Peer Interactions in Teams and their Spill-over Effect: Evidence from a Natural Field Experiment

Batsaikhan, mongoljin and Kamei, Kenju

Middlesex University London, Keio University

20 May 2025

Peer Interactions in Teams and their Spill-over Effect: Evidence from a Natural Field Experiment

Mongoljin Batsaikhan[1] and Kenju Kamei [2]

[1] Business School, Middlesex University in London, United Kingdom. Email: mongoljin.batsaikhan@gmail.com.

[2] Faculty of Economics, Keio University, Japan. Email: kenju.kamei@gmail.com

May 2025

Abstract: Team-based collaboration is integral to education, work, and daily life, fostering ability-driven peer effects through discussions, social comparisons, and knowledge sharing. Despite extensive evidence of peer effects in specific contexts, their broader impacts on comparable but different activities remain underexplored. Our study addresses this gap using a novel dataset from Mongolia that combines a natural field experiment in classrooms, university entrance examination scores, and grade point averages in the university. First-year undergraduate students were randomly paired to collaboratively complete weekly assignments throughout a course. Low-ability students (based on their entrance exam scores) paired with high-ability peers significantly improved their academic performance not only in the specific course but also in other concurrent courses, showing strong spillover effects. The magnitude of the spill-over relative to the direct effect was 0.723. These pairings had no adverse effects on high-ability students. The findings highlight the Pareto efficiency of peer interactions in groups with large ability differences and offer insights into improving productivity and learning through ability-based spillovers.

JEL classification codes: C93, J24, M53, M54, I23

Keywords: peer effects, spillover effects, a natural field experiment, teamwork

Acknowledgement: The authors thank Bolorsuvd Batbold, Gantumur Purevjav, Munkhtsetseg Delgersambuu, Lkhagvajargal Byambasuren, Unurjargal Davaa, Undral Erdenebaatar, and Otgonbold Dorjderem at the University of Economics and Finance in Mongolia for helping run the field experiment in the 2023/24 academic year, and Nomuunzaya Purevbaasan and Anarmaa Maibat for their excellent research assistance. The natural field experiment reported in this paper received a Research Ethics approval by Middlesex University London (approval #: 26586).

1. Introduction

Situations in team production in which its members collaboratively engage in production activities are ubiquitous, whether in education, workplaces, or private lives. This study proposes that peer interactions during collaborations between high- and low-ability individuals can not only invoke ability-based peer effects on the relevant activities (through discussions and social comparisons), but the effects can also *spill over* to the performance *in their other similar activities* of those involved. This study demonstrates this proposition using a natural field experiment in a higher education environment. Specifically, peer interactions between students working in pairs not only improve their academic performance in a course but also enhance their learning in other courses that they are taking simultaneously, as evidenced by a significant increase in the overall grade point averages (GPAs) of the other courses.

Over the last two decades, both field and laboratory experiments have shown that people's behaviors and performances are affected by the presence and behavior of other members of their peer groups (e.g., Sacerdote, 2001; Carrell *et al.*, 2009; Katz *et al.*, 2001; Lyle, 2007, 2009; Mas and Moretti, 2009; Duflo *et al.*, 2011; De Grip and Sauermann, 2012; Feld and Zölitz, 2017; Booij *et al.*, 2017; Azoulay *et al.*, 2010; Jackson and Bruegmann, 2009; Arcidiacono *et al.*, 2017). "Peer effects" are not only due to learning through social interactions, as discussed in the studies above, but are also driven merely by peer pressure or social information (e.g., Falk and Ichino, 2006; Brune *et al.*, 2022; Guryan *et al.*, 2009). Using university environments, recent work has revealed that low-ability (-achieving) members benefit from peer interactions if they are effectively, e.g., are forced to be, paired with high-ability (-achieving) members (e.g., Lyle, 2009; Kamei and Ashworth, 2023; Wu *et al.*, 2023).[1] Peer interactions may not only help low-ability students to revise their study behavior, but they may also encourage knowledge sharing from high- to low-ability students. However, the existing research has focused on the effects of peer interactions on performance in a specific course/peer group.

Despite the significant importance of peer effects, researchers have paid little attention to the potential positive or negative spill-over effects of peer learning activities in one domain on students' performance in different domains. To fill this gap in the literature, this study

[1] The same effect was recently detected in a laboratory experiment setup with high internal validity (e.g., Kimbrough *et al.*, 2022).

investigates how peer interactions in a specific course influence students' performance in that course, and if yes, whether such academic achievement affects their performance in other courses they are concurrently taking. In particular, this study focuses on the effects of peer interactions on the learning outcomes of low-ability students.

Bias due to endogeneity, namely selection effects (peer groups tend to be endogenously formed) and the reflection problem (peer effects can occur in two ways: from individuals to their partners, and from partners to individuals), causes difficulty in identifying peer effects. The data set used in this study addresses this issue in three ways. First, peer learning activities are intentionally created in the first core course ("Principles of Microeconomics") at the University of Economics and Finance in Mongolia during the fall semester of the 2023/24 academic year. Nearly all first-year undergraduate students in six business majors take this course as their first course, as it is a mandatory course that must be taken for graduation. As such, this study classifies students as high- or low-ability/achieving by using their university entrance examination scores as *pre-treatment* achievement measures. Before conducting the experiment, the university agreed to provide the authors with the confidential data for research purposes by anonymizing them (removing any identifiable information). Second, peer interaction activities proceed in two-person pairs. Each student is *randomly* and *exogenously* paired with another student at the onset of the course by the university; and the pairing remains fixed throughout the course. This pairing is treated as an exogenous shock. Hence, it can be used as an instrument (along with another instrument discussed later) to identify spill-over effects. The "Principles of Microeconomics" course has weekly seminar activities where each student discusses and solves questions collaboratively with their pair mate. While "Principles of Economics" is a large-size course, each seminar group has 16 to 30 students and group compositions stay unchanged throughout the semester. Note that subgroups can emerge endogenously when the interaction unit has more than two students or when students are allowed to choose their partners. Prior research suggests that high-ability students may choose to interact with like-minded high-ability students in a peer group, which could in turn hurt learning for low-ability students (e.g., Carrell *et al.*, 2013; Feld and Zölitz, 2017; Booij *et al.*, 2017). However, this study not only uses two-person pairs in collaborative pair work, but the pairing is also exogenous and no pairing change is allowed; therefore, causal effects of pairing on performance can cleanly be measured here. Each week, immediately after the peer interaction activities, their seminar tutors carefully explain the solutions, related concepts, and the materials to all the students in the seminar groups. Therefore, the knowledge levels are maintained the same for all students. Third, the spill-over effects can be evaluated by using the unique data of students' recorded performances in the other courses. The university allowed the authors to access the students' GPA in the same fall semester after removing any identifiable information.[2] Notably, this enables us to keep track of the students' objective outcome variables without attrition. The GPA is the average of the student's scores in all courses taken, using credits as weights. As a student is scored out of 100 in each course, the GPA is also calculated out of 100 marks.

A key feature of the present experiment is the use of a minimum peer group size (two students). While the literature on peer effects in the microenvironment is scarce, a setup with the minimum size provides a methodological advantage that enables us to determine exactly who interacts with whom. There are two closely related branches of prior studies. The first branch uses random assignment of dorms/roommates in universities. Earlier research has successfully measured peer effects by overcoming issues such as the selection and reflection problems. For example, the seminal work of Sacerdote (2001) uses data from Dartmouth College, which has a policy that incoming freshmen are randomly assigned to dorms/roommates. He found that the students' GPAs were significantly and positively correlated with their roommates' GPAs. Following his research, numerous studies have identified peer effects in setups similar to those of Sacerdote (2001); however, the size and significance of peer effects not only depended on the analysis method used, but the results were also mixed (see Sacerdote 2014 for a survey). These ambiguous results may be attributed to differences in the heterogeneous local environments. For instance, students' interactions with each other in a dorm room or within dorm buildings can differ according to physical factors (e.g., location, university, building, rule) and non-market institutions (e.g., culture, norms, customs). Control can also be an issue here, as researchers do not know the exact kinds of social interactions students have inside their dorms or with their roommate(s).

[2] The "Principles of Economics" course had not had the peer learning component prior to the 2023/24 academic year. Responding to the authors' research proposal, the university agreed not only to offer the research ground during that academic year to test the spill-over effects of peer learning activities, but also to provide the authors with anonymized data (entrance examination scores, the data of the students' course performance and other courses' GPA). Following the agreement, the instructor team of the "Principles of Economics" modified the course structure in collaboration with the authors before the semester began.

The second branch is more recent research using academic classroom interventions, which is the closest to the present research. Compared to the first branch, the educational setups in academic classrooms are more homogeneous across locations, and the contents of the interventions are clearly defined, thereby increasing internal validity. The results showed that low-achieving students benefit from being paired with high-achieving students. First, Li *et al.* (2014), Lu and Anderson (2015), and Wu *et al.* (2023) used the unique educational environments in primary and middle schools in China such that each student sits with the same desk mate for the entire semester (two students sit next to each other, where two-person desks are placed).[3] While desk mate assignments in itself had only weak effects (if any) on performance, low-ability students' performance was found to improve substantially when paired with high-ability students, if the high-ability students were given financial incentives to educate their paired low-ability students (as peer interactions then became active). More recently, Kamei and Ashworth (2023) provided evidence of similar effects in higher education. In their natural field experiment in the United Kingdom, each student was randomly paired with another in a course and engaged in peer learning activities during a tutoring session. Kamei and Ashworth found that low-ability students improved their final exam scores by approximately 7.9% (the improvement was statistically significant) when paired with high-ability, rather than low-ability, students in peer learning activities during the semester. In addition, high-ability students did not show perverse reactions to being paired with low-ability students, unlike people's general tendency to dislike being forced to do something (Kamei and Markussen, 2023). The present study aims to take the next significant step by investigating how peer interactions during learning activities in one setup affect academic performance *beyond the setup*. A university provides an ideal research ground to achieve this aim because students take multiple courses and their seats are not decided across courses, unlike in primary or middle school setups in China.

According to a theoretical analysis, peer interaction activities in a course (task) strengthen the academic performance in that course (task) for low-ability students who are paired with a high-ability student, if they are concerned about social factors (such as shame), image, and the like. Through social comparisons, low-ability students may revise their study behavior in the

[3] Li *et al.* (2014) and Wu *et al.* (2023) used primary school students in China, while Lu and Anderson (2015) used middle school students (grade 7) in China. The students take all courses in the same seats throughout the semester.

course to reduce such negative psychological effects. By contrast, the effects of peer interaction experiences on performance in other courses are unclear. On the one hand, improved achievement in the specific topic (task) may help low-ability students improve their academic (work) performance beyond the setup. This is because a deep understanding of a topic, a sense of accomplishment, and the confidence gained, may motivate students to study other subjects as well. The knowledge and skills gained from studying the specific subject (e.g., analytical skills, mathematical thinking) may also be applicable to other subjects. In a theoretical model, this can be modeled as a reduction in their study effort cost in other concurrent courses. If this hypothesis is correct, a positive effect will emerge beyond the course with peer interaction activities. However, factors such as narrow bracketing and mental accounting may hinder their ability to concurrently focus on other courses. For example, achieving or anticipating better academic performance in one course may justify their reduced effort in studying their other courses.

The effects of peer interactions on high-ability students are theoretically ambiguous. Social effects (such as pride) or image concerns may encourage high-ability students to study harder when paired with low-ability students rather than with high-ability students. However, under this circumstance, inequality-averse concerns may cause high-ability students to weaken their study efforts.

In the "Principles of Microeconomics" course, each student had a (randomly assigned) fixed partner throughout the semester in their seminar sessions. Every week, prior to attending seminars, students attempted problem sets individually. In seminars, they sat next to each other, discussed each other's answers, and then jointly wrote single answers as pairs to the problem sets after agreement. The end-of-semester final examination results (answered individually) indicated that low-ability students obtained significantly better scores when with high-ability rather than low-ability students in the weekly peer interaction activities. The effect size was large (5.769 out of 100 points, i.e., 0.325 standard deviation of the exam score). By contrast, the exam performance of high-ability students was not affected by with whom they were paired.

The interactive activities among peers had striking spill-over effects. The academic achievement in the "Principles of Microeconomics" course helped low-ability students improve overall GPA in other courses significantly. The relative magnitude of the spill-over effect to the direct effect was 0.723. Hence, pairing with a greater spread in abilities and achievement is a

Pareto improvement. This implies that it is misleading to gauge peer effects only from relevant activities, and the ability peer effects could comprise direct and spill-over effects. In some sense, this collaborates with prior research findings that peers influence not only students' academic performance but also their social behavior—another type of spill-over effect. Our finding is the first to demonstrate a spill-over effect on other academic activities.

