Peer Effects on Study Effort

4.1. Introduction

Studies of peer effects in education are nothing new. For decades, empirical researches in economics have accumulated rich evidence on the importance of peers regarding students' success. A spectrum of meaningful educational outcomes has been explored: students' academic success (Bettinger, Loeb, & Taylor, 2014; Hanushek, Kain, Markman, & Rivkin, 2003; Hoxby, 2000; Lyle, 2007; Sacerdote, 2001; Zimmerman, 2003); personal well-being (Borsari & Carey, 2001; Eisenberg, Golberstein, & Whitlock, 2014); and other far-reaching decisions (Anelli & Peri, 2015; De Giorgi, Pellizzari, & Redaelli, 2010; Marmaros & Sacerdote, 2002). Interestingly, few papers in economics have addressed the possible influence of peers on college student academic effort, which is clearly linked to their academic success (e.g., Stinebrickner & Stinebrickner, 2008).

The gap in the literature is partially due to the difficulty in obtaining data on students' effort. A few studies attempted to use surveys to solicit students' study time (e.g., Grave, 2011; Stinebrickner & Stinebrickner, 2004, 2008), but such an index is noisy in nature as students' effort is constructed based on a few sampled days. Another related line of research used students' readily available attendance data (e.g., Arulampalam, Naylor, & Smith, 2012; Dobkin, Gil, & Marion, 2010; Kirby & McElroy, 2003; Stanca, 2006); however, such data do not capture study effort outside classroom as well as students' engagement during class.

This study provides novel evidence on the determinants of students' effort by using a unique dataset that captures students' interaction with course management system (CMS) in a "flipped course" at a Chinese regional college. Interaction with the CMS provides reliable information on students' effort in this course. First, in the flipped classroom pedagogy, the CMS captures students' learning of course materials that usually takes place offline during the class[1]. Second, the flipped course is a general education course that introduces rudimentary computer knowledge. The content and homework are fairly straightforward for college students, entailing only minimum effort beyond going through course materials. Thus, even if the CMS does not capture students' outside-class offline effort, such effort should be minimal. Therefore, students' interaction with the CMS represents a larger portion of students' effort than in other settings[2].

In addition to the unique student effort data, the housing arrangement in this college facilitates credible identification of peer effects. Students in this college are randomly assigned to two peer groups—dormitory roommates and social group-mates—conditional on a few known observables. Exogenously formed peer groups preclude the possibility that students' similarities in study effort are due to similarities in unobservable factors (e.g., education aspiration or ability). Nevertheless, we performed several random checks before conducting the analysis to ensure that similar students were not grouped into peer groups.

The study found clear evidence of peer effects on college students' study effort: A pair of randomly assigned peers have more similar study effort patterns than a pair of randomly assigned non-peers, suggesting significant peer effects. To the best of our knowledge, Stinebrickner and Stinebrickner (2008) offered the only other paper that providing credible evidence of peer effects on college students' study effort. In addition to verifying their findings in a different education context, this study contributes to the literature in several import ways. Most importantly, while Stinebrickner and Stinebrickner (2008) elicited students' effort through surveys sent via campus

[1] In a flipped classroom, students are required to self-study via CMS before a weekly gathering with other students and the instructor. And during the gathering, instructors only briefly go through course content in the first 10 minutes and organize students into small groups to solve challenges that have been posted online before class. Therefore, CMS captures students learning of course material that usually takes place in classroom.

[2] However, as the online interaction with CMS does not capture students' offline effort, measurement error still persists in our measurement of students' effort. Fortunately, measurement error in the dependent variable does not impose bias on the estimation of peer effects.

mail, this study relies on objective data collected on the college's course management system. Subjective surveys might be particularly misleading when studying peer effects. For example, the well-cited study by Cardner and Steinbery (2005) showed that the presence of peers would influence individuals' survey answers. In the context of study effort, although a hardworking student has no obvious incentive to lie on the survey, the less diligent roommates might exaggerate their study effort to avoid being considered sluggish. Thus, surveys might introduce a measurement error correlated with roommates' level of effort. The current study overcomes this potential problem by utilizing objective effort data.

Second, whereas Stinebrickner and Stinebrickner (2008) focused on students' use of time, the current study provides new evidence of peer effects on a conceptually better measurement of study effort: students' academic engaging activities, such as the number of times a student views course material and participates in forum discussions. Third, the study compares the effects of being roommates on two students' academic similarity before and after including their effort similarity. We found that the effect size of being roommates decreased considerably after including students' effort similarity in the model, suggesting that effort is likely to mediate peer effects on academic performances. This result adds to the limited empirical work that has studied the mechanism of peer effects on academic success (a few exceptions include Lavy & Schlosser, 2011; Pop-eleches & Urquiola, 2013; Stinebrickner & Stinebrickner, 2008).

The rest of the paper is organized as follows: Section Two reviews the most relevant literature in peer effects and study effort. A detailed description of the data is outlined in Section Three, including: 1) the flipped course in which the data for this study was collected, 2) the measurement of students' effort similarity, and 3) the room assignment rules and random checks. In Section Four and Five, empirical strategies used in estimating peer effects and the results of this study are presented. The final section concludes with a discussion of findings, limitations, and possible policy implications.

4.2. **Peer Effects in Education**

Literature in peer effects experienced a proliferation of credible empirical studies in recent years (e.g., Booij, Leuven, & Oosterbeek, 2017; Carrell, Fullerton, West, 2009; Falk & Ichino, 2006; Foster, 2006; Hanushek, Kain, Markman, and Rivkin, 2003; Hoxby 2000; Kremer & Levey, 2008; Sacerdote 2001; Zimmerman 2003). Researchers generally identified two confounders that impede the identification of peer effects: selection bias and common shocks (Manski, 1993). Natural peer networks are usually formed based on some selection variable which might not be captured by covariates and have a direct impact on students' outcomes. In addition, peer networks might experience an environmental impact that influences everyone in the group, e.g., a good teacher promote every pupil's grade in the class (Nye, Konstantopoulos, & Hedges, 2004). If not captured by the data, this common shock will also confound peer effects.

Various economics studies have proposed a conceptually succinct approach to disentangling selection effects from peer effects. These studies adopted exogenously instead of naturally formed peer groups to study peer effects (e.g., Sacerdote, 2001). The rationale is straightforward: if peer group formation is random, group members were not similar in any way at the beginning of a study; therefore, the clustering of students' outcomes at the end of a study could not be the result of their similar unobservable. In practice, however, it is difficult to find exogenous peer groups because a peer group formation is everything but random (Sacerdote, 2014). Nevertheless, researchers have scouted in multiple educational settings and enumerated several possible candidates, including conditionally randomly assigned college roommates (e.g., Han & Li, 2009; Sacerdote, 2001; Zimmerman, 2003), randomly assigned social groups in military academies (Lyle, 2007, 2009), K-12 cohort female ratio (Hoxby, 2000; Lavy & Schlosser, 2011), randomly assigned middle school or high classmates (De Giorgi et al., 2010; Imberman, Kugler, & Sacerdote, 2009; Kang, 2007), and randomly assigned peers in field experiments (e.g., Booij et al., 2017; Falk & Ichino, 2006).

In comparison to selection bias, common shocks are often less of a problem because many studies have focused on the effect of exogenous peer characteristics on students' success (Lyle, 2007 and 2009; Sacerdote, 2001, Zimmerman, 2003). In cases where it does matter, the most widely used approach to tackle this problem is through fixed effects (e.g., Bayer, Ross, & Topa, 2008; Bettinger, Loeb, & Taylor, 2014; Nanda & Sorensen, 2006) or the instrumental variables method (De Giorgi et al., 2010). The fixed effects approach, however, relies on panel data to control for peer group level fixed effects; in instances where such data is not available, fixed effects only eliminate common shocks that impact multiple peer groups.

After selection bias and common shocks are handled, additional challenges arise based on the choice of model specification. For individual level linear-in-mean models, Manski (1993) illustrated that it was nearly impossible to separate exogenous and endogenous peer effects due to the reflection problem. Furthermore, when exogenous peer effects were assumed to be zero, identification of endogenous peer effects must rely on instrument variables. An alternative to individual level models is to compare the outcomes of pairs of peers with pairs of non-peers. For example, Bayer et al. (2008) studied the effects of living in the same U.S. census block on pairs of individuals' location of employment. The authors found that a pair of individuals living in the same census block was 33 percent more likely to work in the same location than a pair of individuals who lived in the same group of blocks. However, this type of model requires dichotomous response variables to provide a clear interpretation of model coefficients. In addition, it cannot disentangle peer group level common shocks and peer effects. Nevertheless, the model has a very important advantage over linear-in-mean models: it does not require instrumental variables when only endogenous peer effects are present.

