Classroom Peer Effects and Academic Achievement: Quasi-Randomization Evidence from South Korea

Changhui Kang*
Department of Economics
National University of Singapore

August 2006
Journal of Urban Economics, forthcoming

* Address: Department of Economics, National University of Singapore, 1 Arts Link, Singapore 117570, Singapore; E-mail: ecskch@nus.edu.sg, Phone: +65-6516-6830, Fax: +65-6775-2646.

Abstract

Endogenous formation of peer groups often plagues studies on peer effects. Exploiting quasi-random assignment of peers to individual students that takes place in middle schools of South Korea, we examine the existence and detailed structure of academic interactions among classroom peers. We find that mean achievement of one's peers is positively correlated with a student's performance (standardized mathematics test score). Employing IV methods, we show that such a relationship is causal: the improvement in peer quality enhances a student's performance. Quantile regressions reveal that weak students interact more closely with other weak students than with strong students; hence their learning can be delayed by the presence of worst-performing peers. In contrast, strong students are found to interact more closely with other strong students; hence their learning can be improved by the presence of best-performing peers. We also examine the implications of these findings for two class formation methods: ability grouping and mixing.

JEL *Classification* : I20, C20

Keywords : Peer, Randomization, Ability Grouping, Quantile Regression

Classroom Peer Effects and Academic Achievement: Quasi-Randomization Evidence from South Korea

Abstract

Endogenous formation of peer groups often plagues studies on peer effects. Exploiting quasi-random assignment of peers to individual students that takes place in middle schools of South Korea, we examine the existence and detailed structure of academic interactions among classroom peers. We find that mean achievement of one's peers is positively correlated with a student's performance (standardized mathematics test score). Employing IV methods, we show that such a relationship is causal: the improvement in peer quality enhances a student's performance. Quantile regressions reveal that weak students interact more closely with other weak students than with strong students; hence their learning can be delayed by the presence of worst-performing peers. In contrast, strong students are found to interact more closely with other strong students; hence their learning can be improved by the presence of best-performing peers. We also examine the implications of these findings for two class formation methods: ability grouping and mixing.

JEL *Classification* : I20, C20
Keywords : Peer, Randomization, Ability Grouping, Quantile Regression

1 Introduction

The effects of peer groups and social interactions play a vital role in various policy debates. Researchers examine the effects of peer group and neighborhood characteristics on the likelihood of teen pregnancy and high-school completion (Evans et al. [19]), teenagers' criminal behavior (Glaeser et al. [22], Ludwig et al. [43]), children's academic achievement and educational outcomes (Aaronson [1], Hoxby [32], Hanushek et al. [30], Angrist and Lang [2]), college students' grade, choice of major (Sacerdote [57], Zimmerman [65]) and occupational choice (Marmaros and Sacerdote [47]), and students' drug/alcohol use (Gaviria and Raphael [21], Kremer and Levy [38]).

Among various dimensions of peer interactions, the effect of classroom/school peers on a student's own academic performance is at the heart of the diverse debates on educational reform. For example, a certain structure of peer interactions among classmates, schoolmates and friends in the residential neighborhood is either implicitly or explicitly assumed in arguments on ability grouping (Argys et al. [4], Betts and Shkolnik [6], Figlio and Page [20], Kang et al. [36]), school desegregation (Angrist and Lang [2], Guryan [27]), school choice (Epple and Romano [18], Cullen et al. [15]) and school competition (Hoxby [33, 34], Epple et al. [17]). Nonetheless, the existence and nature of academic interactions among students remain controversial. While some studies find no significant (or small) peer effects (Hanushek [28], Angrist and Lang [2], Lefgren [42], Stinebrickner and Stinebrickner [58] and Arcidiacono and Nicholson [3]), other studies report significantly positive effects of peer quality on academic achievement (Hoxby [32], Zimmer and Toma [64], Boozer and Cacciola [8], Sacerdote [57], Hanushek et al. [30], Zimmerman [65], Winston and Zimmerman [63]).

Such lack of consensus on peer influence reflects various empirical challenges confronted by studies on peer effects.[1] In the face of difficulties in consistent estimation, random assignment of peers has considerable appeals. Recently, Sacerdote [57], Zimmerman [65], and Stinebrickner and Stinebrickner [58] used random (or quasi-random) variation of peers in college dormitories to explore the relationship between peers' and own academic achievement. There are, however, few studies that exploit similar random assignment of peers at pre-collegiate levels. Conventional wisdom states that the influence of peers is more important to elementary and secondary school students. Although various studies have examined the effects on non-academic outcomes such as teen childbearing, alcohol use, smoking and crimes for students before college, studies that exploit random assignment of peers in an examination of these issues are rare—let alone those that focus on academic interactions among students.[2]

In this study we exploit unique quasi-randomization that takes place in the allocation system of middle school students (Grades 7 to 9) in South Korea—specifically during the mid-1990s. Using such random assignment of peers, we examine the existence and structure of academic

[1] Such challenges include the following: First, a selection problem plagues the empirical analysis. Students with similar personal and family backgrounds form a peer group and self-select into such a group. As a result, total correlation of academic outcomes across classrooms or schools may not be attributed to the effect of peers. Second, a co-movement in the outcomes of a student and her classmates or schoolmates may take place because both are subject to a common institutional environment. For example, a budget cut at a school may exert negative effects uniformly on all students, while noise from a nearby construction site affects all students in the school. Third, errors are likely to be associated with measuring the true peer group of a student. Which group—classroom or school grade—forms the true peers of the student is often ambiguous in the empirical context. Such errors may produce peer effects that are empirically negligible due to the attenuation bias. Fourth, in a learning context, a student's behavior affects the behavior of her classmates and is reciprocally affected by them as well. Thus, a student's outcome is simultaneously determined with the outcomes of peers (the reflection problem à la Manski [44]).