The remainder of the paper proceeds as follows: Section 2 summarizes the experimental design and logistics, and Section 3 discusses the hypotheses and empirical strategies. Section 4 reports the results. Section 5 discusses and concludes the study.

2. Experimental Design and Procedure

The natural field experiment is conducted in "Principles of Microeconomics" (ECN111), the first economics course, at the University of Economics and Finance (Mongolia) in the fall semester of the 2023/24 academic year. The syllabus is included in Appendix A.1. The university is located in Ulaanbaatar, and is highly ranked in Mongolia with high-quality entrants. The "Principles of Microeconomics" is a compulsory course that undergraduate students in six non-economics majors (accounting, banking, business management, marketing, trade, finance) must take to graduate; almost all students take the course when they are first-year students. The course covers several fundamental topics in consumer and producer theories and equilibrium analysis, such as demand and supply, elasticity, preferences, market structure (e.g., oligopoly), and externality.[4] This course comprises three key activities, and students' final grades are assessed based on multiple factors (see details below). In their official transcripts, students receive both raw total scores (out of 100 marks) and letter grades—A (90-100), B (80-89), C (70-79), D (60-69) or F for failing to obtain the credit when the total is below 60%.

The fall semester began on September 1, 2023.[5] The three key course activities were lectures, seminars, and individual work. First, there were a total of 15 blocks (one block per week), starting from the first teaching week, built on a flipped learning model. Attendance at lectures and seminars was mandatory and accounted for five marks out of the total score in this course. Specifically, attendance score was calculated as: 5 × attendance rate (e.g., a student gets

[4] The course closely follows the textbook by Parkin (2018), *Microeconomics* (8th edition))'s officially translated version in Mongolian. Each lecture (week) covers one chapter of the textbook.
[5] In the 2023/24 academic year, 520 undergraduate students took the "Principles of Microeconomics" course.

4.5 = 5 × 0.9 marks if the attendance rate is 90%). One week before each block, the students were assigned a problem set for teamwork and another non-graded problem set for their own practice. Each block had three sessions in sequence on the same date, lasting three hours with a 10-minute break between sessions. The first session was a 50-minute lecture. The second session comprised 50 minutes of teamwork activities divided into 30 minutes of team study to discuss, write, and submit a joint solution for the teamwork problem set, followed by 20 minutes in which the seminar tutors explained the solution. The third session comprised 50 minutes of another seminar for the tutors to explain the answers to the other pre-assigned seminar questions and end-of-chapter questions related to the lecture materials. The lecture materials (e.g., slides and recorded video lectures) were pre-uploaded on the course's online platform (that uses Moodle). The students were required to watch the recorded lectures independently before attending class. The first 50 minutes (lecture) in each block were dedicated to one-way delivery of an overview of the recorded materials (although the students were free to ask questions during the lecture). All students attended the same lecture in a large lecture hall.

The student record office allocated students to one of 14 seminar groups before the fall seminar began. The course had six seminar tutors, one of whom was assigned to each seminar group before the first seminar session. The seminar group composition remained the same for the entire semester. The same seminar tutor was in charge of their assigned seminar group for the second and third sessions in all blocks. In each seminar group, students were randomly paired with one another, except that one interaction unit was a three-student team if the number of students in the group was odd.[6] The pairing remained the same for all seminar activities.[7] Each block had the 30-minute teamwork study period, except for the examinations and short

[6] Whether a student was assigned to a two-person team or a three-person team was completely random and exogenous in a seminar group when its size was an odd number. Not only does learning experience differ by the team size but it is also not possible to use the same method to define high- or low-ability students per team for teams with different sizes. Therefore, this study uses the recorded data of two-person teams as the data set. As almost all teams had two individuals, this omission has minimal effects on the empirical analysis in Section 4.

[7] The only exception was when a student withdrew from the course. If two students in different pairs dropped out in the same week, the remaining two were matched together. This happened with only 14 students in the experiment. If only one student in a pair dropped out, that student was randomly added to a group of two students in the same seminar. This happened with one student. As this regrouping may have affected students' learning and the timing of dropping out and regrouping differed by case, the data were removed from the analysis. Together with the omission indicated in footnote 6, the number of eligible students in the analysis was 446 undergraduate students (out of which 442 are first-year students).

assignment weeks.[8] The authors were not involved in seminar group allocation or pairing processes.[9] Note that this collaborative teamwork has not been conducted during seminars in the past years.

As discussed above, a set of teamwork questions (see Appendix A.2 for an example) was distributed one week prior to each block. Students were asked to solve the problems individually in a notebook before attending the seminar. This pre-study was the basis for the team discussions to write joint answers to the problem set in class. During the 30-minute teamwork period, the teammates sat next to each other. Seminar tutors monitored and, as necessary, clarified the team task but did not provide hints to solve the problem. All teams were required to submit their joint answers in 30 minutes, after which their seminar tutors carefully explained the solutions. To gauge the levels of their pre-study, when students started to discuss and work on the problem set with their partners, their tutors collected the seminar notebooks from the students individually; the tutors checked and recorded whether students came prepared with their own solutions before class. However, the students' scores for seminar activities were determined solely by their jointly submitted team answers and not by their pre-study levels. The seminar tutors marked each team's jointly submitted script and privately returned it to the team in the following week's block. Two (three) persons in a pair (a three-student team) received the same score, except when they were absent.[10] As students' learning can be affected by their assigned seminar group and/or tutor because they have social interactions inside the group and tutors' instructions may vary, seminar group clustering and seminar tutor dummies are included in the data analysis in Section 4.

Seminar attendance was mandatory for the course. The average attendance rate was 91.8%.[11]

---

[8] Short assignments are shorter versions of examinations in which students would take short quizzes from the end-of-chapter questions independently in a computer room. These were given in a multiple-choice question format similar to the examination format.

[9] The authors of this paper are academic staff members at different universities. They had neither teaching duties at the University of Economics and Finance nor interactions with the students in the course. The setting that a neutral teaching team whose members do not know of the research purpose conducts an experiment is ideal, minimizing possible experimenter demand effects (e.g., Friedman and Sunder, 1994).

[10] Student received zero marks if they were absent in a seminar or if they did not make any attempt for the teamwork question in advance.

[11] As detailed later in this section, students are classified as either high- or low-ability based on their Mathematics score in the university entrance examination. The attendance rates were similar for the two classifications: The average attendance rate of the low-ability students who were matched with low-ability students (high-ability students) was 88.8% (91.9%), and that of the high-ability students who were matched with low-ability students

Finally, students had to solve homework individually and independently every week in the format of a test, outside lecture and seminar periods. These problems were similar to the exercises at the end of each textbook chapter. This activity was tested twice in the "short assignment" weeks (Week 6 and Week 14— see Appendix A.1) in which students were given some random questions from the end-of-chapter questions during the assignment.

The midterm and final examinations were scheduled in the seventh week (October 16-20, 2023) and after the 15th week of teaching, i.e., in the 16th week (December 18-22, 2023), respectively.[12] Both examinations were conducted during seminar class hours using the course's online platform. Each student was assigned a randomly constructed test from the textbook test bank. Appendix A.3 includes examples of these questions. Each student's total score in the course (100 marks) comprised the following six factors: midterm exam (20 marks), final exam (30 marks), seminar score (10 marks), short assignments (30 marks), attendance at lectures (5 marks) and attitude (5 marks). Attitude is each instructor's subjective evaluation of the students' attitudes toward learning and destructive behaviors during class. Almost all students achieved 100% in this case with the average attitude score of 4.94 out of 5 marks.

It should be noted here that collaborative seminar activities were set as usual teaching activities. Hence, the external and internal validity of the intervention was maximized, because students engaged in these activities without knowing about the on-going experiments.

This study classified students from an absolute perspective in the cohort using their university entrance examination scores in Mathematics. The entrance examination scores are ideal to gauge the students' ability/achievement levels prior to the course, because the "Principles of Microeconomics" is the first course that they take after entering the university. Mathematics is a mandatory subject for entrance examinations at the University of Economics and Finance. The rationale for using Mathematics scores in classification is that mathematical ability is known to strongly affect students' performance in economics courses (e.g., Ballard and Johnson, 2004; Arnold and Straten, 2012). Students were divided into two categories based on the entrance examination scores. Students with scores above (below) the average Mathematics score were classified as high-ability/achieving (low-ability/achieving). As students' performance

(high-ability students) was 91.1% (95.4%). The difference in the average attendance rate between any two sets is far from significant (two-sided $p > 0.100$, Mann-Whitney test).

[12] The exact dates of the midterm and final exams as well as short assignments varied by the seminar group.

within their classified ability group may also be affected by their mathematics ability, the Mathematics score can also be used as an instrument along with the exogenous pairing dummies in the analysis (see Section 3.2 for the detail).

In sum, this study uses a novel data set comprising the students' confidential performance information in "Principles of Microeconomics," their entrance examination scores, and the GPAs of the other courses they took in the same semester, following the ethics board's approval and the university's permission.[13]

3. Hypothesis and Empirical Strategy

3.1. Hypothesis

While students solve the problem set jointly as a pair during the first 30 minutes of seminars, they are not only carefully explained by their seminar tutors the solutions and related concepts, but they have also interactive discussions with the tutors, in the following 20 minutes. Thus, all students can acquire similar skills and knowledge of the assigned problem sets, regardless of their scores in the collaborative pair work. However, social comparisons with their pair mates can invoke social (psychological) effects such as shame and pride among students.[14]

A simplified version of the theoretical framework used by Kamei and Ashworth (2023) can be applied to mathematically discuss how pairing two students with different ability levels affects their study effort provision. Suppose that $y_{i\in\{h,l\}}$ is the effort that type i (the high-ability or low-ability student in the pair) puts in "Principles of Microeconomics." Assume, for simplicity, that the payoff function for academic performance in the course is expressed as Equation (1):

$$\pi_i(y_i) = g_i(y_i) - \gamma_i {y_i}^2. \tag{1}$$

$$g_i(y_i) = \alpha_i + \beta_i y_i. \tag{2}$$

[13] The students took a variety of different types of courses, other than "Principles of Microeconomics," such as ethics, languages, humanities and business. The number of the other courses that at least one of the students took was 65.

[14] Further, peer interactions may help correct students' misperceived own performance and standing in the classroom, as people are known to be over-optimistic about their own ability, thus forming biased beliefs about their performance (e.g., Abeler *et al.*, 2011; Svenson, 1981; Langer, 1975; Svenson, 1981; Larkin *et al.*, 2012). While the theoretical analysis discussed below is built on social effects, adding its effects on correcting misperceptions would strengthen the direct effects summarized in Hypothesis 1.

Here, $g_i(y_i)$ denotes the degree of academic performance. g_i is assumed to be increasing in y_i subject to α_i and β_i that satisfy the following conditions: $\alpha_h > \alpha_l > 0$ and $\beta_h > \beta_l > 0$. $\alpha_h > \alpha_l$ means that the high-ability student has higher skills than the low-ability student even before exerting any effort, while $\beta_h > \beta_l$ means that the former is a fast learner, i.e., the high-ability student has a higher unit return from putting effort than the low-ability student. $\gamma_i {y_i}^2$ in Equation (1) is the effort cost that type *i* incurs. It can be assumed that $\gamma_l > \gamma_h > 0$: the low-ability student incurs a higher unit effort cost from study than the high-ability student.

Using the first-order condition, the optimal effort provision each by *i* = *h* and *l* can be derived as follows:

$$y_h{}^* = \frac{\beta_h}{2\gamma_h}. \tag{3}$$

$$y_l{}^* = \frac{\beta_l}{2\gamma_l}. \tag{4}$$

It follows clearly that $y_h{}^* > y_l{}^*$ and thus $g_h(y_h{}^*) > g_l(y_l{}^*)$, as $\alpha_h > \alpha_l$, $\beta_h > \beta_l$ and $\gamma_h < \gamma_l$.

Pair work enables students to compare their academic performance with that of their peers. Psychological effects that are potentially activated are, for example, shame and pride (Bowles and Gintis, 2005), self- and/or social-image concerns (Bénabou and Tirole, 2006 and 2011), or inequality aversion (Fehr and Schmidt, 1999). These psychological effects are kicked in only when two students with different ability levels are paired together. Assume that students incur disutility from shame, image concerns, or inequality aversion, if they realize that their academic performance is worse than that of their paired mate. Each psychological factor decreases the utility of low-ability students. By contrast, the psychological effects on high-ability students are ambiguous. High-ability students may enjoy feelings of pride or positive self/social image, which increases their utility, when they are aware that their performance is better than that of their pair mate. However, in this circumstance, it is also possible that high-ability students experience disutility because of aheadness aversion.