Although the existing literature offers abundant guidance on the identification of peer effects, it only accumulates scant evidence of peer effects on study effort. Stinebrickner and Stinebrickner (2008) might be the only exception. Using housing data from Berea College and a detailed survey of students' time diary, the authors found that if a randomly assigned roommate brought video games to the dorm at the beginning of the freshmen year, his/her roommate would study roughly 0.7 hour (male roommate) and 0.5 hour (female roommate) less every day. In addition, roommates' study time in high school also positively impacted students' use of time in their freshmen year of college. Two other related studies, Carrell and Hoekstra (2010) and Figlio (2007), found that students' academic engagement inside the class is susceptible to peer effects. Specifically, they found that the presence of disruptive peers within classrooms would increase a student's propensity to misbehave and dis-engage during regular class time. Another study by Pop-Eleches and Urquiola (2013) provided further evidence of peer effects on pupils' effort outside classroom. The authors found that attending a secondary school with high-ability peers increased students' frequency of doing homework after school.

This study provides novel evidence of peer effects on study effort using conditionally randomly assigned roommates to eliminate selection bias, and use fixed effects to control for common shocks that affect the entire peer group. In addition, the study adopts Bayer et al.'s (2008) identification strategy to compare peers' effort similarity with non-peers' effort similarity. The result adds to the thin literature that supporting the link between peers and self-study effort.

4.3. Data

We obtained school administrative data from a four-year, tier-two regional college from south-west China. The college enrolls mainly resident students from its home province, with approximately 70-80% from other provinces in China. The college offer 23 majors, yet most students were concentrated in a small number of accounting and finance-related fields of study. For example, over 55% of students were initially enrolled in the most popular three majors.

Our initial sample include all students from the 2014 freshmen class: 2250 in total. Instead of using surveys, study effort indexes were inferred from students' records in the course management system of a compulsory hybrid course. A detailed description about the compulsory hybrid course and the construction of the effort index is provided below. In addition to students' effort data, their background demographics information, first-semester academic performance, dormitory room numbers, as well as semester GPAs are also available.

The final sample excludes students who: 1) did not persist through the first semester; 2) did not take the compulsory course; 3) and lived with previously known friends. Only five students in the initial sample dropped out before the end of the first semester, and 136 students did not take the online course in spring 2015/16 for unknown reasons. As the study effort data for these students are not available, they are not included in the final sample. The last exclusion is 216 students who possibly gamed the random room assignment process and managed to live with previously known friends. These students could have unobserved similarity with their roommates thereby imposing an upward bias on the estimated peer effects. The room assignment and random checks discussed in the later session will details how these students were filtered out and the consequence of excluding them from the final sample. Excluding these three type of students result in a final sample of 1,893 students. Table 1 compares the descriptive statistics for students in the initial sample, the sample with effort data, and the final sample. With the above sample restriction strategy, the effort sample and the final sample only differs in the percentage of students who registered online.

Since the research question concerns the effects of peer-group membership on the effort similarity between two students, the unit of analysis is naturally pairs of students instead of individual students. Therefore, we transformed the final sample of 1,893 students into 38,548 pairs of students who were in the same social groups. A social group is a common administrative unit in Chinese colleges and universities. In the sample, female and male students in adjacent dormitory rooms were grouped into the same social groups, which typically total 40 to 60 students (consequently, roommates were always from the same social group)[3]. Each social group was managed by an academic advisor and a student affair officer, who were responsible for students' social and academic well-being. In addition, almost every school organized extracurricular activities were centered around social groups. Table 2 demonstrates the descriptive statistics for the pairs of students in the final sample.

The online course. The online course titled, Introduction to Computers, was a compulsory general education course for all freshmen studying at the Chinese college. It took place during the spring semester 2015/16 and lasted a total of 18 weeks. This general education course focused on providing students with hands-on experience of the commonly used Microsoft office software, namely Word, Excel, PowerPoints, and Outlook. In addition, it introduced rudimentary computer knowledge, from the functionality of the CPU to the binary number system. The course was available every spring semester and was undertaken by nearly all freshmen; e.g., in 2014, 93.73% of freshmen completed this course.

---

[3] Figure 1 shows the hierarchical structure of cohorts, social groups, and dormitory rooms.

The Introduction to Computers course was a flipped hybrid course. Students watched online videos to learn basic computer knowledge, downloaded and uploaded course exercises, and participated in online forums. Meanwhile, all students need to meet with a teacher once every other week in a classroom. More often than not, lectures focused on the most frequent errors in the uploaded exercises or questions in the online forum, as well as some key concepts in the course. However, students spent most of the class time discussing challenges and questions with other students in groups. This course format is very similar to the flipped-class pedagogy, in which the transferring of knowledge is done by self-pathed learning (usually taking place on an online platform), and the traditional lecture time is used to discuss challenges and solve problems.

Though it was a hybrid course, the majority of the Introduction to Computers course was undertaken online. According to the course syllabus, students were expected to spend around 3 class hours[4] every week studying course material and doing homework on the online platform. Meanwhile, the onsite meeting that was held every other week only lasts for 2 class hours. Accumulatively, students' expected online participation amounted to 50 class hours, while the onsite class participation only amounted to 18 class hours. This suggests that students' online study effort would account for a large proportion of their course study effort. In addition, students who were highly involved in the online platform were likely to be highly involved in the offline classroom as well. With this assumption, students' online effort is a strong proxy for their overall effort in the Introduction to Computers course.

The offline portion of the Introduction to Computers course took place in medium sized classes, each containing 40 to 50 students. Freshmen were assigned to these sessions by the

[4] A class hour is 45 minutes in Chinese colleges and universities.

Office of Academic Affairs at the beginning of the spring semester. All of the classes, 40 in total, were taught by five lecturers. These lecturers used standard course content, syllabus, homework, and exams for the general education course. Therefore, with the exception of differences in teachers' experience and teaching style, all freshmen had more or less the same course experience.

Effort similarity. The effort similarity between a pair of students was constructed in several steps. First, we obtained students' recorded activities in Introduction to Computers. Second, we composited a student level index to measure a student's overall effort in this course. Finally, two students' effort indexes were transformed to construct a single number index that ranges from zero and one to measure the similarity between two students' study effort.

The course management system recorded seven types of students' activities: 1) the total number of course material downloaded, 2) the total number of course video viewed, 3) the total number of slides viewed, 4) the total number of forum posts viewed, 5) the total number of forum posts replied, 6) the total number of forum posts liked, and 7) the total number of forum posts started. All seven items are highly positively skewed as shown in Figure 2.We applied log-transform to remedy this problem[5], and transformed items are internally consistent (alpha=0.817, n=1893).

We then constructed an effort index, y_i, for each student by averaging the seven transformed effort items. The predictive validity of this composite is moderate: the correlation between students' effort index and course grade is 0.247. This correlation coefficient might appear to be low, but it is actually better than the predictive validity of NSSE on GPA (Gordon et al., 2008) and comparable to other academic engagement constructs (e.g., Simons-Morton &

[5] For observations whose values are zeros, they are mapped to zero instead of ln(0).

Chen, 2009). In addition, even the effort index is not a perfect measure of their actual course study effort. Measurement error in the outcome variable does not bias peer effects towards zero but only reduces the power of the analysis.

The final effort similarity between a pair of students was constructed using the Equation (1):

$$S_{ij} = 1 - \frac{abs(y_i - y_j)}{\max(y_m - y_n)} \quad (1)$$

A pair of students' effort similarity, S_{ij}, equals to one minus the absolute effort difference between student i and j divided by the maximum effort difference between any pair of students. Intuitively, S_{ij} is simply a linear transformation of two students' effort difference. This transformation aims to scale S_{ij} between zero and one; it increases as students' effort difference decreases. Therefore, effort similarity approaches one when students i and j have a similar level of study effort, and goes to zero when one student study very hard while the other is highly disengaged. The distribution of S_{ij} in the data is shown in Figure 3.

Room assignment and random checks. We conducted several interviews with the Director of Residential Life at the college to gain a thorough understanding of the room assignment process. We learnt that the room assignment process followed three simple rules with minor twists. First, only students of the same gender are assigned into the same residential halls. Second, students in two special programs: Accounting –The Association of Chartered Certified Accountants (ACCA) and Joint Degree Programs are grouped into the same dormitory rooms[6]. Third, students not in the two special programs may choose to register early online or register onsite right before the academic year begins. For students who register online, the registration system offers them

[6] Unlike most college and universities in China (e.g., Li & Han, 2009), this particular college does not sort same major students into a dormitory room (except for the two aforementioned special programs).

eight random rooms to choose from. For those who register onsite, such choice is not available. Instead, available beds are queued (not randomly), and students are assigned to the next available bed according to their registration sequence. For the 2014 cohort, the housing assignment arranged the total sample of 2250 students into 565 rooms, including 559 fully occupied quadruples, three quadruples with three students, two quadruples with two students, and one quadruple with a single student.

