Reconsidering Optimal Peer Assignments: How do Peer Personality Differences Matter in Knowledge Spillover?[*][†]

Li Li, Eric Mak, Chunchao Wang

May 19, 2019

Abstract

We explored the means of assigning peers among students in a class to facilitate knowledge spillover and found that the result depends on how personalities differ among the relevant peers within a designated peer group. In a field experiment held within Chinese classrooms, we construct a measure of non-cognitive peer difference between a fixed pair of deskmates as peers. Only after this new measurement was considered, optimizing peer assignments within classes yields a $2-6\%$ gain in academic achievement after a semester. Because assimilation between peers may fail, reassignments over time are necessary. Class teachers in China, using peer assignments to manage their large classes, have such non-cognitive concerns.

Keywords: Peer Assignment; Academic Peer Effects; Knowledge Spillover; Non-Cognitive Peer Difference; Class Size

JEL Code: I2, I24, J24

[*]Li: lli.li@utoronto.ca; Mak: honam.mak@utoronto.ca; Wang: twangcc@jnu.edu.cn.

[†]All three authors contribute equally to this work. We sincerely thank Robert McMillan, Philip Oreopoulos, Shouyong Shi, and Aloysius Siow for their invaluable advice. We also benefit much from the comments and suggestions by Richard Freeman, Kala Krishna, Elisa Gianonne, Qi Li, Ruilin Zhou, and the seminar participants in University of Guelph, University of Toronto, Brock University, Harvard University, and Penn State University. In China, where our field experiment was held, we would like to thank Haoren He, Fangwen Lu, and Le Wang for their comments. This paper was previously circulated under the title ”Local Peer Group Optimization: Seat Rearrangements within Large Chinese Classrooms.”

1 Introduction

In classrooms, a low-achieving student could benefit from having a high-achieving student as his peer; if this is the case, could a teacher assign *who peers with whom* among her students to facilitate such knowledge spillover?[1] This *peer assignment problem* (Bhattacharya, 2009; Carrell et al., 2013; Fruehwirth, 2014; Garlick et al., 2018) involves only the reallocation of existing resources—the students to be assigned—and not investments for additional teachers and classrooms.

Owing to its low implementation cost, the peer assignment problem attracts economists to formalize it as follows: a planner assigns peers among students, according to their baseline scores, to optimize an average of their final scores; for any candidate peer assignment, the average final score that can be achieved is forecasted using an estimated academic peer effect model. If successful, this formalization could offer quantitative advice on how peer assignments should be performed and what their potential gains would be. Against expectations, the forecasts used could not correctly predict the outcome of peer assignments implemented by either economists (Carrell et al., 2013) or policy makers (Garlick et al., 2018). In both studies, the designated peer groups—a military squadron or a class—have many students; within each of these designated peer groups, the selection of the truly relevant peers is free and hence endogenous. Both studies found that the pattern of peer selection within each designated peer group also changes because the peer assignments cause changes in the group compositions; such *within-group sorting* invalidates the forecasts used for these peer assignments.[2]

[1]See Hoxby and Weingarth (2005) and Sacerdote (2011) for reviews of academic peer effect models, which are used for estimating knowledge spillovers within classrooms.

[2]Specifically, based on the peer effect estimates obtained from a previous experiment, Carrell et al. (2013) first attempted to optimize the peer assignment of a subsequent cohort of students in a military college; the low-achieving students in the optimized peer groups performed worse than the counterpart in a randomized assignment benchmark. Carrell et al. (2013) reported that "We provide evidence that within our optimally designed peer groups, students avoided the peers with whom we intended them to interact and instead formed more homogeneous subgroups." In Garlick et al.

The above is an instance of the Lucas Critique (Lucas, 1976) in the context of peer effects—reduced-form estimates would change upon the implementation of policies. As concluded by Carrell et al. (2013), "Social processes are so rich and complex that one needs a deep understanding of their formation before one can formulate 'optimal policy'." To this end, a possible solution is to try tracing how the *relevant* peers interact with each other in a microscopic sense.

We offer an attempt as follows. First, our field experiment is held in a setting in which a designated peer group is a *fixed* pair of deskmates within a class, rather than a macroscopic group as in the studies of Carrell et al. (2013) or Garlick et al. (2018); by reducing the designated group size to a mimimum of two, we eliminate within-group sorting. In these classrooms, the seating plan layout and regulations also restrict interactions across the designated peer groups. Second, as our hypothesis, knowledge spillover depends on how peer relationships manifest between a pair of deskmates. In the construction of an operational proxy of peer relationships, we measure how one's own and his peer's Big Five personality measurements differ as an index, which we label as *non-cognitive peer difference* hereafter; we trace its change over time by using surveys. Third, following our predecessors, we conduct peer randomization in our field experiment to eliminate selection bias. With these three means, we estimate a *microscopic* academic peer effect model—one that describes knowledge spillover as a function not only of own and peer baseline scores, but also of non-cognitive peer difference as well. Our model moves one step forward towards capturing a structural peer relationship which is policy-invariant, thereby addressing the Lucas Critique. We use the resulting peer effect estimates to evaluate the peer assignment problem again.

(2018) where the policy-maker is in charge of peer assignments (ability tracking), the estimated peer effect model understate negative treatment effect of ability tracking among low-achieving students. Garlick et al. (2018) reported that "spatial proximity generates peer effects only when students are also socially proximate and likely to interact."

Since the seminal work by Sacerdote (2001), peer randomization has been adopted for the elimination of peer selection based on *own* unobservables, such as one's non-cognitive skill in attracting particular peers.[3] However, we show that *peer* unobservables, represented by a peer error term in our academic peer effect model, would still confound the peer effect estimates despite peer randomization. Most experimental peer effect studies since Sacerdote (2001) interpreted the peer error term as a classical measurement error that is orthogonal to the observables.[4] As we propose that peer relationship matters in knowledge spillover, we have to measure it or else it enters the peer error term, so that orthgonality can no longer be guaranteed.[5]

Our measurement of non-cognitive peer difference is based on the skill formation literature (Cunha and Heckman, 2007, 2008), which finds a strong complementarity between one's own cognitive and non-cognitive inputs among children.[6] Many economists have adopted psychological (non-cognitive) measurements as education inputs (see the review by Almlund et al. (2011)). We add to their results by considering the difference of these psychological measurements among peers as another education input.[7] A recent work by Zarate (2019) finds that having so-

[3]See, for instance, Carrell et al. (2009); Ammermueller and Pischke (2009); Duflo et al. (2011); Booij et al. (2017); Zarate (2019) among others for this experimental approach. To address this selection bias, there are other studies which rely on administrative data. For instance, Burke and Sass (2013) estimate peer effects using all public school students in Florida; Carrell et al. (2018) find long-run impacts in the labor market from disruptive peers during school years.

[4]Angrist (2014) shows that when both own and peer observables (the baseline scores) are subject to measurement errors, peer effect estimates can be biased in any direction. Though if the peer assignment is random, Feld and Zölitz (2017) show that only standard attenuation bias remains.

[5]Fruehwirth (2014) discussed an alternative strategy of substituting out this peer unobservable by the peer endogenous outcome, finding that it is generally invalid for policy purposes. Her discussion did not presume an experimental environment with peer randomization. Our derivations highlight that the lost of identification happens even under peer randomization, and explicitly illustrate how this issue affects the results of the peer assignment problem.

[6]See also Heckman and Rubinstein (2001); Heckman et al. (2006); Cunha et al. (2010); Heckman and Mosso (2014); García et al. (2016); Conti et al. (2016).

[7]The peer effect literature often find large non-cognitive related peer effects, with

ciable peers only helps improve a secondary school student's social skills, but not academic outcomes; but for our younger primary school students, non-cognitive peer difference strongly interacts with peer baseline score in determining a student's academic achievement in our peer effect regressions—as we found, a high-achieving peer could cause harm rather than good to a low-achieving student if their non-cognitive peer difference is relatively large.[8] Only after this interaction term is considered, optimizing the microscopic peer assignment within each class yields a 2% to 6% gain in academic achievement within one semester, depending on the specific subject and class; without considering non-cognitive peer difference, the final scores achieved by an assignment entirely based on baseline scores, is nevertheless indistinguishable from a random assignment benchmark.

We then proceed to characterize non-cognitive peer difference in the optimal assignments. As we found, the optimal assignments generally reduce non-cognitive peer difference relative to that of the randomized benchmark; we also find that a planner would need to adjust the peer assignment over time; otherwise, our evaluation shows that at least half of the gains obtained from a previous optimization of the peer assignment would be lost. The reason is that non-cognitive peer difference is not stable or converging over time as one may expect. As we found, assimilation—a decrease of non-cognitive peer difference over time—may not necessarily succeed; about half of the deskmates did not assimilate but differentiate instead.[9] Upon further investigations, we find

respect to domestic violence (Carrell and Hoekstra, 2010) and criminality (Bayer et al., 2009; Damm and Dustmann, 2014) for instance. See also Murphy (2019) for an application of non-cognitive peer effects in the military.

[8] Specifically, we find that having high-achievers as peers leads to a positive academic peer effect (an elasticity of about 0.1) if and only if non-cognitive peer difference is relatively small (1 SD below mean); a negative academic peer effect of comparable magnitude is found if the non-cognitive peer difference is relatively large (1 SD above mean).

[9]At the macroscopic level, economists have found evidence of immigrants assimilating to the local workforce, as reflected by their respective earnings distributions converging to each other (Chiswick, 1978; Borjas, 1985, 1995, 2015; Abramitzky et al.,

systematic assimilation patterns across gender and age. First, female-female pairs assimilate better than other gender pairs, thereby offering a potential explanation to female peer effects within the classroom (Lu and Anderson, 2015) and within schools (Lavy and Schlosser, 2011).[10] Second, younger, lower grade peers assimilate better than their older, higher grade counterparts. This finding agrees with the skill formation literature that early inputs are more effective than later inputs.

Upon assimilation failure, a student would be 17% less likely to report his deskmate as a friend after a semester, given that he has done so in the baseline survey. This result supports our use of non-cognitive peer difference—a measure of psychological differences between peers—as a proxy of peer relationships in the peer assignment problem.

Our last question is whether microscopic peer assignments can be put to actual practice. In the field site of our experiment—primary school classrooms in China—it is common to have more than 40 students constitute a class; some largest classes have a size of more than 60. Such large class sizes are common in China due to her rapid urbanization and lack of education resources. As the "Bad Apple Principle" (Lazear, 2001) states, local disruptions—possibly due to assimilation failure between peers—would increase with the class size. Since reducing class size[11] is difficult, as a countermeasure most classrooms in China are managed by having the seats assigned by "class teachers;" each class teacher would recurrently adjust her peer assignment microscopically for the resolution of peer conflicts between a pair of deskmates in a class. In order to understand this traditional routine, we organized our class teachers to reshuffle their seating plans at will after our main experiment. We found that the class teachers who reduced non-cognitive peer differ-

2014). Economists also discuss cultural assimilation in various contexts (Lazear, 1999; Meng and Gregory, 2005; Abramitzky et al., 2016).

[10] Our finding concerning assimilation between female peers also agrees with a recent literature that finds that female teachers are more effective in promoting academic and non-cognitive outcomes of female students (Gong et al., 2018).

[11] See, for instance, Hanushek (1998); Krueger (1999, 2003); Chetty et al. (2011).

ence more upon the reshuffle were more successful in raising academic achievements. Using this data, we also found that an old deskmate, separated by the reshuffle, would no longer exhibit peer effects. Together with insignificant estimates of general neighbor peer effects in our main peer effect regressions, we can ensure that peer effects are local within deskmates as our designated peer group.

Taken together, these pieces of evidence suggest that bringing down the peer assignment problem to a microscopic level can help understand the issue; the Chinese classrooms, with a pair of deskmates as a well-defined designated peer group, provide us such an opportunity. In terms of implementation, we note that sometimes macroscopic assignments (such as ability tracking across classes) are being restricted.[12] As our results suggest, microscopic peer assignments could serve as a viable alternative, and indeed they are prevalently employed by the Chinese class teachers as actual policy makers.

The rest of this paper is organized as follows. Section 2 describes the practical context of peer assignments within Chinese classrooms, our field experiment design, and the measurement of non-cognitive peer difference. Section 3 describes our academic peer effect model and the corresponding econometric arugments. Section 4 reports the academic peer effect estimates. Section 5 discusses the results of our optimization attempt. Section 6 reports the assimilation patterns. Section 7 provides a description of the seating plan reshuffle. Section 8 concludes our paper.

[12] Our result relates to a much larger literature on ability tracking which concerns macroscopic peer assignments across classes. Ability tracking is a hotly debated topic in education. See, for instance, Hoxby (2000); Hoxby and Weingarth (2005); Hanushek et al. (2003); Ding and Lehrer (2007); Figlio (2007); Aizer (2008); Carrell and Hoekstra (2010); Figlio and Page (2002); Epple et al. (2002); Fu and Mehta (2018); Garlick et al. (2018). Carrell et al. (2013), whose peer groups are military squadrons, also belong to this case. In China, ability tracking is explicitly banned by law for compulsory education. See Article 22, 57 of the Education Law, People's Republic of China.

2 Context, Field Experiment, and Data

As explained in our introduction, we choose to study academic peer effects in classrooms in China because of their particular relationship to microscopic peer assignments.

