Peer Effects
in Academic Performance

Ryohei Hayashi

September 2016

Peer Effects in Academic Performance[*]

Ryohei HAYASHI[†]

September 20, 2016

Abstract

This paper deals with the data of dormitory students in National Institute of Technology, Kagoshima College to demonstrate the existence of peer effects in academic performance. The data have unique advantages to avoid the difficulties of the self-selection problem and reflection problem. The data shows freshmen's academic performance and previous year's junior high school records, and roommate's previous year's academic performance for using an instrumental variable method. The results of my findings suggest that peer's academic performance does not have any effects on freshmen's. In spite of considering the asymmetric relationship between roommates, self-selection bias when choosing subjects, and nonlinearities of effects, there is no significance in any models.

JEL classifications code: J13, J44, C26.

Keywords: learning, reflection problem, instrumental variable method, self-selection bias.

1 Introduction

Does dormitory students' academic performance vary with their peer roommates? No one can have any choice but to accept the given conditions in which they were born: sex, height, inherent abilities, family background, race, parental income and many others. However, a peer is one of the sparse selective options. You can choose your peers by yourself, and also your parents, teachers, matrons of dormitories and managers of workplaces can arrange an opportunity to let you choose your peers by assigning the best peers next to you (Carrell, et al., 2013).

If your performance depends on your peers' performance, you may be more careful to choose your peers. The hypothesis that peers' performance affects one's performance has been defined as peer effects. People, especially parents and teachers, believe this hypothesis strongly. Then parents let their children avoid bad company.

Not only parents but also educational researchers have a keen interest in peer effects[1]. School choice, selective admissions, affirmative action, distance learning, class disruption and many other matters may be affected by peer effects. Educational researchers understand peer effects as the key issue to solve the problem of equality of education opportunity. It is because of Coleman Report (1966). While he was attempting to describe the relation of achievement to school characteristics, Coleman unexpectedly found out the fact that "a pupil's achievement is strongly related to the educational backgrounds and aspirations of the other students in the school[2]." After this report, numerous research papers which mention peer effects were published (Foster, 2003; Kremer and Levy, 2003; Sacerdote, 2001; Stinebrickner and Stinebrickner, 2005; Winston and Zimmerman, 2004; Zimmerman, 2003). These papers discuss whether peer effects are statistically significant or not. If they exist, then peer effects can be used as the corrective action effectively for gender, family income, or race disparity.

Peer effects are also important issues in labor economics studies. Economists are interested in the sign of peer effects, positive or negative, and also in the magnitude of peer effects. Economists understand peer effects as one kind of externality. Market with externality will fail. Policy makers, including classroom teachers, matrons of dormitories and managers of workplaces, have to fix that failure to put the effective equilibrium into practice. If peers' performance affects others' performance positively, policy makers need to support peers by giving some positive incentive like praising peers, publishing peers' svirtues, giving special treatment to assign a single room, or giving money. Otherwise, if peers' performance affects others' performance negatively, policy makers need to punish peers by a negative incentive like imposing tax, suspension from school, summoning parents, or giving

---

[1] See Sacerdote (2011)'s survey.
[2] Coleman (1966) also pointed out as follows:

> An important part of a child's school environment consists not of the physical facilities of the school, the curriculum, and the teachers, but of his fellow-students. A child's fellow-students provide challenges to achievement and distractions from achievement; they provide the opportunities to learn outside the classroom, through association and casual discussions (p.183).

additional homework. Moreover, for the internalization of the spillover of peer effects, the amount of the incentive must be the same as that of the spillover of peer effects. That is the reason why economists are interested in the sign and a magnitude of peer effects. They are also interested in the nature and the process of affecting peer effects. It could be possible for peer effects to affect the result directly, indirectly, or both ways. There are many economists suspecting the nonlinearities in the peer effects (Duncan et al, 2005; Sacerdote, 2001; Hoxby and Weingarth, 2005).

As detailed in Brock and Durlaf (2001), social interaction, which includes peer effects, is roughly divided into two structures. One is *local interaction* where individuals face of one or a small number of appropriately defined peers. The other is *global interaction* where individuals face the mean of a reference group. In school classmate data, there is a tendency to show a strong significant relationship between global interaction models[3]. But on the contrary, there is a tendency to show a weak significant relationship or an insignificant relationship between local interaction models.