In a nutshell, the dormitory room assignment process in this college is only conditionally random: students who have the same registration type, similar registration sequence, same gender, or are enrolled in Accounting (ACCA) or Joint Degree Programs are more likely to be assigned to the same dormitory rooms by the school. After controlling for these five factors, students should be randomly distributed to dormitory rooms in theory. However, after extensive communication with the Director of Residence Life, we found that two system loop holes exist. First, since students who register online are given eight random rooms to choose from, friends can potentially select into these rooms. Second, for those who register onsite, the room assignment is highly dependent on the registration sequence. As students are more likely to attend the registration in pairs, they are more likely to end up in the same room. As a result, students might take advantage of these system flaws to select roommates. If this roommates' selection process is based on an unobservable which has an impact on students' study effort, the clustering of similar students in dormitory rooms will bias the peer effects.

An easy check for the potential selection is to test the correlation of observables between roommates after taken the housing assignment process into consideration. Table 3, column one, demonstrates moderate sorting of students into dormitory rooms: students with similar College

Entrance Exams[7] (CEE) percentiles are more likely to become roommates. Therefore, at least a small portion of students likely exploited the system flaws to select their roommates. Recalling that the room assignment process is either online or onsite for a short period of time, students do not have much opportunity to collect information on each other nor build social connections. It is therefore reasonable to assume that most sorting did not happen between new acquaintances formed during the registration process but between former established social connections. Therefore, if pre-college social connections could be identified with the available data, the selection problem could be solved by excluding the self-selecting roommates from the sample.

We used students' high school ID as a proxy for students' pre-college social connections. To check the validity of this proxy, we tested if the clustering of students from the same high school in dormitory rooms parallels the sorting of friends across dormitory rooms. A Monte Carlo Simulation with 3,000 times repetition was performed to simulate a complete random process of assigning same gender students into rooms. This process indicates the range of rooms housing multiple same high school students under random assignment. The distribution of the rooms is illustrated in Figure 4. Under perfect random assignment, the number of rooms housing multiple same high school students does not exceed 20 in any of the 3,000 simulations. However, the data capture 55 such rooms, which is highly unlikely under random assignment. Therefore, as expected, students from the same high schools are indeed significantly clustered in the dormitory rooms, suggesting that high school ID is a good proxy for pre-college friendship.

Deleing the 55 rooms housing multiple same high school students indeed eliminate any significant correlation between roommates. As shown in Table 3, column three, the previous random check shows no significant correlation between individual and roommates' average

[7] College Entrance Exams in China is the equivalent of SAT in the U.S.

characteristics after dropping the 55 rooms. Moreover, as a double check, we performed a random analysis on the dropped sample to test the similarities among these self-selected students. As expected, we found multiple significant correlations between roommates' observables in the dropped sample: roommates are likely to have a similar CEE percentile, both being fresh high school graduates, having the same high school academic track, and sharing the same initial major (Accounting—ACCA Direction).

To summarize, the above random checks suggest that a small amount of students gamed the housing rules in order to select their dormitory rooms. The selection is most likely to happen among previously known friends instead of new acquaintances. We used students' high school ID as a proxy of their pre-college friendship and marked all rooms housing multiple same high school students as possible self-selections. These possible selections indeed have highly correlated observables, and dropping them from the sample seemly exclude the small amount of self-selected roommates.

We performed the main analysis on the final sample that dropped rooms housing multiple same high school students. This reduction resulted in a sample size of 1893 students in a total 510 rooms, including 394 fully occupied quadruples, 88 quadruples with three students, 25 quadruples with two students, and three quadruples with one students (not used in the analysis). Though sample selection is a bad practice in general, the benefits outweighs the cost in this situation. By deleting possible self-selected roommates from the sample, we could exclude selection bias while estimating peer effects, thereby increasing the credibility of the results. Moreover, though the final sample is significantly different from the initial sample, the difference size is negligible. As shown in Table 1, the far right column, only the percentage of online registered students are statistically different between the initial sample and the final sample.

4.4. Empirical Strategy

The first research question focused on whether two students are more likely to have similar effort patterns if they live in the same dormitory room. We used the following baseline model[8] to estimate the difference in effort similarity between roommates and social group mates:

$$S_{\mathrm{ij}} = \alpha_s + \beta_0 R_{ij} + \theta C_{ij} + \delta X_{ij} + \gamma Z_{ij} + \varepsilon_{ij} \quad (2)$$

where α_s stands for the average effort similarity within each social group s. R_{ij} is a dummy variable that equals to one only when students i and j were peers. C_{ij} measures the common shocks that happen to both students, e.g., a common instructor in the hybrid course. X_{ij} represented students i and j's pair characteristics, e.g., both i and j are female students, Z_{ij} represents students i and j's conditional random assignment factors: e.g., students' registration type and registration sequence. The independent variable R_{ij} is of main interest in this study because its coefficient measures peer effects on students' effort. The direction and significance of peer effect is indicated by β_0: If being peers significantly increase students i and j's effort similarity, then peer effects on effort must be positive and significant.

The major goal of this study is to establish a causal link between peer status, β_0, and students' effort similarity, S_{ij}. This entails disentangling the selection effect and common shocks from peer effects. Conceptually, selection effects describe the phenomena that individuals select peers based on mutual similarities which usually contribute to their similar performances. Common shocks represent environmental variables that not only happen to all peers but also

---

[8] The baseline model is inspired by Bayer et al (2008), which studied peer effects on individuals' place of work.

explains their similar performances. A classical example of common shock is teacher effects on students' academic performance (Nye et al., 2004). Finally, peer effect is a complex phenomenon stemming from social interactions. It might take forms of peer teaching (Loke & Chow, 2007), peer pressure (Borsari & Carey, 2001), social norms (Unger et al., 2002), and shared information (Ryan, 2000). The usual way of estimating peer effect in economics is through eliminating selection bias and common shocks.

To illustrate how selection effects and common shocks might confound peer effects in a technical manner, let's break down ε_{ij} in Equation (2) into three parts:

$$\varepsilon_{ij} = u_{ij} + v_{ij} + w_{ij} \tag{3}$$

where u_{ij} captures the unobserved individual similarities between i, j that do not change over time. v_{ij} captures the common shocks that happen to students i and j after college enrollment. w_{ij} is the true noise. In the presence of selection, students i and j's peer status may depends on their unobserved individual similarities u_{ij}. In that case, not controlling u_{ij} in the model will result in omitted variable bias. Fortunately, the peer group formation in this study is randomly conditioned on Z_{ij} in Equation (2). Therefore, unobserved similarities u_{ij} are not correlated with peer status R_{ij}. Besides selection bias, common shocks v_{ij} may also be correlated with students i and j's peer group status. For example, Table 4 shows that a student is more likely to be taught by the same instructor as their roommates. As an instructor has a significant influence on students' academic engagement level (Skinner & Belmont, 1993), excluding them from the model will create a bias on estimated peer effect, β_0. To remedy this problem, we controlled detailed social and academic activities that affect students i and j, including their social group fixed effects, course instructor fixed effects, and the percentage of same course sessions in the semester.

The baseline model in Equation (2) can be easily extended to test heterogeneous peer effect:

$$S_{ij} = \alpha_s + \beta_1 R_{ij} + \beta_2 D_{ij} R_{ij} + \theta C_{ij} + \delta X_{ij} + \gamma Z_{ij} + \varepsilon_{ij} \qquad (4)$$

where D_{ij} measures the demographic similarity between students i and j, e.g., i and j both have the same ethnicity. D_{ij} are all centered to avoid changing the meaning of β_1. Therefore, β_1 in Equation (4) still captures the difference in effort similarity between an 'average' pair of roommates and an 'average' pair of non-roommates. Current literature suggests that peer effects are stronger among similar students than their counterparts (Bayer et al., 2008; Hoxby, 2000). Therefore β_2 is hypothetically positive.

The aforementioned models are useful to test the existence of peer effects on study effort but fail to provide interpretable evidence on the size of peer effects. An easy remediation would be characterizing students' effort into different categories and using a multinomial logistic model to check if being roommates increase their chance of falling into a particular effort category. Following this path, we first divided students into three categories-- diligent, average, and lazy--- based on their manifested online effort relative to the cohort. A diligent student's online effort is larger than the 75th percentile of the sample effort distribution. An average student's online effort is between the 25th and 75th sample effort percentiles. Finally, a lazy student's online effort is less than the 25th percentile. Then we used the Equation (5) to measure the size of peer effects:

$$\mathrm{Ln}\left(\frac{P(Y_{ij}=r)}{P(Y_{ij}=0)}\right) = \alpha_s + \beta_r R_{ij} + \theta C_{ij} + \delta X_{ij} + \gamma Z_{ij} + \varepsilon_{ij} \qquad (5)$$

Where Y_{ij} divides a pair of students into four groups: both diligent, both average, both lazy, or have different effort pattern. Students having different effort pattern are coded as zero and used as the reference group. R_{ij} is again the variable of interest, and other independent variables have the same meaning as in Equation (2). β_r measures the size of peer effects and is hypothetically larger than zero.