[2] Among others, Boozer and Cacciola [8] exploit random assignment of peers in a US experimental study of class-size reduction (Project STAR) to examine academic peer effects among kindergarten and elementary school students. Hanushek [29], however, questions the true randomness of this experiment.

interactions among students within classroom. The South Korean system of student placement to middle schools and the classrooms during the mid-1990s is particularly interesting. From 1969, when the "leveling policy" was first introduced in secondary education in South Korea, the law requires that elementary school graduates be *randomly* (by lottery) assigned to middle schools—either public or private—in the relevant residence-based school district. In addition, the grouping of students by ability and achievement levels within school is extremely rare due to parental objections and the government's traditional leveling policy of secondary education. As a result, the quality of the student body of each school is fairly similar *within* a school district, while there exist variations in it *across* school districts; the non-grouping (or ability mixing) in school exposes students to a classroom peer group that is nearly exogenously and randomly determined.[3]

Using this randomization in student placement, we present that mean achievement of classroom peers is positively correlated with a student's performance (standardized mathematics test score). Using instrumental variables (IV) methods, we also show that the relationship is causal: the improvement in peer quality enhances a student's performance. In addition, quantile regressions reveal that weak students interact more closely with other weak students than with strong students; hence their learning would be more greatly affected and delayed by the presence of worst-performing peers. In contrast, strong students are found to interact more closely with other strong students; hence their learning can be improved by the presence of best-performing peers. We also examine the implications of these findings for two class formation methods: ability mixing and grouping.

In addition to our empirical findings, we argue that the conventional interpretation about peer effects may mislead a class formation policy (ability mixing and grouping) based on peer interactions. We show below that the conventional interpretation based on a ceteris-paribus experiment fails to be informative to the debate on ability grouping, since it draws a variation in outcome when either average or dispersion of peer quality alone changes partially, keeping the other constant. If a student is relocated across ability mixed and ability grouped classes, she experiences simultaneous changes in both average and dispersion of peer quality. We suggest that a more relevant experiment to the class formation policy is how one divide a *given* pool

[3]Another merit of the South Korean situation is that classrooms are homogeneous in terms of race and ethnicity. In contrast, there may be confounding factors in US studies in which different ability combinations of students often accompany different racial combinations.

of students into classes of different formats, allowing simultaneous changes in both average and dispersion of peer quality. In addition, we argue that, rather than directly supporting ability mixing or grouping, the coefficient of the dispersion of peer quality reveals the relative strength of the influence of weak and strong students on the individual student. The positive coefficient implies a benefit of ability grouping to a strong student; the negative coefficient that of ability mixing to a weak student. Our interpretation of the empirical results gives rise to the policy implication that is directly opposite the conventional one.

The rest of the paper is organized as follows. We review the related literature in Section 2. Section 3 discusses the institutional background. We specify the empirical model in Section 4, and describe the data in Section 5. Section 6 presents the estimation results and their implications for ability mixing and grouping in education. Section 7 concludes the paper.

2 Academic Interactions among Students and Ability Grouping in Education

2.1 Existing Literature on Peer Interactions

Studies concerning the effects of peers on a student's academic achievement have been relatively small, but they recently receive growing attention. For example, interest in social interactions in various arenas is increasing (Brock and Durlauf [9], Moffitt [51]); the class/school formation policy (classroom tracking/ability-grouping/streaming versus untracking/ability-mixing) is frequently debated in the context of educational reform (Argys et al. [4], Betts and Shkolnik [6], Figlio and Page [20]). Early as well as recent studies, however, report mixed results about the presence and magnitude of the effects of peer groups on the individual student.

Hanushek [28] shows there exists little effect of peer group quality on a student's academic outcome. In contrast, Summers and Wolfe [60] and Henderson et al. [31] report significant impacts of average peer quality on a student's achievement. These studies, however, are usually vulnerable to various sources of bias (Moffitt [51]). To overcome limitations arising from endogenous peer-group formation, recent studies exploit either randomized (or semi-randomized) assignment of peers to individual students, or extremely comprehensive data sets that contain rich information on one's educational background (e.g., the characteristics of school, teacher, peer groups and family, a history of one's growth in test scores, etc).

Based on random assignment of roommates in college dormitories, Sacerdote [57] and Zimmerman [65] show a significantly positive association between roommates' quality and a student's own outcome. In contrast, Stinebrickner and Stinebrickner [58] report only limited evidence of the positive association. And, in their study that exploits random assignment of class-level peers arising from Project STAR, Boozer and Cacciola [8] report significantly positive peer influence on elementary school students in Tennessee.

Instead of random assignment of students, several studies employ rich information of data sets to explore the effects of peers. Employing a unique data set on Texas elementary school students, Hanushek et al. [30] find a strong influence of the academic quality of peers (measured by previous academic performance) on a student's own outcome. Hoxby [32] also relies on an exogenous variation of peer groups in Texas elementary schools and finds significant positive effects of peers' achievement. Based on data from San Diego Unified School District, Betts and Zau [7] show positive and significant effects of peers on own achievement growth, while Vigdor and Nechyba [62] report mixed results (either positive or non-existing peer influence) on the basis of North Carolina public schools. Employing an international data base from the International Association for the Evaluation of Educational Achievement (IEA), Zimmer and Toma [64] present significantly positive effects of the quality of peers on a student's own math test score. On the contrary, examining students attending 124 US medical schools, Arcidiacono and Nicholson [3] find no significant association between own and peers' achievements, once the school fixed effects are controlled for in the estimation.

2.2 Connection with Ability Grouping in Education

The existence of peer effects in education illustrates the presence of social interactions.[4] When there is a social interaction, positive (negative) externality that arises from it expands (shrinks) the overall social effect beyond (below) individual marginal effects. Social interactions in education have great policy implications for designing the education system of a nation and class

[4]Academic interactions among school/classroom peers may take place through several channels. High-achieving peers can instruct the classmates and help their learning. Low-achieving peers may interrupt learning of others by disruptive behavior. And, good (or bad) performance and behavior of peers can influence a student's preference in such a way that she desires to resemble the group taking a particular action. For example, strong (weak) peers may set a standard a student expects to achieve (avoid). For a broad discussion of various forms of peer effects and social interactions, see Manski [45] and Glaeser and Scheinkman [23].

organization within a school. For example, in the US as in many other countries, whether to track or detrack classrooms in primary and secondary public school is at the heart of recent debates over educational reform (Argys et al. [4], Betts and Shkolnik [6], Figlio and Page [20]).[5] Here, the structure of peer interactions often serves as a guiding light of the debate.[6]