Peer effect is a familiar topic, but there are difficulties to demonstrate it. To say "demonstration of peer effect" inevitably connotes demonstrating the causal relationship of peer effect in which peers affect others. There are four well-known difficulties[4]. One is called an endogenous problem[5], in which we could not know whether an individual's outcome is affected by peer's outcome or vice versa from correlated data. Another is called an exogenous problem, in which individual's outcome is affected by peer's background characteristics[6] (e.g. sex, height, family background, race, and parental income). A third is called a contextual effect, in which both individual's outcome and peer's outcomes are simultaneously affected by their shared social and environmental factors (e.g. neighborhood, classroom teacher, school curriculum, social norm). The other is called a selection problem, in which individuals generally self-select peers. To avoid these problems, a few methods have been suggested. Steglich et al. (2010) show the stochastic actor-oriented model using dynamic network data. Others are experimental methods of using random assignment data (Sacerdote, 2001; Falk and Ichino, 2006) or partial assignment of random interventions (Duflo and Saez, 2003; An, 2011). The stochastic actor-oriented model relies on strong parametric assumptions, and the contextual confounding problem of biased network. The experimental methods are problematic when experiments are unethical to conduct randomization.

The results of literature are confusing. Most of the literature insist that there are strong effects of peers (Sacerdote, 2001; Hasan and Bagde, 2013; Angrist and Lang, 2004; Carrell at el., 2009; Rumberger and Palardy, 2005) or small limited effects (Stinebbrickner and Stinebrickner, 2006;

[3] The effects of peer-group characteristics on individual outcomes are easily confounded with the effects of students' unobserved, individual characteristics (Evans et al., 1992; Moffitt, 2001).
[4] An (2015) detailed these problems effectively by using directed acyclical graphs (DAGs) figure.
[5] This problem is also called as a reflection problem.
[6] Betts and Morell (1999) report that there are significant relations between students' grade point averages and their gender, ethnicity, potential income and SAT score.

Zimmerman, 2003). Unlike them, McEwan and Soderberg (2006) and Angrist (2014) show there are no effects.

The Instrumental variable (IV) methods have been argued to be useful as alternatives in identifying endogenous peer effects. The advantage of IV's can be used to obtain consistent estimators in the presence of omitted variables to solve the errors-in-variables problem (Woodlridge, 2015).

There have only been a few studies using IV methods to estimate peer effects (Duncan et al., 1968; Bramoulle et al., 2009; O'Malley et al., 2014; An, 2014; Hoxby, 2000). However, these few studies suggest that the former methods of peer effects studies were overestimated because of unobserved variables influence on the analysis. For example, Angrist (2014) reanalyzed Sacerdote (2011)'s data in the IV method and pointed out that Sacerdote's estimate was likely to be statistically insignificant.

In this paper, peer effects of *local interaction* using dormitory students' data is demonstrated. I use previous year's data, like school records of junior high school or previous year's GPA, to control one's and peers' unobserved ability. And my data have the advantage to avoid a reflection problem because they are limited to use two-pair rooms with one freshman and one sophomore student. Two-pair rooms dataset is also good to define one's peer because a freshman has only a sophomore in his/her room. These controls make it easier to demonstrate clean peer effects of *local interaction*.

2 Empirical Strategy

2.1 Data and Dataset

I use the academic records of dormitory students' in National Institute of Technology, Kagoshima College in Japan. The advantage of this data is that the students are assigned to two-pair rooms with a freshman and a sophomore at random, and also that the students have almost no choice in selecting their subjects.

Randomly assigned data is good for avoiding the self-selection problem. I use only the data of the rooms with two students; one student is a freshman and the other is a sophomore. The dorm warden uses the following protocol in assigning rooms. First, the students must submit application forms and the warden decides to accept or reject before the new semester begins. According to the policy of this college, all the freshmen are supposed to stay in the dormitory[7], which the warden does not reject. Thus, totally around 200 freshmen enter the dormitory every year[8]. Second, the warden makes a decision whether to accept or reject the sophomores by using the data of students' *penalty points list*[9]

---

[7] The requirement for all freshmen to stay in the dormitory is good for avoiding the selection problem. In the case of using college dormitory data, it is possible that dormitory students are potentially biased, for the students who apply for entering the dormitory do not mind staying with other roommates. It follows that dormitory students consist of students who have cooperative characteristics.

[8] As of May 2015, freshman students in the dormitory are 195 (91.5% of all 213 freshman students including 7 repeated students) and include 17 female students (73.9% of 23 female freshman students including 0 repeated students).