In terms of national average, China ranks first worldwide in average class size—a clear outlier with 37 students per class relative to the OECD average of 21. Furthermore, the typical class size in many schools located in the suburban areas of China is substantially above 40. In our sample (to be described in detail below) the class size ranges from 37 to 64.[13] These figures are beyond the class size studied in the previous class size literature. For instance, the well-known STAR program reduces the class size from the standard size of 22 to a smaller size of 15.[14] As discussed in the introduction, such large classes in China present an interesting case that is apparently inconsistent with the "Bad Apple Principle" by Lazear (2001), which motivates the reduction of class size worldwide.

Education researchers have attributed the large classes in Chinese primary schools to several reasons (Jin and Cortazzi, 1998; Teng and Liang, 2012). First, due to the rapid urbanization in China, most ru-

[13]The class size in some locations could be surprising. In a provincial study of the education system in Henan Province (adjacent to Hubei Province in our study), Teng and Liang (2012) found that the class size in a "key-point" (elite) primary school amounted to 133 students in 2011. More generally in the Henan province, approximately 22% of the primary schools have 46-55 students per class; another 25% have 56-65 students per class, and there are 14% with 66 or more students per class. In its extreme, large class sizes could lead to safety problems. A stampede occurred in a primary school in Henan province when students rushed for the toilet before a morning exam: https://www.telegraph.co.uk/news/2017/03/22/deadly-stampede-chinese-school-reports/. It should be mentioned, though, that Henan province represents an extreme case because of its relatively slow economic development.

[14]For other well-known references, Angrist and Lavy (1999) adopted a regression continuity design to study the effect of class size with the upper threshold being 40 students; the cutoff in a similar Swedish study (Fredriksson et al., 2012) is 30. Among developing countries, Urquiola (2006) examined the case of Bolivia, where the class size is around 30 to 40 students.

ral children tend to live in county towns (mostly the county capitals) instead of their surrounding villages.[15] Second, local governments have been consolidating education resources by closing down village primary schools. Third, parents are aware of the better education quality of primary schools located in the county towns and thus sending their children to these county primary schools despite village primary schools are located nearby. Given these push and pull factors, the primary schools located in county towns are typically oversized, and there is not much choice in reducing their size.[16]

Instead of reducing class size, China solves the class management problem through the so-called "class teacher system." A class teacher, who is also the teacher of a particular subject, specializes in managing a class and, in particular, is responsible for designing seating plans.[17] While the actual practice varies, class teachers typically reassign the seats within their classes whenever they observe significant conflicts between a pair of deskmates. Such microscopic adjustments to the peer assignment continue throughout the semester. In our discussion with the class teachers, we identify several common patterns of their peer assignments:

1. Initially, the class teacher cannot easily identify whether deskmates can get along with each other. Therefore, the initial seating plan is often based on the plan used in the previous academic year; otherwise, the students are randomly assigned. Subsequently, class teachers rearrange the seats through trial and error depending on their observations on students.

[15] Since 1978, urban population has increased from 18% to almost 60% by 2015, the period of our study. See Henderson et al. (2009) for a recent general review of China's urbanization.

[16] A recent policy discussion of reducing class size in rural China has been concerned to reduce class sizes of over 100 to a more reasonable number, say, 60. Nevertheless, this target is still far beyond the world avarage.

[17] Given the lack of a direct English equivalent of the concept of class teacher, we adopt our own translation.

2. Seats near a teacher's desk are generally preferred by students. Therefore, parents usually ask teachers to have their children seated in these preferred seats. This finding implies the importance of a student's location in a classroom in general. Consequently, we control for location fixed effects in our regressions.

3. To even out the disadvantages of some seats (some columns are located further away from the teacher's desk), some teachers rotate the seats periodically by rotating columns in pairs. As such, the deskmates pairs are preserved. We do not regard these rotations as seat reassignments in this paper.

Figure 1 shows an example seating plan of a typical class in China. Each classroom is organized as an rectangular array. The number of students within a classroom can be maximized by using a rectangular array arrangement. A rectangular array arrangement can fit in the largest number of students within a classroom, which explains why it is the standard seating arrangement found in most Chinese classrooms. Confusions in seating is minimized by assigning each student to a fixed seat. For easy access to every seat in the classroom, every two columns is separated by an aisle. As such, students sit in pairs, and that each student has a fixed deskmate unless reassigned by a class teacher.

Teacher's Desk

	Column 1	Column 2		Column 4	Column 5		Column 7	Column 8		Column 10	Column 11
Row 1	(1,1)	(1,2)		(1,4)	(1,5)		(1,7)	(1,8)		(1,10)	(1,11)
Row 2	(2,1)	(2,2)		(2,4)	(2,5)		(2,7)	(2,8)		(2,10)	(2,11)
Row 3	(3,1)	(3,2)		(3,4)	(3,5)		(3,7)	(3,8)		(3,10)	(3,11)
Row 4	(4,1)	(4,2)		(4,4)	(4,5)		(4,7)	(4,8)		(4,10)	(4,11)
Row 5	(5,1)	(5,2)	**Aisle**	(5,4)	(5,5)	**Aisle**	(5,7)	(5,8)	**Aisle**	(5,10)	(5,11)
Row 6	(6,1)	(6,2)		(6,4)	(6,5)		(6,7)	(6,8)		(6,10)	(6,11)
Row 7	(7,1)	(7,2)		(7,4)	(7,5)		(7,7)	(7,8)		(7,10)	(7,11)
Row 8	(8,1)	(8,2)		(8,4)	(8,5)		(8,7)	(8,8)		(8,10)	(8,11)
Row 9	(9,1)	(9,2)		(9,4)	(9,5)		(9,7)	(9,8)		(9,10)	(9,11)

Figure 1: A Sample Seating Plan

In our sample of 21 classes in three schools in Hubei, China, we record the x-y coordinates of the seats to identify deskmates. We identify the deskmates of 892 out of 1005 students. The remaining students sit alone so that they do not have deskmates.[18] The classrooms in our sample vary in their dimensions, with some having more rows than the others. The average classroom size has 7 rows and 11 columns (including aisles). Table 1 presents the class-level summary statistics of our sample. The smallest class has a class size of 37, whereas the largest class has a size of 64; the mean class size is 47.9.

[18]For example, the student seating in the second row, fourth column has the coordinates $(2,4)$ in our data-set, while his deskmate has the coordinates $(2,5)$. One aisle occupies a column (e.g., the leftmost aisle after the first and second rows is labelled column 3 (missing)). This setup yields the following simple algorithm to identifying the deskmate of student i:

1. The deskmate shares the same row number $x[i]$ as i.
2. The deskmate is the neighbor in columns, i.e. the mapping is $1 \to 2, 2 \to 1, 4 \to 5, 5 \to 4, 7 \to 8, 8 \to 7, 10 \to 11, 11 \to 10$, with columns $3, 6, 9$ as aisles.

Table 1: Class-Level Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Year of Birth (Class Mean)	21	2007	0.683	2005	2007
Year of Birth (Class Std. Dev.)	21	0.531	0.106	0.380	0.776
Baseline Height (Class Mean)	21	137.400	3.706	130.100	143.800
Baseline Height (Class Std. Dev.)	21	7.658	2.005	3.126	11.470
Proportion of Male	21	0.560	0.062	0.451	0.692
Class Size	21	47.860	6.850	37	64
Number of Rows	21	7.190	1.030	6	9
Number of Columns (including aisles)	21	11.000	0.000	11	11

In this classroom environment, students interact with their deskmates for a substantial amount of time per day. The three primary schools in our sample have similar daily schedules, starting early at around 7 AM with morning recitations, followed by a breakfast at school, and then by regular classes. The first three 40-minute regular classes take place from 9 AM to 11:30 AM with a break in between. After lunch, the fourth and fifth class take place from 2 PM to 4 PM. The next 45 minutes are allocated for extra-curricular activities. During these five regular classes, students stay in their seats and are strictly forbidden to walk around the classroom. Given that most students live far away from their schools, they usually eat their breakfast and lunch in the classrooms in their own seats. This classroom environment provides us an unusual advantage in identifying exactly who interacts with whom that a more flexible peer environment (e.g. a squadron in Carrell et al. (2013), dormitory in Sacerdote (2001)) cannot offer. Specifically, this classroom environment eliminates the possibility of within-group sorting and also prevents communication across the designated peer groups.

In our field experiment, we requested the three schools to randomize the seat arrangements for all their Grade 3,4, and 5 classes.[19] The three

[19] At the beginning of the first semester, we randomize the seating plan of each class. However, unlike Lu and Anderson (2015), we enforce the randomization to the entire class without allowing for preferential treatments as they did. Being aware that teachers and parents might have a tendency to overrule our randomization, our monitoring team regularly invigilated these schools. To this end, we formed a monitoring team

schools provided us with the baseline scores (the score in the semester before our experiment), and the midterm and final scores for each semester. We focused on Chinese Language and Math because the English baseline scores for most (3/4) of the sample are unavailable. These exam scores are critical in determining which high school a student can enter in the future (there is significant sorting across high schools; see e.g. Ding and Lehrer (2007)). As such, these exams can be considered high-stake tests of cognitive ability.

In proxying peer relationships between deskmates, we measure own and peer personalities in a survey. Specifically, we use the Big Five, a widely-used taxonomy that encompasses most of the relevant psychological traits (Costa Jr and McCrae, 1992). The Big Five is measured by a vector of Likert-style questionnaire items (an ordinal score in the range of 1–5). In this paper, we adopt a Chinese version of the Revised NEO Five-Factor Inventory (NEO-FFI) (Costa and MacCrae, 1992), which consists of 60 items selected from a longer version of the questionnaire because of their strong correlations with their associated factor scores. See Almlund et al. (2011) for a review. We use this information to construct our metric of non-cognitive peer difference.

We do not claim that non-cognitive peer difference is the best measure for peer relationships. Instead, with operationality in mind, we note that Big Five questionnaires are relatively easy to systematically administer, so that actual planners can potentially use them to make decisions on peer assignments.

Our approach has some limitations. As an important concern, the answers of our Big Five questionnaire items are self-reported even no obvious reason for mis-reporting is present; there is no stake involved in answering the questionnaire. Perhaps a bigger concern is that given

consisted of our graduate students, and staff from the Hubei Education Bureau. As an administrative order, the teachers in each class are asked to provide the finalized seating plan and take photographs each week during the semester, making sure that the seating plan was followed through. Due to confidentiality, we do not include these photos in this paper but they are available upon request.

that the students are young in age, they may not fully understood the questionnaire items and think about the best answer seriously before answering. To address these concerns, each class teacher guided her students to fill in the questionnaire items in small groups, explaining what each item meant. Surely, we cannot rule out measurement error as a potential issue; we address it in the appendix.

In the survey, we also collected information about students' self-reported friendships, which we shall use for verifying the validity of using the Big Five.

3 Peer Effect Model

3.1 The Standard Academic Peer Effect Model

In this section, we first prove that although peer randomization would eliminate the selection of peers, it cannot distinguish between observable and unobservable peer effects; an identification problem would result. Formally, let $i \in \mathbb{S} = \{1, 2, \ldots, N\}$ index individuals; let $t = 0$ be the baseline period where peer randomization takes place, and let $t = 1$ be the time where the current academic achievement is measured. We denote the academic achievement for individual i at time $t \in \{0, 1\}$ by y_{it}.

A classroom with fixed seats defines a known peer function $p : \mathbb{S} \to \mathbb{S}$ such that $p(i) \in \mathbb{S}$ indicates the deskmate of individual $i \in \mathbb{S}$. We write y_{i0} to represent the baseline test score of individual i, and we write $y_{p(i)0}$ to represent the baseline test score of $p(i)$. Both y_{i0} and $y_{p(i)0}$ are observable to the researcher.

An academic peer effect model can be expressed as follows:

$$y_{i1} = f(y_{i0}, y_{p(i)0}) + u_{i0} + \lambda u_{p(i)0} \tag{1}$$

where u_{i0} is the own unobservable error term, and $u_{p(i)0}$ is the peer counterpart. $f(.)$ is a non-linear function that depends on own and peer baseline score. $\lambda \in \mathbb{R}$ is a parameter capturing the strength of unobservable peer influences. The peer effect model (1) is symmetric in the treatment of observable and unobservable inputs in the sense that own and peer components are both explicitly stated. In most peer effect models, the unobservable peer inputs are subsumed into a correlated effect—a common factor shared by both i and $p(i)$ (see e.g. Manski (1993)). This would be a natural approach when $p(i)$ constitutes of many peers. Here, because the peer relationship is microscopic where $p(i)$ is a particular individual in the class, we naturally introduce u_{i0} and $u_{p(i)0}$ as two separate unobservable inputs to highlight their separate importance.[20]

Peer randomization is important in the estimation of causal peer effects. Peer randomization at $t = 0$ implies that for any measurement, X_{i0} and $X_{p(i)0}$ are i.i.d. draws from the same underlying population X, whether X is observable or not. Therefore,

$$(y_{i0}, u_{i0}) \perp\!\!\!\perp (y_{p(i)0}, u_{p(i)0}) \tag{2}$$

Conceptually, consider the following group comparison:

$$E[y_{i1}|y_{i0} = y^*, y_{p(i)0} = y_H] - E[y_{i1}|y_{i0} = y^*, y_{p(i)0} = y_L] \tag{3}$$

which is to compare, among students having the same own baseline score y^*, the current academic achievement y_{i1} of two comparison groups, namely, 1) those who have peers with a higher baseline score y_H, and 2) those who have peers with a lower baseline score y_L. According to the

[20] As another remark, most of the peer effect literature discusses a static setup. In these static peer effect models, the concurrent influence of peer outcomes on own outcomes leads to the reflection problem (Manski, 1993). Nevertheless, in many academic peer effect papers, e.g. Sacerdote (2001), data on baseline score are available. Consequently, these academic peer effect papers directly consider a one-period lag specification where the main independent variable is peer baseline score instead of peer current academic performance. In this case, the reflection problem does not apply. We follow the same approach in this work. In the context of academic performance, a strictly simultaneous peer effect on current academic achievement—the exam scores—cannot exist, under the assumption that no cheating has taken place during the exams. During the exams, seats are separated from each other to prevent cheating.

peer effect model (1), this between-group comparison groups identifies the following:

$$
\begin{aligned}
& f(y^*, y_H) - f(y^*, y_L) \\
& + E[u_{i0} | y_{i0} = y^*, y_{p(i)0} = y_H] - E[u_{i0} | y_{i0} = y^*, y_{p(i)0} = y_L] \\
& + \lambda E[u_{p(i)0} | y_{i0} = y^*, y_{p(i)0} = y_H] - \lambda E[u_{p(i)0} | y_{i0} = y^*, y_{p(i)0} = y_L]
\end{aligned} \tag{4}
$$

In Equation (4), $f(y^*, y_H) - f(y^*, y_L)$ is the marginal effect of changing the peer baseline score, ceteris paribus. The sum of the two terms in the second line is zero because the own unobservable inputs for the two comparisons groups are balanced due to peer randomization at $t = 0$, which eliminates the selection effect. However, given that the average peer unobservable inputs are not balanced by the peer randomization, the sum of the terms in the third line of (4) is generally non-zero.