In the empirical analysis of peer effects, researchers often regress a student's own achievement against the (leave-out) mean outcome of classroom/school-grade peers to explore the relationship between average peer quality and own achievement. In order to examine the effect of heterogeneity of peer quality, they also control for the variance (or standard deviation) of peers' outcomes in addition to average peer quality. Given the estimates, studies interpret that a non-linear concave-form association between average quality of one's peers and own outcome—student performance rising with average quality of peers at a decreasing rate—is evidence supporting ability mixing in education, while a convex-form association implies the superiority of ability grouping.[7] As for the parameter of the variance in peer quality, studies suggest that a positive coefficient implies mixing over grouping, while a negative coefficient grouping over mixing. Zero coefficient means the zero-sum nature of regrouping students.[8]

Unlike such conventional interpretations, we argue that the shape of partial association between own achievement, on the one hand, and the average and dispersion of peer quality, on

[5] Another interesting example of a unique large-scale education reform implemented on a nationwide scale would be the "leveling policy" of secondary education in South Korea, which started from 1969. Under the leveling regime, middle schools and general (non-vocational) high schools—both private and public—which had been strongly stratified by students' ability in the pre-leveling period, are assigned students with varying abilities within a school district. With minor revisions since the inception, South Korea presently maintains the traditional leveling policy in placing students to middle and high schools. The implicit assumption of this reform is that mixing students of varying ability levels within school and classroom would generate a net gain at the national level. For details of the reform and its evaluation for high school, see Chung [14] and Kang et al. [36].

[6] Benabou [5] shows that the efficient and outcome-maximizing method of class formation depends upon the degree of complementarity of a student's own ability and that of his/her peers. Lazear [41] shows that the total size of educational outcome is maximized under student grouping by the likelihood of behaving (an ability measure), when there exists no peer influence on his/her behavior. If a behaving student encourages non-behaving students into behaving and the non-behaving students do not interrupt the learning of the behaving students in a classroom, he points out, the mixed classroom is preferred. Epple and Romano [18] call for empirical evidence of the extent of complementarity of peer ability and own ability in educational production in order to measure the total effect of competition between private and public schools.

[7] Henderson et al. [31], Summers and Wolfe [60], Zimmer and Toma [64] and Zimmerman [65] find concavity in peer effects. Although he does not use an academic outcome as peer characteristics, Glewwe [24] finds a convex relationship and suggests tracking may be a better method. In contrast, Hoxby [32] does not find non-linear peer effects.

[8] Zimmer and Toma [64] and Vigdor and Nechyba [62] show a positive association between dispersion in peer quality and own average achievement. They interpret that the finding supports ability mixing in education. Hanushek et al. [30] find insignificant effects of dispersion in peer quality on average achievement growth and infer that a regrouping of students will have little impact on overall average. In addition to interpretation issues, there may be a concern for consistent estimation for the parameter of heterogeneity of peer quality. When a study uses data from a grouping-based education system, the estimate may be subject to a bias due to endogenous decisions (by parents and schools) to place a student to mixed or grouped classrooms (Hanushek et al. [30, p.541]).

the other, has little direct relevance to the overall effectiveness of ability mixing and grouping. Conventional interpretations attempt to measure the effectiveness of a class format by drawing a ceteris-paribus change in outcome when a student is relocated from one peer group to another that is partially different. For instance, a positive (negative) coefficient on the variance of peer quality is often interpreted as supporting ability mixing (grouping), since a student's movement from a peer group with low heterogeneity—grouping—to another with high heterogeneity—mixing—leads to an increase (decrease) in outcome. However, it is necessary to note that a change in class formats between ability mixing and grouping accompanies that in *average* value of peer quality as well as its variance. Therefore, an evaluation of ability mixing versus grouping requires simultaneous consideration of many features of peer quality distribution—not only average peer quality and its dispersion, but the fractions of weak and strong peers in the classroom (see Figure 1).

In addition, we argue that, rather than directly supporting ability mixing or grouping, the coefficient of heterogeneity of peer quality reveals the relative strength of the influence of weak and strong students on the individual student—complementarity and substitutability according to Benabou [5]. Holding its average fixed, rising (falling) heterogeneity in peer quality means that the student has an increasing (decreasing) degree of contact with both worst-performing and best-performing peers, provided that the distribution of peer quality spreads symmetrically. If the student interacts more closely with best-performing peers than with worst-performing peers, the coefficient will be positive. If a student is more affected by worst-performing peers than by best-performing peers, it will be negative. This line of interpretation leads to policy implications that are the exact opposite of those made by the conventional interpretations. We discuss the interpretation issues in greater detail in section 6.2.

3 Institutional Background

As briefly discussed in the Introduction, since 1969, the year when the leveling policy was first introduced to secondary education, the South Korean system of student allocation to middle schools has been fairly simple and straightforward. A pool of elementary school graduates, who have to attend middle school under the nine-year compulsory education law, are randomly by lottery assigned to either a private or public middle school within a residence-based school district.[9] (As private middle schools are heavily subsidized by the government, they are little different from public schools in administration and curriculum in South Korea.) Ordinary middle schools are not allowed to select incoming students, but only entrusted to educate those assigned to them.[10] A regular school district is usually the size of a municipality; there exist a total of 179 middle school districts in the nation.[11] Once assigned, students are not permitted to change schools within the same district. Those transferring across different school districts due to residential relocations are also required to be randomly assigned to a new school in the new district.

In addition to such a district-level placement of students, within-school grouping according to student ability has been largely avoided by both private and public schools due to parents' strong resistance and the long tradition of the government's egalitarian leveling policy in secondary education. In middle schools, students are placed to classes in the manner of ability mixing at the beginning of the academic year (March 1st). Ability mixing in each classroom are usually attained by organizing each class in a grade in balance of students' achievement of the previous year or through randomization.[12] Once placed, students stay in the classroom throughout the year so that the entire class experiences the same nationally-recommended curriculum over the year. This procedure of student placement causes each class of the same grade to have an almost identical class size and a fairly similar average academic quality of students at the beginning of the year.

As for teacher appointment, a homeroom teacher is placed to each class so as to perform administrative and disciplinary duties, while individual subject teachers visit the classroom to instruct in respective fields. In the following year, students in a grade are reshuffled (again

[9]For the lottery system and an overview of secondary education in South Korea, see Marlow-Ferguson and Lopez [46, South Korea] and OECD [53, Chapters 1 and 2].

[10]Physical education middle schools—unique middle schools for special education—are able to give selective admissions to students. There existed only four such middle schools in Korea that admitted a total of 202 students in 1995. (Since the period of interest in this study is the mid-1990's, statistics are reported for 1995.) Students with special needs are educated in separate classes in ordinary middle schools. These special classes account for only a small part of middle school education. In the entire nation, there were 671 such classes in a total of 51,523 middle-school classes in 1995 (*Yearbook of Educational Statistics 1995, National Statistical Office*).