[9] The dorm warden records the points of student's life and study attitude. Warden increase the points

of the previous year[10]. Third, the warden assigns one freshman and one sophomore of the same engineering course to one room. Both freshmen and sophomores could not express the preference of their roommate to the warden. There is no possibility of the self-selection problem occurring in this condition. Two-pair assigned data which have a freshman and a sophomore is good for avoiding the reflection problem and the endogenous problem. I use sophomores' records of the previous year as instrumental variables. Freshmen and sophomores were not acquainted with each other in the previous year. Therefore, there is no reason to correlate freshmen's records of this semester and of the previous year, though the sophomores' records of the previous year indirectly affect those of this semester.

The students have no choice in selecting their subjects because the 1st and 2nd grade subjects are fixed by school curriculums except art or music credits[11]. This condition is good for avoiding the self-selection problem. Consider that there is an unusually easy class in which all the students take 100 points. The sophomores took this class in the previous year, tell the freshmen how this class is easy, and then some freshmen will take this class. In this case, freshmen's and sophomores' previous year's average points and GPA will be correlated. This correlation is not from academic performance but the selection of their subjects. The data of this paper do not allow this bias.

This paper deals with the model of *local interaction*. In the *local interaction* model, it is necessary to clarify one's appropriately defined peers (Brock and Durlauf, 2001). The data define peers easily. The freshman's peer is a sophomore student in a room. In the dormitory, the students have to sit at their desks and concentrate on studying from 8 p.m. to 11 p.m. except for 40 minutes' break during the school days, 5 days a week. And no one is allowed to leave the dormitory from 8 p.m. to 6 a.m. These rules make the roommates very close to each other. The freshmen often ask sophomores, their roommates, about the study, because the sophomores already passed that class in the previous year. Thus there are good opportunities to communicate with each other.

The summary of variables for the dataset used is shown in Table 1. The data for the classes of 2008 through 2014 are used. I use only freshmen's and sophomores' data[12]. In the beginning, there are a

---

when the students break the rules. For example, absence from the roll call, bringing prohibited goods, entering the prohibited area, late staying, bringing non-dormmates into the dormitory, stealing, drinking, smoking, violence, leaving the room door unlocked while at school, wasting electricity, incorrect waste segregation, breaking properties, carelessness of fire, etc. In the study attitude, leaving from their own rooms, chatting, dispersion, visiting other students' rooms in the study hour, etc. Dormitory students will be forced to leave the dormitory when these points exceed a standard. And permission to enter the dormitory will be given to students with low points.

[10] As of May 2015, sophomore students in the dormitory are 132 (67.0% of all 197 sophomore students including 3 repeated students) and include 12 female sophomore students (66.6% of 18 female sophomore students including 0 repeated students).

[11] Students can select from Music or Art.

[12] There are freshman students to 5th grade students, and post graduate students in this dormitory. As for over the 3rd grade, 55.9% of all the students are away from the dormitory, and staying at boarding single houses or come from their parents' home. In addition, over the 3rd grade dormitory students can choose their pair students to be group leaders in the dormitory. It could not able to avoid a self-selection problem. Thus, I do not use the data over 3rd grade students.

4

total of 2,275 freshmen and sophomores with 1,527 unique students. And a total of 1,143 pairs with 1,143 unique pairs exist. To set the condition, I remove the rooms whose peer students' data are not obtained (147 cases). Then I remove the single rooms data (7 cases) and the three pair rooms data (110 cases). After that, I also remove the cases of the rooms which have paired with over 3rd grade students (125 cases), the same grade students (180 cases) and repeated students (0 cases). Then next, I remove the cases of the rooms with another engineering courses' students (9 cases). Now the dataset is filled with pure cases of freshmen and sophomores of the same engineering course (565 cases). Finally, I remove the cases of leaving from the dormitory or exchanging pairs in midterm (11 cases). I use 554 two-pair rooms: ones with a freshman and a sophomore and ones where both are in the same engineering course.

Table 2 presents the estimation result of the robustness check. To make sure that freshmen's academic performance is not related with sophomores' before they became roommates, I regress the estimation model as

$$P_i^{PreviousScore} = \alpha_0 + \alpha_1 P_j^{PreviousScore} + \alpha_2 X_i + u_i$$

$$(i=1, \dots, N, j=1, \dots, N).$$

The dependent variable $P_i^{PreviousScore}$ is previous year's school records points of junior high school for freshman i. $P_j^{PreviousScore}$ is the 2-year-ago school records points of junior high school for sophomore j who is a freshman's peer. X_i is the vector of freshman's covariates which include a department dummy and year dummy. And u_i is a random error term. Table 2 shows that there is statically no significance between freshmen's performance and sophomores' before they became roommates. This result means that roommates are assigned randomly.