We now discuss the consequences of this confounding in the peer assignment problem. For exposition, we first illustrate a simplified version of a microscopic peer assignment problem. Suppose that in a class there are two high-achieving students H_1, H_2 and two low-achieving students L_1, L_2; the policy-maker wants to exploit knowledge spillover by pairing up the two high-achieving students with the two low-achieving students. The two possible assignments, $\{(H_1, L_1), (H_2, L_2)\}$ and $\{(H_1, L_2), (H_2, L_1)\}$, are shown below by solid and dashed arrows:

$$
\begin{tikzcd}
H_1 \arrow[r] \arrow[dr, dashed] & L_1 \\
H_2 \arrow[r] \arrow[ur, dashed] & L_2
\end{tikzcd}
$$

The policy-maker decides between these two feasible peer assignments according to an objective function, typically the average final achievement among the low-achieving students L_1, L_2.

Similar to most welfare problems, some subjectivity is involved in deciding the welfare weights. Policy-makers typically want to place a higher welfare weight on low-achieving students for equality reasons.

In our discussion, we focus only on the low-achieving students as we find that the adverse effects to high-achieving students are almost zero and statistically insignificant. Carrell et al. (2013) considered a similar objective function, although their problem is macroscopic.

In full, we consider the following optimization problem. Let $\mathbb{S}_{low} \subset \mathbb{S}$ be a set of students whose current academic achievement is to be maximized in expected terms, and let $\mathbb{S}_{high} = \mathbb{S} \backslash \mathbb{S}_{low}$ be its complement. $\mathbb{S}_{low}$ is defined as the set of students whose baseline score is below the median. For simplicity, suppose that n is even so that $\mathbb{S}_{low}$ and $\mathbb{S}_{high}$ are equal sized. As such, the problem becomes one that searches for a mapping $q : \mathbb{S}_{low} \to \mathbb{S}_{high}$ that maps a deskmate from $\mathbb{S}_{high}$ to each student from $\mathbb{S}_{low}$. We denote the corresponding space of feasible mappings by $\mathcal{Q}$.

The standard academic peer effect model leads to the following peer assignment problem:

$$\max_{q \in \mathcal{Q}} \frac{1}{n/2} \sum_{i \in \mathbb{S}_{low}} f(y_{i0}, y_{q(i)0}) \tag{5}$$

Note that f must be non-linear. Otherwise, the solution of the peer assignment problem is indeterminate because all assignments $q \in \mathcal{Q}$ would lead to the same objective value.[21]

As discussed above, f cannot be identified due to the confounding problem. However, the optimization algorithm assumes so. To illustrate the consequences of this issue, consider a multiplicative interaction ex-

[21]The ideal peer assignment problem should be one that incorporate the own and peer unobservables, that is:

$$\max_{q \in \mathcal{Q}} \frac{1}{n/2} \sum_{i \in \mathbb{S}_{low}} [f(y_{i0}, y_{q(i)0}) + u_{i0} + \lambda u_{q(i)0}] \tag{6}$$

Given that the unobservable component $\sum_{i \in \mathbb{S}_{low}} [u_{i0} + \lambda u_{q(i)0}]$ is invariant to the choice of the assignment q, it can be eliminated from the objective function, resulting in (5). This result critically relies on the linearity of the unobservable component. Given that the observable component f must be non-linear, such asymmetric treatment to the unobservable component could be rather unnatural if the researcher believes that another skill of importance is within the unobservable component.

ample:

$$f(y_{i0}, y_{q(i)0}) = \beta_0 + \beta_1 y_{i0} + \beta_2 y_{q(i)0} + \beta_3 y_{i0} y_{q(i)0}$$

This is a typical specification used in the empirical literature. In this example, the estimated marginal effect of raising peer baseline score, holding own baseline score fixed at $y_{i0} = y^*$ can be expressed as

$$\beta_2 + \beta_3 y^* + \lambda \frac{\partial \mathbb{E}[u_{p(i)0} | y_{i0} = y^*, y_p]}{\partial y_p}$$

instead of $\beta_2 + \beta_3 y^*$.

If $\beta_3 > 0$, then f is supermodular. Therefore, the true optimal assignment would be positive assortative matching (PAM), i.e. suppose that for $i, j \in \mathbb{S}_{low}$ such that if $y_{i0} > y_{j0}$, then $y_{q(i)0} > y_{q(j)0}$ optimally. This result is proven as follows: if $q(i)$ and $q(j)$ are swapped, the total achievement among the low-achievers would change by $\beta_3(y_{i0}y_{q(j)0} + y_{j0}y_{q(i)0} - y_{i0}y_{q(i)0} - y_{j0}y_{q(j)0})$ which is less than zero. However, given that $\beta_2 + \beta_3 y^* + \lambda \frac{\partial \mathbb{E}[u_{p(i)0} | y_{i0} = y^*, y_p]}{\partial y_p}$ is being identified instead, the "theoretically" optimal assignment would generally be different from the true optimal one unless a strictly increasing function g exists such that

$$g\left(\beta_3 y^* + \lambda \frac{\partial \mathbb{E}[u_{p(i)0} | y_{i0} = y^*, y_p]}{\partial y_p}\right) = \beta_3 y^* \tag{7}$$

for all y^*, that is, the unobservable peer effects would not alter the ordering of peer effects among the low-achieving students. This monotonicity condition is less demanding than exogeneity, i.e. shutting down unobservable peer effects:

$$\lambda \frac{\partial \mathbb{E}[u_{p(i)0} | y_{i0} = y^*, y_p]}{\partial y_p} = 0$$

Nevertheless, this monotonicity condition has no guarantee to be satisfied in the data, thereby likely resulting in a sub-optimal peer assignment.

As an important remark, the above is a microscopic peer assignment model which concerns the peer relationship between two (or more) individuals within a class. By constrast, in a macroscopic peer assignment model, a designated peer group consist of many students. Within such a group, the selection of the truly relevant peer $p(i)$ remains free and hence endogenous, decided by a process which is possibly determined by the unobserved non-cognitive factors of all group members. Because a macroscopic peer assignment changes group compositions, indirectly $p(i)$ would be changed. Ultimately, this within-group sorting results in a change in the reduced-form relationship estimated at the group level. In our paper, peer assignments are done at the microscopic level, so that $q(i)$ (corresponding to $p(i)$ in estimation) is exogenously controlled by the planner. This is essential in addresing the Lucas Critique.[22]

3.2 Introducing Non-Cognitive Peer Difference

We now study the case where the policy-maker can perform peer assignments based on non-cognitive peer difference. The own and peer personalities at time 0 are denoted by $n_{i0}, n_{p(i)0}$ respectively. We define a metric of non-cognitive peer difference which measures the distance between $n_{i0}, n_{p(i)0}$:

$$d_{i0} \equiv |n_{i0} - n_{p(i)0}|$$

In our work, $n_{i0}, n_{p(i)0}$ are measured by Big Five measurements so that they are multi-dimensional. We correspondingly define the metric as that induced by the L_1 norm (sum of absolute distances across items).

[22]We can extend the model to include general neighbor(s). Let $p'(i)$ denote another peer of individual i. Peer randomization at time 0 guarantees that both $p(i)$ and $p'(i)$ are randomly assigned, such that:

$$(y_{i0}, d_{i0}, u_{i0}) \perp\!\!\!\perp (y_{p(i)0}, d_{p(i)0}, u_{p(i)0}) \perp\!\!\!\perp (y_{p'(i)0}, d_{p'(i)0}, u_{p'(i)0}) \tag{8}$$

holds, i.e. the own, deskmate, and general neighbor variables are independent to each other, whether observable or not. The identification analysis is analogous to the above, which we shall not repeat. To keep the notation simple, we continue our discussion while suppressing general neighbor terms.

The resulting peer effect model is:

$$y_{i1} = f(y_{i0}, y_{p(i)0}, d_{i0}) + u_{i0} + \lambda u_{p(i)0} \tag{9}$$

Some discussion concerning functional form is warranted. We construct the metric d_{i0} rather than letting $n_{i0}, n_{p(i)0}$ to enter separately in (9). There are two reasons behind this choice. First, in this paper, we are primarily interested in modelling peer relationships as a determinant of knowledge spillover, which naturally leads to our specification which take d_{i0} as a variable capturing between-peer differences in personality attributes. Second, the Big Five, while being comprehensive, does not yield a univariate ranking. In personality psychology, people are considered "different from each other" rather than being "better or worse" under the Big Five, which covers multiple dimensions.[23] If $n_{i0}, n_{p(i)0}$ enters separately, one immediate question would be whether $f(.)$ should be monotonic with respect to these two *vectors* of arguments. As another remark, peer randomization guarantees that even if n_{i0} as an own variable is not being controlled by the regression, the estimates remain unbiased due to peer randomization.

Analogous to the above, we consider the following peer assignment problem:

$$\max_{q \in \mathbb{Q}} \frac{1}{n/2} \sum_{i \in \mathbb{S}_{low}} [f(y_{i0}, y_{q(i)0}, |n_{i0} - n_{q(i)0}|)] \tag{10}$$

which requires $n_{i0}, n_{q(i)0}$ to be observable to be evaluated. In our empirical section, we adopt the following functional form with three-way

[23]Economists do consider conscientiousness as a skill and indeed it is correlated with individual success, but here we do not emphasize this point.

interactions:

$$
\begin{aligned}
& f(y_{i0}, y_{q(i)0}, |n_{i0} - n_{q(i)0}|) \\
\equiv\ & \beta_0 + \beta_1 y_{i0} + \beta_2 y_{q(i)0} + \beta_3 |n_{i0} - n_{q(i)0}| + \\
& + \beta_4 y_{i0} y_{q(i)0} + \beta_5 y_{i0} |n_{i0} - n_{q(i)0}| + \beta_6 y_{q(i)0} |n_{i0} - n_{q(i)0}| \\
& + \beta_7 y_{i0} y_{q(i)0} |n_{i0} - n_{q(i)0}|
\end{aligned} \tag{11}
$$

Considering non-cognitive peer difference can break down the PAM assignment because changing the assignment $q(i)$ would affect both $y_{q(i)0}$ and $|n_{i0} - n_{q(i)0}|$. Whether the two effects offset or reinforce each other depends on two factors: 1) the joint distribution of peer baseline score and non-cognitive peer differences and 2) how non-cognitive peer difference affects the total peer effect. Either way, we should expect that after considering non-cognitive peer difference, a different optimal peer assignment should result; indeed, this is the case. We explore the corresponding empirical findings in Section 5.

4 Empirical Results

4.1 Academic Peer Effect Estimates

We report our major peer effect regression results in this section. We run ordinary least square regressions of Chinese Language test scores on the own Chinese Language baseline score, deskmate baseline score, and baseline non-cognitive peer difference d_{i0}. We take logs of all test scores, so that the peer effect estimates are elasticities.[24] We also run the corresponding regressions for Math. We are also interested in comparing deskmates and more general neighbors. Therefore, following two previous within-classroom peer effect studies (Lu and Anderson, 2015;

[24] A few students have a score of zero. To avoid taking log of zeros, we add 1 to the scores before taking logs. Alternatively, we drop the zeros. The results from the two approaches are almost identical.

Hong and Lee, 2017), we construct the neighbor-4 measure corresponding to each variable, including the two neighbors in the front and the two neighbors at the back.[25] In this set of regressions, all non-cognitive peer difference variables are standardized to have a zero mean and unit variance. Based on our previous arguments in Section 2, we control for several confounding factors in the peer effect regressions. The list includes class fixed effects, exact location fixed effects, and gender pair fixed effects.

The regression results for Chinese and Math are presented in Tables 2 and 3, respectively. For presentation purposes, here we do not run regressions with full three-way interactions, which are difficult to interpret. Instead, we present three variants of the peer effect regressions that use the whole sample, the observations with own baseline score above and below the median respectively. This reduces three-way interactions to two-way interactions, so that the interactions shown in the tables are those between peer baseline scores $y_{p(i)0}$ and non-cognitive peer difference $|n_{i0} - n_{p(i)0}|$.

