between lecture group and quiz score for this study? Explain in context of the study.

3. In 2016, the Bozeman Daily Chronicle reported the average cost of rent per month for a two-bedroom apartment in Bozeman was $1022 per month, but a Stat 216 student that had been looking for housing found that hard to believe. Based on their search, the Stat 216 student believes that rent costs much more per month. The student decided to test the Chronicle's claim and took a random sample of 40 two-bedroom apartments for rent. The average rent cost in the student's sample was $1520 per month with a standard deviation of $465.
   a. [4 pts] State the Stat 216 student's null and alternative hypothesis for conducting a hypothesis test in **words**.
   H_0:

   H_a:

   b. [5 pts] To run a hypothesis test, the student decided to use the applets to simulate a null distribution. To simulate one sample in the null distribution with cards, the student would have to do the following (fill in each blank with the missing value and circle the correct word in each set of brackets):
   The student would need _____ cards. On each card, the student would write the rent costs $\begin{Bmatrix} plus \\ minus \end{Bmatrix}$ _____ . Next the student would draw _____ cards $\begin{Bmatrix} with \\ without \end{Bmatrix}$ replacement. The $\begin{Bmatrix} proportion \\ mean \end{Bmatrix}$ of those values would be calculated and plotted.

   c. [4 pts] The p-value obtained from simulation was smaller than 0.0001. **Interpret** this p-value in context of the problem. *Do not evaluate strength of evidence!*

   d. [4 pts] What conclusion can the Stat 216 student make from these results of their study? Be sure to write your conclusion in the context of the problem, and include the appropriate scope of inference.

4. [3 pts] Noxious weeds are invasive plants that quickly spread and deplete the soil of nutrients that are required for all aspects of farming and ranching. With much at stake, the state of Montana has hired a team of researchers to explore the issue. They want to know the average number of noxious weeds per acre in North Central Montana. To do so, researchers divided the area of interest into 1,000,000 acres and randomly selected 24 of these acres. For each selected

acre, they counted and removed the noxious weeds. Can we use this study to generalize our results to the population of interest? Circle one.

(A) No, since we only observed 24 acres.

(B) Yes, since we randomly selected the 24 acres from all 1,000,000 acres.

(C) No, since this is an observational study.

(D) Yes, since our results were statistically significant.

5. [2 pts] Discuss as a group how you studied for Exam 2, including what seemed to work and what didn't and how you changed your study habits from Exam 1. Write down suggestions for how to study better for the final exam.