2.2 Estimation Models

Instrumental variable (IV) methods have been argued to be useful in identifying peer effects. Using IVs is one of the effective ways to cope with omitted variable bias, simultaneity bias, and measurement error (An, 2015). To demonstrate peer effects in the dormitory, I specify the estimation model as

$$P_i = \beta_0 + \beta_1 P_j + \beta_2 X_i + u_i$$

$$(i=1, \dots, N, j=1, \dots, N).$$

The dependent variable P_i is the academic performance for freshman i which was measured as a standardized average score, GPA of the summer semester, or percentile in his/her class. β_0 is a constant variable. P_j is the academic performance of the sophomore student who is a freshman's peer. X_i is the vector of freshman's covariates which include school record points of junior high school, a department dummy value, a dorm building dummy and a year dummy. And u_i is a random error term. The coefficient β_1 is the interested measure of peer effect.

In this model, the regression errors u_i are assumed to be correlated with P_j because of endogenous repressors. This correlation leads to the OLS estimator inconsistent for β. Therefore, I

estimate the model by two stage least squares (2SLS). At the first stage, all the covariates of j (X_j) and j's previous year's academic performance ($P_j^{PreviousYear}$) are used to predict the sophomore student's academic performance (P_j). $P_j^{PreviousYear}$ logically satisfies the assumption that $E(u_i|P_j^{PreviousYear})$=0. At the second stage, the predicted sophomore student's academic performance ($\overline{P_j}$) is used to predict the freshman student's academic performance instead of P_j and X_j. $\bar{P}_j$ satisfies the assumption that $E(u_i|\bar{P}_j)$=0. Shown formally,

the first stage equation is: $\bar{P}_j = \gamma_0 + \gamma_1 P_j^{PreviousYear} + \gamma_2 X_j + v_j$

$$(j=1, \dots, N),$$

where X_j is the vector of sophomore's covariates.

And the second stage equation is: $P_i = \beta_0 + \beta_1 \bar{P}_j + \beta_2 X_i + \mu_i$

$$(i=1, \dots, N, j=1, \dots, N).$$

In summary, I regress freshman's academic performance on sophomore's performance using instruments sophomore's previous year's academic performance.

3   Results

Table 3 reports the results of regression in OLS and 2SLS for four different measures of academic performance. The coefficients of $\bar{P}_j$ are not statistically significant at the 10% level in every measure. And also the signs of coefficients point in inconsistent directions. The larger value in Average Point (AVG), Standardized Average Point (STDAVG), and Grade Point Average (GPA), the higher in academic performance. On the other hand, the larger value in relative rank in his/her class (Percentile), the lower in academic performance. Every coefficient of sophomore's performance is positive in the results of regression. These results suggest there were no peer effects from sophomore to freshman.

The coefficient of peer's performance significantly affects himself/herself in most previous research. For example, in the study of Dartmouth College by Sacerdote (2001), they showed strong evidence for the existence of peer effects in student outcomes. And also Zimmerman (2003) used the data from Williams College, to show the results of suggesting that peer effects are not large, but statistically significant in their models. Against these studies, my data presented in this paper, there were no significant coefficients of peer's performance.

I thought out the reason sophomores' influences are negligible for freshmen. For example, in the case of the sophomores' helps, such as giving the past questions, giving the lecture notes, advice on the evaluated attitude in the class, and information on teachers' hidden classification criteria, they seem not to be used in the year. To make this possibility clear, I divided the subjects into common and uncommon ones. Common subjects are taught both freshmen and sophomores by the same teachers, while uncommon subjects are taught by different teachers, even though they have the same subject titles. The summary of variables is shown in Tables 4 and 5. Table 6 shows that the regression result of 2SLS for common subjects and uncommon subjects. Both common subjects and uncommon

subjects, the coefficients of $\bar{P_j}$ are not statistically significant at the 10% level except uncommon subject's GPA. In the same way as Table 3, the signs of coefficients point in inconsistent directions.