[11]For example, the Seoul metropolitan region, in which there were a total of 352 (boys-only, girls-only and co-ed) middle schools and 180,698 incoming students in 1995, has eleven school districts for middle school allocation (*Yearbook of Educational Statistics 1995*). Each district on average contains 32 schools to which an average of 16,427 students were randomly assigned. As a result, each school was placed a total of 513 incoming students on average.

[12]For example, if there are 100 students in grade 7 and 10 classes to be formed in grade 8, the student who performed best in the end-of-grade-7 exam is assigned to Class 1, the second-best student to Class 2, the 10th student to Class 10, the 11th student to Class 1 (or 10) and so forth.

in the manner of ability mixing) to meet a new peer group and teachers. This institutional structure and the class formation procedure yield the exogenous formation of one's classroom peer group during the entire middle-school period in South Korea.

With a slight revision in 1996, most of the school assignment and class formation procedures described previously are also practiced nowadays. Staring from 1996, the government slightly modifies the student placement in a district in such a way that it accommodates the needs of individual students and parents that had been ignored in the old system: the government allows the student to indicate her preference for a few (two to five) schools within the relevant school district; the attending school is likewise randomly determined among those indicated by the student. Traditional ability mixing within middle school, however, has changed little since the recent revision of the district-level placement. Our study is not affected by this revision in 1996, because it examines students who attended middle schools during 1995.

4 Empirical Framework

In this paper we examine a model of educational production given by:

$$y_{ij} = \beta_0 + X_i\beta_1 + P_i\beta_2 + Z_i\beta_3 + T_{ij}\beta_4 + \tau_j + u_{ij}. \tag{1}$$

Here, y_{ij} is the standardized math score of student i in school j; X_i is a vector of i's personal and family background; P_i is a vector of i's peer-group variables related to math score (specified below in detail); Z_i is a vector of i's exogenous peer-group variables other than P_i; T_{ij} is a vector of i's teacher, class and school characteristics; τ_j is a school fixed effect of j[13]; and u_{ij} is the random error term.[14] We suppose that u_{ij} is further decomposed into ϵ_{1ij} and ϵ_{2ij}. ϵ_{1ij}

[13] T_{ij} and τ_j control for the association of peer achievements that may arise because students in a school share common environments and resources—the correlated effects (Manski [44]). Z_i controls for peer interactions that can be ascribed to background variables of peers that are exogenously determined (e.g., family income and education level of peers' parents)—the contextual (or exogenous) effects (Manski [44]).

[14] The current specification faces two limitations imposed by the data. First, due to the lack of a measure available in the data, this specification omits the pre-determined ability of the student that is popularly used in value-added models. Thus, if a student's ability is correlated with P_i, our estimate for P_i suffers from bias. (Studies of peer effects and social interactions often use contemporary measures of peers when there is no pre-determined variable available (e.g., Case and Katz [10], Gaviria and Raphael [21]).) Nonetheless, as long as the assignment of peers is independent of a student's ability, which is the basis of this study, potential bias will be relatively small in this study (see Table 2 which shows little correlation between peer quality and a student's own observable traits). Second, again in the absence of information, we do not control for the effect of a homeroom teacher—a teacher who is in charge of administrative and disciplinary matters in the class. Such omission may lead to an overstatement of classroom peer effects. Some related discussions are given in section 4 and footnote

represents such part of u_{ij} that may be correlated with P_i, for example, due to the reflection problem or common class factors, and ϵ_{2ij} is the remainder.

We employ two different sets of P_i to examine the existence and detail structure of peer interactions. First, following many studies on peer effects, we use the average value of math scores of classroom peers excluding own score ($\bar{y}_{-i}$), its square term ($\bar{y}^2_{-i}$) and their standard deviation ($\tilde{y}_{-i}$).[15] (Their coefficients are defined as γ_1, γ_2 and γ_3, respectively.) Adding a square term of the mean value captures potential non-linearity of peer effects. And adding the standard deviation is to examine the effect of dispersion of peer quality on the student's achievement.

The second set of P_i is designed to explicitly distinguish between different views on peer interactions. There are four different (not exhaustive) views on peer interactions: (1) the view that the strong inspire the learning of both the strong and weak; (2) the view that the strong inspire the strong's learning alone, not the weak's; (3) the view that the weak interrupt the weak's learning, not the strong's; (4) the view that the weak interrupt the learning of both the strong and weak. These views are related to complementarity versus substitutability of learning among/between strong and weak students according to Benabou [5]: the more important the lower (upper) tail of the distribution is in shaping the outcome, the greater the *complementarity* (*substitutability*) between individual contributions (p.589).[16] We produce two variables in order to contrast the four views. One is the proportion of weak peers (excluding oneself) within a classroom who are below the 25th percentile of the nationwide math score distribution, and the other is the proportion of strong peers (excluding oneself) who are above the 75th percentile. The proportion of weak peers reveals the degree of complementarity, while that of strong peers the degree of substitutability. We test statistical significance of the two parameters and compare

(17). Of course, the needs to include the pre-determined ability of the student and consider the effect of a homeroom teacher is a possible extension for future research.

[15] Empirical studies have been employing two different units as a reference group of peers: classroom-level peers and grade-level peers. Summers and Wolfe [60], Hoxby [32], Hanushek et al. [30] and Angrist and Lang [2] use grade-level peers, while Zimmer and Toma [64], Boozer and Cacciola [8] and Lefgren [42] rely on classroom-level peers. Both classroom-level and grade-level peers are examined in Betts and Zau [7] and Vigdor and Nechyba [62]. No consensus seems to exist as to which reference group is more important to a student's achievement and which one should be used in empirical analysis. For our analysis, we use classroom-level peers, not grade-level peers. The reason is that (1) the nature of ordinary middle school classrooms in South Korea—students staying in a classroom, while teachers visiting it for instruction—dictates the selection, and that (2) only classroom peers are available from the data.

[16] For example, different methods of producing a same commodity are substitutes for each other, since the best innovation is incorporated into the next generation's know-how; a pair of shoes are complements, since the entire quality is determined by that of a worse side. In a similar vein, Zimmer and Toma [64] and Winston and Zimmerman [63] associate the different views on peer interactions with "peer distance"—how far apart the peers are in their behavior. As explained below in greater detail, the coefficient (γ_3) of the dispersion of peer quality used in the first set of P_i also shows the relative importance of complementarity and substitutability in learning.

their magnitudes to examine which direction of peer interaction is stronger.