To show the roles of these non-material preferences, assume the utility functional forms for each type, $\theta_{i\epsilon\{h,l\}}$ as follows for simplicity:

$$\theta_h(y_h, y_l) = \pi_h(y_h) - \mu_h f(g_h(y_h) - g_l(y_l)). \tag{5}$$

$$\theta_l(y_l, y_h) = \pi_l(y_l) - \mu_l f(g_h(y_h) - g_l(y_l)). \tag{6}$$

Here, μ_h and μ_l are utility weights on feelings of pride or shame, image concerns, and inequality aversion, by high- and low-ability students, respectively. $\mu_l > 0$ always, while $\mu_h < 0$ ($\mu_h > 0$) if the total positive effects of pride and image concerns are stronger (weaker) than a utility loss due to aheadness aversion. *f* is a function that explains how the intra-pair academic performance difference affects the payoff. *f* is increasing in the argument (i.e., *f'* > 0); its form can be convex, linear, or concave. The optimality conditions are reduced to Equations (7) and (8) for high- and low-ability students, respectively:

$$\beta_h - 2\gamma_h y_h{}^{**} - \mu_h \beta_h f'\big(g_h(y_h{}^{**}) - g_l(y_l{}^{**})\big) = 0. \quad (7)$$

$$\beta_l - 2\gamma_l y_l{}^{**} + \mu_l \beta_l f'\big(g_h(y_h{}^{**}) - g_l(y_l{}^{**})\big) = 0. \quad (8)$$

Equation (8) suggests that a low-ability student exerts more effort due to psychological effects when matched with a high-ability student than otherwise (i.e., $y_l{}^{**} = \frac{\beta_l}{2\gamma_l} + \frac{\mu_l \beta_l}{2\gamma_l} f'(g_h(y_h{}^{**}) - g_l(y_l{}^{**}) > \frac{\beta_l}{2\gamma_l}$). However, as shown in Equation (7), the impact on the high-ability student's effort provision can be positive or negative depending on μ_h; and no directional predictions are possible.

Hypothesis 1: *Low-ability students achieve better academic performance in the "Principles of Microeconomics" course, when they are matched with a high-ability student rather than a low-ability student in the pair work.*

The effects of peer interactions on other courses that students concurrently took are ambiguous. First, a deep understanding of the materials in "Principles of Microeconomics" and a sense of accomplishment there may motivate low-ability students to study other subjects more diligently. For instance, a student who finds the study of economic theory interesting for analyzing economic phenomena may also become interested in studying languages logically, driven by their enhanced positive mental state and increased confidence. This effect can be modeled as a reduction in study effort cost. Denote a low-ability student's effort in the other courses and his/her belief about the average performance levels of other students there as e_l and $\bar{q}^b$, respectively. As assumed previously, if the student anticipates that his/her performance in the other courses is worse than that of the other students on average, his/her utility is expressed by $h_l(e_l, \bar{e}^b)$ below.

$$h_l(e_l, \bar{e}^b) = \varphi_l(e_l) - \mu_l f(\bar{q}^b - q_l(e_l)). \quad (9)$$

where $\varphi_l(e_l) = q_l(e_l) - \gamma_l e_l{}^2$. $q_l(e_l) = \delta_l + \rho_l e_l$ is the student's academic performance in the other courses ($\delta_i > 0$, $\rho_i > 0$, and $\gamma_i > 0$). As derived previously, the first-order condition reduces as follows:

$$\rho_l - 2\gamma_l e_l + \mu_l \rho_l f'\left(\bar{q}^b - q_l(e_l)\right) = 0.$$

Here, $\mu_l \rho_l f'\left(\bar{q}^b - q_l(e_l)\right) > 0$. Hence, low-ability students put more effort into other courses if they have lower unit effort cost γ_l thanks to our intervention.

Hypothesis 2: *Low-ability students improve academic performance in the other courses they are concurrently taking when they are paired with high-ability students rather than low-ability students in "Principles of Microeconomics."*

Notably, the peer interaction activities in "Principles of Microeconomics" may negatively affect performance in other courses they are taking simultaneously. For example, the above analysis did not consider heuristics in decision-making. Spending more time in studying the materials of "Principles of Microeconomics" to avoid feeling shamed or achieving better academic performance in that course may mean that low-ability students reduce study effort in the other courses they are taking, owing to mental accounting (e.g., Thaler, 1999) and narrow bracketing (e.g., Camerer *et al.*, 1997). For instance, students may have their targeted study effort or targeted academic performance when deciding whether to work or enjoy leisure activities. Increasing study effort in one course or achieving better performance in that course may justify their reduced effort to study their other courses. These possibilities lead to the following alternative hypothesis to Hypothesis 2:

Hypothesis 2': *Low-ability students achieve poorer academic performance in the other courses they are concurrently taking when they are paired with a high-ability student rather than a low-ability student in "Principles of Microeconomics."*

3.2. Empirical Strategy

As pairing in the seminar activities is random and exogenous, this study adopts a simple empirical strategy to estimate the treatment effects. Denote y_i and GPA_i as the academic performance in "Principles of Microeconomics" and the GPA of the other courses for student *i*, respectively. The effect of peer interactions on y_i (direct effect) can be estimated by using a linear regression with the exogenously determined pairing as an independent variable. In the regression, their Mathematics score in the university entrance examination is also included as an independent variable, since mathematics ability is known to affect performance in economics courses. Setting the regression to estimate the direct effect as the first-stage regression, the effect of peer interactive experience on GPA_i (the spill-over effect) can be estimated by using two-stage least squares (2SLS).

As discussed in Section 2, each student in "Principles of Microeconomics" is classified as high-ability (low-ability) if the student's Mathematics score in the university entrance examination is above (below) the average. Appendix Figure B.1 reports a histogram of the scores. Treatment effects are estimated separately for high- or low-ability students.

Let us discuss the estimation procedure for low-ability students here. Partner *j* of low-ability student *i* is either a high- or low-ability student. Thus, the first-stage regression equation can be expressed as Equation (10):

$$y_{ig} = \alpha + \beta \cdot 1_{\{j=\text{high}\}} + \delta \cdot math_i + \boldsymbol{x}_{ig}\boldsymbol{\gamma} + \varepsilon_{ig}. \quad (10)$$

In this equation, $1_{\{j=\text{high}\}}$ is an indicator variable that equals 1 if partner *j* is a high-ability student (the reference group comprises low-ability students who are matched with a low-ability student). $math_i$ is student *i*'s Mathematics score in the entrance examination. Subscript g denotes the seminar group ID. $\boldsymbol{x}_{ig}$ is the set of other control variables. This includes *i*'s gender dummy, *i*'s partner *j*'s gender dummy, the interaction between the two gender dummies, *i*'s major dummy, and a dummy indicating the instructor of the seminar group to which *i* and *j* belong. Notice that gender composition in pairs may affect the size of peer effects (e.g., Black *et al.*, 2013; Lavy and Schlosser, 2011; Lu and Anderson, 2015). Teaching ability and methods may differ by seminar instructor, although they are instructed to teach in a certain manner. The error term ε_{ig} is assumed to be correlated within seminar group g (i.e., $Cov[\varepsilon_{ig}, \varepsilon_{jg}] \neq 0$), as seminar activities other than teamwork may be somewhat interactive.

A common concern in randomized controlled trials is non-compliance. This concern is absent in this study, as "Principles of Microeconomics" is a compulsory course, and any changes in pairings are not allowed in the peer interaction activities. This makes the estimation in the first stage straightforward, as discussed above.

The equation to be estimated for the spill-over effect (second-stage regression) can be written as in Equation (11):

$$GPA_{ig} = \theta + \mu y_{ig} + \boldsymbol{x}_{ig}\boldsymbol{\varphi} + v_{ig}. \tag{11}$$

As peer discussions are limited to only 30 minutes per week, it can be assumed that the peer interactions do not produce strong enough effects to change low-ability students' general study behavior. Instead, it is reasonable to suppose that the experience affects their study effort locally due to factors such as the social effects discussed in Section 3.1. It can then be assumed that better achievement in "Principles of Microeconomics" (an increase in y_{ig}) affects their academic performance in other courses, due to a reduction in effort cost there. The issue here is that y_{ig} is treated as endogenous in Equation (11). With the ordinary least squares, y_{ig} may be correlated with the error term v_{ig}. For example, the spill-over effect μ is subject to upward (downward) bias if unobserved characteristics that have positive effects on GPA_{ig} are positively (negatively) correlated with y_{ig}.[15] The exclusion of the $1_{\{j=\text{high}\}}$ and $math_i$ variables in Equation (11) is valid, as partner assignment is completely random in the course, and mathematics ability is known to affect students' performance in economics studies more strongly than in other non-economics studies.[16] While this study uses the $1_{\{j=\text{high}\}}$ dummy and Mathematics score (entrance examination) as instruments to identify the spill-over effect, the validity of these instruments (i.e., tests of underidentification and weak identification) are checked using Kleibergen and Paap's (2006) rk LM and Wald statistics.

The direct and spill-over effects are estimated for high-ability students in the same manner as for low-ability students just discussed.

[15] There are numerous examples. For instance, students' IQ that cannot be captured by mathematics ability (e.g., verbal ability) may positively affect both y_{ig} and GPA_{ig}, which could inflate the estimate of μ.

[16] If we use only the $1_{\{j=\text{high}\}}$ dummy as an instrument and include the Mathematics score as a control for both the first- and second-stage regressions, the Mathematics score obtains a strongly significant and positive coefficient at the 1% level in the first-stage regression; but it has a coefficient estimate of almost zero (far from significant) in the second-stage regression. This also empirically justifies in our data set that mathematics ability matters substantially (matters little) in "Principles of Economics" (in the other courses) in the University of Economics and Finance.

4. Results

4.1. Overview of the Data

The number of eligible students (those whose pairings remained the same for the entire semester) was 446—see footnote 7. As shown in Appendix Figure B.1, students' Mathematics scores in the university entrance examination varied greatly. The mean score was 620.5 out of 800. Of the 446 students, 230 were classified as low-ability (their Mathematics scores were less than the mean) and 216 were classified as high-ability (their Mathematics scores were higher than the mean). As a result of random pair formation executed by the university, the number of pairs whose members were both high-ability students—high-high pairs, hereafter—was 66 (132 students), that of pairs whose members were both low-ability students—low-low pairs, hereafter—was 73 (146 students), and that of pairs in which a high-ability student was paired with a low-ability student—high-low pairs, hereafter—was 84 (168 students).

As attending seminars was mandatory, the students' attendance rates were very high. The overall attendance rate was 91.81%.

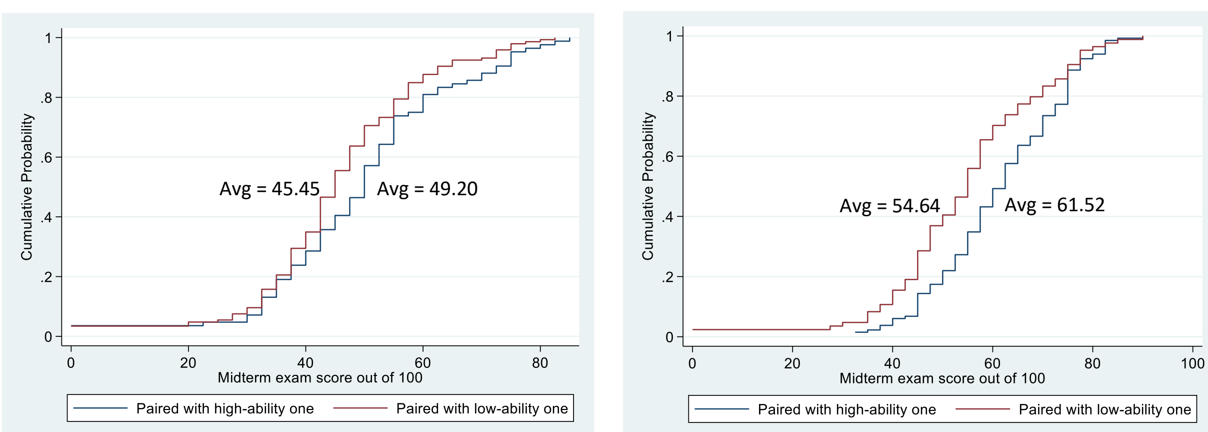
On average, each assessment component was better for high-ability students than for low-ability students. A close look at the data shows that each ability type performed much better when paired with a high-ability student rather than with a low-ability student in the pair work of the seminar. Figure 1 reports the cumulative distributions of students' midterm exam scores as an interim achievement measure, their final exam scores as a final achievement measure, and their aggregate total scores as an overall evaluation measure for the course (Panels A to C).[17] Irrespective of ability classification, students' performances were higher when they were paired with a high-ability student rather than a low-ability student. The positive effects of pairing appear to spill over to students' performance in the other courses that they took concurrently (Panel D). Section 4.2 investigates whether the direct and spill-over effects are significant, following our empirical strategy (Section 3.2).