Our second research question focus on whether effort mediate the effects of peers on students' academic performance. The mediation effect can be tested using Equation (6) and (7):

$$A_{\mathrm{ij}} = \alpha_s + \beta_3 R_{ij} + \theta C_{ij} + \delta X_{ij} + \gamma Z_{ij} + \varepsilon_{ij} \tag{6}$$

$$A_{\mathrm{ij}} = \alpha_s + \beta_4 R_{ij} + \rho S_{ij} + \theta C_{ij} + \delta X_{ij} + \gamma Z_{ij} + \varepsilon_{ij} \tag{7}$$

where A_{ij} denotes students' similarity in academic performance, constructed the same way as S_{ij}: one minus the absolute difference between students' academic performance divided by the maximum possible academic difference between any pair of students. We operationalized academic performance as students' spring semester GPA in the freshmen year[9] and their course grade in Introduction to Computers. $S_{ij}, C_{ij}, X_{ij}, and\ Z_{ij}$have the same meaning as in previous equations.

β_3 in Equation (6) captures the comprehensive peer effects on students' academic outcomes. By comprehensive peer effects, we mean all direct and indirect effects of peers on students' academic performance, including, but not limited to, choice of courses, eliciting of efforts, peer teaching, etc. β_3 is a causal measure of comprehensive peer effects because the inclusion of Z_{ij} eliminate selection bias and C_{ij} handles after-enrollment common shocks that apply to all roommates. β_4 in Equation (7) represents comprehensive peer effects on academic performances excluding the part mediated through students' effort. Hypothetically, β_4 should be considerably smaller than β_3 if the mediation effect of student effort is large enough.

All the aforementioned models study peer effects on a pair of students' effort similarity. Comparing to linear-in-mean models that directly regress students' effort on roommates' effort,

[9] This is the semester when students take the introduction computer course.

these models avoid the simultaneity bias problem. However, the parameters in these models are difficult to interpret and impossible to compare with other studies that directly investigate peer effects on a student's level of effort. Nevertheless, the sign and significance of the parameters would provide evidence for the existence and direction of peer effects on students' effort. Because without positively significant peer effects, roommates' study effort could not be more similar than non-roommates' study effort after controlling selection and common shocks.

All models in Equation (2)-(4) and Equation (6)-(7) are estimated using OLS. Since the peer group formation is conditionally random and after-enrollment common shocks are controlled, the OLS estimator of peer effects are likely to be unbiased. However, as Bayer et al. (2008) illustrated, errors in Equation (2) to Equation (7) are clustered thereby OLS and maximum likelihood likely underestimates the standard errors.[10] To remedy this, we used bootstrap to calculate the proper standard errors for all the models.

4.5. Peer Effects on Effort Similarity

Table 5 illustrates the results of the baseline model in Equation (2), where a pair of students' effort similarity is explained by their peer status and a basket of controls. We gradually added in controls from left to right, so the decrease of peer effect size is observable after controlling selection bias and common shocks. The first column presents the simple regression which only uses students' peer status to explain their effort patterns. Without other controls, being roommates significantly increases students' effort similarity. This effect, as expected, plunges from 0.0376 to 0.0210 after adding in control variables that set up the random assignment condition. Furthermore, we added a basket of demographic and academic controls to

[10] For example, if $S_{ab} = 1, S_{ac} = 1$, then S_{bc} must also equal one. Therefore, pairs of students' outcomes are clustered together.

check if the specification in column two removes all selection bias. As shown in column three, the estimated peer effect is stable in comparison to the previous model (0.0210 vs. 0.0205), implying an absent of selection bias. Finally, column four demonstrates the preferred specification in Equation (2), where after-enrollment common shocks are controlled by adding fixed effects in social groups and course teachers, as well as the percentage of shared course sessions. The estimated peer effects decrease slightly from 0.0205 to 0.0180, suggesting that a small part of roommates' effort similarity is explained by their common college experience. Taken together, this table illustrates that selection bias and common shocks explain about half of the manifested peer effort similarity (0.0187 out of 0.0367). Nevertheless, after controlling these two main sources of confounders, being peers still significantly increases students' effort similarity by 0.0180.

Table 6 presents the results of the baseline model in Equation (2) on different gender and different types of effort.[11] Each cell in Table 6 represents the estimated peer effects on a particular type of effort for a particular gender. For all students, peer effects on forum participation is considerably stronger than that on individual learning: being peers increases students' forum participation similarity by 0.0204 while the effect size on individual learning similarity is only 0.0108. The difference is not surprising as forum discussions are much more cooperative in nature than individual learning activities. For example, a forum post initialized by a roommate is likely to elicit comments by other peers while individual learning activities, such as watching course videos, is unlikely to have the same effect. Though not shown in Table 6,

[11] Overall course effort is divided into two mutually exclusive parts: individual learning and forum participation. Individual learning is constructed using three indicators: 1) the total number of course material downloaded, 2) the total number of course video viewed, and 3) the total number of slides viewed. Forum participation is constructed using the following indicators: 1) the total number of forum posts replied, 2) the total number of starred forum posts, and 3) the total number of starred forum posts.

students' similarity on covariates also explain part of their similarity in effort. Specifically, students from the same province, sharing a large portion of courses, both from urban or rural areas, and having similar College Entrance Exam scores or first semester GPA, all present similarly effort patterns.

The majority of the findings on the pooled sample extend to the two gender subgroups. On average, roommates' effort similarity is stronger than non-roommates' for both female and male students. Furthermore, peer effects are again stronger on forum discussion than individual learning. Peer effect size ranges from 0.00990 to 0.0205. The effect is slightly stronger for female students than male students on all aspects. However, peer effects on individual learning similarity is not significant for male students. This is probably due to a much smaller sample size in comparison to female students (5,118 male pairs vs. 20,630 female pairs); the point estimator for male students is close to that of the female students.

To examine the heterogeneous specification, we allowed the peer effects to vary conditioning on students' demographic similarity as shown in Equation (4). Table 7 presents the estimated results. None of the demographic similarities moderate peer effects on students' effort similarity in any model. This is somewhat contradictory with the existing literature, in which peers' social-demographic characteristics are found to moderate peer effects. For example, in Bayer et al. (2008), neighbor effects on working location was stronger for neighbors with children and who were of similar age than less similar neighbors. In addition, in Hoxby (2002), the increment of black students had the worst detrimental effect on other black students in the classroom. The most likely explanation is that peers with similar social-demographic characteristics formed a stronger tie in the abovementioned studies in comparison to this one. This may be because those studies used larger peer groups; social connections were likely to happen within rather than across social-demographic groups. However, despite possible demographic difference, all roommates are likely to have strong social interactions due to mandatory sharing of living space. Therefore, social-demographic difference is not likely to explain variation in social interactions between roommates thereby having no moderating effect on peer effects.

So far, Table 6 and 7 have enriched the baseline model by exploring possible heterogeneous specifications across gender, effort types, and demographics. Table 8 would advance the baseline model in a different direction: its interpretability. We divided a pair of students into four categories: both low effort, both middle effort, both high effort, and in different efforts. Table 8 suggests that in general, being roommates increase a pair of students' odds of being in the same effort category: 31.65% for the both low effort category, 13.66% for the both middle category, and 35.58% for the both high category. The effect of being roommates persist across different effort types (individual learning and forum particiaption), and the effect ranges from 8.55% to 47.85% in terms of odds ratios. Interestingly, the size of peer effect again appears to be larger on forum participation than on individual learning activities.

Lastly, we examined the possible mediation effect of effort patterns for peer effects on students' academic outcomes. The result is illustrated in Table 9. Column 1 and 4 show the OLS estimation of comprehensive peer effects on students' academic outcomes, without controlling the effort similarity between a pair of students. As expected, being roommates significantly increases students' similarity in academic performance, both in the computer introduction course and overall semester GPA. This finding is coherent with the majority of literature that investigated peer effects on students' academic outcomes (Han & Li, 2009; Hanushek et al., 2003; Hoxby, 2000; Lyle, 2007; Sacerdote, 2001; Zimmerman, 2003). To further examine the mediation effect, we added students' overall effort similarity in column 2 and 5. Interestingly, after controlling study effort, estimated peer effects drop by almost half: from 0.00650 to 0.00386 on semester GPA and 0.00605 to 0.00302 on course grade. Peers' direct effect on semester GPA (0.00386) is still significant at the 0.05 level but not on course grade (0.00302).[12] In addition, students' effort similarity becomes the single most powerful predictor of students' similarity in academic performance: 0.151 for semester GPA and 0.171 for course grade. Recalling that previous tables suggest that peers influence each other's effort patterns, this drop in peer effects size supports our initial hypothesis that at least a portion of peer effects on academic performance is mediated through its effects on effort patterns. The last two columns, 3 and 6, further break down study effort to forum participation and individual learning activities. Estimated direct peer effects on academic performance is comparable to those in column 2 and 5. Further, both types of effort have significant and similar effect size on students' academic performance.