However, the group leader dummy shows a meaningful result. The group leader dummy is a proxy variable of peer's influence power. The group leader plays a part in role-calling once every morning and twice every night. And he/she also gives orientation for the rules of the dormitory, for example, separation of garbage and time schedules. The group leaders can be influential as stronger peers than ordinary students. In Table 6, the coefficients of group leader dummy on AVG, STDAVG, and Percentile are statistically significant at not less than 5% level only about a common subject. This result suggests that group leaders play the role of the middleman to exchange information. To sum up, to be sophomores is negligible on freshmen's performance, but to be group leaders is influential enough to affect freshmen's performance.

Next, I suspect that unobserved variables brought about overestimation because of the reflection problem. Though the student i's performance affects j's, vice versa at the same time. These two phenomena are not distinguishable in view of one-period data. In case i and j affect each other, those variables correlate more closely, compared with the case in which j affects i one way. However, the reverse causation effect of i on j must be omitted to demonstrate peer effects. To do that, additional data, such as j's previous year's performance record as an instrumental variable, are required.

In Scaredote (2001)'s framework, he assumed that i and j's background ability is not measured with error. He did not use instrumental variables and estimated only one-period data. Then, Angrist (2014) pointed out that Sacerdote's estimation possibly had been overestimated. Angrist used additional data from an entrance examination, SAT data, as instrumental variables to control potential ability of j's. According to his re-estimation using data of Dartmouth College, a strong correlation in roommate GPAs seems to be driven solely by common variance components in outcomes. As a result of the IV method, there was statistically no significance in the coefficient of peer effects.

In the same way as Angrist's IV method, I used school record points of junior high school to control freshman's potential ability, and previous year's academic performance to control sophomore student's potential ability in our model. Our case that there was no significance in the coefficient of peer effects supports Angrist's argument. Similarly, at Wellesley College MacEwan and Soderberg (2006) reports there were no peer effects using the additional data of freshmen's SAT scores.

As the third point, the present literature are used as co-worker data. For example, classmates in the same class and roommates of the same grade. However, in my case, the two students are not co-workers, but one is an assistant, and the other is assisted. This asymmetric relationship possibly makes peer effects weaker.

To confirm this hypothesis, I made one more dataset that both roommates are freshmen. According to the same formula shown above, sophomore's performance is changed into peer freshman's school record points of junior high school. Tables 7 and 8 show the summary of variables and results,

respectively. In spite of the hypnosis, there are not significant coefficients of peer effects. The signs of coefficients point in inconsistent directions.

At last, peer effects might affect nonlinearities from peers. McEwan and Soderberg (2006) report that there are no peer effects on students' grade point averages in linear specifications. And there is some evidence that students' SAT scores have nonlinear effects on their roommates' achievement, but the results are not robust.

Then I separate data bottom 25%, top 25% and middle 50% by using data of the entrance examination. To separate categories, I calculate freshman's percentile rank in his/her own class by school record points of junior high school. And also I calculate sophomore's percentile rank in his/her class by previous year's average point. The combined result of the analysis is shown in Table 9. In every case, there were not significant peer effects. I do same analysis in the cases of {5%, 90%, 5%}, {10%, 80%, 10%}, there were not significant peer effects in any case.

4   Conclusion and discussion

Roommate peer effects in the *local interaction* model are not statistically significant in academic performance in this study. Also, nonlinear roommate peer effects were not found in this study. The data used are strictly controlled to demonstrate clean peer effects. And students have good opportunities to affect each other. However, I couldn't find out the evidence of the existence of peer effects. This result does not coordinate with most of the former literature.

It is suspected that the former literature overestimated peer effects because of self-selection problem and reflection problem. Random assigned room data is not enough to avoid the self-selection problem. Because students arbitrarily choose subjects after they are randomly assigned to their rooms. The effect of roommates, when choosing the subjects, can be regarded as one kind of peer effects. However, it does not affect their academic performance but the value of the index to measure their academic performance. My data have the advantage of not contaminating this self-selection bias.

As Anglist (2014) pointed out that it is required to use IV method to avoid the reflection problem. I strictly control the reflection problem using previous year's data. I assert that the analysis of the former literature should be demonstrated again with control the reflection problem.

The data used in this paper was obtained only for a short period of time, one semester (5 months). To elucidate the peer effects, especially measuring learning effects, a longer period of time may be required for investigation.

References

An, W., (2011). "Models and Methods to Identify Peer Effects," In: Scott, John, Carrington, Peter J. (Eds.), *The Sage Handbook of Social Network Analysis*, The Sage Publications, London, 514-532.

An, W., (2015). "Instrumental Variables Estimates of Peer Effects in Social Networks," *Social Science Research*, 50, 382-394.