---

[17] Unlike their exam performances and aggregate total scores, the students' attendances rates were affected by the pairing only to modest degree, due to the fact that seminar attendance was mandatory. The average attendance rate of high-ability students was 95.41% in the high-high pairs, whose percentage was somewhat higher than when they belonged to the high-low pairs (91.16%). The average attendance rate of low-ability students was 91.97% in the high-low pairs and 88.84% in the low-low pairs.

It is essential to check whether key demographic variables are balanced before conducting an econometric analysis. As previously discussed, the research team obtained the following individual characteristics of students from university administration data: gender, majors (finance-related majors or not),[18] and the Mathematics scores in the university entrance examination. Appendix Table B.1 reports the test results regarding whether these variables are balanced between those paired with a low-ability student and those paired with a high-ability student. This revealed that all variables are statistically balanced for low-ability students (Panel A of Table B.1). Hence, the clear performance differences shown in Figure 1 (Panels A.i, B.i, C.i, and D.i) are not due to a possible imbalance in the characteristics. However, this is not the case for high-ability students. While gender variables are balanced, by chance, high-ability students who were paired with a high-ability student had significantly higher entrance examination scores compared with those who were paired with a low-ability student. In addition, the distributions of majors are not balanced among high-ability students. Thus, the observation that the performance of high-ability students differed by pairing (Panels A.ii, B.ii, C.ii, and D.ii) may be due to unbalanced individual characteristics. Thus, it is crucial to control for these characteristics when estimating direct and spill-over effects for high-ability students.

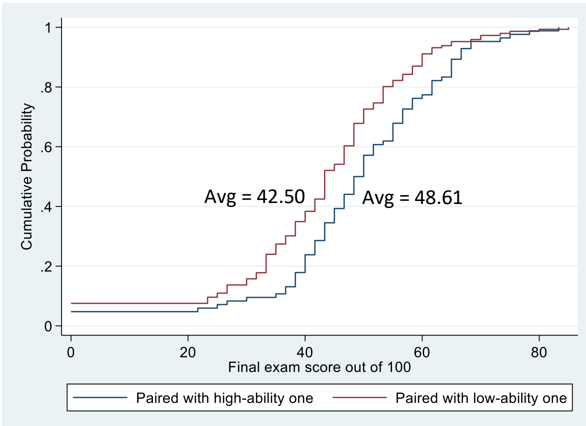
Figure 1: *Cumulative Distributions of Students' Performance by Pairing Outcome*



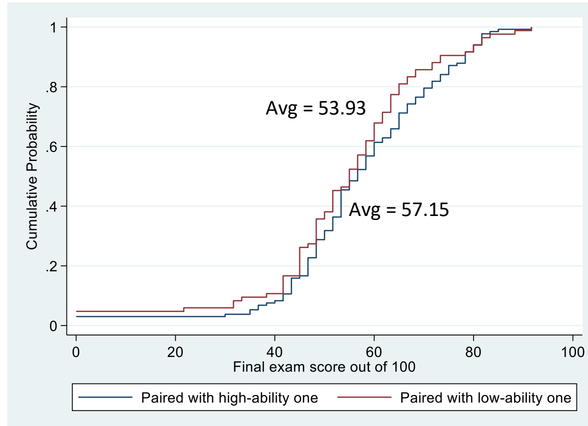
(i) Low-ability students (ii) High-ability students

A. Midterm Exam Score

[18] All the students enrolled are business majors (non-economics). As finance is the closest subject to economics among business majors, the major information may be correlated to their performance in an economics course.

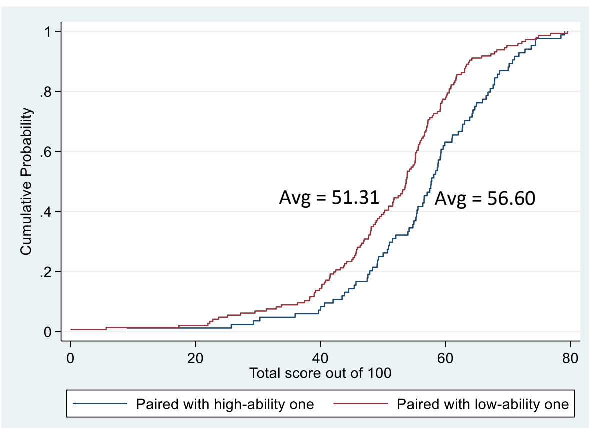


(i) Low-ability students

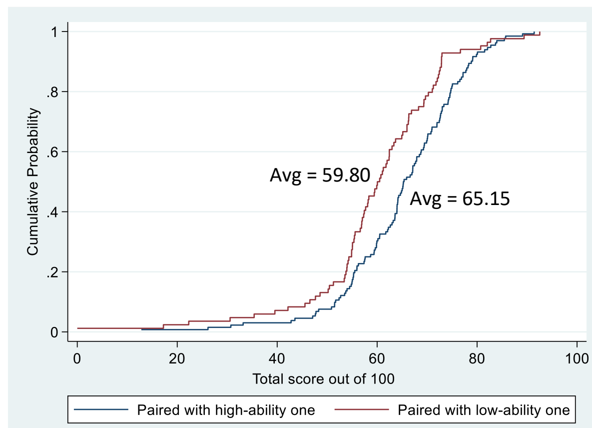


(ii) High-ability students

B. Final Exam Score

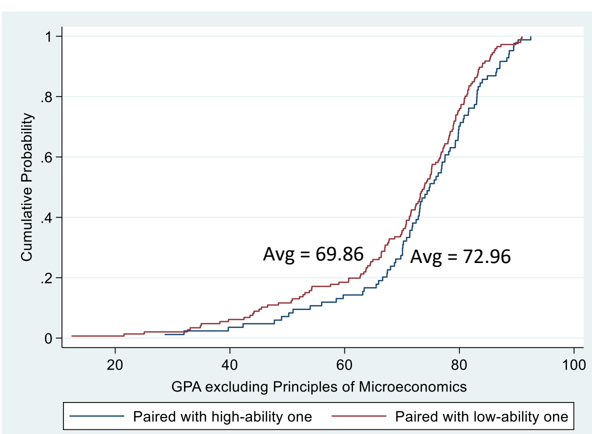


(i) Low-ability students

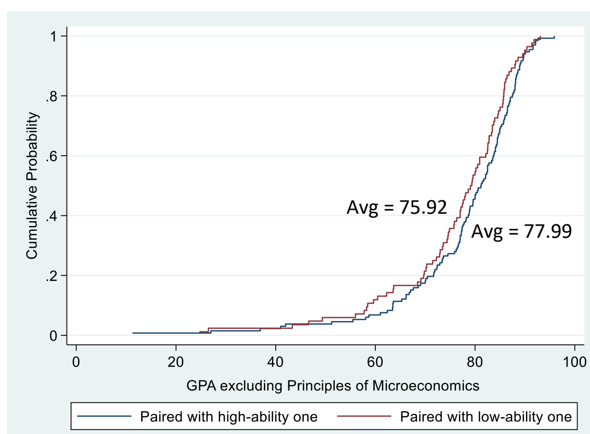


(ii) High-ability students

C. Total Score



(i) Low-ability students



(ii) High-ability students

D. GPA of the Other Courses

Notes: All four measures (midterm exam score, final exam score, total score, and GPA of the other courses) were calculated out of 100. The midterm and final exam scores accounted for 20% and 30% of the total scores, respectively.

4.2. Identifying the Direct and Spill-over Effects of Peer Interactions

The potential gains that students enjoy from their paired high-ability counterparts can be identified using OLS, as discussed in Section 3.2. Table 1 reports the regression results when using three outcome measures as dependent variables: midterm exam scores (Models 1 and 2), final exam scores (Models 3 and 4), and aggregate total scores of the course (Models 5 and 6).[19] While the odd-numbered models used only the high-low (high-high, for Panel B) pair dummy and the Mathematics score as independent variables, the even-numbered models used both these two variables and other controls as independent variables. The results revealed a clear contrast between high- and low-ability students. First, the low-ability students not only made significantly stronger achievements at the end (Models 3 and 4), but they also obtained better total scores in the course (Models 5 and 6), when they were paired with a high-ability student than otherwise. The results for the strong direct effects are robust regardless of whether controls are included (Models 3 to 6). These impacts correspond to an increase of 0.325 S.D. (standard deviation) for their final achievement, and an increase of 0.341 S.D. for their total scores, if we use the estimation results in Models 4 and 6. Note that the effects of pairing were weak at the interim stage, as gauged by the midterm exam scores (Models 1 and 2). This may indicate that a sufficiently long duration of peer interactions is required to generate an effect.

Second, the direct effects were weak for high-ability students (see Panel B). On the one hand, they suffered in the short run when a low-ability student rather than a high-ability student was assigned as a partner. The negative effect was detected as significant, at least at the 10% level (Models 1 and 2). However, in the long run, their final achievement levels (Models 3 and 4) were not affected by the pair assignment, perhaps because of their ability to learn the content by themselves. Owing to this, their total course scores were also not affected by pair assignments (Models 5 and 6). Note that the high-high pair dummy obtains a significant and positive coefficient (Model 5), but this is due to an imbalance in individual characteristics. When the characteristics are controlled for, the dummy variable fails to reach significance (Model 6).

[19] The total score is an accumulated score comprising the midterm exam (20%), final exam (30%), short assignments (15% × 2), seminar (10%), attendance (5%), and attitude (5%). Since seminar scores were shared between pairs, low-ability students might benefit merely from being paired with a high-ability peer simply due to the score sharing. Appendix Table B.4 presents estimation results in which the seminar score is excluded from the total score. The results are qualitatively almost identical to those reported in Tables 1 and 2.

Result 1 (direct effect): *(a) Low-ability students improved performance in Principles of Microeconomics more strongly when paired with high-ability students, rather than low-ability students. (b) High-ability students' performances (whether achievement or evaluation) were not affected by being forcedly paired with low-ability students.*

Table 1: *Direct Effects of Being Paired with High-Ability Students*

A. Low-ability students

Dependent variable: / Independent variable:	Midterm exam score		Final exam score		Total score	
	(1)	(2)	(3)	(4)	(5)	(6)
High-low pair dummy {=1 if the partner is a high-ability student}	3.031 (2.080)	2.790 (2.022)	5.551*** (1.781)	5.769*** (1.808)	4.744** (1.734)	4.731** (1.878)
Math score (entrance exam)	0.098*** (0.027)	0.096*** (0.030)	0.076*** (0.025)	0.080*** (0.027)	0.074*** (0.015)	0.074*** (0.017)
Finance-related major dummy	----	3.219 (2.737)	----	-0.068 (1.643)	----	1.744 (1.570)
Female dummy {=1 for female}	----	-4.760 (4.669)	----	0.483 (4.442)	----	0.663 (3.333)
Female partner dummy {= 1 if paired partner is female}	----	-10.159** (3.943)	----	-5.548 (3.341)	----	-5.191** (2.134)
Interaction: Female dummy × Female partner dummy	----	7.510 (5.770)	----	0.187 (5.157)	----	1.396 (3.671)
Constant	-9.039 (15.566)	-4.210 (17.890)	0.152 (13.961)	-1.335 (14.788)	9.728 (9.081)	9.905 (9.772)
Observations	230	230	230	230	230	230
R-squared	0.079	0.135	0.065	0.108	0.096	0.145

B. High-ability students

Dependent variable: / Independent variable:	Midterm exam score		Final exam score		Total score	
	(1)	(2)	(3)	(4)	(5)	(6)
High-high pair dummy {=1 if the partner is a high-ability student}	6.019** (2.185)	5.409* (2.881)	2.486 (2.147)	0.325 (3.246)	4.722*** (1.543)	2.893 (2.315)
Math score (entrance exam)	0.049*** (0.016)	0.048** (0.020)	0.042** (0.016)	0.014 (0.014)	0.037** (0.014)	0.020* (0.011)
Finance-related major dummy	----	3.288 (3.163)	----	11.418*** (2.838)	----	7.244*** (2.284)
Female dummy {=1 for female}	----	5.646* (2.997)	----	2.826 (3.099)	----	4.186* (2.253)
Female partner dummy {= 1 if paired partner is female}	----	-1.513 (3.227)	----	-6.854*** (2.095)	----	-4.408* (2.492)
Interaction: Female dummy × Female partner dummy	----	-0.958 (3.925)	----	4.469 (4.459)	----	2.855 (3.497)
Constant	21.541*	18.172	25.536**	38.610***	35.051***	40.266***

	(10.446)	(12.794)	(11.565)	(8.916)	(9.876)	(6.984)
Observations	216	216	216	216	216	216
R-squared	0.089	0.134	0.027	0.179	0.062	0.213

Notes: A linear regression with robust standard errors clustered at the seminar group level. Numbers in parentheses are standard errors. Instructor dummies are also included in the even-numbered models (the estimates are omitted to conserve space). All observations of low-ability (high-ability) students were used in the estimations in Panel A (Panel B). *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The next question is whether the strong direct effect detected for low-ability students can spill over to their academic performance in other courses that they took concurrently. Table 2 reports the results of the 2SLS following our empirical strategy (Section 3.2). The dependent variable is the GPA of the other courses a student took concurrently. Either their final exam score (i.e., a final achievement measure) or total score in the course (i.e., a final evaluation measure) was used as each student's performance measure of the "Principles of Microeconomics" course. The first-stage regression results were precisely the ones included in Panel A of Table 1. The high-low pair dummy and Mathematics scores (entrance exam) are valid instruments, as evidenced by the tests of underidentification and the Kleibergen-Paap Wald-F statistic.