Several features emerge from these peer effect estimates. First, the main effects of peer baseline score in all specifications are all almost zero and being statistically insignificant. Due to the normalization of non-cognitive peer difference, the interpretation of these main effects is that if non-cognitive peer difference is at its mean, then there is no peer effect. This is consistent with previous research which typically finds small, statistically insignificant unconditional academic peer effects.

Second, the final achievement of high-achieving students are robust to the assignment of low-achieving peers: their corresponding interaction terms are statistically zero. This is consistent with the idea of knowledge spillover as a one-directional transfer of knowledge from high-achieving students to low-achieving students. This result provides a jus-

[25]For those students seated in the first row or the last row, some neighbors could be missing. Correspondingly, we define the neighbor measures by averaging over the available neighbors.

tification to our objective function of only focusing on the low-achieving students in optimizing the peer assignment.

For low-achieving students, we find much larger and significant interaction effects. For example, take the Chinese final interaction coefficient of -0.10 (for students with own baseline score below the median). This regression coefficient implies that, the elasticity of peer baseline score for low-achieving students with respect to Chinese final is:

- ~ 0.10 if the non-cognitive peer difference is $-1 s.d.$ relative to its mean.
- ~ 0 if the non-cognitive peer difference is at its mean; and
- ~ -0.10 if the non-cognitive peer difference is $+1 s.d.$ relative to its mean.

A similar intepretation can be applied for other interaction coefficients in the tables. To conclude, we find that for low-achieving students, assigning a high-achiever as his peer has opposite signed effects, depending on whether the peer is a close peer or a distant peer in terms of personalities. That is, if a student has a high-achieving and close peer, he would benefit from his peer by getting a higher final academic achievement; but if the peer is high-achieving yet distant in personality, the result could be detrimental.

We have two measurements over time (midterm and final) and two subjects. Over time, the strength of peer interaction increases as expected. The Math interaction terms are larger than those of Chinese in terms of magnitude. One possible reason is that the gain from knowledge spillover is stronger for Math—a student could benefit from his peer who knows how to solve a well-defined mathematics problem, while the benefit of having a high-achieving peer in improving one's language ability is less obvious. And since knowledge spillover depends on non-cognitive peer difference, we should observe larger interaction coefficients for Math. This larger interaction coefficient for Math, however, does not necessarily mean that the planner should weigh more on non-cognitive peer difference in the peer assignment problem for math. We have to actually evaluate the peer assignment problem to draw conclusions.[26]

Our results relate to some well-known classroom peer models (Hoxby and Weingarth, 2005). Among them, the "Boutique model" proposes that having homogeneous peers are beneficial to own achievement, while the "Shining Light" model proposes that high-achieving students set good examples to their low-achieving peers. Generally, the set of scenarios of peer interactions that could be justified by theoretical peer effect models is rather large, so that empirical studies must be performed to find out which among these scenarios are true. The situation could become more complicated when more than one dimension—baseline score and non-cognitive peer difference—needs to be considered. Our results suggest that having a high-achieving student as a peer is beneficial only if she is close to the low-achieving student to be assigned in terms of personalities; as such, our results are consistent with the Boutique model along the non-cognitive dimension, and the Shining Light Model along the cognitive dimension. Furthermore, a high-achieving peer who is also

[26] Suppose that in the optimal assignment, two particular low-achieving students i_1, i_2 has their high-achieving peers being j_1, j_2. As necessary condition for the optimum, reassigning j_2 to i_1 and j_1 to i_2 must yield a lower total output:

$$\begin{aligned}
&[f(y_{i_10}, y_{j_10}, |n_{i_10} - n_{j_10}|) + f(y_{i_20}, y_{j_20}, |n_{i_20} - n_{j_20}|)] \\
&-[f(y_{i_10}, y_{j_20}, |n_{i_10} - n_{j_20}|) + f(y_{i_20}, y_{j_10}, |n_{i_20} - n_{j_10}|)] \\
&\equiv \beta_3(|n_{i_10} - n_{j_10}| + |n_{i_20} - n_{j_20}| - |n_{i_10} - n_{j_20}| - |n_{i_20} - n_{j_10}|) + \\
&+\beta_4(y_{i_10}y_{j_10} + y_{i_20}y_{j_20} - y_{i_10}y_{j_20} - y_{i_20}y_{j_10}) \\
&+\beta_5(y_{i_10}|n_{i_10} - n_{j_10}| + y_{i_20}|n_{i_20} - n_{j_20}| - y_{i_10}|n_{i_10} - n_{j_20}| - y_{i_20}|n_{i_20} - n_{j_10}|) \\
&+\beta_6(y_{j_10}|n_{i_10} - n_{j_10}| + y_{j_20}|n_{i_20} - n_{j_20}| - y_{j_20}|n_{i_10} - n_{j_20}| - y_{j_10}|n_{i_20} - n_{j_10}|) \\
&+\beta_7(y_{i_10}y_{j_10}|n_{i_10} - n_{j_10}| + y_{i_20}y_{j_20}|n_{i_20} - n_{j_20}| \\
&-y_{i_10}y_{j_20}|n_{i_10} - n_{j_20}| - y_{i_20}y_{j_10}|n_{i_20} - n_{j_10}|) > 0
\end{aligned} \tag{12}$$

Clearly, β_7 that corresponds to the interaction term is not the only factor that affects whether this condition holds. Even β_3, which corresponds to a linear rather than an interaction term, would matter because different peer assignments would imply different total non-cognitive distance between peers.

a distant peer can potentially discourage own achievement, consistent with the “Invidious Comparison” model.

It should be noted that the aforementioned classroom peer effect models in Hoxby and Weingarth (2005) concern macroscopic peer effects, whereas our results focus on peer effects that are local between deskmates only. Notably, we do not find any statistically significant result for general neighbors, whose peer effect estimates are an order of magnitude smaller than the corresponding ones for deskmates. This is important in our microscopic peer assignment problem since if the peer effects are global instead, then peer assignments between deskmates would have little or no effect.

Table 2: Peer Effect (Chinese)

	First Sem. Midterm			First Sem. Final		
	(1)	(2)	(3)	(4)	(5)	(6)
Own Baseline	0.767***	0.759***	0.778***	0.607***	0.573***	0.843***
	(0.046)	(0.045)	(0.073)	(0.053)	(0.052)	(0.100)
Deskmate1 Baseline	−0.026	−0.037	−0.019	0.016	−0.020	0.015
	(0.026)	(0.035)	(0.012)	(0.021)	(0.018)	(0.018)
Non-Cog Diff (Deskmate1)	−0.000	0.008	−0.000	0.292***	0.398***	−0.010
	(0.092)	(0.109)	(0.034)	(0.062)	(0.096)	(0.040)
Neighbor1 Baseline	−0.023	−0.082	−0.023	−0.037	−0.053	−0.020
	(0.029)	(0.082)	(0.020)	(0.060)	(0.124)	(0.020)
Non-Cog Diff (Neighbor1)	−0.056	−0.197	−0.022	−0.159*	−0.153	−0.069
	(0.155)	(0.502)	(0.068)	(0.096)	(0.192)	(0.061)
Interaction (Deskmate1)	−0.006	−0.016	−0.001	−0.072***	−0.103***	0.001
	(0.023)	(0.031)	(0.008)	(0.017)	(0.029)	(0.009)
Interaction (Neighbor1)	0.019	0.057	0.006	0.042*	0.047	0.016
	(0.039)	(0.123)	(0.015)	(0.022)	(0.046)	(0.014)
Constant	1.246***	1.550***	1.129***	1.800***	2.203***	0.654
	(0.190)	(0.372)	(0.335)	(0.401)	(0.734)	(0.467)
class FE	Yes	Yes	Yes	Yes	Yes	Yes
location FE	Yes	Yes	Yes	Yes	Yes	Yes
gender pair FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	882	444	452	879	442	451
R^2	0.681	0.695	0.495	0.708	0.737	0.551
Adjusted R^2	0.634	0.596	0.332	0.665	0.652	0.406

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
Interactions: between peer baseline and non-cognitive peer differerence.
Standard errors are obtained by wild bootstrap and clustered at class level.
All test scores are in logs.
(1),(4): Whole sample
(2),(5): Own baseline below the median
(3),(6): Own baseline above the median
Baseline non-cognitive peer differences are standardized.

Table 3: Peer Effect (Math)

	First Sem. Midterm			First Sem. Final		
	(1)	(2)	(3)	(4)	(5)	(6)
Own Baseline	0.843***	0.818***	1.212***	0.839***	0.878***	0.953***
	(0.061)	(0.068)	(0.088)	(0.076)	(0.075)	(0.079)
Deskmate1 Baseline	−0.019	−0.043	−0.038*	0.060	0.068	−0.029**
	(0.028)	(0.039)	(0.023)	(0.047)	(0.068)	(0.015)
Non-Cog Diff (Deskmate1)	0.119	0.327***	−0.076	0.376*	0.649**	−0.035
	(0.112)	(0.119)	(0.085)	(0.216)	(0.282)	(0.050)
Neighbor1 Baseline	0.014	−0.008	−0.023	0.085	0.158	0.030
	(0.037)	(0.052)	(0.047)	(0.057)	(0.106)	(0.036)
Non-Cog Diff (Neighbor1)	0.112	−0.127	0.241*	−0.149	−0.347	0.189
	(0.081)	(0.183)	(0.128)	(0.229)	(0.298)	(0.181)
Interaction (Deskmate1)	−0.022	−0.066**	0.019	−0.085*	−0.147**	0.009
	(0.026)	(0.027)	(0.019)	(0.049)	(0.064)	(0.011)
Interaction (Neighbor1)	−0.028	0.023	−0.055*	0.036	0.083	−0.041
	(0.019)	(0.043)	(0.029)	(0.053)	(0.069)	(0.042)
Constant	0.462*	0.666**	−0.880	−0.026	−0.516	0.110
	(0.273)	(0.287)	(0.536)	(0.602)	(0.868)	(0.373)
class FE	Yes	Yes	Yes	Yes	Yes	Yes
location FE	Yes	Yes	Yes	Yes	Yes	Yes
gender pair FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	882	465	442	879	463	441
R^2	0.760	0.787	0.554	0.648	0.681	0.476
Adjusted R^2	0.725	0.721	0.406	0.595	0.583	0.301

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
Interactions: between peer baseline and non-cognitive peer differerence.
Standard errors are obtained by wild bootstrap and clustered at class level.
All test scores are in logs.
(1),(4): Whole sample
(2),(5): Own baseline below the median
(3),(6): Own baseline above the median
Baseline non-cognitive peer differences are standardized.

5 The Peer Assignment Problem

Given the causal estimates reported in Section 4, we are now in a position to discuss the peer assignment problem and report the associated results.

The size of the set of feasible assignments $\mathcal{Q}$ is $(n/2)!$. Given the large class sizes in our sample, brute force computation is not feasible. As one well-known way of attack, we relax the peer assignment problem by allowing for fractional assignment, thereby transforming the problem into the following linear programming problem[27]:

$$\max_{m \in \mathcal{M}} \frac{1}{n/2} \sum_{i \in \mathbb{S}_{low}} \sum_{j \in \mathbb{S}_{high}} f(y_{i0}, y_{j0}, |n_{i0} - n_{j0}|) m(i,j) \tag{13}$$

such that $m : \mathbb{S}_{low} \times \mathbb{S}_{high} \to \mathbb{R}_+$ is the fraction of matches between type $i \in \mathbb{S}_{low}$ and $j \in \mathbb{S}_{high}$. The function m is to be chosen is restricted to be within a set $\mathcal{M}$ defined by the following non-negativity and accounting constraints:

$$m(i,j) \geq 0, \forall i \in \mathbb{S}_{low}, \forall j \in \mathbb{S}_{high} \tag{14}$$

$$\sum_{i \in \mathbb{S}_{low}} m(i,j) = 1, \forall j \in \mathbb{S}_{high} \tag{15}$$

$$\sum_{j \in \mathbb{S}_{high}} m(i,j) = 1, \forall i \in \mathbb{S}_{low} \tag{16}$$

In particular, the two accounting constraints (15), (16) govern that the number of matches must match the mass of agents of each type. Since each type in a deskmate matching problem is a particular seat occupied by exactly one student, the total mass is 1 for each type.[28]

[27] In the language of optimal transport, our original problem is the Monge problem and our relaxed problem is a Kantorovich problem. See Villani (2008) and Galichon (2016).

[28] Noting that our problem is discrete, m can be regarded as a doubly stochastic matrix of dimension $(n/2) \times (n/2)$. Linear programs of this size can be solved almost instantaneously by using non-commercial solvers, so that computation is a non-issue.

We first run a set of peer regressions with full three-way interactions among own baseline score, peer baseline scores, and non-cognitive peer difference. Here we do not discretize the own baseline score like we did previously in Section 4, which we did only for presentation purposes. We evaluate the academic achievement given the following kinds of peer assignments:

1. (Full Interaction) Optimizing the peer assignment based on the full interaction model.

2. (Univariate) An assignment based on optimizing a mis-specified model with only the two-way interaction between own and peer baseline scores; non-cognitive peer difference is ignored. This mis-specified model is the microscopic version of model considered in Carrell et al. (2013) and most academic peer effect papers. For a fair comparison against our full interaction model, this univariate model is best-fitted to the same data.

3. (Minimizing NCPD) An assignment entirely based on minimizing non-cognitive peer difference (NCPD). It represents the other extreme relative to the univariate optimization.

4. (Maximizing NCPD) An assignment entirely based on maximizing non-cognitive peer difference.

5. (Worst Assignment) The value obtained by the worst assignment (minimizing the objective function).