Angrist, J. D., Lang, K., (2004). "Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program," *The American Economic Review*, 94(5), 1613-1634.

Angrist, J. D., (2014). "The Perils of Peer Effects," *Labour Economics*, 30, 98-108.

Betts, J. R., Morell, D., (1999). "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects," *The Journal of Human Resources*, 34(2), 268-293.

Brock, W. A., Durlauf, S. N., (2001). "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68, 235-260.

Carrell, S. E., Fullerton, R. L., West, J. E., (2009). "Does Your Cohort Matter? Measuring Peer Effects in College Achievement," *Journal of Labor Economics*, 27(3), 439-464.

Carrell, S. E., Sacerdote, B. I., West, J. E., (2013). "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometorica*, 81(3), 855-882.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., York, R. L., (1966). *Equality of Educational Opportunity*, U. S. Department of Health, Education, and Welfare, Office of Education, Government Printing Office.

Duncan, O. D., Haller, A. O., Portes, A., (1968). "Peer Influences on Aspirations: A Reinterpretation," *American Journal of Sociology*, 74 (2), 119-137.

Duncan, G. J., Boisjoly, J., Kremer, M., Levy, D. M., Eccles, J., (2005). "Peer Effects in Drug Use and Sex among College Students," *Journal of Abnormal Child Psychology*, 33 (3), 375-385.

Esther, D., Emmanuel, S., (2003). "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, 118(3), 815-842.

Evans, W. N., Oates, W. E., Schwab, R. M., (1992). "Measuring Peer Group Effects: A Study of Teenage Behavior," *The Journal of Political Economy*, 100(5), 966-991.

Falk, A., Ichino, A., (2006). "Clean Evidence on Peer Effects," *Journal of Labor Economics*, 24(1), 39-57.

Forster, G., (2006). "It's Not Your Peers, and It's Not Your Friends: Some Progress Toward Understanding the Educational Peer Effect Mechanism," *Journal of Public Economics*, 90, 1455-1475.

Hasan, S., Bagde, S., (2013). "The Mechanics of Social Capital and Academic Performance in an Indian College," *American Sociological Review*, 78(6), 1009-1032.

Hoxby, C., (2000). "Peer Effects in the Classroom: Learning from Gender and Race Variation," *NBER Working Papers* 7867, National Bureau of Economic Research, Inc.

Hoxby, C. M., Weingarth, G., (2005). "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects," *Mimeograph*.

Kremer, M., Levy, D., (2008). "Peer Effects and Alcohol Use among College Students," *Journal of Economic Perspectives*, 22(3), 189-206.

McEwan, P. J., Soderberg, K. A., (2006). "Roommate Effects on Grades: Evidence from First-year Housing Assignments," *Research in Higher Education*, 47(3), 347-370.

Moffit, R. A. (2001). "Policy Interventions, Low Level Equilibria, and Social Interactions," In: Durlauf, S. N., and Peyton Y., H. (eds.), *Social Dynamics*, MIT Press, Cambridge, MA, 45-82.

O'Malley, A. J., Elwert, F., Rosenquist, J. N., Zaslavsky, A. M., Christakis, N. A., (2014). "Estimating Peer Effects in Longitudinal Dyadic Data Using Instrumental Variables," *Biometrics*, 70, 506-515.

Rumberger, R. W., Palardy, G. J., (2005). "Test Scores, Dropout Rates, and Transfer Rates as Alternative Indicators of High School Performance," A*merican Educational Research Journal*, 42(1), 3-42.

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

Sacerdote, B., (2011). "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" In: Hanushek, E., Machin, S., Woessmann, L. (Eds.), *Handbook of the Economics of Education*, First edition, 3. Elsevier.

Steglich, C., Snijders, T. A. B., Pearson, M., (2010). "Dynamic Networks and Behavior: Separating Selection from Influence," *Sociological Methodology*, 40(1), 329-393.

Stinebrickner, T. R., Stinebrickner, R., (2005). "What Can Be Learned About Peer Effects Using College Roommates? Evidence from New Survey Data and Students from Disadvantaged Backgrounds," *CIBC Working Paper Series*, Working Paper # 2005-4.

Winston, G. C., Zimmerman, D. J. (2004). "Peer Effects in Higher Education," In: Hoxby, C. (eds.), College Choices: *The Economics of Where to Go, When to Go, and How to Pay for it*, University of Chicago Press, Chicago.

Wooldridge, J. M., (2015). *Introductory Econometrics: A Modern Approach*, 6th Edition, South-Western College Pub.