There are three major difficulties for consistent estimation. First, it is possible that $E(P_i \tau_j) \neq 0$. This arises because students in a school share similar characteristics that are not measured by controlled variables. It may also take place due to sorting by parents into neighborhoods and school districts. Second, there may be unobservable factors that commonly affect all students in the same class. For example, a good math teacher raises the achievement of a student as well as her classmates. Although we attempt to control for teacher effects by means of observable variables (T_{ij}), they may not be sufficient. Recent research shows that unobservable teacher effects matter in student performance (Rivkin et al. [54], Rockoff [56]). Third, if there truly exist peer interactions, a student's own outcome is affected by the performance of peers, which is also affected by her own behavior (the reflection problem, Manski [44]). This problem especially arises for the relationship between y_{ij} and $\bar{y}_{-i}$, creating simultaneous determination of own outcome and the outcome of peers. This leads us to non-identifiability of the true (causal) parameter γ_1 (Moffitt [51]).

For estimation of peer interactions, we first use OLS methods. To handle non-zero correlation between P_i and τ_j, we employ the fixed-effects method: we control for τ_j by assigning a dummy variable for each j (Hanushek et al. [30], Lefgren [42], Arcidiacono and Nicholson [3]). To the extent that overall differences in achievement across schools are accounted for by school fixed effects, peer effects in our models are identified by between-classroom variation in peer quality within a school. As long as $E(P_i \epsilon_{2ij} | \tau_j) = 0$, which would be a reasonable assumption given the randomization within a school and a school district, we obtain an estimate for β_2, which is not contaminated by the problems due to correlated unobservables and sorting into a school district—two major sources of potential bias in peer effects studies.

Although OLS methods under quasi-randomization of peers has its own merits, the estimate $\widehat{\gamma}_1$ for $\bar{y}_{-i}$ may have yet to show the true causal relationship between peers' and own outcomes for other two reasons ($E(\bar{y}_{-i} \epsilon_{1ij}) \neq 0$): unobservable factors common in a class and the reflection problem. In order to handle such problems, we employ instrumental variables (IV) methods. Specifically, exploiting the fact that both math and science test scores are available for each student in the data, we use the mean science score of peers ($\bar{y}^s_{-i}$) as an IV for the mean math score of peers ($\bar{y}^m_{-i}$). If school fixed effects are controlled for and class-specific factors such as the quality of subject teachers are independent between math and science classes within school,

$\bar{y}^s_{-i}$ will be highly correlated with $\bar{y}^m_{-i}$, but is not likely to influence i's math achievement and its error term (u_{ij}) in equation (1). To the extent that $\bar{y}^s_{-i}$ is related to i's math achievement solely through $\bar{y}^m_{-i}$, it serves as a good IV and enables a causal interpretation for the estimate $\widehat{\gamma}_1$.[17] In addition to $\bar{y}^s_{-i}$, we also add the standard deviation of science scores of peers ($\tilde{y}^s_{-i}$) as an IV for $\bar{y}^m_{-i}$ in order to generate an over-identified model.

In preceding estimations, we specify that the relationship between P_i and y_{ij} are common for every student. On the contrary, the impact of peers is likely to be heterogeneous and vary according to the achievement level of the student. For instance, improvement (deterioration) in peer quality may have a larger positive (negative) impact on weak students than on strong students. And the presence of weak students is likely to be more detrimental for weak students than for strong students. To investigate heterogeneous effects of peer interactions over students with differing ability levels, we employ a (conditional) quantile regression technique. This method highlights the correlation between P_i and y_{ij} at different quantiles of the math-score distribution. To deal with the endogeneity of $\bar{y}^m_{-i}$ in quantile regressions, we employ instrumental quantile regression (IVQR) methods proposed by Chernozhukov and Hansen [12, 13].

5 Data

5.1 Description of the Data

For the empirical analysis, we employ data from the Third International Mathematics and Science Study (TIMSS)—the tests conducted internationally by the International Association for the Evaluation of Educational Achievement (IEA) for 41 countries in 1994 and 1995.[18] For the countries participating in the study, mathematics and science tests were administered to

[17] We thank a referee for suggesting $\bar{y}^s_{-i}$ as an IV for $\bar{y}^m_{-i}$. As the referee points out, a limitation of using $\bar{y}^s_{-i}$ as an IV is that the IV estimate $\widehat{\gamma}_1$ may be biased if there are learning shocks that affect all subjects—for example, capability of the homeroom teacher or poor lighting in the classroom. In the data, we have no good tools to consider them. In addition to $\bar{y}^s_{-i}$, one might want to use exogenous peer variables in Z_i as IVs for $\bar{y}^m_{-i}$, since they may also be correlated with $\bar{y}^m_{-i}$ and exogenous to u_{ij}. In fact, their OLS estimates of equation (1) are not significant if school fixed effects are controlled for (see Table 4 below). There are, however, two problems with this approach. First, the insignificant OLS estimate $\widehat{\beta}_3$ does not necessarily imply the exogeneity of Z_i in (1). The consistency of $\widehat{\beta}_3$ depends upon the exogeneity of other explanatory variables and that of $\bar{y}^m_{-i}$ in particular, the violation of which is the very reason to invoke IV methods. Second, the bulk of empirical research shows that contextual factors have independent effects on student outcome (Jencks and Mayer [35], Chase-Lansdale and Gordon [11], Leventhal and Brooks-Gunn [39]). Recent research on random assignment to a neighborhood also raises questions about independence between Z_i and children's educational outcomes (Kaufman and Rosenbaum [37], Currie and Yelowitz [16], Leventhal and Brooks-Gunn [40]).