Table 10 replicates the mediation test on gender subgroups. The overall pattern is similar to the ones in the pooled sample: for both genders, peers' direct effects on academic performance is higher when students' study effort is absent and drop considerably when effort is accounted for. Therefore, study effort likely mediates at least part of peer effects on students' academic performance for both genders. In addition, female students appear to be more susceptible to peer effects than male students: while peer status generally has significant direct impact on female students' similarity in academic performances, being roommates does not significantly increase

[12] This does not necessary mean peers still have a direct effect on overall semester GPA after controlling their study effort. This is because we only controlled peer effort similarity in the computer course, which does not fully represent students' overall study effort during the academic semester. Therefore, the appearing direct peer effect on overall semester GPA could be due to uncontrolled effort similarity in the other courses.

male students' similarity in academic performance at all. This result may echo Han and Li's (2009) finding that only female students responded to peer effects on academic performance. However, this conclusion is not final. An alternative interpretation of this difference is the large sample gap between female and male students. The insignificant peer effects on the male sample might simply be due to the larger standard error that comes with a smaller sample size. As a matter of fact, the point estimator of peer effects on course grade is larger for male students than female students (0.00831 vs. 0.00487).

As a robust check, we re-estimated peer effects in Table 5-Table 10 by three different specifications. In specification one, we added the same high school students who share dormitory rooms, which increased the sample from 38,548 to 47,633 pairs of students; in specification two, we used the absolute effort difference between a pair of students as the outcome variable instead of students' effort similarity; in specification three, we used the average of non-transformed students' activities on the online platform as the effort index instead of the average of log-transformed activities. The robust checks yielded similar outcomes with minor twists.[13] Table 11 presents the outcomes on the baseline model. Further results are available upon request.

4.6. Discussion.

To understand whether peers exert influence on a student's effort at school, this study compare roommates' and non-roommates' effort similarity in a compulsory hybrid course in a Chinese College. After controlling selection bias and common shocks, roommates' effort

[13] First, the estimated peer effects are stronger in specification 1. This is expected as excluded roommates share considerable similarities thereby tend to have more similar effort pattern than an average pair of roommates. Second, peer effects are significantly negative on effort difference in specification 2. This is expected as effort difference is the opposite of effort similarity. And lastly, peer effects are of similar magnitude in specification 3, suggesting that transformation of effort indicator does not produce false significance in the result.

patterns are still significantly more similar than non-roommates' effort patterns, suggesting that peer effects on effort is positive and significant. In addition, the study find that the strength of peer effects is conditioned on the type of effort, with students' forum participation activities being consistently more susceptible to peer effects than individual learning activities across all models.

Importantly, these effects on effort appear to mediate peer effects on students' academic outcomes. Peers' academic quality has been found to be related to a student's academic performance (e.g, Sacerdote, 2001; Zimmerman, 2003). However, the mechanism of this effect is always left as a black box, partially due to the lack of students' detailed behavioral data. A few exceptions include Lavy and Schlosser (2011) and Agirdag, Houtte, and Avermaet (2012), in which the authors found less disruptive behavior and a sense of futility mediated peer effects on students' academic performances. The current study theorizes that at least part of peer effects on academic performance takes place through altering a student's effort pattern. This hypothesis is supported by the empirical results. First, roommates' effort patterns are significantly more similar than non-roommates. Second, though being roommates has a significant positive impact on students' similarity in academic performance, the estimated effect drop to nearly half of its original size after accounting for students' effort.

The study also find that peer effects on effort is homogenous across demographic subgroups. The result is not consistent with some existing literature in which peer effects were found stronger within than across demographic subgroups (Bayer et al., 2008; Hoxby, 2000). The most likely explanation for this inconsistency is that social connection is more likely to occur within demographic groups when peer groups are large, as in classes and neighborhoods, but not in small peer groups such as dormitory groups. In addition, the study finds that female and male students respond more or less similarly to peer effects when it comes to study effort. However, they respond differently to peer effects on academic performance.

Overall, the current study finds robust evidence of positive peer effects on study effort. By controlling selection bias and common shocks, this study adds more credible evidence to prior works in this area. But readers might concern why can peers influence a student's effort in the first place? Though the current study does not directly investigate this mechanism, previous researches offer some insights.

First, residential peers' bad habits might distract a student' from putting more effort into learning. This explanation resembles the "bad apple" model (Hoxby & Weingarth, 2005). For example, Stinebrickner and Stinebrickner (2008) found that if a roommate brought video games to the school, a student was likely to spend less time on studying. In addition, some studies in K-12 settings found disruptive peers impede a student's engagement during class (Carrell et al., 2010; Figlio, 2007). Second, hardworking peers might serve as role models that inspire a student to put more effort into study, or the "shining light" model in Hoxby and Weingarth (2005)'s terminology. Stinebrickner and Stinebrickner (2006) made a similar argument. They contended that the only possible way roommates could affect students' first-year achievement was by inspiring them to spend more time on study. Lastly, Bishop (2004) argued that, when the reward of learning depends on the relative rank related to other students, peers tend to discourage the "nerds" to work less hard. We found the opposite equally possible: When the reward of learning depends on the relative rank, individuals might be forced to put more effort into study when one's peers work hard.

The readers might also concern the representativeness of our sample as it only includes a single regional college in China. The size of peer effects on effort might differ in other Chinese universities, let alone colleges and universities in the US. The lack of representation is an endemic issue for studies that exploit randomly assigned roommates to estimate peer effects, as it is very difficult to secure housing data from multiple colleges. Nevertheless, as the natural experiment setting guaranteed high internal validity for this study, its external validity can be tested by further research employing the same empirical setting in other schools. Although every single one of these studies only has limited external validity, together they will be able to depict a relatively representative picture of peer effects.

With these limitations in mind, the current study adds novel evidence to the causal link between peer and study effort. Such a link has two important implications. First, the link suggests that peer study habits might be a better indicator of peer quality than academic ability when studying residential peer effects in the U.S. (as argued in Stinebrickner & Stinebrickner, 2004). When using academic ability as a measure of peer quality, the empirical results in the U.S. are mixed and weak (e.g., Foster 2006; Lyle 2007; Sacerdote 2001; Zimmerman, 2003). The most convincing explanation is that roommates in the U.S. settings do not interact regularly on academic matters (Stinebrickner & Stinebrickner, 2004). Therefore, roommates' academic ability does not directly affect one's first-year achievement. Several others studies support this argument. For example, Han and Li (2009) found strong residential peer effects in China, where roommates share the same majors and thus are likely to interact on academic issues. Poldin, Valeeva, and Yudkevich (2013) found college study partners have stronger explanation power on an individuals' GPA than college friends. Parker (2012) found that having dorm-mates taking the same economic class benefit students (as they may have academic interaction), while other dorm-mates do not have such impact. Therefore, future studies on residential peer effects should use peer study habits as a measure of peer quality.

Second, as study effort is often found to have a positive impact on a student's academic performance (e.g., Grave, 2011; Stinebrickner & Stinebrickner, 2008), its determinants, especially malleable ones, are of particular interest. The causal link between peers and study effort suggests that policymakers could increase one's effort by manipulating the student's peer group. However, such manipulation can be precarious at times. Carrell et al.'s (2011) experiment at West Point suggested that peer group manipulation may not lead to expected results. In their study, low-achievers and high-achievers select into sub-social networks based on academic ability and did not interact much across these sub-structures. As a result, low-achievers assigned to high-achieving peers performed worse than low-achievers assigned to a random peer group. This is not to say peer group manipulation is impossible. A recent research by Li, Han, Zhang, and Rozelle (2014) revealed that peer group manipulation accompanied with group incentives can increase low-achievers' academic performance. This is probably because the group incentives (e.g., monetary reward) encouraged meaningful academic interaction between the low-achieving student and high-achieving student in their study.

At last, this study provides some tentative policy suggestion. Schools can increase student effort by introducing hard working students into his or her peer network. This is increasingly possible with the proliferation of course management systems, and hybrid and online courses. With these technology infrastructures, students' effort level could be easily inferred from their online data. Starting from there, course instructors could purposely mix lowly engaged students and highly engaged students into the same study groups and provide incentives for them to interact both inside and outside of classrooms. Across courses, academic advisors could mix diversely engaged students who share multiple courses into a single learning community. Increased social interactions with hard working students will likely benefit the low-achievers.