Yann, B., Habiba, D., Bernard, F., (2009). "Identification of Peer Effects through Social Networks," *Journal of Econometrics*, 150 (1), 41-55.

Zimmerman, D. J., (1999). "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *Williams Project on the Economics of Higher Education DP*-52, Department of Economics, Williams College.

Zimmerman, D. J., (2003). "Peer Effects in Academic Outcomes: Evidence from a Natural Experiment," *The Review of Economics and Statistics*, 85(1), 9-23.

Table 1: Summary of variables.

Variable		Obs.	Mean	Std. Dev.	Min	Max
Year[15]		554	2012	1.69	2009	2014
Dormitory		554	3.41	2.81	1	9
Freshmen	Class	554	2.91	1.37	1	5
	Summer semester average points	554	80.1	6.83	37.6	95.7
	Summer semester standardized average points	554	-0.00403	0.66	-3.29	1.56
	Summer semester GPA	554	1.55	0.39	0	2.48
	Summer semester percent rank in class	554	0.51	0.287	0.0222	1
	School records points of junior high school	554	74	7.59	55	90
Sophomores	Summer semester average points	554	77.8	8.04	48	96.9
	Summer semester standardized average points	554	0.163	0.649	-1.82	1.87
	Summer semester GPA	554	1.39	0.454	0.36	2.58
	Summer semester percent rank in class	554	0.454	0.275	0.0222	1
	School records points of junior high school	554	75	7.87	52	90
	Previous year's average points	554	81.3	6.19	61.2	95.6
	Previous year's standardized average points	554	0.12	0.62	-2.11	1.52
	Previous year's GPA	554	1.61	0.372	0.56	2.57
	Previous year's percentile rank in class	554	0.46	0.283	0.0233	1
	Group leader dummy (leader to 1, otherwise 0)	554	0.00361	0.06	0	1

---

[15] N.B. The whole data is used from 2008 to 2014. The data in 2008 are used as instrumental variables. This summary is analyzed intended for freshmen. Therefor, the data in 2008 is not shown in this summary, because of a time lag.

Table 2: The result of regression before freshman and his/her peer became roommates.

	Freshman's School records points of JH
Sophomore's School records points of JH	0.047
Control	
Department	yes
Year	yes
constant	68.361 ***
N	554
r2	0.088

Standard errors in parentheses (* p<0.1, ** p<0.05, *** p<0.01) , cluster

robust by year, class.

Table 3: The results of regression using four measures of academic performance.

	OLS				2SLS			
Variable	AVG	STDAVG	GPA	Percentile	AVG	STDAVG	GPA	Percentile
Freshman								
School records points of JH	0.45 ***	0.0472 ***	0.0426 ***	-0.0189 ***	0.45 ***	0.0472 ***	0.0426 ***	-0.0189 ***
Sophomore								
Performance	0.0332	0.0254	0.0308	0.0246	0.0309	0.0296	0.0299	0.0296
Group leader dummy	-2.66	-0.217	-0.243	0.153 ***	-2.59	-0.211	-0.242	0.153 ***
Control								
Department	Yes	yes	yes	yes	yes	yes	yes	yes
Dorm building	yes	yes	yes	yes	yes	yes	yes	yes
Year	yes	yes	yes	yes	yes	yes	yes	yes
Constant	42.5 ***	-3.44 ***	-0.828 **	1.87 ***	42.7 ***	-3.45 ***	-0.817 **	1.87 ***
N	554	554	554	554	554	554	554	554
r2	0.324	0.285	0.34	0.239	0.325	0.285	0.34	0.239

Standard errors in parentheses (* $p<0.1$, ** $p<0.05$, *** $p<0.01$), cluster robust by year, class

Table 4: Summary of variables calculated about only common subjects.

Variable		Obs.	Mean	Std. Dev.	Min	Max
Freshmen	Summer semester average points	554	80.4	6.73	35.8	96.5
	Summer semester standardized average points	554	-0.00543	0.679	-3.38	1.85
	Summer semester GPA	554	1.63	0.394	0	2.56
	Summer semester percent rank in class	554	0.509	0.288	0.0222	1
Sophomores	Previous year's average points	554	81.4	6.2	58.5	95.1
	Previous year's standardized average points	554	0.125	0.658	-2.47	1.76
	Previous year's GPA	554	1.68	0.368	0.75	2.56
	Previous year's percentile rank in class	554	0.455	0.28	0.0233	1

Table 5: Summary of variables calculated about only uncommon subjects.