The regression results indicate significant and positive spill-over effects. Regardless of the performance measure used (final exam or total score), it has a significantly positive coefficient. This effect is economically large. The average total score of the low-ability students in "Principles of Microeconomics" was 5.29 points (= 56.60 – 51.31) higher in the high-low pairs than in the low-low pairs—see Panel C of Figure 1. The estimation result for Model 4 in Table 2 suggests that this boosted the GPA of the other courses they took by 5.29 × 0.723 = 3.82 marks. The S.D. (standard deviation) of the GPA of the other courses was 14.01 marks. Thus, this effect corresponds to an increase of 0.273 S.D. of the GPAs. If we instead use their final exam scores as the medium of spillover, the impact is calculated as (48.61 – 42.50) × 0.638 = 3.90 marks—see Panel C of Figure 1 and Model 2 of Table 2.[20]

Result 2 (spill-over effect): *Low-ability students' improvements of academic performance in "Principles of Microeconomics" positively affected their learning in the other courses they*

[20] Another way to classify students as high- or low-ability is to use the median Mathematics score of the sample. The median Mathematics score in the university entrance examination was 615, only slightly different from the mean (620.5). Appendix Table B.2 reports the 2SLS results when those whose Mathematics score was above (below or equal to) the median is classified as high-ability (low-ability). It shows qualitatively similar results to those of Tables 1 and 2 in the paper.

simultaneously took. The relative magnitude of the spill-over effect to the direct effect was large at 0.723.

Table 2: *Spill-Over Effects for Low-Ability Students*

Independent variable: / Performance indicator in the Principles of Micro:	Final exam score [final achievement] (1)	(2)	Total score [final evaluation] (3)	(4)
Final exam score (out of 100)	0.661***	0.638***	----	----
	(0.173)	(0.164)		
Total score (out of 100)	----	----	0.719***	0.723***
			(0.124)	(0.131)
Finance-related major dummy	----	-1.278	----	-2.557
		(1.723)		(1.665)
Female dummy {=1 for female}	----	6.442***	----	6.241***
		(2.026)		(2.034)
Female partner dummy {= 1 if paired partner is female}	----	0.981	----	1.220
		(2.660)		(2.506)
Interaction: Female dummy × Female partner dummy	----	0.145	----	-0.762
		(2.665)		(2.548)
Constant	41.439***	38.403***	32.685***	29.053***
	(8.210)	(7.458)	(7.213)	(7.138)
Observations	230	230	230	230
R-squared	0.235	0.350	0.471	0.557
Underidentification test:				
a. Kleibergen-Paap rk LM statistic	11.190	10.645	10.457	9.756
b. *p*-value	0.0037***	0.0049**	0.0054***	0.0076***
Weak identification test: Kleibergen-Paap rk Wald F statistic	9.942	10.106	12.980	10.707
Overidentification test of all instruments				
a. Hansen J statistic	0.781	0.007	0.770	0.055
b. *p*-value	0.3768	0.9337	0.3802	0.8142

Notes: 2SLS with robust standard errors clustered at the seminar group level. Numbers in parentheses are standard errors. The dependent variable is the GPA of the other courses a student took in the same semester. The high-low pair dummy and Mathematics score (entrance exam) were used as instruments for the final exam score and total score variables. Columns (3), (4), (5), and (6) of Panel A in Table 1 shows the first-stage regression results for Columns (1), (2), (3), and (4) in Table 2, respectively. Instructor dummies are also included in the even-numbered models (the estimates are omitted to conserve space). All observations of the low-ability students were used in the estimation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

To estimate the direct and spill-over effects of peer interactions, we classified students based on their Mathematics scores in the university entrance examination. As the Mathematics scores varied widely from student to student (Figure B.1), one may ask whether the direct and spill-over effects are affected by the absolute size of the intra-pair ability difference. For example, if the abilities of students in a team differ greatly, low-ability students may not improve their performance because they are unable to communicate effectively with each other owing to, for instance, fundamentally different study attitudes or the lacks of basic skills, or because they just give up after being discouraged by their poor own performance. Alternatively, the poorer knowledge and skills a low-ability student possesses compared with the high-ability student, the more the former may benefit, as they have more room for improvement. As a supplementary analysis, the same 2SLS analysis was conducted by using the high-low pair dummy, their Mathematics scores, the intra-pair difference in the Mathematics score, and its squared term as instruments in the first-stage regression.

Appendix Table B.3 reports the first- and second-stage regression results. It reveals that, in the first-stage regression, while the intra-pair difference in the Mathematics score consistently has a negative coefficient estimate for low-ability students' performances in "Principles of Microeconomics," its absolute size is relatively small; for example, it is 0.031 for their final exam score. The average difference in the Mathematics score between the high- and low-ability students was 107, which translates into a negative effect of $-0.031\times107 = -3.317$ points (out of 100 points) in the final examination. By contrast, the high-low pair dummy obtains a much larger positive coefficient estimate than the intra-pair difference variable. For example, the coefficient estimate is 8.991 points when the dependent variable is their final exam score. In addition, the squared term of the intra-pair difference in the Mathematics score does not have significant coefficient estimates for any outcome measure. Hence, it can be concluded that a large intra-pair ability difference does not inhibit less able from learning effectively or changing their study behavior.

5. Conclusions

Peer interactions characterize many aspects of economic and social life. Recent research in education has proposed that positive peer effects may be expected in a team with a greater spread in ability because low-ability individuals benefit from being paired with high-ability individuals. For example, Li *et al.* (2014), Lu and Anderson (2015), and Wu *et al.* (2023) demonstrated that desk mate assignments affect the academic performance of primary and middle schools in China. Kamei and Ashworth (2023) also showed that in the United Kingdom's higher education, low-ability students improved their academic performance if they were given peer learning activities with high-ability students, while high-ability students did not show perverse reactions to being forcedly paired with low-ability students. Further, prior research in personnel economics and human resource management suggests that teams with a greater spread in ability are more productive in the workplace (e.g., Hamilton *et al.*, 2003). The present study took a significant step by investigating for the first time how peer interactions in one setup affect performance *beyond the setup*.

This study combined a new natural field experiment in Mongolia with confidential university grade data to answer this question. The results revealed that low-ability students improve their academic performance in a course when they engage in peer learning activities with high-ability students, rather than with low-ability students. The effect was economically large; it corresponds to an increase of 0.325 S.D. of their final exam score. Strikingly, such engagement experience enabled low-ability students to improve their GPA in other courses they were taking concurrently. The spill-over effect was strong. The relative magnitude of the spill-over effect to the direct effect was 0.723. The results also indicated that paired high-ability students did not suffer from such pairing, which means that introducing peer learning interactions with a greater spread in ability leads to a Pareto improvement.

In economics, instructors rarely formally use group work in their courses; however, this should be encouraged in similar settings, as it is highly cost-effective. Notably, in this study, collaborative pair work was incorporated into the "Principles of Microeconomics," which was taught based on a standard introductory textbook by Parkin (which is used globally in many courses). This textbook has close similarity with other worldwide textbooks, such as Gregory Mankiw's widely used "Principles" textbooks in that it has similar material organization, detailed official slides, and an abundant number of questions in the test bank. Therefore, our collaborative work is easily applicable to many universities in various countries and managing collaborative teamwork in a course requires only small additional burdens on instructors.

Although the results are clear, this study is only the first step toward examining the spill-over phenomenon of peer effects. For example, while the field experiment in this study was conducted in Mongolia, new experiments will be required to judge the validity of the findings in other societies, because not only people's preferences but also other factors, such as norms, culture, institutions, or environments, differ by country and society. In addition, it is unclear how the findings detected for university students can be extended to people of different ages, such as primary school students or mid-careers in recurrent education. Of course, it would be interesting to explore how the spill-over phenomenon prevails in other contexts, such as in the workplace, sports, and daily life. Further research, whether experimental or empirical, is definitely needed to establish the generalizability of the findings in different subject pools and in other contexts.

Reference:

Arcidiacono, Peter, Josh Kinsler, and Joseph Price, 2017. Productivity Spillovers in Team Production: Evidence from Professional Basketball. *Journal of Labor Economics* 35(1): 191-225.

Ariely, Dan, Anat Bracha, and Stephan Meier, 2009. Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially. *American Economic Review* 99: 544-555.

Arnold, Ivo and Jerry Straten, 2012. Motivation and Math Skills as Determinants of First-Year Performance in Economics. *Journal of Economic Education* 43(1): 33-47.

Azoulay, Pierre, Joshua Graff Zivin, and Jialan Wang, 2010. Superstar Extinction. *Quarterly Journal of Economics* 125(2): 549-589.

Ballard, Charles, and Marianne Johnson, 2004. Basic Math Skills and Performance in an Introductory Economics Class. *Journal of Economic Education* 35(1): 3-23.

Bénabou, Roland, and Jean Tirole. 2006. Incentives and Prosocial Behavior. *American Economic Review* 96(5): 1652-1678.

Bénabou, Roland, and Jean Tirole. 2011. Identity, Morals, and Taboos: Beliefs as Assets. *Quarterly Journal of Economics* 126(2): 805-855.

Black, Sandra, Paul Devereux, and Kjell Salvanes, 2013. Under Pressure? The Effect of Peers on Outcomes of Young Adults. *Journal of Labor Economics* 31(1): 119-153.

Booij, Adam, Edwin Leuven, and Hessel Oosterbeek, 2017. Ability Peer Effects in University: Evidence from a Randomized Experiment. *Review of Economic Studies* 84(2): 547-578.

Bowles, Samuel, and Herbert Gintis, 2005. "Prosocial emotions," in L. Blume, S. Durlauf (Eds.), *The Economy as a Complex Evolving System III: Essays in Honor of Kenneth Arrow*, Oxford University Press, Oxford: 337-367.

Brune, Lasse, Eric Chyn, and Jason Kerwin, 2022. Peers and Motivation at Work Evidence from a Firm Experiment in Malawi. *Journal of Human Resources* 57(4): 1147-1177.

Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler, 1997. Labor Supply of New York City Cabdrivers: One Day at a Time. *Quarterly Journal of Economics* 112(2): 407-441.

Carrell, Scott, Bruce Sacerdote, and James West, 2013. From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica* 81(3): 855-882.

Falk, Armin, and Andrea Ichino, 2006. Clean Evidence on Peer Effects. *Journal of Labor Economics* 24(1): 39-57.

Fehr, Ernst, and Klaus Schmidt, 1999. A Theory of Fairness, Competition, and Cooperation. *Quarterly Journal of Economics* 114(3): 817-868.

Feld, Jan, and Ulf Zölitz, 2017. Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. *Journal of Labor Economics* 35(2): 387-428.

Friedman, Daniel, and Shyam Sunder, 1994. *Experimental Methods: A Primer for Economists*. Cambridge University Press.

Hamilton, Barton, Jack Nickerson, and Hideo Owan, 2003. Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation. *Journal of Political Economy* 111(3): 465-497.

Jackson, Kirabo, and Elias Bruegmann, 2009. Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers. *American Economic Journal: Applied Economics* 1(4): 85-108.

Kamei, Kenju, and John Ashworth, 2023. Peer Learning in Teams and Work Performance: Evidence from a Randomized Field Experiment. *Journal of Economic Behavior & Organization* 207: 413-32.

Kamei, Kenju, and Thomas Markussen, 2023. Free Riding and Workplace Democracy—Heterogeneous Task Preferences and Sorting. *Management Science* 69(7):3884-3904.

Kimbrough, Erik, Andrew McGee, and Hitoshi Shigeoka, 2022. How Do Peers Impact Learning? An Experimental Investigation of Peer-to-Peer Teaching and Ability Tracking. *Journal of Human Resources* 57: 304-339.

Kleibergen, Frank, and Richard Paap, 2006. Generalized Reduced Rank Tests using the Singular Value Decomposition. *Journal of econometrics* 133(1): 97-126.