Tables and Figures

Table 1: Descriptive Across Samples

	Initial Sample		Effort Sample		Final Sample		T-test Statistics
	mean	s.d.	mean	s.d.	mean	s.d.	
% of Female Students	0.706	(0.456)	0.709	(0.454)	0.703	(0.457)	-0.645
% of Fresh High School Graduate	0.244	(0.430)	0.241	(0.428)	0.251	(0.433)	0.913
% From Host Province	0.769	(0.422)	0.755	(0.430)	0.747	(0.435)	-0.842
% of Racial Minority	0.031	(0.173)	0.032	(0.175)	0.033	(0.179)	0.363
% Registered Online	0.376	(0.484)	0.373	(0.484)	0.343	(0.475)	-2.729**
% of Science Track in High School	0.430	(0.495)	0.450	(0.498)	0.460	(0.499)	0.887
# of same major room mates	1.157	(1.136)	1.217	(1.141)	1.193	(1.130)	-0.917
Dormitory Room Size	3.991	(0.123)	3.991	(0.125)	3.990	(0.132)	-0.340
CEE Percentiles	0.800	(0.110)	0.801	(0.110)	0.798	(0.112)	-0.898
Majoring in Accounting(CPA Direction)	0.508	(0.500)	0.540	(0.499)	0.535	(0.499)	-0.477
Majoring in Construction Cost	0.089	(0.285)	0.095	(0.293)	0.095	(0.293)	-0.041
Majoring in Finance(Assets Management)	0.076	(0.265)	0.080	(0.272)	0.077	(0.267)	-0.490
# of times accessing course material			26.305	(25.925)	25.968	(25.470)	-0.577
# of times downloading documents			28.440	(28.516)	28.466	(28.650)	0.041
# of times completely watched a course video			3.360	(4.651)	3.279	(4.571)	-0.770
# of forum posts replied			27.192	(32.855)	27.086	(32.403)	-0.142
# of forum posts reviewed			66.462	(68.852)	66.144	(67.568)	-0.205
# of forum posts initialed			2.310	(4.939)	2.244	(4.746)	-0.601
# of forum posts liked			9.565	(42.691)	9.225	(40.717)	-0.364
Sample Size	2250		2019		1893		

Note: T-test Statistics compares means between effort sample and final sample; * p<0.05, **p<0.01, *** p<0.001

Table 2: Descriptive for Pairs of Students

	mean	s.d.
Overall Effort Similarity	0.783	(0.165)
Individual Learning Similarity	0.808	(0.143)
Forum Participation Similarity	0.796	(0.156)
Roommates	0.067	(0.250)
% of Shared Courses in the Second Semester	0.234	(0.281)
Both from the Same Province	0.576	(0.494)
Both Non-fresh High School Graduates	0.064	(0.244)
Both Fresh High School Graduates	0.559	(0.496)
Both Minority	0.005	(0.068)
Both Majority	0.931	(0.254)
Both Science Track in High School	0.215	(0.411)
Both Humanities Track in High School	0.286	(0.452)
Both from the Same Major	0.405	(0.491)
First Semester Std GPA Difference<0.1	0.463	(0.499)
CEE Percentile Difference<0.1	0.463	(0.499)
Age Difference<0.5	0.354	(0.478)
Both Attend High School in Urban Cities	0.365	(0.481)
Both Attend High School in Rural Areas	0.181	(0.385)
Both Female	0.539	(0.499)
Both Male	0.133	(0.339)
Both Registered Online	0.204	(0.403)
Both Registered Offline	0.532	(0.499)
Both Majored in ACCA	0.028	(0.164)
Both enrolled in Study Abroad Programs	0.028	(0.165)
Both Taught by Instructor One	0.117	(0.322)
Both Taught by Instructor Two	0.051	(0.219)
Both Taught by Instructor Three	0.087	(0.282)
Both Taught by Instructor Four	0.064	(0.245)
Both Taught by Instructor Five	0.063	(0.243)
Sample Size	38548	

Table 3: Correlation between Self and Roommates Observables

	Effort Sample	Dropped Sample	Final Sample
CEE Percentile	0.0489*	0.220**	0.0284
	(0.0227)	(0.0759)	(0.0240)
Fresh Graduate	0.0185	0.349***	-0.0215
	(0.0355)	(0.0987)	(0.0380)
Science Track	0.0631	0.304**	0.0252
	(0.0351)	(0.105)	(0.0374)
Minority	-0.0193	-0.0872	-0.0194
	(0.0359)	(0.123)	(0.0378)
Majoring in Accounting(CPA Direction)	0.0277	0.280**	-0.00616
	(0.0352)	(0.104)	(0.0374)
Majoring in Construction Cost	0.0148	0.0473	0.00770
	(0.0363)	(0.114)	(0.0383)
Majoring in Finance(Assets Management)	0.0417	0.193	0.0116
	(0.0351)	(0.111)	(0.0371)
Attend High School in Urban Cities	0.0611	0.186	0.0421
	(0.0342)	(0.103)	(0.0363)
Control Registration Sequence	Yes	Yes	Yes
Gender	Yes	Yes	Yes
ACCA Class	Yes	Yes	Yes
International Class	Yes	Yes	Yes
Observations	2109	216	1893

Standard errors in parentheses, * p<0.05, ** p<0.01, ***p<0.001

Table 4: Correlation between Self and Roommates' Instructors in Introduction to Computers, Linear Probability Model

	Taught by Instructor 1	Taught by Instructor 2	Taught by Instructor 3	Taught by Instructor 4	Taught by Instructor 5
# of Peers Taught by Instructor 1	0.201***				
	(0.00938)				
# of Peers Taught by Instructor 2		0.0640***			
		(0.0128)			
# of Peers Taught by Instructor 3			0.184***		
			(0.0102)		
# of Peers Taught by Instructor 4				0.144***	
				(0.0113)	
# of Peers Taught by Instructor 5					0.0720***
					(0.0128)
Random controls	Yes	Yes	Yes	Yes	Yes
Observations	1876	1876	1876	1876	1876
Rsq	0.260	0.087	0.177	0.114	0.085

Standard errors in parentheses; *p<0.05, ** p<0.01, *** p<0.001

Table 5: Baseline Model Result, Peer Effects on Effort Similarity

	(1)	(2)	(3)	(4)
Roommates	0.0367***	0.0210***	0.0205***	0.0180***
	(0.00286)	(0.00301)	(0.00300)	(0.00303)
Both From the Same Province			-0.00236	0.00331
			(0.00172)	(0.00177)
Both Former High School Graduates			0.000379	0.0000409
			(0.00358)	(0.00358)
Both Fresh High School Graduates			-0.00473**	-0.00476**
			(0.00179)	(0.00181)
Both Minority			-0.0903***	-0.00765
			(0.0179)	(0.0237)
Both Majority			0.00246	0.00377
			(0.00332)	(0.00337)
Both Science Track in High School			-0.00785**	-0.00641*
			(0.00272)	(0.00268)
Both Humanities Track in High School			-0.00599*	-0.00655*
			(0.00285)	(0.00278)
Both from the Same Major			0.000487	-0.0168***
			(0.00178)	(0.00306)
First Semester Std GPA Difference<0.1			0.0129***	0.0128***
			(0.00163)	(0.00165)
CEE Percentile Difference<0.1			0.0114***	0.0109***
			(0.00254)	(0.00252)
Age Difference<0.5			0.00369*	0.00248
			(0.00178)	(0.00171)
Both Attend High School in Urban Cities			0.00312	0.00243
			(0.00193)	(0.00187)
Both Attend High School in Rural Areas			0.00956***	0.00945***
			(0.00226)	(0.00228)
Constant	0.781***	0.758***	0.751***	0.745***
	(0.000863)	(0.00192)	(0.00404)	(0.0106)
Random controls	No	Yes	Yes	Yes
Group fixed effect	No	No	No	Yes
Teacher fixed effects	No	No	No	Yes
% of Shared Credits	No	No	No	Yes
N	38548	38548	38548	38412
R-sq	0.003	0.029	0.034	0.066
	0.003	0.029	0.034	0.066

Standard errors in parentheses; Bootstrapped Standard Errors; *p<0.05, **p<0.01, ***p<0.001

Table 6: Peer Effects in Baseline Model by Gender and Different Types of Effort

	All Pairs	Female Pairs	Male Pairs
Overall Effort	0.0180***	0.0187***	0.0159*
	(0.00303)	(0.00340)	(0.00619)
Individual Learning Activities	0.0108***	0.0107***	0.00990
	(0.00279)	(0.00299)	(0.00552)
Forum Participation	0.0204***	0.0205***	0.0194***
	(0.00288)	(0.00322)	(0.00573)
Random controls	Yes	Yes	Yes
Group fixed effect	Yes	Yes	Yes
Teacher fixed effects	Yes	Yes	Yes
% of shared credits	Yes	Yes	Yes
N	38412	20630	5118

Standard errors in parentheses; Bootstrapped Standard Errors; Also include in models are covariates in Table 5 column 4; *p<0.05, ** p<0.01, *** p<0.001