6. (Random Assignment) The achievement averaged over 100 random assignments.

We then compute the value of the true objective function (13) given these assignments. As a remark, the location fixed effects enter the objective function linearly. The total contribution of location fixed effects

Even though we consider the relaxed problem, the optimal assignments remain pure (each student is being assigned to a unique peer).

to the objective function would be invariant to a change in the peer assignment, so we do not discuss them here.

We use random assignments as a benchmark and compare the other optimization results against it. For academic achievements, we subtract the optimization results from that obtained from the randomized assignment for each class; then, we compute the first quartile, the median, and the third quartile of this difference. To be precise, let $a \in \{1, 2, 3, 4, 5\}$ where $1, 2, 3, 4, 5$ label the optimal assignment, univariate, minimizing non-cognitive peer diference, maximizing non-cognitive peer difference, and worst assignment respectively; let c denote a particular class; let $\bar{y}_c^a$ denote the computed average score of low-achieving students in class c under assignment a. let $\bar{y}_c^{random}$ denote the average of average scores under 1000 random assignments. Our exercise is to evaluate the 25th, 50th (median), and 75th quantiles of $\bar{y}_c^a - \bar{y}_c^{random}$ for $a \in \{1, 2, 3, 4, 5\}$.

We repeat this procedure separately for Chinese mid-term and final, and the math equivalents. Since the dependent variables are log scores, the figures as log differences are all interpreted as percentage gains; for instance, the median achievement gains for Chinese midterm and Chinese final by optimizing according to our extended peer effect model are about 4%; the counterparts for Math midterm and Math final are slightly less. These figures are modest yet expected since this is only a one semester experiment.

The univariate model is not performing much differently from the randomized benchmark, as shown by its corresponding figures in Table 4 which are close to zero for both subjects. A few classes actually perform worse than the randomized benchmark, which may explain the result in Carrell et al. (2013). Whereas if the planner minimizes non-cognitive peer difference, she would be able to get most of the achievement gains for Chinese but not for Math. For instance, the median class can gain about 2% for Chinese; the corresponding figures for Math are almost zero. While if she maximizes non-cognitive peer difference instead, the result would be close to the worst assignment for Chinese but not for

Math—in fact for Math, the academic achievement obtained by either maximizing or mimimizing non-cognitive peer difference is almost the same as the randomized benchmark. These results show that managing non-cognitive peer difference within a class is more important for improving Chinese relative to Math.

Table 4: Optimized Achievement Scores Under Various Assignments

		Chinese Midterm	Chinese Final	Math Midterm	Math Final
Full Interaction	1st Quartile	0.031	0.028	0.016	0.017
	Median	0.043	0.039	0.020	0.025
	3rd Quartile	0.060	0.061	0.026	0.030
Univariate	1st Quartile	0.013	-0.001	0.007	0.002
	Median	0.016	0.005	0.009	0.003
	3rd Quartile	0.021	0.012	0.013	0.004
Minimizing NCPD	1st Quartile	0.016	0.019	-0.015	0.007
	Median	0.024	0.024	-0.008	0.013
	3rd Quartile	0.035	0.037	-0.002	0.019
Maximizing NCPD	1st Quartile	-0.022	-0.023	-0.001	-0.033
	Median	-0.017	-0.018	0.006	-0.014
	3rd Quartile	-0.008	-0.011	0.014	-0.005
Worst Assignment	1st Quartile	-0.064	-0.037	-0.045	-0.035
	Median	-0.039	-0.029	-0.030	-0.029
	3rd Quartile	-0.029	-0.018	-0.024	-0.013

Notes: Results are percentage gains relative to random assignment.

In Table 5, we report the average non-cognitive peer difference achieved under the various assignments. We report the figures as ratios against the randomized benchmark (instead of difference). We find that to optimize the assignment for the sake of Chinese midterm and Chinese final, the optimal assignments should reduce the non-cognitive peer difference by about 20% relative to the randomized benchmark. In fact, the figures are very close to those that are being obtained by smallest attainable non-cognitive peer difference, as shown by the figures for minimizing non-cognitive peer difference.

Agreeing with our result that non-cognitive peer difference is relatively unimportant for Math, we find that in our model, the level of non-cognitive peer difference is relatively higher than the randomized benchmark. For Math midterm and Math final, the level of non-cognitive peer difference are not lower than that of the randomized benchmark; in fact for Math midterm the corresponding ratio (median) is greater than 1, and is rather close to the one obtained by maximizing non-cognitive peer difference. Whereas since the univariate model does not consider non-cognitive peer difference, its corresponding figures are close to 1.

To summarize, depending on the academic outcome considered, the optimal assignment may not always minimize non-cognitive difference, even though generally this is the case. Specifically, the result would depend on the relative strength of knowledge spillover and non-cognitive peer difference. Depending on the circumstances, one force may dominate the other.

Table 5: Non-Cognitive Peer Difference Under Various Assignments

		Chinese Midterm	Chinese Final	Math Midterm	Math Final
Full Interaction	1st Quartile	0.661	0.688	1.147	0.733
	Median	0.813	0.774	1.182	0.832
	3rd Quartile	0.850	0.828	1.237	0.895
Univariate	1st Quartile	0.980	0.965	0.997	0.978
	Median	1.004	0.983	1.012	1.008
	3rd Quartile	1.027	1.018	1.031	1.024
Minimizing NCPD	1st Quartile	0.615	0.616	0.639	0.633
	Median	0.716	0.719	0.734	0.731
	3rd Quartile	0.782	0.783	0.791	0.789
Maximizing NCPD	1st Quartile	1.172	1.170	1.188	1.179
	Median	1.214	1.217	1.208	1.212
	3rd Quartile	1.281	1.288	1.282	1.271
Worst Assignment	1st Quartile	0.978	0.989	0.967	0.977
	Median	0.996	1.014	1.007	1.017
	3rd Quartile	1.022	1.038	1.030	1.041

Notes: All results are relative to the results obtained by random assignment (as ratios).

We then perform a counterfactual exercise where we replace the baseline non-cognitive peer difference $|n_{i0} - n_{q(i)0}|$ with that after a semester $|n_{i1} - n_{q(i)1}|$, and examine if the optimal assignment significantly changes. If non-cognitive peer difference is stable over time, then the optimal assignment should be stable as well. However, this is not the case. Deferring the discussion of how non-cognitive peer difference changes over time in the next session, we first discuss the consequence of this instability here: we find that apart from a very small number of exceptions, the optimal assignment $q(i)$ changes for all i (the percentage is $\geq 95\%$ for all classes); the loss in academic achievement if the re-optimization is not being done is about $1-3\%$, which is more than half of the correponding gains obtained by previous optimizations. This hypothetical assignment demonstrates the need for readjustments of peer assignments over time, which agrees with the traditional routine of Chinese class teachers.

Table 6: Loss in academic achievement when readjustment is not performed

	First Quartile	Median	Third Quartile
Chinese Midterm	0.023	0.029	0.044
Chinese Final	0.025	0.033	0.053
Math Midterm	0.009	0.011	0.016
Math Final	0.010	0.012	0.024

6 Assimilation

Assimilation—the convergence in non-cognitive peer difference between peers—is perhaps a natural outcome to expect. We distinguish between homophily (sorting of peers given characteristics, see e.g. Currarini et al. (2009)) and assimilation (peers influencing each other, given the network). Like immigrants facing locals, the Chinese classroom is a context where assimilation is relatively important than homophily due to immobility—seats are fixed unless reassignment occurs, so that in case assimilation fails, peer conflicts cannot be resolved by sorting.

The discussion of microscopic assimilation is mostly found in the Group Socialization Theory in psychology (Harris, 1995), which suggests the importance of peer groups, rather than parental figures, in affecting personality and behaviors. Within groups, there is assimilation through imitation. Notably, the theory also suggests the possibility of differentiation—peers become less similar over time. As the theory claims, differentiation is particular likely when the peers are from different categories (e.g. by gender).

In this section, we examine assimilation in detail. We report the temporal change in non-cognitive peer difference $d_{i1} - d_{i0}$. For convenience, here we define assimilation as $d_{i1} - d_{i0} \leq 0$ and differentiation as $d_{i1} - d_{i0} > 0$. Figure 2 plots the distribution of $d_{i1} - d_{i0}$. In the plot, 51% of the observations assimilate, while the other 49% differentiate. The distribution of $d_{i1} - d_{i0}$ is unimodal, with a mean of -0.03, which is statistically significant at the one percent significance level.

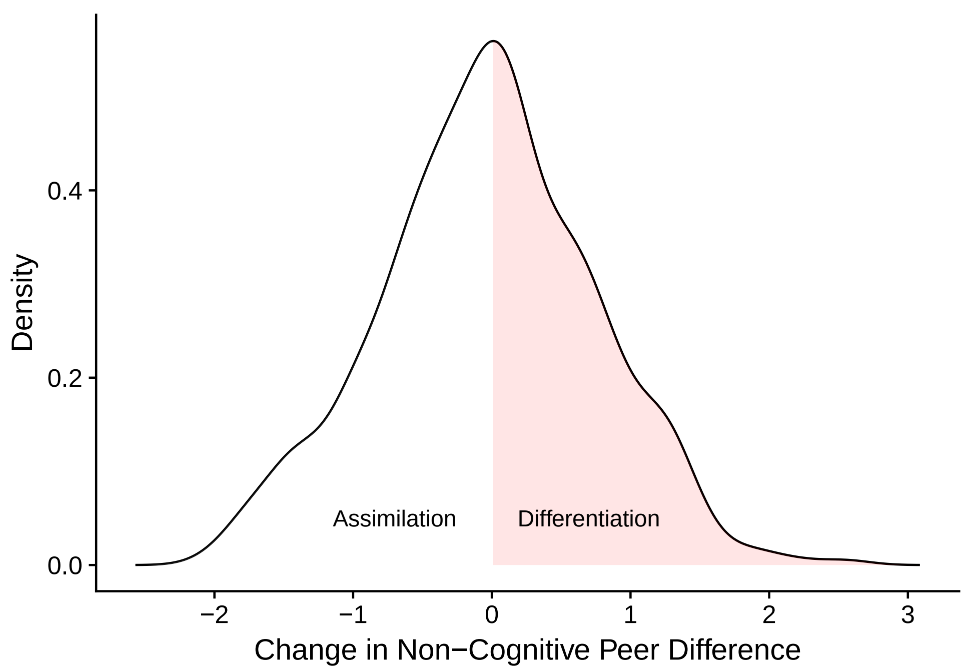

Figure 2: Distribution of Change in Non-Cognitive Peer Difference (Round 2 - Round 1)

While measurement error is clearly a possible cause of this result—for instance, the respondents may not understand the questionnaires and fill in random answers each time—it cannot explain the systematic patterns with respect to gender and age that we are going to report below. Attenuation bias would imply that all estimates are biased towards zero, which is not the case.

6.1 By Gender

We find that the success of assimilation highly depends on own and peer gender. Table 7 reports how gender explains the differences of the non-cognitive skill between deskmates. Specifically, we run an OLS regression of d_{it} for $t = 0, 1$ and $d_{i1} - d_{i0}$ on possible gender pairs, i.e. male-female, male-male and female-female, with male-male as the omitted group.

At the baseline, there is almost zero level difference in d_{i0} across gender pairs. After a semester, we find that $d_{i1} - d_{i0}$ (Column 3) is heterogeneous across gender pairs. Specifically, female-female pairs have a significantly negative coefficient of -0.1747, which is about 6 times relative to the mean assimilation of 0.03. While this figure is only $1/5$ of the standard deviation of $d_{i1} - d_{i0}$, it is still relatively large. We do not find similar effects for the other gender pairs. In other words, the degrees of assimilation of mixed gender pairs and male-male pairs are similar.

These results relate to Lu and Anderson (2015), who find that female-female pairs improve academic achievement relative to other possible gender pairs. Together with our result that non-cognitive peer difference complements the peer baseline score effect, the significantly better assimilation of female-female pairs can potentially explain their results. Lavy and Schlosser (2011) also show that adding female students to a class can improve the overall academic performance. As the authors point out, a higher proportion of female peers "lowers the level of classroom disruption and violence, improves inter-student relationships." While their results are global and our gender assimilation result is local (between deskmates), our findings are consistent with each other.

Table 7: Explaining Non-Cognitive Peer Differences (By Gender)

	Baseline	After the First Semester	Change
	(1)	(2)	(3)
Male-Female	−0.0112	−0.0911*	−0.0799
	(0.0452)	(0.0510)	(0.0672)
Female-Male	−0.0038	−0.0995*	−0.0957
	(0.0450)	(0.0508)	(0.0669)
Female-Female	−0.0164	−0.1911***	−0.1747**
	(0.0555)	(0.0626)	(0.0825)
class FE	Yes	Yes	Yes
location FE	Yes	Yes	Yes
sd of Dep. Variable	0.5706	0.6449	0.7736
Observations	892	892	892
R^2	0.4745	0.4773	0.3695
Adjusted R^2	0.4125	0.4157	0.2952

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

6.2 By Grade (Age)

Similarly, we examine the non-cognitive peer difference by grade. The omitted group in Table 8 is Grade 3. We do not introduce class dummies because they would absorb all relevant variation.

At the baseline, the average non-cognitive peer differences of Grade 4 and Grade 5 are smaller than that of Grade 3, as indicated by their negative coefficients. This result suggests that initially the higher grades are more homogenous than the lower grades.

In terms of change, the positive Grades 4 and 5 coefficients suggest that the non-cognitive peer differences in higher grades are more difficult to be reduced relative to those of Grade 3. Similar to our interpretation for gender, the magnitudes of the coefficients (greater than 0.2 for both Grade 4 and Grade 5) are large with respect to the baseline non-cognitive peer difference with unit variance.