[18] See Martin and Kelly [49] and Gonzalez and Smith [25] for details of the database.

three different populations toward the end of the school year: (1) Population 1: 9-year-old students, (2) Population 2: 13-year-old students, (3) Population 3: students in the final year of secondary education. In addition to the tests, TIMSS gathered detailed background information through separate student, teacher and school-principal questionnaires. This information is combined with an individual student's test outcomes. For the empirical analysis, we focus on the mathematics test scores collected for the 13-year-old students who were attending middle school (grades 7 and 8) at the time of the study.[19]

The TIMSS data were constructed by a three-level sampling: school, class and student. And appropriate weights were assigned at each level of sampling.[20] In case of South Korea, just like other ordinary participating countries, a total of 150 middle schools were selected (by the stratified sampling) nationwide, and two classes (one from grades 7 and 8 each) were sampled at random from each school. For other TIMSS-participating countries, all students of each sampled classroom were tested. For South Korea, however, not all students in a sampled classroom participated in the tests. About a third of all students were randomly selected for the tests, and the actual number of tested students varied from 15 to 20. Each student in a sampled classroom was assigned the weight in order to reflect the population of the classroom. In this study, we inflate the raw class-level data by the within-class student weight for estimation.[21] Although we are mainly interested in what happens within a classroom rather than a school or a nation in this study, we also use other weights, when they are more appropriate to show a national-level trend in educational characteristics (e.g., Table 3).

Since the math test was administered to two different grades, we standardize the math score using its grade-specific mean and standard deviation.[22] In order to maintain homogeneity

[19]Two reasons can be offered for our focus on mathematics test scores. First, a large body of literature reports that mathematics test scores have a significant bearing on a student's labor market performance (Grogger and Eide [26], Mumane et al. [52]). Hence mathematics among other subjects is likely to be a major focus of concern by schools, parents and students. Second, as a rule in South Korea, mathematics is instructed by one mathematics teacher to a whole class, while science subjects are taught by several teachers of different fields. Therefore, the influence of teachers, if any, can be better understood in mathematics than in science subjects.

[20]The three different weight variables are the following. The first is School Weighting Factor (WGTFAC1) and its adjustment (WGTADJ1), whose multiplication produces the sampling weight for the school. The second is the Class Weighting Factor (WGTFAC2), which reflects the selection probability of the classroom within the school. The third is Student Weighting Factor (WGTFAC3) and its adjustment (WGTADJ3), whose product shows the selection probability of the individual student within a classroom. Obtained from these weight variables is Total Student Weight (TOTWGT), which shows the sampling weight of an individual student in a country's entire population.

[21]Specifically, we use the integer part of [Student Weighting Factor($WGTFAC3$) $\times$ Student Weighting Adjustment ($WGTADJ3) + 0.5$] to inflate the raw class-level data.

[22]The weighted (by Total Student Weight(TOTWGT)) mean and standard deviation of the scores are 580.0 and 104.1 for grade 7, respectively. The corresponding figures for grade 8 are 610.2 and 108.5, respectively.

of schools, we restrict our analysis to non-rural schools. Due to this restriction, we exclude about 17.4 percent ($= 100 \times \frac{5{,}827-4{,}813}{5{,}827}$) of observations from the unweighted sample and about 13.4 percent ($= 100 \times \frac{15{,}698-13{,}598}{15{,}698}$) from the weighted (by within-class student weight) sample. Descriptive statistics of the main sample used for the analysis are shown in Table 1.

INSERT TABLE 1 HERE.

The overall mean of the standardized math scores is slightly higher than zero, as the sample is restricted to non-rural schools, and the overall standard deviation is nearly one. The average age of students is 13.6, reflecting the design of TIMSS, and the share of male students is 0.56. Around 70 percent of parents are educated at least up to the upper secondary level, and half of the students own more than 100 books in the home. About 40 percent of students have a computer at home. Students are equally divided between the two grades. As for the teacher's characteristics, a half of students are taught by male teachers, and 10 percent is taught by teachers holding a post-graduate degree. The average experience of teachers is 12.3 years. The average class size is 55.8, which is quite high by international standard, though it reflects the corresponding national figure 48.2 (*Yearbook of Educational Statistics, 1995*). (Higher class size of the sample is due to the restriction to non-rural schools.) South Korean middle schools are divided by gender. Students attending boys-only schools occupy 37 percent of the sample, while those attending girls-only and coed schools account for 34 and 29 percent, respectively. Among non-rural school students, 65 percent go to schools close to the center of a town/city.

5.2 Evidence of Randomization

In order to empirically verify the random and exogenous determination of peer groups, we first run regressions of peer variables (P_i) on a student's personal and family traits and her math teacher's characteristics. Note that, as the regressions control for school fixed effects, the coefficients show the average partial correlation between the peer variable and various characteristics within school. If assignment of peers is truly random within school, peer variables should have no systematic association with both observable and unobservable characteristics of a student, as long as the school effects are controlled for. Here, however, we check the correlation between the peer variable and a student's observable traits alone, assuming that no such association implies randomization along the unobservable dimension. Table 2 presents fairly convincing evidence of randomization of peers within South Korean middle school.

INSERT TABLE 2 HERE.

The mean ($\bar{y}_{-i}$) and standard deviation ($\tilde{y}_{-i}$) of math scores of peers have little significant correlation with a student's personal and family traits. The only exception is the dummy for both parents being present in the peer-mean equation, which shows a positive association. Teacher variables are not systematically correlated with peer quality as well. F-tests confirm that the controlled student and teacher characteristics are not jointly systematically associated with the mean and dispersion of peer quality.

The proportions of weak and strong peers within classroom also appear to be uncorrelated with observable student and teacher variables. Exceptions are the dummies of both parents being present and father's education level in the weak-peers-proportion equation. However, father's education level shows an unexpected direction given that educated parents are less likely to put their child in a class with a high proportion of weak peers. Overall, the assignment of peers in South Korean middle school appears to be nearly random as a result of the unique system of student allocation in a school district and the paucity of ability grouping in middle school.

Given that South Korea adopts a unique system of student placement, it would be interesting to see the features of its classrooms from an international perspective. In Table 3 we calculate two indices to show the ability-mixed nature of South Korean middle school and their classroom. We compare the indices of South Korea with those of other Asian and Western TIMSS-participating countries.

INSERT TABLE 3 HERE.

The first index is the proportion of overall variance of math scores that is attributed to within-class (or within-school) variance as opposed to between-class (or between-school) variance.[23] When ability mixing is widespread in a country, we expect a high fraction of within-class

[23] We decompose the total weighted (by Total Student Weight (TOTWGT)) variance of math test scores into the within- and between-classroom (or school) variances for each country, as follows:

$$\sigma^2 = \sum_j F_j \sigma_j^2 + \sum_j F_j (m_j - \bar{m})^2$$

(or school) variance in total national variance of test scores. In contrast, when ability grouping is widely applied, the proportion of within-class (or school) variance will be relatively low, since similar-ability students will be gathered within a classroom (or a school).