Table 7: Heterogeneous Model, By Gender and Different Types of Effort

	All Pairs	Female Pairs	Male Pairs
Overall Effort			
Roommates	0.0180***	0.0188***	0.0161**
	(0.00291)	(0.00330)	(0.00606)
High School Location* Roommates	0.00431	0.00612	0.00469
	(0.00596)	(0.00676)	(0.0120)
Home Province* Roommates	-0.00184	-0.00280	0.000166
	(0.00607)	(0.00699)	(0.0125)
Age* Roommates	0.00467	0.00858	-0.00598
	(0.00601)	(0.00668)	(0.0121)
Race* Roommates	-0.0203	-0.00689	-0.0303
	(0.0123)	(0.0151)	(0.0243)
Individual Learning			
Roommates	0.0108***	0.0108***	0.00974
	(0.00269)	(0.00305)	(0.00570)
High School Location* Roommates	0.000717	-0.00232	0.0116
	(0.00548)	(0.00614)	(0.0114)
Home Province* Roommates	-0.00396	-0.00672	0.00928
	(0.00543)	(0.00608)	(0.0112)
Age* Roommates	0.00731	0.00721	0.000549
	(0.00548)	(0.00627)	(0.0113)
Race* Roommates	-0.00190	0.000508	-0.0186
	(0.0115)	(0.0132)	(0.0239)
Forum Participation			
Roommates	0.0203***	0.0205***	0.0194***
	(0.00291)	(0.00330)	(0.00577)
High School Location* Roommates	0.00277	0.00409	0.00369
	(0.00554)	(0.00663)	(0.0113)
Home Province* Roommates	0.00273	-0.000531	-0.00113
	(0.00579)	(0.00667)	(0.0116)
Age* Roommates	0.00288	0.00683	-0.00262
	(0.00563)	(0.00663)	(0.0116)
Race* Roommates	-0.00893	0.00795	-0.0160
	(0.0112)	(0.0136)	(0.0222)
N	38412	38412	38412

Standard errors in parentheses; Bootstrapped Standard Errors; Also include in models are covariates, common shock variables, and random control variables in Table 5 column 4; *p<0.05, ** p<0.01, *** p<0.001

Table 8: Multinomial Logistic Model; Dependent is Four Effort Categories; Odds Ratio Presented

	Overall Effort		
	Both Low	Both Middle	Both High
Roommates	1.317***	1.137*	1.326***
	(3.329)	(2.490)	(3.725)
	Individual Learning		
	Both Low	Both Middle	Both High
Roommates	1.254**	1.0854	1.239**
	(2.776)	(1.586)	(2.719)
	Forum Participation		
	Both Low	Both Middle	Both High
Roommates	1.392***	1.237***	1.478***
	(4.002)	(4.269)	(4.981)
Random controls	Yes	Yes	Yes
Group fixed effect	Yes	Yes	Yes
Teacher fixed effects	Yes	Yes	Yes
% of shared credits	Yes	Yes	Yes
N	38412	38412	38412

T-statistics in parentheses; Bootstrapped Standard Errors; Also include in models are covariates in Table 5 column 4; *p<0.05, ** p<0.01, *** p<0.001

Table 9:Mediation Effect; Dependent is Academic Performance Similarity (Semester GPA or Course Grade)

	Semester GPA			Course Grade		
Roommates	0.00650***	0.00386*	0.00324*	0.00605*	0.00302	0.00235
	(0.00168)	(0.00160)	(0.00160)	(0.00245)	(0.00239)	(0.00239)
Overall Effort Similarity		0.151***			0.171***	
		(0.00244)			(0.00367)	
Individual Learning Similarity			0.113***			0.121***
			(0.00285)			(0.00430)
Forum Participation Similarity			0.101***			0.119***
			(0.00262)			(0.00394)
Random controls	Yes	Yes	Yes	Yes	Yes	Yes
Group fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Teacher fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
% of Shared Credits	Yes	Yes	Yes	Yes	Yes	Yes
N	38412	38412	38412	38190	38190	38190
R-sq	0.143	0.220	0.221	0.102	0.150	0.150

Standard errors in parentheses; Bootstrapped Standard Errors; Also include in models are covariates in Table 5 column 4; *p<0.05, ** p<0.01, *** p<0.001

Table 10:Mediation Effect By Gender; Dependent is Academic Performance Similarity (Semester GPA or Course Grade)

Semester GPA	Female Pairs			Male Pairs		
Roommates	0.00760***	0.00572***	0.00563***	0.00474	0.00327	0.00205
	(0.00148)	(0.00143)	(0.00144)	(0.00367)	(0.00354)	(0.00357)
Overall Effort Similarity		0.0878***			0.141***	
		(0.00275)			(0.00781)	
Individual Learning Similarity			0.0632***			0.102***
			(0.00322)			(0.00896)
Forum Participation Similarity			0.0646***			0.0872***
			(0.00281)			(0.00862)
N	20630	20630	20630	5118	5118	5118
Course Grade						
Roommates	0.00487*	0.00366	0.00295	0.00831	0.00668	0.00483
	(0.00224)	(0.00219)	(0.00222)	(0.00560)	(0.00547)	(0.00550)
Overall Effort Similarity		0.0885***			0.170***	
		(0.00423)			(0.0121)	
Individual Learning Similarity			0.0732***			0.118***
			(0.00496)			(0.0139)
Forum Participation Similarity			0.0576***			0.118***
			(0.00432)			(0.0133)
N	20598	20598	20598	5038	5038	5038

Standard errors in parentheses; Bootstrapped Standard Errors; Also include in models are covariates, common shock variables, and random controls in Table 5 column 4; *p<0.05, ** p<0.01, *** p<0.001

Table 11:Robust Check for Baseline Model Under Three Alternative Specifications; Dependent is Effort Similarity

	Specification One	Specification Two	Specification Three
Roommates	0.0184***	-0.0881***	0.00616***
	(0.00269)	(0.0143)	(0.00124)
Both From the Same Province	0.00366*	-0.0162	-0.00271***
	(0.00157)	(0.00847)	(0.000697)
Both Former High School Graduates	0.000379	-0.000201	-0.00327
	(0.00325)	(0.0174)	(0.00181)
Both Fresh High School Graduates	-0.00444**	0.0233**	-0.000797
	(0.00159)	(0.00878)	(0.000686)
Both Minority	-0.00994	0.0375	0.00556
	(0.0227)	(0.115)	(0.00584)
Both Majority	0.00311	-0.0185	-0.00320*
	(0.00307)	(0.0168)	(0.00128)
Both Science Track in High School	-0.00728**	0.0315*	0.000796
	(0.00246)	(0.0131)	(0.000955)
Both Humanities Track in High School	-0.00474	0.0321*	0.00130
	(0.00261)	(0.0136)	(0.00110)
Both from the Same Major	-0.0119***	0.0823***	-0.000832
	(0.00275)	(0.0152)	(0.00128)
First Sem Std GPA Difference<0.1	0.0118***	-0.0628***	0.00414***
	(0.00146)	(0.00815)	(0.000670)
CEE Percentile Difference<0.1	0.00948***	-0.0537***	0.0000352
	(0.00233)	(0.0126)	(0.000915)
Age Difference<0.5	0.00233	-0.0122	0.00401***
	(0.00153)	(0.00836)	(0.000663)
Both Attend High School in Urban Cities	0.000508	-0.0119	0.00133
	(0.00170)	(0.00964)	(0.000700)
Both Attend High School in Rural Areas	0.00903***	-0.0463***	-0.00161
	(0.00200)	(0.0112)	(0.00107)
Constant	0.752***	1.253***	0.943***
	(0.00973)	(0.0496)	(0.00553)
N	47497	38412	38412
R-sq	0.066	0.066	0.061

Standard errors in parentheses; Bootstrapped Standard Errors; Also included are common shocks and random control variables in Table 5 column 4*p<0.05, ** p<0.01, *** p<0.001