Our results agree with the skill formation literature, which shows that non-cognitive skill is more malleable at younger ages (and hence lower grades). While the findings in the skill formation literature mostly concern school and family inputs, here we show that the same conclusion holds with respect to peer inputs.[29]

Table 8: Explaining Non-Cognitive Peer Differences (by Grade)

	Baseline	After the First Semester	Change
	(1)	(2)	(3)
Grade 4	-0.080^{*}	0.134^{***}	0.214^{***}
	(0.043)	(0.050)	(0.058)
Grade 5	-0.180^{***}	0.108^{**}	0.288^{***}
	(0.043)	(0.049)	(0.058)
location FE	Yes	Yes	Yes
sd of Dep. Variable	0.5706	0.6449	0.7736
Observations	892	892	892
R^2	0.095	0.088	0.109
Adjusted R^2	0.015	0.006	0.029

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

6.3 By Age Difference

We also examine within each class (and hence grades) how age difference affect the evolution of non-cognitive peer difference. The regression results are shown in Table 9. The first column shows that at the baseline, due to peer randomization, a larger age difference does not statistically predict a larger peer non-cognitive difference. However, after a

[29] In the skill formation literature, while peer inputs are often mentioned, the data sets used are usually representative panels whose observations are i.i.d. across individuals. Peer inputs, if present, would be measured indirectly by interviewing the respondent rather than his peer. For our case, we have direct, separate measures for own and peer.

semester, one extra year in peer age difference predicts a 6% increase in non-cognitive peer difference, suggesting that pairs with a larger age difference face more difficulties in assimilation.

Comparing with the gender results presented in Table 7, the effect of age difference on assimilation is not as large, although still notable and strongly significant. In Chinese rural schools, due to a number of reasons (mostly due to deferrals, but also include gifted children who are promoted to a higher class early), the age variation within classes is rather large. As reported in the summary statistics in Table 1, some classes have a within-class standard deviation of birth year exceeding 1 year, thereby implying that some pairs can have a large age difference relative to other pairs.

Table 9: Explaining Non-Cognitive Peer Differences (by Age)

	Baseline	After the First Semester	Change
	(1)	(2)	(3)
Age Difference	0.021^{*}	0.059^{***}	0.038^{**}
	(0.012)	(0.014)	(0.018)
class FE	Yes	Yes	Yes
location FE	Yes	Yes	Yes
sd of Dep. Variable	0.5706	0.6449	0.7736
Observations	892	892	892
R^2	0.478	0.483	0.371
Adjusted R^2	0.418	0.423	0.299

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

6.4 Assimilation and Friendship

As we constructed, non-cognitive peer difference is a psychological metric. Here we examine whether it corresponds to peer relationships as we intended. Specifically, we ask each student to provide the names of three male friends and three female friends, six persons in total. Using this information, we check whether the deskmates are among the self-reported friends in each round, and check if the reporting pattern relates to our results on assimilation. Initially, about 30% of the deskmates are named as one of the friends, which suggests that a student is quite likely to causally write down the name of his randomly assigned deskmate as one of his friends, given that there are six slots in total. One interesting issue is then what causes a student to stop writing his first semester deskmate again as a friend in the second semester, given that he did so in the first semester. If this happens, we expect that this behavior is a strong indicator of failed assimilation; indeed, thi is the case. As Table 10 shows, non-cognitive peer difference does not predict whether a student would report his deskmate as a friend in the first semester. However, a regression shows that the probability of reporting a first semester deskmate as a friend in the second semseter, conditioned on that she did so in the first semester, is 17% lower if the assimilation fails (differentiation occurs).[30]

[30]It should be noted that for the purpose of peer assignments, we cannot use this friendship data, even though it may appear to be more direct than using non-cognitive peer difference as a proxy.

Table 10: Friendship and Deskmates

	Deskmate1 is Friend in Round 1	Deskmate1 is Friend in Round 2
	(1)	(2)
Non-Cog Diff (Round1)	−0.004	−0.005
	(0.014)	(0.012)
Assimilation		0.014
		(0.027)
Deskmate1 is Friend in Round 1		0.343***
		(0.035)
Interaction (2)*(3)		0.169***
		(0.053)
Constant	0.241***	0.088***
	(0.014)	(0.019)
Observations	892	892
R^2	0.000	0.124
Adjusted R^2	−0.001	0.120
Notes:	***Significant at the 1 percent level.	
	**Significant at the 5 percent level.	
	*Significant at the 10 percent level.	
	Non-cognitive peer difference in both rounds are standardized.	

7 A Reshuffle of Seating Plans

Before economists can recommend on policies, it could be informative to know about how actual policy-makers are currently doing their job; the same applies to peer assignments, especially microscopic ones which are new to economists. As another concern, the peer randomization that took place in the first semester for our experiment was regarded disruptive to the classes. Some class teachers expressed to us their wish to regain their power to assign peers, which indicates their strong beliefs in the importance of peer assignments, and in their ability to do the job well. Indeed, class teachers have better information than economists as they observe students in real time, and they may observe things that a

Big Five questionnaire cannot capture. In this experiment, we observed non-cognitive peer difference only twice, since it was not possible to ask students to fill in questionnaires too frequently. With prudence in mind, we chose to observe what class teachers do in the second semester. By doing so, we gained some information on how actual peer assignments are performed as a traditional routine within classrooms.

In the second semester, we asked the class teachers to reshuffle the seating plan. Unlike Round 1 where we enforce our randomization strictly, in the second semester we did not monitor the process. Effectively, this undid our previous randomization while allowed the class teachers to reveal their own preferences in the resulting peer assignments. As we found, this reshuffle was throughoutly done. We found that 95% of the students have changed seats, i.e. their (x, y) coordinates become different in Round 2. Specifically, we evaluated:

$$|n_{i1} - n_{p(i,1)1}| - |n_{i1} - n_{p(i,0)1}|$$

where $p(i, 1)$ is the new deskmate of student i and $p(i, 0)$ is the old deskmate. The above was the change of non-cognitive peer difference achieved by changing the deskmate. Then we calculated class avarages of this statistic. We also evaluate, for each class, the change of average score from the first semester final to the second semester midterm (for Chinese and Math, respectively). Putting them together, we found that change in non-cognitive peer difference was negatively related to change in academic performance at the class level, whose result is shown in Table 11.

Table 11: Change in NCPD and Change in Grades after Reshuffle

	Change in NCPD	Change in Chinese	Change in Math
Change in NCPD	1.000	-0.263	-0.261
Change in Chinese	-0.263	1.000	0.034
Change in Math	-0.261	0.034	1.000

These correlations of changes indicate that some, but not all, class teachers could improve academic achievement in their classes by reducing non-cognitive peer difference. Because all classes were randomized in the same way in the first semester, this correlation of changes is likely to be caused by the variation among class teachers in their way of performing peer assignments rather than across-class differences in student quality. As we observed, most class teachers based their peer assignments on their own intuition and experience, rather than on a systematic theory; this could lead to a large heterogeneity in results. While suggestive, if economists are able to offer systematic advice based on quantitative methods and scientific measurements, it could be useful to class teachers or actual planners in general.

As the second exercise, given that the deskmates for most students change after the reshuffle, we used this opportunity to conduct a placebo test. If peer effects were local, then the first semester deskmate variables should not correlate to the second semeseter outcomes. We cannot reject the null that the first semester peers have zero effect. These results are shown in Table 12. As expected, the estimated interaction effects are much smaller in magnitude, have mixed signs, and are all statistically insignificant. Together with the evidence obtained from our main peer regressions that general neighbors do not matter, we argue that within-classroom peer effects are local in nature. This property is important for our policy recommendations. If within-classroom peer effects are global, then seat reassignments would be irrelevant since the peer effects linger even after a reassignment.[31]

[31]Given the independence between deskmates and more general neighbors, the Round 1 specifications can omit the general neighbor variables.

Table 12: Local Peer Effect Placebo Test (Round 2 Outcomes, Round 1 Peers)

	Chinese Midterm	Chinese Final	Math Midterm	Math Final
	(1)	(2)	(3)	(4)
Own Baseline	0.680***	0.701***	0.939***	0.982***
	(0.086)	(0.109)	(0.077)	(0.074)
Deskmate1 Baseline	−0.007	−0.002	−0.023	0.029
	(0.022)	(0.024)	(0.040)	(0.046)
Non-Cog Diff (Deskmate1)	0.047	−0.085	0.139	0.212
	(0.060)	(0.063)	(0.278)	(0.279)
Neighbor1 Baseline	−0.066	−0.059	0.039	−0.011
	(0.077)	(0.084)	(0.056)	(0.064)
Non-Cog Diff (Neighbor1)	0.083	−0.152	−0.054	0.212
	(0.130)	(0.106)	(0.160)	(0.200)
Interaction (Deskmate1)	−0.018	0.014	−0.030	−0.049
	(0.015)	(0.016)	(0.064)	(0.063)
Interaction (Neighbor1)	−0.015	0.038	0.007	−0.054
	(0.029)	(0.025)	(0.037)	(0.047)
Constant	1.400**	1.651**	−0.058	−0.222
	(0.630)	(0.703)	(0.533)	(0.588)
class FE	Yes	Yes	Yes	Yes
location2 FE	Yes	Yes	Yes	Yes
gender pair FEs	Yes	Yes	Yes	Yes
Observations	877	874	877	876
R^2	0.704	0.668	0.689	0.623
Adjusted R^2	0.660	0.618	0.642	0.567

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
Standard Errors are obtained by wild bootstrap and clustered at class level.
All Test Scores are in logs.
Baseline Non-Cognitive peer differences are standardized.

8 Conclusions

In this paper, we reconsider the peer assignment problem which is still open—the optimization of peer assignments could possibly backfire even if the peer effect estimates are obtained from experiments. As discovered by the literature, one reason behind this puzzle is within-group sorting. We address within-group sorting using three means: restricting the designated peer group size to a microscopic level, measuring non-cognitive peer difference, and peer randomization. All three means are important to microscopic peer assignments. As our results show, non-cognitive peer difference is one major determinant in optimizing microscopic peer assignments. Assimilation, as its change over time, has rich patterns which imply that microscopic peer assignments have to be dynamic. As a traditional routine, the Chinese class teachers do consider such non-cognitive issues among deskmates in deciding their microscopic peer assignments, agreeing with our findings.

Our setting, where students have fixed deskmates, provides us the possibly cleanest environment to eliminate within-group sorting, although external validity becomes a concern. Though we note that the fixed seat system is rather common in some other regions such as Singapore, Hong Kong, India, South Korea, and Japan. We are attracted by Chinese classrooms because of their practical need in using microscopic peer assignments for class management due to their large class size. We further note that a fixed seat system, as a prerequisite of applying microscopic peer assignments, is rather common in workplaces with limited office space; although in such context, the problem becomes more complicated since workers have different roles so that knowledge spillover becomes more difficult to define.

The peer assignment problem is deep, which much left to be explored. Echoing Carrell et al. (2013), one may need to further study what the policy-maker can potentially do and what they actually do, before attempting to recommend on "optimal" peer assignments as the final goal.

In this paper, we lay some groundwork.

References

ABRAMITZKY, R., L. P. BOUSTAN, AND K. ERIKSSON (2014): "A nation of immigrants: Assimilation and economic outcomes in the age of mass migration," *Journal of Political Economy*, 122, 467–506.

——— (2016): "Cultural assimilation during the age of mass migration," Tech. rep., National Bureau of Economic Research.

AIZER, A. (2008): "Peer effects and human capital accumulation: The externalities of ADD," Tech. rep., National Bureau of Economic Research.

ALMLUND, M., A. L. DUCKWORTH, J. HECKMAN, AND T. KAUTZ (2011): "Personality psychology and economics," in *Handbook of the Economics of Education*, Elsevier, vol. 4, 1–181.

AMMERMUELLER, A. AND J.-S. PISCHKE (2009): "Peer effects in European primary schools: Evidence from the progress in international reading literacy study," *Journal of Labor Economics*, 27, 315–348.

ANGRIST, J. D. (2014): "The perils of peer effects," *Labour Economics*, 30, 98–108.

ANGRIST, J. D. AND V. LAVY (1999): "Using Maimonides' rule to estimate the effect of class size on scholastic achievement," *Quarterly Journal of Economics*, 114, 533–575.

BAYER, P., R. HJALMARSSON, AND D. POZEN (2009): "Building criminal capital behind bars: Peer effects in juvenile corrections," *The Quarterly Journal of Economics*, 124, 105–147.

BHATTACHARYA, D. (2009): “Inferring optimal peer assignment from experimental data,” *Journal of the American Statistical Association*, 104, 486–500.

BLACK, D. A. AND J. A. SMITH (2006): “Estimating the returns to college quality with multiple proxies for quality,” *Journal of Labor Economics*, 24, 701–728.

BOOIJ, A. S., E. LEUVEN, AND H. OOSTERBEEK (2017): “Ability peer effects in university: Evidence from a randomized experiment,” *The Review of Economic Studies*, 84, 547–578.

BORJAS, G. J. (1985): “Assimilation, changes in cohort quality, and the earnings of immigrants,” *Journal of labor Economics*, 3, 463–489.

——— (1995): “Assimilation and changes in cohort quality revisited: what happened to immigrant earnings in the 1980s?” *Journal of Labor Economics*, 13, 201–245.

——— (2015): “The slowdown in the economic assimilation of immigrants: Aging and cohort effects revisited again,” *Journal of Human Capital*, 9, 483–517.