For South Korea, over 95 percent of the overall variance of math scores in non-rural schools is explained by within-classroom variance. In other words, less than 5 percent of the overall variance is explained by between-classroom variance. Within-variance occupies the same proportion in overall variance when we use the school as a unit. This implies that in South Korea, non-rural middle schools and their classrooms are attended by students of varying abilities due to the unique placement system and rare ability grouping of students. South Korea's proportion of the within-classroom (or school) variance is the highest among TIMSS-participating countries. To compare with other Asian countries, Japan shows similar features to Korea, while Thailand, Singapore, and Hong Kong display relatively low within-classroom/school variances. Among Western countries, Denmark displays high within-classroom/school variances, while Netherlands, Germany, and U.S.A show low within variances.

The second index is based on the school questionnaire that was administered to school principals to ask whether students followed the same course of study in mathematics. Same (different) courses of study imply ability mixing (grouping). We calculate the weighted (by Total Student Weight (TOTWGT)) proportion of students who were educated under the same course of mathematics as another index of ability mixing for each country. Such a proportion for South Korea and Japan is equal to one, implying that every student experience the same course of math study under little grouping. Singapore among Asian countries, and U.K., U.S.A, and Netherlands among Western countries have low rates of the same course in math, which suggests that students experience grouping according to their different math capabilities.

In sum, the preceding two indices show the ability-mixed nature of middle school and its classroom in South Korea. Institutionally, this stems from nearly random assignment of students to schools and classrooms.

---

where σ^2 is the overall variance of math test scores, F_j is the fraction of students in classroom (or school) j, m_j and σ_j^2 are the mean and variance, respectively, of the test score within j, and $\bar{m}$ is the overall mean. The proportion of the within-variance is given by the ratio of $\sum_j F_j \sigma_j^2$ to σ^2. Here we employ the non-rural sample of each country, and the statistics are weighted by Total Student Weight(TOTWGT) in order to reflect the reality of a country.

6 Estimation Results

6.1 Average Effects of Peers

6.1.1 Association between Mean Score of Peers and Own Average Outcome

INSERT TABLE 4 HERE.

Table 4 shows the estimation results of OLS regressions for equation (1).[24] Columns (1)-(3) do not control for school fixed effects, while columns (4)-(6) include them as explanatory variables. Column (4) is employed in order to indirectly show within-school randomization of peers: when peers are exogenously assigned to a student within school and school fixed effects are controlled for in the estimation, the estimates for P_i should be close across different specifications whether or not student and teacher variables are included.

When we do not control for school effects ($\tau_j = 0$ for all j) and the contextual components of peer effects ($\beta_3 = 0$), the mean value of math scores of peers has a positive correlation with a student's own math score. According to column (2) of Table 4, the estimate $\widehat{\gamma}_1$ implies that a one standard deviation (SD) greater mean math score of peers is associated with a 0.55 SD higher own score. In terms of raw scores, a 106.5-point greater mean score of peers is associated with around a 58.6-point higher math score of a student. Such an association is significantly different from zero. In addition, there is no evidence that the effect of peers' mean outcome has a nonlinear structure as far as the average outcome of a student is concerned. The dispersion of peers' scores also fails to have a significant impact on a student's own average achievement.

When the contextual components (Z_i) are additionally controlled for in column (3), the degree of peer interactions rises and the estimate $\widehat{\gamma}_1$ becomes 0.653 (s.e. 0.032). The proportion of male and other background variables of classroom peers are jointly significant. Although students are randomized across classrooms within school, they may not be randomly assigned across school districts due to endogenous residential sorting. This concern invokes the possibility of $E(P_i\tau_j) \neq 0$ and calls for the control for school fixed effects.

When school fixed effects are included in the regressions, the magnitude of association between peers' mean outcome and own achievement falls by more than a half. When no student

[24] Since there are multiple observations for each school and each class, the standard errors for OLS and IV estimates are adjusted for robustness and clustering at the classroom level. When the school level is used for clustering adjustment, the results remain qualitatively similar.

and teacher variables are controlled for in the regression, the estimate $\widehat{\gamma}_1$ is 0.344 (s.e 0.058) in column (4). This amount is fairly close to and statistically indistinguishable from the corresponding estimates (in columns (5) and (6)) obtained when student and teacher variables are included in the estimation. Such closeness of the estimates indirectly shows near randomization of peers within South Korean middle school. When student and teacher variables are controlled for in the regressions, the estimate $\widehat{\gamma}_1$ remains between 0.258 and 0.267, whether or not contextual influences, which become negligible once the school effects are controlled for, are considered.[25] The estimates imply that a one SD greater mean score of peers is associated with a 0.26 to 0.27 SD higher own score.[26] And they are significantly different from zero. As previously discussed, these amounts may not be interpreted as causal due to unobservable factors common in a class and the reflection problem. Nonetheless, they are informative of true peer interactions because the estimates are less vulnerable to residential sorting by parents and student grouping in school.

Similar to the results without school fixed effects, there is no evidence suggesting a nonlinearity of peer effects with regard to average math score of a student. In addition, a student's own outcome is not significantly affected by the dispersion of peers' scores within classroom. The finding that $\gamma_1 > 0$, $\gamma_2 = 0$ and $\gamma_3 = 0$ is often interpreted as suggesting that regrouping of students by the achievement level may be a zero-sum game where gains and losses between weak and strong students are offset without generating extra overall benefits. We, however, argue that the implication of this finding for the mixing-versus-grouping controversy requires further consideration, as the change in class format accompanies the shift in the entire distribution of peer quality. We revisit this issue when we discuss quantile regression results.

[25] Insignificant contextual peer impacts are in contrast with studies reporting their (positive) effects on a student's academic outcome (Vandenberghe [61], McEwan [50], Robertson and Symons [55]). In view of our study, their estimates appear to be subject to an upward bias as they fail to completely remove unobservable components shared by peers.