Langer, Ellen, 1975. The Illusion of Control. *Journal of Personality and Social Psychology* 32(2): 311-328.

Larkin, Ian, Lamar Pierce, and Francesca Gino, 2012. The Psychological Costs of Pay-for-Performance: Implications for the Strategic Compensation of Employees. *Strategic Management Journal* 33(10): 1194-1214.

Lavy, Victor, and Analia Schlosser, 2011. Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics* 3(2): 1-33.

Li, Tao, Li Han, Linxiu Zhang, and Scott Rozelle, 2014. Encouraging Classroom Peer Interactions: Evidence from Chinese Migrant Schools. *Journal of Public Economics* 111: 29-45.

Lu, Fangwen, and Michael Anderson, 2015, Peer Effects in Microenvironments: The Benefits of Homogeneous Classroom Groups. *Journal of Labor Economics* 33(1): 91-122.

Lyle, David, 2009. The Effects of Peer Group Heterogeneity on the Production of Human Capital at West. *American Economic Journal: Applied Economics* 1(4): 69-84.

Mas, Alexandre, and Enrico Moretti, 2009. Peers at Work. *American Economic Review* 99(1): 112-145.

Sacerdote, Bruce, 2001. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *Quarterly Journal of Economics* 116(2): 681-704.

Sacerdote, Bruce, 2014. Experimental and Quasi-Experimental Analysis of Peer Effects: Two Steps Forward? *Annual Review of Economics* 6: 253-72.

Svenson, Ola, 1981. Are We All Less Risky and More Skillful than our Fellow Drivers? *Acta Psychologica* 47: 143-48.

Thaler, Richard, 1999. Mental Accounting Matters. *Journal of Behavioral Decision Making* 12(3): 183-206.

Wu, Jia, Junsen Zhang, and Chunchao Wang, 2023. Student Performance, Peer Effects, and Friend Networks: Evidence from a Randomized Peer Intervention. *American Economic Journal: Economic Policy* 15 (1): 510-42.

NOT FOR PUBLICATION

Appendix A: Additional Information for the “Principles of Microeconomics”

A.1. Syllabus (translated in English)

Week (block)	Content/Chapter in the textbook
1	Introduction to Economics: What Is Economics? The Economic Problem
2	Demand and Supply
3	Elasticity
4	Efficiency and Equity
5	Markets in Action
6	Utility and Demand; Possibilities, Preferences, and Choice
7	Organizing Production; Midterm Examination
8	Output and Costs
9	Perfect Competition
10	Monopoly
11	Oligopoly
12	Regulation and Antitrust Law
13	Externalities
14	Public Goods and Common Resources
15	Markets for Factors of Production
16	Final Examination

Assessment Component	Timing	Content
Short Assignment 1	Week 6	Demand and Supply, Elasticity, Efficiency and Equity, and Market in Action
Short Assignment 2	Week 14	Perfect Competition, Monopoly, Oligopoly, Regulation and Antitrust Law, and Externalities
Midterm Examination	Week 7	Introduction to Economics, Demand and Supply, Elasticity, Efficiency and Equity, Market in Action, and Utility and Demand
Final Examination	Week 16	All content covered in the class

A.2. An example for seminar questions

[An example of multiple-choice question:]

In a market, at the equilibrium price,

a) neither buyers nor sellers can do business at a better price.

b) buyers are willing to pay a higher price, but sellers do not ask for a higher price.

c) buyers pay the minimum price they are willing to pay for any amount of output and sellers charge the maximum price they are willing to charge for any amount of production.

d) none of the above is true.

[An example of short answer problems:]

The following table presents the demand and supply schedules for comic books.

Price (per comic book)	Quantity demanded (per month)	Quantity supplied (per month)
$2.5	14,000,000	8,000,000
$3.0	13,000,000	10,000,000
$3.5	12,000,000	12,000,000
$4.0	11.000,000	13,000,000
$4.5	10,000,000	14,000,000

a) Graph these demand and supply schedules below. What is the equilibrium price? What is the equilibrium quantity?

b) What is the marginal benefit received by the consumer of the 12,000,000th comic book?

c) What is the minimum price required for a producer to produce the 12,000,000th comic book?

d) Suppose that the price of a movie, a substitute for comic books, rises so that at every comic book price, consumers now want to buy 2,000,000 more comic books than before. For example, at a price of $2.50, consumers now buy 16,000,000 comics. Plot this new demand curve in the figure you draw in question a. What is the new equilibrium price? What is the new equilibrium quantity?

A.3. Examples of the midterm/final examination questions

Both the midterm and final exams were created using the test bank of the textbook. The following is an example of a midterm exam:

* * * * * * * * * * * * * * * * *

Exam

Name_______________________________

MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question.

1) During the summer, you have made the decision to attend summer school, which prevents you from working at your usual summer job in which you normally earn $6,000 for the summer. Your tuition cost is $3,000 and books and supplies cost $1,300. In terms of dollars, the opportunity cost of attending summer school is

A) $4,300. B) $10,300. C) $6,000. D) $3,300.

2) When the government chooses to use resources to build a dam, these sources are no longer available to build a highway. This choice illustrates the concept of

A) a market mechanism. B) a fallacy of composition.
C) opportunity cost. D) macroeconomics.

3) Scarcity requires that people must

A) compete. B) make choices. C) trade. D) cooperate.

4) Which of the following is <u>NOT</u> a factor of production?

A) a university professor B) 100 shares of Microsoft stock
C) mineral resources D) an apartment building

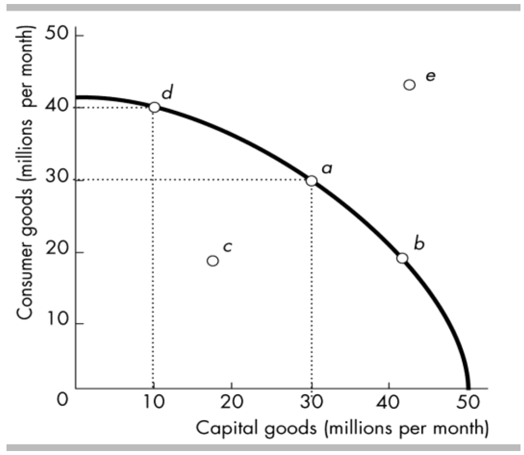
5) When operating on its *PPF*, a country can produce 2 tons of butter and 200 cars OR 3 tons of butter and 150 cars. The opportunity cost of 1 ton of butter is ______cars per ton of butter.

A) 0.75 B) 200 C) 300 D) 50

Point	Production of grain (tons)	Production of cars (cars)
A	0	30
B	2	28
C	4	24
D	6	18
E	8	10
F	10	0

6) The table above lists six points on the production possibilities frontier for grain and cars. Given this information, which of the following combinations is unattainable?

A) 7 tons of grain and 10 cars B) 4 tons of grain and 26 cars
C) 2 tons of grain and 27 cars D) 6 tons of grain and 18 cars



7) Refer to the production possibilities frontier in the figure above. Which production point is unattainable?

A) Point *a* B) Point *b* C) Point *c* D) Point *e*

8) Any point on a production possibilities frontier (*PPF*) itself is

A) inefficient. B) production efficient. C) unattainable. D) equitable.

9) If the marginal benefit of a good exceeds its marginal cost

A) we should produce less to achieve the allocatively efficient use of resources.
B) we've achieved efficient resource use.
C) we cannot tell if more or less should be produced to achieve the allocatively efficient use of resources.
D) we should produce more to achieve the allocatively efficient use of resources.

10) A factor market is a market in which

A) firms sell goods and services.
B) firms sell the services of the factors of production.
C) households buy goods and services.
D) households sell the services of the factors of production they control.

Price (dollars per pound)	Quantity supplied (pounds)	Quantity demanded (pounds)
3	1	7
4	2	5
5	4	4
6	5	2
7	6	1

11) The above table shows the demand schedule and supply schedule for chocolate chip cookies. What is the equilibrium quantity and equilibrium price for chocolate chip cookies?

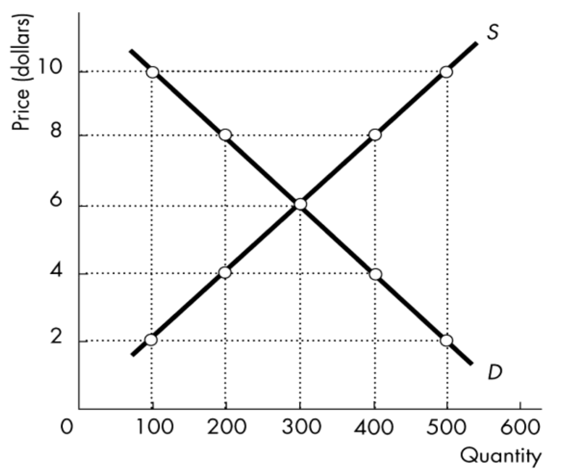
A) 7 pounds, $3.00 per pound B) 2 pounds, $6.00 per pound
C) 4 pounds, $5.00 per pound D) 2 pounds, $3.00 per pound

12) Suppose that people find out that eating more fish improves their health, leading them to increase their demand for fish. As a result, the equilibrium price of fish __and the equilibrium quantity ____.

A) rises; decreases B) falls; decreases C) falls; increases D) rises; increases

13) If the price of a candy bar is $1 and the price of a fast food meal is $5, then the

A) relative price of a candy bar is 5 fast food meals per candy bar.
B) relative price of a fast food meal is 5 candy bars per fast food meal.
C) money price of a candy bar is 1/5 of a fast food meal per candy bar.
D) money price of a fast food meal is 1/5 of a candy bar per fast food meal.



14) At a price of $10 in the above figure, there is

A) a surplus of 400 units.
B) a surplus of 200 units.
C) a shortage of 400 units.
D) a shortage of 200 units.

15) When the price of a pizza decreases from $12 to $10, it is definitely the case that the

A) substitution effect means people buy more pizza.
B) income effect means people buy less pizza.
C) quantity demanded of pizza will not change.
D) None of the above answers is correct.

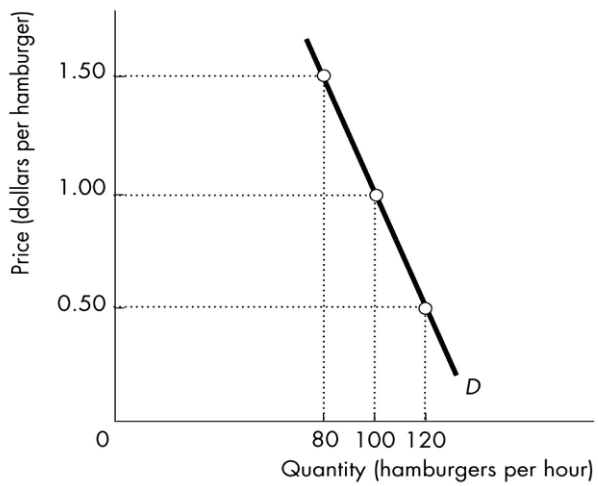
16) The quantity demanded is

A) independent of the price of the good.
B) independent of consumers' buying plans.
C) the amount of a good that consumers plan to purchase at a particular price.
D) always equal to the equilibrium quantity.

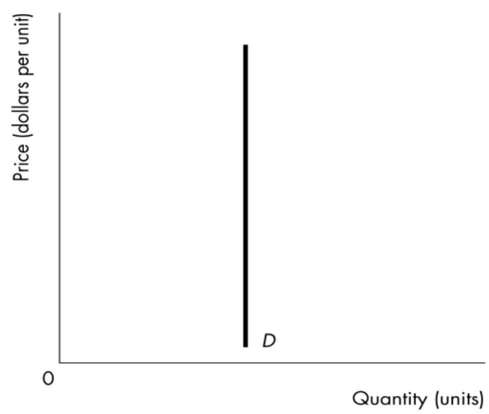
	Price (dollars per bushel)	Quantity demanded (bushels)
A	10	0
B	8	4
C	6	8
D	4	12
E	2	16

17) The table above gives the demand schedule for peas. Between point A and point B, the price elasticity of demand equals

A) 0.22.
B) 0.11.
C) 0.50.
D) 9.09.



18) The figure illustrates the demand for hamburgers. When the price is $1.00 a hamburger, the elasticity of demand is _______________and a 1 percent increase in the price will __the quantity of hamburgers demanded by ____percent.
A) 1.00; decrease; 0.40 B) 0.40; decrease; 0.40 C) 5.00; decrease; 5.00 D) 2.50; increase; 2.50



19) The demand curve in the figure above illustrates the demand for a product with
A) a price elasticity of demand that is different at all prices.
B) zero price elasticity of demand at all prices.
C) unit price elasticity of demand at all prices.
D) infinite price elasticity of demand.