BURKE, M. A. AND T. R. SASS (2013): “Classroom peer effects and student achievement,” *Journal of Labor Economics*, 31, 51–82.

CAMERON, A. C. AND D. L. MILLER (2015): “A practitioner’s guide to cluster-robust inference,” *Journal of Human Resources*, 50, 317–372.

CARRELL, S. E., R. L. FULLERTON, AND J. E. WEST (2009): “Does your cohort matter? Measuring peer effects in college achievement,” *Journal of Labor Economics*, 27, 439–464.

CARRELL, S. E., M. HOEKSTRA, AND E. KUKA (2018): “The long-run effects of disruptive peers,” *American Economic Review*, 108, 3377–3415.

CARRELL, S. E. AND M. L. HOEKSTRA (2010): "Externalities in the classroom: How children exposed to domestic violence affect everyone's kids," *American Economic Journal: Applied Economics*, 2, 211–28.

CARRELL, S. E., B. I. SACERDOTE, AND J. E. WEST (2013): "From natural variation to optimal policy? The importance of endogenous peer group formation," *Econometrica*, 81, 855–882.

CHETTY, R., J. N. FRIEDMAN, N. HILGER, E. SAEZ, D. W. SCHANZENBACH, AND D. YAGAN (2011): "How does your kindergarten classroom affect your earnings? Evidence from Project STAR," *The Quarterly Journal of Economics*, 126, 1593–1660.

CHISWICK, B. R. (1978): "The effect of Americanization on the earnings of foreign-born men," *Journal of political Economy*, 86, 897–921.

CONTI, G., J. J. HECKMAN, AND R. PINTO (2016): "The effects of two influential early childhood interventions on health and healthy behaviour," *The Economic Journal*, 126, F28–F65.

COSTA, P. T. AND R. R. MACCRAE (1992): *Revised NEO personality inventory (NEO PI-R) and NEO five-factor inventory (NEO-FFI): Professional manual*, Psychological Assessment Resources, Incorporated.

COSTA JR, P. T. AND R. R. MCCRAE (1992): "Four ways five factors are basic," *Personality and Individual Differences*, 13, 653–665.

CUNHA, F. AND J. HECKMAN (2007): "The technology of skill formation," *American Economic Review*, 97, 31–47.

CUNHA, F. AND J. J. HECKMAN (2008): "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation," *Journal of Human Resources*, 43, 738–782.

CUNHA, F., J. J. HECKMAN, AND S. M. SCHENNACH (2010): "Estimating the technology of cognitive and noncognitive skill formation," *Econometrica*, 78, 883–931.

CURRARINI, S., M. O. JACKSON, AND P. PIN (2009): "An economic model of friendship: Homophily, minorities, and segregation," *Econometrica*, 77, 1003–1045.

DAMM, A. P. AND C. DUSTMANN (2014): "Does growing up in a high crime neighborhood affect youth criminal behavior?" *American Economic Review*, 104, 1806–32.

DING, W. AND S. F. LEHRER (2007): "Do peers affect student achievement in China's secondary schools?" *The Review of Economics and Statistics*, 89, 300–312.

DUFLO, E., P. DUPAS, AND M. KREMER (2011): "Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya," *American Economic Review*, 101, 1739–74.

EPPLE, D., E. NEWLON, AND R. ROMANO (2002): "Ability tracking, school competition, and the distribution of educational benefits," *Journal of Public Economics*, 83, 1–48.

FELD, J. AND U. ZÖLITZ (2017): "Understanding peer effects: on the nature, estimation, and channels of peer effects," *Journal of Labor Economics*, 35, 387–428.

FIGLIO, D. N. (2007): "Boys named Sue: Disruptive children and their peers," *Education Finance and Policy*, 2, 376–394.

FIGLIO, D. N. AND M. E. PAGE (2002): "School choice and the distributional effects of ability tracking: does separation increase inequality?" *Journal of Urban Economics*, 51, 497–514.

FREDRIKSSON, P., B. ÖCKERT, AND H. OOSTERBEEK (2012): "Long-term effects of class size," *Quarterly Journal of Economics*, 128, 249–285.

FRUEHWIRTH, J. C. (2014): "Can achievement peer effect estimates inform policy? A view from inside the black box," *Review of Economics and Statistics*, 96, 514–523.

FU, C. AND N. MEHTA (2018): "Ability tracking, school and parental effort, and student achievement: A structural model and estimation," *Journal of Labor Economics*, 36, 923–979.

GALICHON, A. (2016): *Optimal transport methods in economics*, Princeton University Press.

GARCÍA, J. L., J. J. HECKMAN, D. E. LEAF, AND M. J. PRADOS (2016): "The life-cycle benefits of an influential early childhood program," Tech. rep., National Bureau of Economic Research.

GARLICK, R. ET AL. (2018): "Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment," *American Economic Journal: Applied Economics*, 10, 345–369.

GONG, J., Y. LU, AND H. SONG (2018): "The Effect of Teacher Gender on Students Academic and Noncognitive Outcomes," *Journal of Labor Economics*, 36, 743–778.

GURYAN, J., K. KROFT, AND M. J. NOTOWIDIGDO (2009): "Peer effects in the workplace: Evidence from random groupings in professional golf tournaments," *American Economic Journal: Applied Economics*, 1, 34–68.

HANUSHEK, E. A. (1998): "Improving Student Achievement: Is Reducing Class Size the Answer?" Tech. rep., Policy Briefing, Progressive Policy Institute, Washington, DC, June.

HANUSHEK, E. A., J. F. KAIN, J. M. MARKMAN, AND S. G. RIVKIN (2003): "Does peer ability affect student achievement?" *Journal of Applied Econometrics*, 18, 527–544.

HARRIS, J. R. (1995): "Where is the child's environment? A group socialization theory of development." *Psychological review*, 102, 458.

HECKMAN, J. J. AND S. MOSSO (2014): "The economics of human development and social mobility," *Annual Review of Economics*, 6, 689–733.

HECKMAN, J. J. AND Y. RUBINSTEIN (2001): "The importance of noncognitive skills: Lessons from the GED testing program," *American Economic Review*, 91, 145–149.

HECKMAN, J. J., J. STIXRUD, AND S. URZUA (2006): "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor economics*, 24, 411–482.

HENDERSON, J. V., J. QUIGLEY, AND E. LIM (2009): "Urbanization in China: Policy issues and options," Tech. rep., Unpublished manuscript, Brown University.

HONG, S. C. AND J. LEE (2017): "Who is sitting next to you? Peer effects inside the classroom," *Quantitative Economics*, 8, 239–275.

HOXBY, C. (2000): "Peer Effects in the classroom: Learning from Gender and Race variation," Tech. rep., National Bureau of Economic Research.

HOXBY, C. M. AND G. WEINGARTH (2005): "Taking race out of the equation: School reassignment and the structure of peer effects," NBER Working paper.

JIN, L. AND M. CORTAZZI (1998): "Dimensions of dialogue: Large classes in China," *International Journal of Educational Research*, 29, 739–761.

KRUEGER, A. B. (1999): "Experimental estimates of education production functions," *Quarterly Journal of Economics*, 114, 497–532.

——— (2003): "Economic considerations and class size," *The Economic Journal*, 113.

LAVY, V. AND A. SCHLOSSER (2011): "Mechanisms and impacts of gender peer effects at school," *American Economic Journal: Applied Economics*, 3, 1–33.

LAZEAR, E. P. (1999): "Culture and language," *Journal of Political Economy*, 107, S95–S126.

——— (2001): "Educational production," *The Quarterly Journal of Economics*, 116, 777–803.

LU, F. AND M. L. ANDERSON (2015): "Peer effects in microenvironments: The benefits of homogeneous classroom groups," *Journal of Labor Economics*, 33, 91–122.

LUCAS, R. (1976): *Econometric policy evaluation: A Critique.*, Amsterdam, North Holland.

MANSKI, C. F. (1993): "Identification of endogenous social effects: The reflection problem," *Review of Economic Studies*, 60, 531–542.

MENG, X. AND R. G. GREGORY (2005): "Intermarriage and the economic assimilation of immigrants," *Journal of Labor economics*, 23, 135–174.

MOULTON, B. R. (1986): "Random group effects and the precision of regression estimates," *Journal of Econometrics*, 32, 385–397.

MURPHY, F. X. (2019): "Does Increased Exposure to Peers with Adverse Characteristics Reduce Workplace Performance? Evidence from a Natural Experiment in the US Army." *Journal of Labor Economics*, 37, 435–466.

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

——— (2011): "Peer effects in education: How might they work, how big are they and how much do we know thus far?" in *Handbook of the Economics of Education*, Elsevier, vol. 3, 249–277.

TENG, M. AND H. LIANG (2012): "A Survey of Current Conditions of Large-Number Classes in the Primary and Secondary Schools in Dengzhou City, Henan Province," in *Annual Report on China's Education, Beijing, China*, ed. by D. Yang, Social Sciences Acadmic Proess (China), in Chinese.

URQUIOLA, M. (2006): "Identifying class size effects in developing countries: Evidence from rural Bolivia," *Review of Economics and Statistics*, 88, 171–177.

VILLANI, C. (2008): *Optimal transport: old and new*, vol. 338, Springer Science & Business Media.

ZARATE, R. A. (2019): "Social and Cognitive Peer Effects: Experimental Evidence from Selective High Schools in Peru," MIT Job Market Paper.

Online Appendix

A Measurement Errors and Standard Errors

A.1 Measurement Error

Measurement error is a well-known problem in psychometrics, especially given the fact that non-cognitive traits are typically more difficult to measure than cognitive ones. A common way to tackle this problem is to combine many measurements to form an index, in order for the measurement errors in each questionnaire item to cancel each other. We form this index by simple averaging.

While useful, simple averaging is not going to eliminate all measurement errors, and hence our estimates suffer from attenuation bias. Though because we already find large estimates, we prefer to leave them as they are.

The literature on measurement errors discusses alternative methods other than simple averaging. The first is to optimize the weights in minimizing the attenuation bias. If one is willing to make the stronger asumption of independence between measurement errors across questionnaire items, as most factor models do, then

$$\varepsilon_{itk} \perp\!\!\!\perp \varepsilon_{itk'}, \varepsilon_{j(i,t)tk} \perp\!\!\!\perp \varepsilon_{j(i,t)t,k'} \forall k \neq k'$$

This assumption implies that the latent factor account for all cross-measurement correlations. Given this assumption, simple average tends to average out the measurement errors, although the weights may not be optimal because the measurement errors may not necessarily have the same variances.

The econometric problem of optimal weighting is discussed in Black and Smith (2006), who measured an unobserved college quality with multiple noisy proxies. Black and Smith (2006) estimated an optimized

δ, found by minimizing the resulting expected mean deviation from the true latent factor. Nevertheless, even after optimizing the weights, measurement error still exists. The motivation of Black and Smith (2006) to optimize the weighting is that the index obtained from simple averaging contains too much measurement error so that it is not predictive. After optimizing the weights, they find that the index now has statistically significant effects of college quality on various individual outcomes.

Alternatively, the cross-measurement independence yields enough moments to achieve point identification of the factor loadings. This approach has also been employed in the skill formation literature. Heckman et al. (2006); Cunha et al. (2010) examined the joint identification of their model without calculating a factor score for each individual. Here we show the point identification of our peer effect model (for a linear case).

Conditional on $Y_{i0}, Y_{j(i,0)0}$, we compute the following partial variance-covariance matrix:

$$\begin{bmatrix} \tilde{var}(Y_{i1}) & \tilde{cov}(Y_{i1}, \hat{d}_{i01}) & \dots & \tilde{cov}(Y_{i1}, \hat{d}_{i0K}) \\ \tilde{cov}(\hat{d}_{i01}, Y_{i1}) & \tilde{var}(\hat{d}_{i01}) & \dots & \tilde{cov}(\hat{d}_{i01}, \hat{d}_{i0K}) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{cov}(\hat{d}_{i0K}, Y_{i1}) & \tilde{cov}(\hat{d}_{i0K}, \hat{d}_{i01}) & \dots & \tilde{var}(\hat{d}_{i0K}) \end{bmatrix}$$
$$= \begin{bmatrix} \beta_3^2 \tilde{var}(\hat{d}_{i0}) + \tilde{var}(u_{i0}) + \tilde{var}(u_{j(i,0)0}) & \beta_3 \rho_{01} \tilde{var}(\hat{d}_{i0}) & \dots & \beta_3 \rho_{0K} \tilde{var}(\hat{d}_{i0}) \\ \beta_3 \rho_{01} \tilde{var}(\hat{d}_{i0}) & \rho_{01}^2 \tilde{var}(d_{i0}) + 2\tilde{var}(\varepsilon_{i01}) & \dots & \rho_{01}\rho_{0K} \tilde{var}(\hat{d}_{i0}) \\ \vdots & \vdots & \ddots & \vdots \\ \beta_3 \rho_{0K} \tilde{var}(d_{i0}) & \rho_{0K}\rho_{01} \tilde{var}(d_{i0}) & \dots & \rho_{0K}^2 \tilde{var}(d_{i0}) \end{bmatrix} \tag{17}$$

Standardize $\rho_{01} = 1$, such that the latent non-cognitive skill N_{it} shares the same unit as the first measurement $\hat{d}_{i01}$. Then the remaining factor loadings $\rho_{02}, \rho_{03}, \dots, \rho_{0K}$ are identified. For example,

$$\rho_{02} = \frac{\tilde{cov}(Y_{i1}, \hat{d}_{i02})}{\tilde{cov}(Y_{i1}, \hat{d}_{i01})} \tag{18}$$

Given that $\tilde{cov}(\hat{d}_{i01}, \hat{d}_{i02}) = \rho_{02} var(d_{i0})$, $\tilde{cov}(Y_{i1}, \hat{d}_{i02}) = \beta_3 \rho_{02} \tilde{var}(d_{i0})$

$$\beta_3 = \frac{\tilde{cov}(Y_{i1}, \hat{d}_{i02})}{\tilde{cov}(\hat{d}_{i01}, \hat{d}_{i02})} \tag{19}$$

i.e. β_3 is identified as the instrumental variable estimator, by using the second measurement to instrument the first. After β_3 is identified, $\beta_0, \beta_1, \beta_2$ can be identified by regressing $Y_{i1} - \beta_3 \hat{d}_{i0}$ on $Y_{i0}, Y_{j(i,0)0}$. Given that other measurements are available, the system is over-identified.