Figure 1: Relationship Between Cohort, Social Group, and Dormitory Room

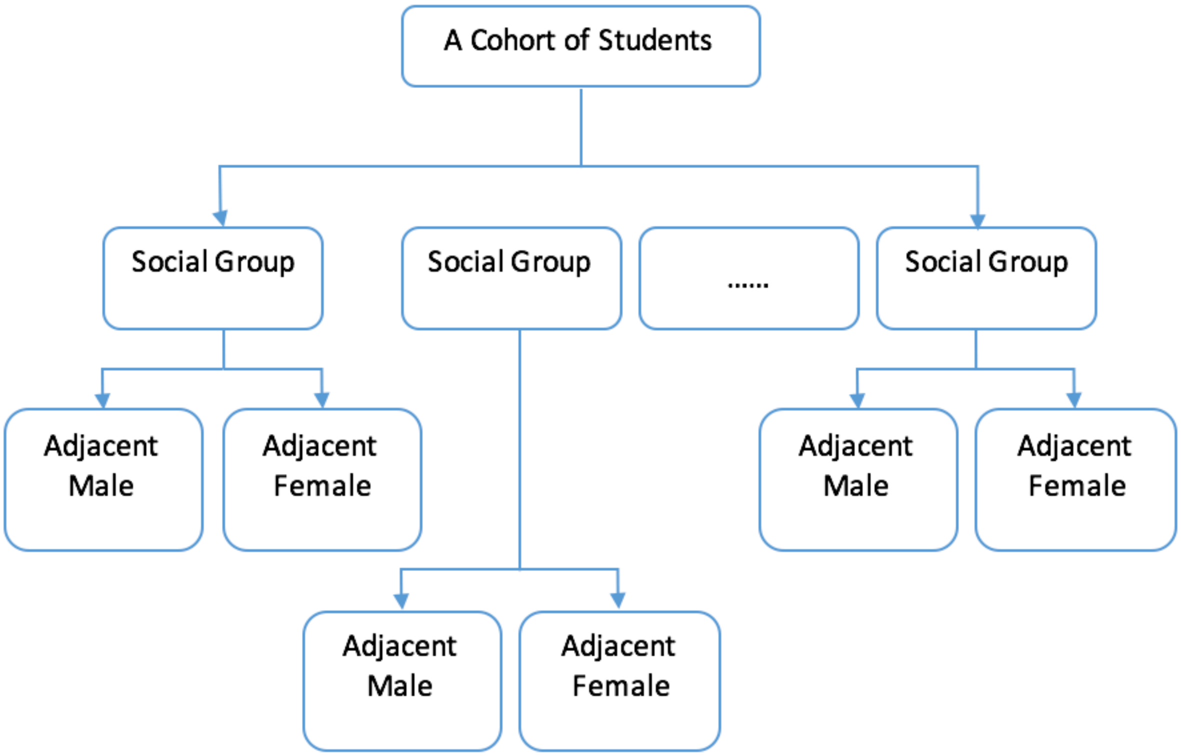

Figure 2:Histograms for Raw Effort Indicators

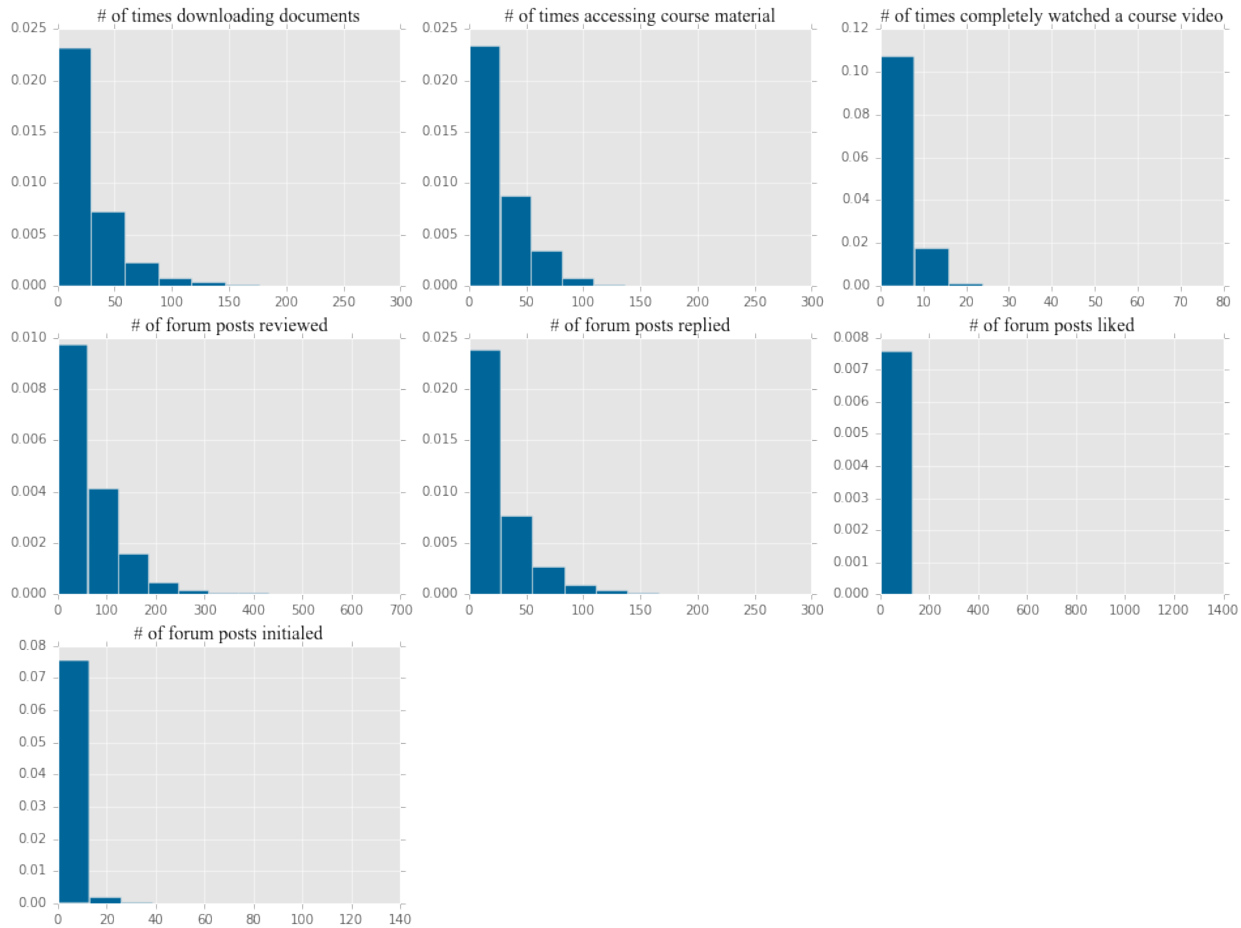

Figure 3:Distribution of the Effort Similarity between a Pair of Students

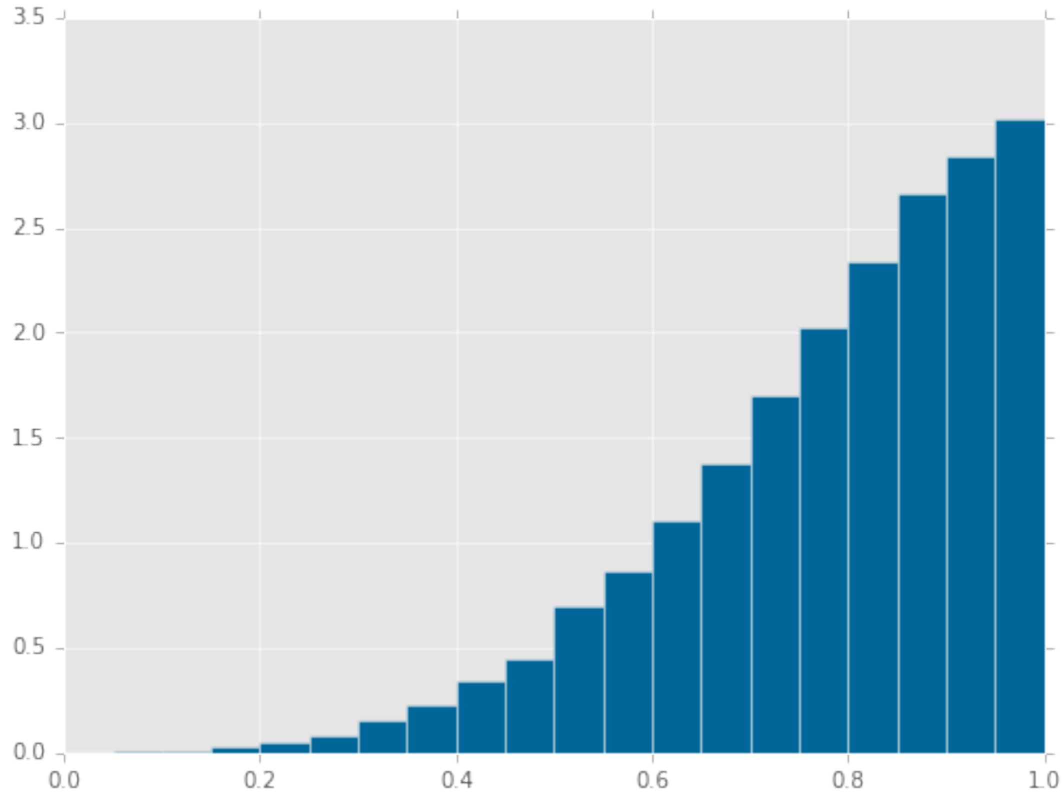

Figure 4:Simulated Number of Rooms Hosing Multiple Same High School Students