20) If there is an increase in the price of broccoli because of disastrous weather that destroys about half of this year's spinach crop, which of the following could be true?
A) The demand curve for spinach has shifted to the leftward and the cross elasticity of demand between spinach and broccoli is positive.
B) The demand curves for spinach and broccoli have become more cross inelastic.
C) The cross elasticity of demand between spinach and broccoli is negative 1.25.
D) The demand curve for broccoli has shifted rightward and the cross elasticity of demand between spinach and broccoli is positive.

21) If a small percentage decrease in the price of chocolate causes a larger percentage decrease in the

quantity supplied, the

A) demand for chocolate is elastic.
B) supply of chocolate is elastic.
C) demand for chocolate is inelastic.
D) supply of chocolate is inelastic.

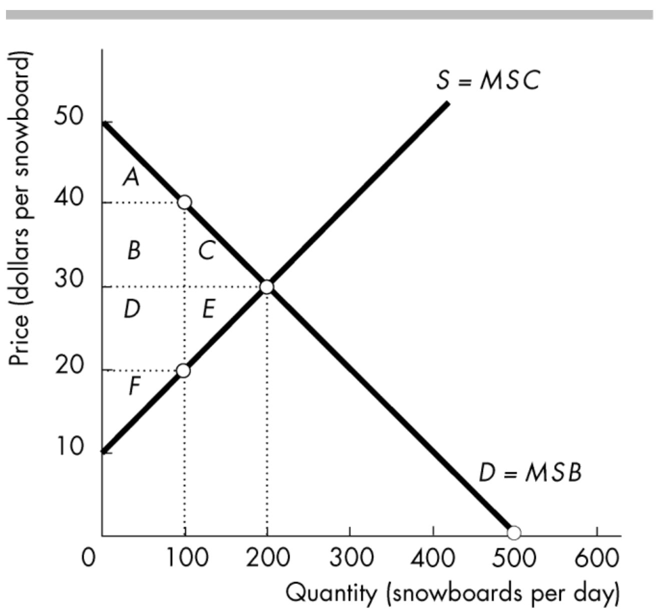
22) Suppose the price of burgers increases from $2 to $3 each. The degree to which quantity demanded responds to this price increase depends on the

A) cross elasticity of demand.
B) price elasticity of demand.
C) income elasticity of demand.
D) the price elasticity of supply.

Quantity (DVDs per week)	Marginal social benefit (dollars per DVD)	Marginal social cost (dollars per DVD)
1	24	16
2	22	18
3	20	20
4	18	22
5	16	24

23) The schedules in the above table give the marginal social benefit and marginal social cost of a DVD. If the number of DVD produced is cut to 2 a week, then the _____.

A) value of the second DVD is $20
B) opportunity cost of the second DVD is $22
C) minimum supply-price of the second DVD is $18
D) price is $18 a DVD



24) In the above figure, what is the amount of producer surplus at the efficient quantity?

A) $0
B) $4,000
C) $1,000
D) $2,000

25) When 2,000 hamburgers per day are produced, the marginal social benefit is $1.50 and the marginal social cost is $1.00. When 7,500 hamburgers per day are produced, the marginal social benefit is $1.00 and the marginal social cost is $1.50. The efficient production quantity of hamburgers is ___a day.

A) between 2,000 and 7,500 B) 7,500 C) more than 7,500 D) 2,000

Price (dollars per can)	Quantity demanded	Quantity supplied
0.30	1000	100
0.40	900	300
0.50	750	400
0.60	600	600
0.70	400	800
0.80	200	1000
0.90	100	1200
1.00	50	1400

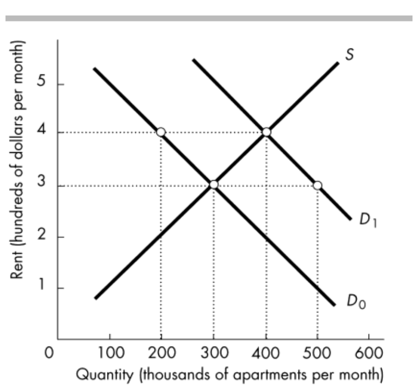
26) The above table gives the market demand and market supply schedules for soda. What is the maximum price consumers are willing to pay for the 400th can of soda?
A) $0.70 per can B) $0.50 per can C) $0.60 per can D) $0.80 per can

27) The resource allocation method that is used to allocate scarce resources between private use and government use is
A) personal characteristics. B) lottery.
C) first-come, first-served. D) majority rule.

28) A country produces only pencils and erasers. Pencil production is allocatively efficient if the marginal ____ of a pencil equals the marginal __of _____.
A) cost; benefit; an eraser B) benefit; cost; a pencil
C) cost; cost; an eraser D) benefit; benefit; an eraser

29) A rent ceiling set above the equilibrium rent
A) has no effect.
B) restricts both the quantity demanded and the quantity supplied.
C) restricts the quantity demanded but not the quantity supplied.
D) restricts the quantity supplied but not the quantity demanded.

30) Suppose the equilibrium wage is $10 per hour. A minimum wage is a __and affects employment if it is set at _____.
A) price ceiling; $10 per hour B) price ceiling; $12 per hour
C) price floor; $8 per hour D) price floor; $12 per hour



31) In the figure above, the initial demand curve is D_0. There are no rent ceilings nor rent floors. The

equilibrium monthly rent is

A) $200 per month. B) $400 per month. C) $300 per month. D) $100 per month.

32) If salt has a ________, then __pay most of any tax levied on salt.

A) high elasticity of demand; buyers
B) low elasticity of supply; buyers
C) low elasticity of demand; buyers
D) high elasticity of supply; sellers

33) The amount of a tax paid by the sellers will be larger the more ________the demand and the more __the supply.

A) inelastic; elastic B) elastic; inelastic C) inelastic; inelastic D) elastic; elastic

34) Which of the following statements is true about taxes?

A) Government revenue from a tax is always greater than the loss of producer surplus and consumer surplus.
B) Taxes always create more deadweight loss than do price ceilings and price floors.
C) Taxes decrease both consumer surplus and producer surplus while creating a deadweight loss.
D) Both answers A and C are correct.

35) Jeannie's marginal utility from her 4th book in a month is 50. Her marginal utility from her 5th book

A) might be more than, less than, or equal to 50 but more information is needed.
B) is less than 50.
C) is greater than 50.
D) equals 50.

36) Tom spends $20 a month on CDs and magazines. The price of a CD is $6 and the price of a magazine is $4. When Tom maximizes his utility, the marginal utility from CDs is _______the marginal utility from newspapers.

A) 1/2 B) the same as C) 20 times D) 1 1/2 times

37) In consumer equilibrium, Harold consumes pizza, sodas, and other goods. Pizza and soda are complements for Harold. The price of a pizza rises while his income remains the same. Harold then consumes

A) more pizza and more soda.
B) more pizza and less soda.
C) less pizza and less soda.
D) less pizza and more soda.

38) According to marginal utility theory, a rise in income will

A) increase consumption of all goods.
B) increase the marginal utility of all goods.
C) increase a consumer's total utility.
D) None of the above answers is correct.

39) Utility is best defined as

A) the practical usefulness of a good.
B) the amount one is willing to pay for a good.
C) the satisfaction from consuming a good.
D) the price of a good.

40) Tom's consumption possibilities are defined by

A) his preferences for consumption of the goods that he consumes.
B) his income only.
C) his income and the prices of the goods that he consumes.
D) the prices of the goods that he consumes only.

Appendix B: Additional Figures and Tables

Figure B.1. *Histogram of the Students' Math Scores in the University Entrance Examination*

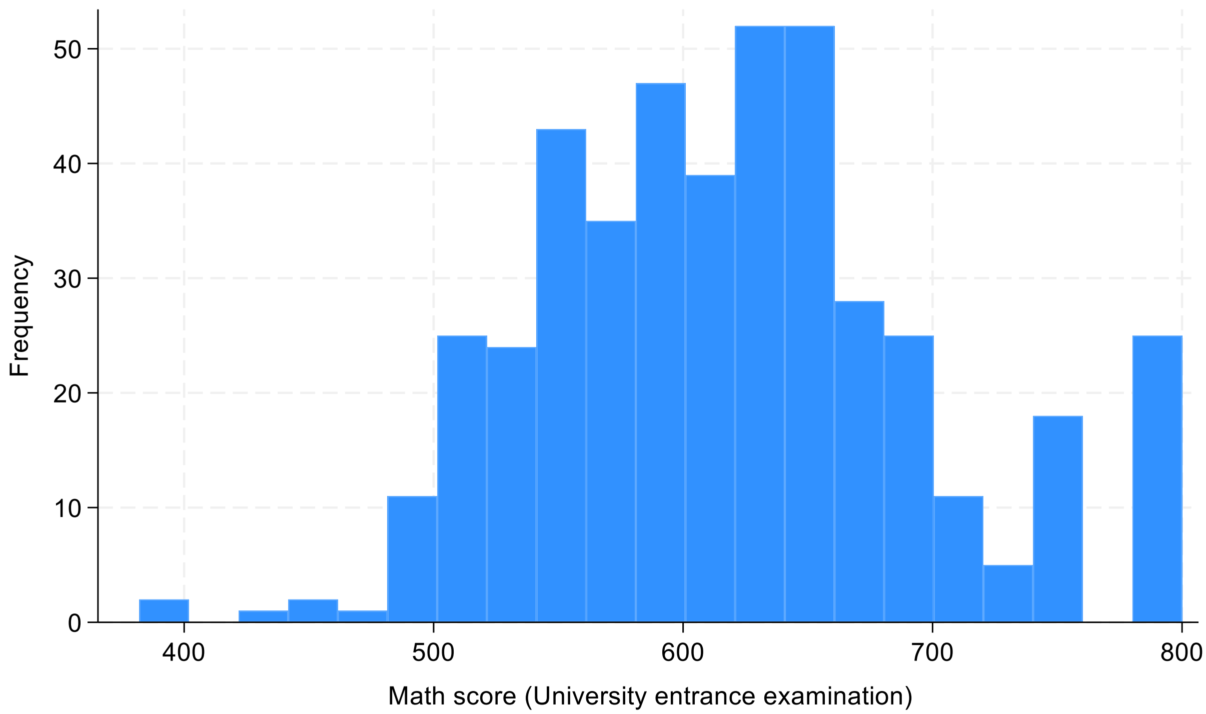


Table B.1. *Randomization Check for Pairing*

A. Low-Ability Students

	Matching outcome		Two-sided *p*-value for H_0: (*i*) = (*ii*)[#1]
	(i) Paired with a low-ability student	(ii) Paired with a high-ability student	
% of female students	0.651 (0.478)	0.595 (0.494)	0.4783
% of those who were paired with a female	0.651 (0.478)	0.691 (0.465)	0.5647
% of finance-related majors	0.473 (0.501)	0.524 (0.502)	0.4947
Average Math score (entrance exam)	559.6 (43.20)	565.9 (37.30)	0.1913

B. High-Ability Students

	Matching outcome		Two-sided *p*-value for H_0: (*i*) = (*ii*)[#1]
	(i) Paired with a low-ability student	(ii) Paired with a high-ability student	
% of female students	0.691 (0.465)	0.712 (0.454)	0.7613
% of those who were paired with a female	0.595 (0.494)	0.712 (0.454)	0.1030
% of finance-related majors	0.560 (0.499)	0.750 (0.435)	0.0046***
Average Math score (entrance exam)	673.1 (50.07)	690.5 (57.45)	0.0241**

Notes: Numbers in parentheses are standard deviations. [#1] *t* tests for comparisons of the Math scores; fisher exact tests to compare proportions.