Variable		Obs.	Mean	Std. Dev.	Min	Max
Freshmen	Summer semester average points	554	80.2	8.47	42	98.8
	Summer semester standardized average points	554	0.00751	0.742	-3.05	1.67
	Summer semester GPA	554	1.48	0.469	0	2.71
	Summer semester percent rank in class	554	0.501	0.289	0.0222	1
Sophomores	Previous year's average points	554	81.3	7.36	54.8	96.8
	Previous year's standardized average points	554	0.115	0.641	-1.74	1.49
	Previous year's GPA	554	1.59	0.443	0.4	2.75
	Previous year's percentile rank in class	554	0.468	0.283	0.0233	1

Table 6: The results of regression separated for common and uncommon subjects using four measures of academic performance.

Variable	Common Subject (2SLS)				Uncommon Subject (2SLS)			
	AVG	STDAVG	GPA	Percentile	AVG	STDAVG	GPA	Percentile
Freshman								
School records points of JH	0.436 ***	0.0473 ***	0.0435 ***	-0.0184 ***	0.511 ***	0.0476 ***	0.0279 ***	-0.0171 ***
Sophomore								
Performance	0.0131	0.0295	0.0219	0.0224	0.0661	0.0448	0.0835 *	0.0127
Group leader dummy	-3.85 **	-0.304 ***	-0.281	0.135 ***	1.83	0.038	0.155	-0.0945
Control								
Department	yes	yes	yes	yes	yes	yes	yes	yes
Dorm building	yes	yes	yes	yes	yes	yes	yes	yes
Year	yes	yes	yes	yes	yes	yes	yes	yes
constant	45.3 ***	-3.48 ***	-0.6 *	1.83 ***	34.9 ***	-3.46 ***	-1.46 ***	1.72 ***
N	554	554	554	554	554	554	554	554
r2	0.301	0.269	0.323	0.219	0.334	0.239	0.353	0.204

Standard errors in parentheses (* $p<0.1$, ** $p<0.05$, *** $p<0.01$), cluster robust by year, class.

Table 7: Summary of variables of freshman-freshman pair rooms data.

Variable		Obs.	Mean	Std. Dev.	Min	Max
Year		290	2011	1.62	2009	2014
Dormitory		290	3.56	2.75	1	9
One self	Class	290	3.02	1.43	1	5
	Summer semester average points	290	79.5	7.09	40.6	93.2
	Summer semester standardized average points	290	-0.0806	0.692	-3.08	1.49
	Summer semester GPA	290	2.52	0.657	0	3.83
	Summer semester percent rank in class	290	0.548	0.286	0.0238	1
	School records points of junior high school	290	73.2	8.24	28	90
Peer	Same as above.					

Table 8: The results of regression using freshman-freshman rooms data.

Variable	OLS				2SLS			
	AVG	STDAVG	GPA	Percentile	AVG	STDAVG	GPA	Percentile
Freshman								
School records points of JH	0.456 ***	0.0466 ***	0.0427 ***	-0.0188 ***	0.456 ***	0.0465 ***	0.0426 ***	-0.0187 ***
Peer Freshman								
School records points of JH	0.00317	-0.00297	0.00182	0.000912	0.00696	-0.0636	0.0426	-0.0486
Control								
Department	yes	yes	yes	yes	yes	yes	yes	yes
Dorm building	yes	yes	yes	yes	yes	yes	yes	yes
Year	yes	yes	yes	yes	yes	yes	yes	yes
constant	44 ***	-3.18 ***	-0.954	1.82 ***	43.7 ***	-3.39 ***	-0.913	1.91 ***
N	290	290	290	290	290	290	290	290
r2	0.338	0.313	0.349	0.306	0.339	0.305	0.356	0.301

Standard errors in parentheses (* p<0.1, ** p<0.05, *** p<0.01), cluster robust by year, class.

Table 9: Interaction between one's own background and roommate's background. (OLS)

Own academic index[17]	Sophomore's academic index[16]		
	Bottom 25%	Middle 50%	Top 25%
Bottom 25%	-0.306	0.349	-0.572
Middle 50%	-0.561	-0.083	0.272
Top 25%	-0.329	0.0269	-0.35

Standard errors in parentheses (* p<0.1, ** p<0.05, *** p<0.01),

cluster robust by year, class.

---

[16] Percentile rank in his/her class by previous year's average point.
[17] Percentile rank in his/her own class by school record points of junior high school.