As a result of peer randomization, $N_{i0} \perp\!\!\!\perp N_{j(i,0)0}$ and so are their respective measurements. Finally, with the factor loadings identified, a standard application of the Kotlarski's lemma identifies the distribution of d_{i0}. While this approach yields exact identification, it assumes that the measurement errors are orthogonal to each other.

A.2 Clustered Standard Errors and Wild Bootstrap

Unlike standard representative samples, peer data sets are inherently non i.i.d. because peers (classmates) interact with each other. Since a peer effect model asserts that both own and peer observable inputs would jointly determine outcomes, the same needs to be considered for unobservable inputs as well. This consideration would imply that the composite error terms $u_{i0} + \lambda_p u_{p(i)0}$ are correlated across individuals within the same class. Specifically, given that the deskmate of one's deskmate is oneself, a mechanical correlation between the composite error term across individuals exists even if $u_{i0} \perp\!\!\!\perp u_{p(i)0}$.

The violation of the i.i.d. assumption implies that the OLS standard error is incorrect, often understating the truth (Moulton, 1986). A well-known solution to this problem is to cluster the standard error by class. One problem in practice is that the cluster-robust standard errors (CRSE) are derived by assuming asymptotics with respect to the number of clusters (rather than individuals), while in field experimental studies the number of clusters is often small. In Lu and Anderson (2015), only

12 classes are available, while in our study we have 21.[32] For a small-cluster correction, we follow Cameron and Miller (2015) and use cluster wild bootstrap standard error in our regressions, which is a way to produce bootstrap re-samples by swapping residuals within clusters.

Nevertheless, we find that the cluster wild bootstrap standard errors are rather close to CRSE, such that the qualitative conclusions of testing tne null hypothesis of zero peer effects are unaffected. Meanwhile the OLS standard errors are about half of the cluster-bootstrap standard errors, thereby implying that our corrections are necessary.

B Randomization Check

As discussed above, randomization guarantees the independence between own and peer variables, whether observable or not. For observables, this condition is testable. Given that the randomization in this study is done by the schools following our instructions, the purpose of this randomization check is to confirm whether the teachers have gone through the randomization procedure as instructed.

Following Sacerdote (2001), we check whether the own and deskmate baseline non-cognitive skills are uncorrelated given the baseline randomization, conditional on location (row $\times$ column) and class fixed effect, since the randomization is conditional on these variables. The corresponding regression results are organized in Table 13. All Big Five proxies have passed the randomization check, with the p-values all greater than 0.05. We also present in the same table the results for baseline test scores and birth year; these variables also pass the randomization check.

As the only exception, the male dummy fails the randomization check, with a coefficient of -0.33 and a standard deviation of 0.03. This result

[32]In principle, the cluster unit should be the randomization blocks. However, the blocks include a very limited number of observations, thereby rendering the estimation of cluster-specific variances unreliable.

implies that the deskmate is more likely of the opposite gender than of the same gender.

Table 13: Randomization Check

	deskmate coeff	deskmate (se)	deskmate (p-value)
Openness	-0.05	0.04	0.18
Conscientiousness	-0.07	0.03	0.05
Extraversion	0.04	0.04	0.23
Agreeableness	-0.05	0.04	0.15
Neuroticism	-0.03	0.04	0.42
Birth Year	-0.06	0.04	0.07
Male	-0.33	0.03	0.00
Chinese Language (Baseline)	0.03	0.04	0.46
Math (Baseline)	0.03	0.04	0.33

This abnormality warrants a discussion. As one possibility, it could be due to a ”small urn problem” as discussed in Guryan et al. (2009) — Specifically, given own gender, the remaining pool of students in the same block are more likely to be of the opposite gender because the sampling is without replacement. The resulting regression coefficient with respect to baseline gender would be negative. In our case, each block is relatively small, mostly with less than 20 students. Therefore, this problem is even more acute than that in Guryan et al. (2009), which studies golfer teams of bigger size. Guryan et al. (2009) proposed a fix to this small urn problem, that is, to add an additional control of the proportion of males within a block (i.e., the eligible peers) excluding own. As Guryan et al. (2009) pointed out, this solution requires significant block size variation to work. In our case, we find insufficient variation in block size within a class. Therefore, we do not attempt to use this solution.

The second possibility is that the randomization is, in fact, conditional on gender given the strong preference of teachers for mixed gender pairs. In any case, we wish to identify peer effects other than the gender effects found in Lu and Anderson (2015). Therefore, for all peer effect regressions in Section 4, we control for gender pairs.

C Consistency of d_{i0} Across Traits

As a robustness check, we examine whether our conclusions critically depends on the definition of the non-cognitive peer difference. This is especially important given that our measure is a composite one that mixes the Big Five. In response, we compute a correlation matrix between the trait-specific $d_{i1} - d_{i0}$. The high correlations (about 0.8) in Table 14 indicate that inter-peer non-cognitive differences are consistent across the Big Five traits. If one pair has large peer difference relative to the average with respect to one Big Five trait, e.g. openness to experience, then a have large peer difference with respect to other traits (e.g. conscientiousness) is likely to be observed as well.

Table 14: Correlation Matrix of Peer Differences in Factor-Specific Indexes (Baseline)

	O (Diff)	C (Diff)	E (Diff)	A (Diff)	N (Diff)
O (Diff)	1.00	0.82	0.84	0.81	0.80
C (Diff)	0.82	1.00	0.80	0.78	0.75
E (Diff)	0.84	0.80	1.00	0.81	0.80
A (Diff)	0.81	0.78	0.81	1.00	0.80
N (Diff)	0.80	0.75	0.80	0.80	1.00

Table 15 presents an analogue correlation matrix for the second round. The correlations unambiguously increase by about 0.05, thereby suggesting that the non-cognitive peer differences become more consistent across factors over time. Given that the second round survey is done after the re-randomization of seats, our findings reject an alternative hypothesis that such high correlations is due to some deskmate pairs copying answers from each other.

Consequently, our peer effect estimates do not depend on whether a specific trait is used. Given that most economists pay special attention on conscientiousness, we report the same peer effect regressions using the conscientiousness peer differences instead in Table 16 and 17 respectively. The signs and magnitudes of the estimates are consistent to our main results.

Table 15: Correlation Matrix of Peer Differences in Factor-Specific Indexes (Second Round)

	O (Diff)	C (Diff)	E (Diff)	A (Diff)	N (Diff)
O (Diff)	1.00	0.86	0.89	0.86	0.85
C (Diff)	0.86	1.00	0.85	0.84	0.85
E (Diff)	0.89	0.85	1.00	0.86	0.87
A (Diff)	0.86	0.84	0.86	1.00	0.85
N (Diff)	0.85	0.85	0.87	0.85	1.00

As another robustness check, this appendix section provides the academic peer effect regression estimates if we replace the Big Five measurements with only conscientiousness. The results are similar to our main regressions.

Table 16: Peer Effect (Chinese, Conscientiousness)

	First Sem. Midterm			First Sem. Final		
	(1)	(2)	(3)	(4)	(5)	(6)
Own Baseline	0.774***	0.765***	0.777***	0.592***	0.552***	0.824***
	(0.047)	(0.043)	(0.072)	(0.058)	(0.058)	(0.102)
Deskmate1 Baseline	−0.029	−0.038	−0.021	0.016	−0.020	0.013
	(0.025)	(0.038)	(0.013)	(0.021)	(0.017)	(0.019)
Non-Cog Diff (Deskmate1)	−0.027	−0.055	−0.017	0.307***	0.484***	−0.001
	(0.103)	(0.133)	(0.036)	(0.062)	(0.140)	(0.035)
Interaction (Deskmate1)	0.003	0.003	0.004	−0.072***	−0.117***	0.001
	(0.026)	(0.037)	(0.008)	(0.015)	(0.038)	(0.008)
Interaction (Neighbor1)	1.110***	1.114***	1.039***	1.682***	2.026***	0.673
	(0.190)	(0.235)	(0.336)	(0.199)	(0.253)	(0.486)
class FE	Yes	Yes	Yes	Yes	Yes	Yes
location FE	Yes	Yes	Yes	Yes	Yes	Yes
gender pair FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	891	450	455	888	448	454
R^2	0.632	0.642	0.492	0.689	0.717	0.544
Adjusted R^2	0.579	0.532	0.334	0.644	0.629	0.402

Notes:
***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
Standard errors are obtained by wild bootstrap and clustered at class level.
All Test Scores are in logs.
(1),(4): Whole sample
(2),(5): Own baseline below the median
(3),(6): Own baseline above the median
Baseline non-cognitive peer differences are standardized.

Table 17: Peer Effect (Math,Conscientiousness)

	First Sem. Midterm			First Sem. Final		
	(1)	(2)	(3)	(4)	(5)	(6)
Own Baseline	0.840***	0.817***	1.212***	0.839***	0.874***	0.981***
	(0.058)	(0.070)	(0.101)	(0.079)	(0.085)	(0.076)
Deskmate1 Baseline	−0.023	−0.053	−0.037*	0.057	0.051	−0.027**
	(0.034)	(0.049)	(0.019)	(0.049)	(0.076)	(0.013)
Non-Cog Diff (Deskmate1)	0.055	0.266*	−0.073	0.362*	0.721**	−0.009
	(0.091)	(0.160)	(0.071)	(0.208)	(0.362)	(0.038)
Neighbor1 Baseline	−0.008	−0.054	0.017	−0.081*	−0.162**	0.003
	(0.020)	(0.036)	(0.016)	(0.046)	(0.082)	(0.008)
Non-Cog Diff (Neighbor1)	0.556**	0.674**	−0.987**	0.350	0.232	0.094
	(0.236)	(0.279)	(0.450)	(0.468)	(0.538)	(0.360)
class FE	Yes	Yes	Yes	Yes	Yes	Yes
location FE	Yes	Yes	Yes	Yes	Yes	Yes
gender pair FEs	Yes	Yes	Yes	Yes	Yes	Yes
Observations	891	469	448	888	467	447
R^2	0.758	0.784	0.539	0.637	0.669	0.469
Adjusted R^2	0.724	0.720	0.392	0.585	0.571	0.300

Notes:

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.
Standard errors are obtained by wild bootstrap and clustered at class level.
All Test Scores are in logs.
(1),(4): Whole sample
(2),(5): Own baseline below the median
(3),(6): Own baseline above the median
Baseline non-cognitive peer differences are standardized.

D Peer Assignment with respect to Gender

In Table 18, we check whether the optimal assignment tends to assign same-gender pairs. Like Table 5, we compare these figures to the corresponding ones under random assignment and report the corresponding ratios. The results are mixed: for Chinese midterm and Math final, our extended peer effect model reports figures that are much lower than 1, thereby indicating that the planner wants to reduce the number of same-sex pairs; however, for Chinese final and Math midterm, the reverse is true.

Nevertheless, one robust conclusion is that the worst assignment is always moving in the opposite direction. This result is expected since the peer assignment problem is linear with respect to the gender pair effects. If one type of gender pair is better than another, then the planner would tend to form the first type of gender pair whenever possible; if the planner minimize the final achievement instead, then the first type of gender pair would be avoided whenever possible. Such forces are constrained if non-cognitive peer difference is considered, but interestingly if the planner is purely optimizing (either maximizing or minimizing) non-cognitive peer difference in class, the results are rather similar to random assignment (the corresponding figures are close to 1). To summarize, it appears that gender pair considerations are almost orthogonal to that concerning non-cognitive peer difference, so that (almost) corner solutions with respect to gender are likely.

Table 18: Number of Same-Sex Pairs Under Various Assignments

		Chinese Midterm	Chinese Final	Math Midterm	Math Final
Full Interaction	1st Quartile	0.204	1.068	1.462	0.382
	Median	0.254	1.134	1.606	0.442
	3rd Quartile	0.443	1.210	1.725	0.594
Univariate	1st Quartile	0.227	0.221	1.657	1.614
	Median	0.262	0.260	1.798	1.773
	3rd Quartile	0.443	0.465	1.865	1.867
Minimizing NCPD	1st Quartile	0.781	0.798	0.725	0.720
	Median	0.853	0.878	0.924	0.934
	3rd Quartile	1.030	1.037	1.138	1.124
Maximizing NCPD	1st Quartile	0.837	0.837	1.006	0.960
	Median	0.962	0.965	1.063	1.050
	3rd Quartile	1.152	1.130	1.146	1.134
Worst Assignment	1st Quartile	1.485	1.517	0.188	0.184
	Median	1.640	1.602	0.250	0.248
	3rd Quartile	1.774	1.757	0.403	0.407
Notes:	All results are relative to the results obtained by random assignment (as ratios).				