Table B.2: *Effects of Being Paired with High-Ability Students*

– Using Median to Classify Students as High- or Low-Ability (A Robustness Check) –

(supplementing Tables 1 and 2 of the paper)

[Direct Effect:]

A. Low-ability students

Dependent variable: / Independent variable:	Midterm exam score (1)	(2)	Final exam score (3)	(4)	Total score (5)	(6)
High-low pair dummy {=1 if the partner is a high-ability student}	3.139	2.658	6.249***	6.297***	5.004**	4.775**
	(2.150)	(2.053)	(1.693)	(1.735)	(1.762)	(1.917)
Math score (entrance exam)	0.105***	0.103***	0.075***	0.081***	0.079***	0.078***
	(0.027)	(0.030)	(0.025)	(0.027)	(0.016)	(0.017)
Finance-related major dummy	----	3.601	----	-0.068	----	1.937
		(2.655)		(1.648)		(1.540)
Female dummy {=1 for female}	----	-4.761	----	0.387	----	0.648
		(4.639)		(4.398)		(3.280)
Female partner dummy {= 1 if paired partner is female}	----	-9.131**	----	-5.537	----	-4.654*
		(4.087)		(3.438)		(2.539)
Interaction: Female dummy × Female partner dummy	----	6.496	----	0.278	----	0.876
		(5.958)		(5.226)		(4.011)
Constant	-12.993	-8.119	0.041	-1.982	7.396	7.252
	(15.640)	(17.919)	(14.305)	(15.346)	(9.260)	(10.394)
Observations	228	228	228	228	228	228
R-squared	0.093	0.148	0.073	0.113	0.109	0.155

B. High-ability students

Dependent variable: / Independent variable:	Midterm exam score (1)	Midterm exam score (2)	Final exam score (3)	Final exam score (4)	Total score (5)	Total score (6)
High-high pair dummy {=1 if the partner is a high-ability student}	6.514***	6.262**	2.587	0.523	5.038***	3.451
	(1.946)	(2.463)	(2.102)	(3.100)	(1.462)	(2.098)
Math score (entrance exam)	0.053***	0.052**	0.043**	0.015	0.039**	0.022*
	(0.017)	(0.020)	(0.016)	(0.014)	(0.014)	(0.011)
Finance-related major dummy	----	2.826	----	11.218***	----	6.926***
		(3.203)		(2.826)		(2.306)
Female dummy {=1 for female}	----	5.616*	----	2.864	----	4.173*
		(2.943)		(3.083)		(2.216)
Female partner dummy {= 1 if paired partner is female}	----	-2.782	----	-7.256***	----	-5.257*
		(3.737)		(2.088)		(2.730)
Interaction: Female dummy	----	0.372	----	4.808	----	3.731
× Female partner dummy		(4.014)		(4.268)		(3.382)
Constant	18.280	14.944	24.873**	37.972***	32.964***	38.173***
	(11.102)	(12.553)	(11.164)	(8.805)	(9.597)	(6.931)
Observations	218	218	218	218	218	218
R-squared	0.099	0.148	0.029	0.180	0.070	0.216

Notes: A linear regression with robust standard errors clustered at the seminar group level. Numbers in parentheses are standard errors. Instructor dummies are also included in the even-numbered models (the estimates are omitted to conserve space). All observations of low-ability (high-ability) students were used in the estimations of Panel A (Panel B).

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

[Spill-Over Effect for Low-Ability Students:]

Independent variable: / Performance indicator in the Principles of Micro:	Final exam score [final achievement]		Total score [final evaluation]	
	(1)	(2)	(3)	(4)
Final exam score (out of 100)	0.647***	0.635***	----	----
	(0.162)	(0.156)		
Total score (out of 100)	----	----	0.718***	0.727***
			(0.118)	(0.126)
Finance-related major dummy	----	-1.191	----	-2.591
		(1.757)		(1.682)
Female dummy {=1 for female}	----	6.596***	----	6.329***
		(2.044)		(2.077)
Female partner dummy {= 1 if paired partner is female}	----	1.313	----	1.258
		(2.636)		(2.493)
Interaction: Female dummy × Female partner dummy	----	-0.342	----	-0.882
		(2.627)		(2.582)
Constant	42.208***	38.415***	32.828***	28.820***
	(7.709)	(7.160)	(6.954)	(6.835)
Observations	228	228	228	228
R-squared	0.244	0.348	0.464	0.550
Underidentification test:				
a. Kleibergen-Paap rk LM statistic	12.676	11.580	11.112	9.971
b. *p*-value	0.0018***	0.0031***	0.0039***	0.0068***
Weak identification test: Kleibergen-Paap rk Wald F statistic	11.232	11.051	13.669	10.727
Overidentification test of all instruments				
a. Hansen J statistic	2.023	0.203	1.598	0.017
b. *p*-value	0.1550	0.6520	0.2061	0.8964

Notes: 2SLS with robust standard errors clustered at the seminar group level. Numbers in parentheses are standard errors. The dependent variable is the GPA of the other courses a student took in the same semester. The high-low pair dummy and the Mathematics score (entrance exam) were used as instruments for the final exam score and total score variables. Columns (3), (4), (5), and (6) of Panel A in Table B.2 show the first-stage regression outcomes for Columns (1), (2), (3), and (4) in the above table, respectively. Instructor dummies are also included in the even-numbered models (the estimates are omitted to conserve space). All observations of low-ability students were used for the estimation.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.3: *Effects of Being Paired with High-Ability Students*

– Does the Size of the Intra-pair Ability Difference Matter? (A Robustness Check) –

(supplementing Tables 1 and 2 of the paper)

[Direct Effect:]

A. Low-ability students

Dependent variable: / Independent variable:	Midterm exam score (1)	Midterm exam score (2)	Final exam score (3)	Final exam score (4)	Total score (5)	Total score (6)
High-low pair dummy {=1 if the partner is a high-ability student}	2.768	3.070	9.104***	8.991**	7.193***	6.943***
	(3.071)	(2.796)	(2.838)	(3.164)	(2.002)	(1.851)
Math score (entrance exam)	0.100**	0.094*	0.042	0.048	0.051**	0.051**
	(0.047)	(0.047)	(0.026)	(0.034)	(0.022)	(0.023)
Intra-pair difference in the Math Score	-0.003	-0.003	-0.039**	-0.031	-0.030**	-0.025*
	(0.020)	(0.020)	(0.016)	(0.021)	(0.011)	(0.013)
Square of the intra-pair difference in the Math score	0.000	0.000	0.000	0.000	0.000*	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Finance-related major dummy	----	3.248	----	0.296	----	1.960
		(2.672)		(1.690)		(1.581)
Female dummy {=1 for female}	----	-4.824	----	-0.306	----	0.193
		(4.598)		(5.004)		(3.613)
Female partner dummy {= 1 if paired partner is female}	----	-10.161**	----	-5.724	----	-5.111**
		(3.931)		(3.562)		(2.291)
Interaction: Female dummy × Female partner dummy	----	7.570	----	0.959	----	1.812
		(5.709)		(5.679)		(3.955)
Constant	-10.592	-2.633	18.664	16.768	22.389*	22.401
	(26.891)	(27.666)	(14.307)	(19.615)	(12.428)	(13.414)
Observations	230	230	230	230	230	230
R-squared	0.080	0.135	0.076	0.114	0.109	0.152

B. High-ability students

Dependent variable: / Independent variable:	Midterm exam score (1)	(2)	Final exam score (3)	(4)	Total score (5)	(6)
High-high pair dummy {=1 if the partner is a high-ability student}	-4.461	-3.950	3.567	3.021	-0.730	-0.802
	(2.771)	(3.289)	(4.038)	(4.183)	(2.509)	(2.795)
Math score (entrance exam)	0.061***	0.060**	0.083***	0.042	0.064***	0.037**
	(0.017)	(0.024)	(0.028)	(0.024)	(0.018)	(0.013)
Intra-pair difference in the Math Score	0.006	0.008	0.041*	0.023	0.026	0.016
	(0.014)	(0.012)	(0.020)	(0.015)	(0.015)	(0.010)
Square of the intra-pair difference in the Math score	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Finance-related major dummy	----	3.068	----	10.872***	----	6.881***
		(3.110)		(2.648)		(2.183)
Female dummy {=1 for female}	----	5.302*	----	2.175	----	3.851*
		(2.815)		(3.000)		(2.153)
Female partner dummy {= 1 if paired partner is female}	----	-1.747	----	-7.332***	----	-4.676*
		(3.208)		(2.161)		(2.571)
Interaction: Female dummy × Female partner dummy	----	-0.606	----	5.183	----	3.253
		(3.866)		(4.490)		(3.608)
Constant	19.711	15.407	0.157	20.537	21.327	31.841***
	(11.800)	(15.804)	(20.146)	(16.314)	(12.931)	(8.828)
Observations	216	216	216	216	216	216
R-squared	0.092	0.136	0.047	0.186	0.077	0.217

Notes: A linear regression with robust standard errors clustered at the seminar group level. Numbers in parentheses are standard errors. Instructor dummies are also included in the even-numbered models (the estimates are omitted to conserve space). All observations of low-ability (high-ability) students were used in the estimations of Panel A (Panel B).

[Spill-Over Effect for Low-Ability Students:]

Independent variable: / Performance indicator in the Principles of Micro:	Final exam score [final achievement] (1)	Final exam score [final achievement] (2)	Total score [final evaluation] (3)	Total score [final evaluation] (4)
Final exam score (out of 100)	0.769***	0.727***	----	----
	(0.176)	(0.181)		
Total score (out of 100)	----	----	0.814***	0.809***
			(0.121)	(0.137)
Finance-related major dummy	----	-1.321	----	-2.745
		(1.733)		(1.700)
Female dummy {=1 for female}	----	6.378***	----	6.162***
		(2.292)		(2.196)
Female partner dummy {= 1 if paired partner is female}	----	1.487	----	1.679
		(2.741)		(2.592)
Interaction: Female dummy × Female partner dummy	----	0.092	----	-0.915
		(2.882)		(2.664)
Constant	36.588***	34.465***	27.664***	24.591***
	(8.326)	(7.883)	(7.027)	(7.256)
Observations	230	230	230	230
R-squared	0.160	0.303	0.471	0.561
Underidentification test:				
a. Kleibergen-Paap rk LM statistic	12.538	11.552	10.920	10.613
b. *p*-value	0.0138**	0.0210**	0.0275**	0.0313**
Weak identification test: Kleibergen-Paap rk Wald F statistic	7.521	6.454	14.906	10.320
Overidentification test of all instruments				
a. Hansen J statistic	3.847	3.181	5.000	4.489
b. *p*-value	0.2785	0.3645	0.1718	0.2132

Notes: 2SLS with robust standard errors clustered at the seminar group level. Numbers in parentheses are standard errors. The dependent variable is the GPA of the other courses a student took in the same semester. The high-low pair dummy, the Mathematics score (entrance exam), intra-pair difference in the Mathematics score, and its squared term were used as instruments for the final exam score and total score variables. Columns (3), (4), (5), and (6) of Panel A in Table B.2 show the first-stage regression outcomes for Columns (1), (2), (3), and (4) of Table B.3, respectively. Instructor dummies are also included in the even-numbered models (the estimates are omitted to conserve space). All observations of low-ability students were used for the estimation.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table B.4: *Effects of Being Paired with High-Ability Students for Low-Ability Students – Using Total Scores excluding Seminar Scores (A Robustness Check)–*
(supplementing Tables 1 and 2 of the paper)

Columns (5) and (6) of Table 1 in the paper use students' total scores as an evaluation metric for the course. As seminar scores accounted for 10 percent of the total score and were shared between pairmates, low-ability students might benefit from being paired with a high-ability peer simply due to the score sharing, even without improving their own performance. To supplement the earlier results, a robustness check was conducted by excluding seminar scores from the total score. The results (below) show strong direct and spill-over effects for low-ability students, suggesting significant academic improvement.

Stage: / Dependent variable: / Independent variable:	First-stage regression Total score[#1] (1)	 (2)	Second-stage regression GPA of other courses (3)	 (4)
High-low pair dummy {=1 if the partner is a high-ability student}	4.088**	4.129**	----	----
	(1.538)	(1.660)		
Math score (entrance exam)	0.070***	0.071***	----	----
	(0.014)	(0.015)		
Total score excluding seminar score (out of 90) [#1]	----	----	0.792***	0.778***
			(0.139)	(0.153)
Finance-related major dummy	----	1.434	----	-2.391
		(1.548)		(1.845)
Female dummy {=1 for female}	----	-0.245	----	6.889***
		(3.046)		(2.017)
Female partner dummy {= 1 if paired partner is female}	----	-4.926**	----	1.318
		(1.951)		(2.517)
Interaction: Female dummy × Female partner dummy	----	1.451	----	-0.892
		(3.301)		(2.531)
Constant	6.282	6.819	33.657***	30.028***
	(8.361)	(8.945)	(7.183)	(7.483)
Observations	230	230	230	230
R-squared	0.100	0.143	0.419	0.506
Underidentification test:				
a. Kleibergen-Paap rk LM statistic	----	----	10.296	9.758
b. *p*-value	----	----	0.0058***	0.0076***
Weak identification test: Kleibergen-Paap rk Wald F statistic	----	----	13.232	11.300
Overidentification test of all instruments				
a. Hansen J statistic	----	----	0.469	0.281
b. *p*-value	----	----	0.4935	0.5961

Notes: 2SLS with robust standard errors clustered at the seminar group level. Numbers in parentheses are standard errors. Columns (1) and (2) show the first-stage regression outcomes for Columns (3) and (4), respectively. Instructor dummies are also included in the even-numbered models (the estimates are omitted to conserve space). All observations of low-ability students were used for the estimation. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. [#1] Accumulated score comprising the midterm exam (20%), final exam (30%), short assignments (15% × 2), attendance (5%), and attitude (5%).