[26] In the math score term, our estimates indicate that a 1-point greater mean score of peers is associated with about a 0.26-point higher own score. This amount of peer interaction is slightly lower than, but comparable to the estimates of other studies using US elementary schools and an international data set. As for the results based on US elementary schools, Hoxby [32] presents a 0.1 to 0.55-point increase in own score in association with a 1-point increase in mean score of peers. Boozer and Cacciola [8, Table 4] show roughly a 0.6-point greater own math score in association with a 1-point increase in mean score of classroom peers. Hanushek et al. [30, Table I] show about a 0.4-point increase in own math score in relation to a 1-point increase in mean score of peers. Vigdor and Nechyba [62, Table 6] report a 1-point increase in peers' mean score is associated with a 0.07-point increase in own math score. From an international data set, Zimmer and Toma [64, Table 1] show that a 1-point higher mean score of peers leads to a 0.6-0.8 point increase in own math score.

6.1.2 Causal Effect of Peer Quality on Student Outcome

INSERT TABLE 5 HERE.

Here we examine the results of IV estimations explained in section 4. OLS results of the reduced form equations and IV results of the structural equations are presented in Table 5. For simplicity of analysis, only the linear term $\bar{y}_{-i}$ is controlled for in estimation, while school fixed effects are also included.

Similar to that in column (6) of Table 4, the OLS estimate $\widehat{\gamma}_1$ is 0.254 (s.e. 0.059) in column (1) of Table 5. When the mean science score of peers ($\bar{y}^s_{-i}$) is included as an extra regressor, its estimate in column (2) is not significantly different from zero, while the estimate $\widehat{\gamma}_1$ slightly falls. If SD of peers' science scores ($\tilde{y}^s_{-i}$) is added, its estimate in column (3) is not significantly different from zero as well. Although examining the significance of the estimates for $\bar{y}^s_{-i}$ and $\tilde{y}^s_{-i}$ in OLS regressions is not a formal test for exogeneity of instruments—the consistency of their OLS estimates depends upon the exogeneity of other explanatory variables, the findings may be informative of the possibility that both $\bar{y}^s_{-i}$ and $\tilde{y}^s_{-i}$ are uncorrelated with u_{ij} in equation (1).

The first-stage results of the reduced form equations for $\bar{y}^m_{-i}$ are given in columns (4) and (5). SD of peers' science scores ($\tilde{y}^s_{-i}$) is not included in column (4), while included in column (5). When $\bar{y}^s_{-i}$ alone is used as an IV for $\bar{y}^m_{-i}$ as in column (4), the estimate for $\bar{y}^s_{-i}$ implies that a one SD greater mean science score of peers is associated with a 0.64 SD higher mean math score of peers. And the correlation is significantly different from zero. Since the F-statistic for the significance of $\bar{y}^s_{-i}$ is 168, which is far above a rule-of-thumb threshold, 10 (Stock et al. [59]), $\bar{y}^s_{-i}$ is not a weak instrument for $\bar{y}^m_{-i}$. When $\tilde{y}^s_{-i}$ is added as another IV for $\bar{y}^m_{-i}$ as in column (5), the estimate for $\bar{y}^s_{-i}$ has similar size to that in column (4) and it is significant. The estimate for $\tilde{y}^s_{-i}$ is, however, statistically insignificant while it is negatively related with $\bar{y}^m_{-i}$. In column (5) $\bar{y}^s_{-i}$ and $\tilde{y}^s_{-i}$ are jointly strong instruments, since the F-statistic for joint significance of these two variables is 85. Thus, as long as $\bar{y}^s_{-i}$ as well as $\tilde{y}^s_{-i}$ is exogenous to u_{ij}, they are legitimate IVs for $\bar{y}^m_{-i}$.

IV estimates $\widehat{\gamma}_1$ are shown in columns (6) and (7). The estimates in column (6) are based on the first-stage specification in column (4), while those in column (7) are based on column (5). IV estimates $\widehat{\gamma}_1$ suggest that an improvement in peer quality reflected by a one SD increase in mean math score of peers enhances a student's own math score by a 0.31 SD. In TIMSS scale, a 106.5-point increase in mean math score of peers leads to a 33-point improvement in own math score. Furthermore, such a relationship is quite precisely estimated. From the results in column (7) we can apply an over-identification test; the test statistic and its p-value are 0.309 and 0.579, respectively. Thus the test does not reject the exogeneity of both instruments $\bar{y}^s_{-i}$ and $\tilde{y}^s_{-i}$; since they are strongly correlated with $\bar{y}^m_{-i}$, this implies that both instruments are valid IVs for $\bar{y}^m_{-i}$.

Once IV estimates are obtained, we can also apply Hausman tests for the exogeneity of $\bar{y}^m_{-i}$ in equation (1). From the results in columns (6) and (7), the tests do not reject the exogeneity of $\bar{y}^m_{-i}$. Therefore, we conclude that the endogeneity of $\bar{y}^m_{-i}$ in equation (1) is not a great concern for consistent estimation, as long as school fixed effects are controlled for in estimation.

6.1.3 Effect of Strong and Weak Peers on Own Average Outcome

While one acknowledges the presence of peer academic interactions within classroom, she may be interested in relative importance of strong and weak peers to a student. Strong students are likely to interact more closely with other strong students than with weak students, in which case their performance will be little affected by weak peers—learning is substitutable among strong students. Alternatively, weak students may be more severely (than strong students) affected by weak peers within classroom—learning is complementary among weak students. Although a full investigation of this issue requires quantile regressions, it is necessary to check the overall patterns with respect to the average achievement of a student. Table 6 shows the estimation results.

INSERT TABLE 6 HERE.

When we do not control for school effects, both proportions of strong peers and weak peers are shown to have a strong correlation with a student's own outcome. From columns (2) and (3), a 10-percentage-point higher proportion of weak peers within classroom is associated with a 0.07 to 0.09 SD lower own math score. A 10-percentage-point higher proportion of strong peers is associated with a 0.07 to 0.08 SD higher own math score. Both of these results are statistically significant at one percent level. And a test does not reject symmetric impacts of strong and weak peers on a student. This implies that a change in own outcome due to an increasing proportion of weak peers is offset by the same amount of a decreasing proportion of strong peers.

When school effects are considered, the patterns remain unaltered, although the magnitude declines by more than a half. The proportion of weak peers decreases a student's own outcome, while that of strong students increases it. And the assignment of peers appears nearly random, as the coefficients of P_i are fairly close across different specifications. Whether or not student and teacher variables are controlled for, a 10-percentage-point increase in weak (strong) peers is associated with roughly a 0.03-0.04 SD lower (higher) own math score. Statistical significance, however, remains at the borderline if student and teacher variables are included in the regression. There also exist symmetric impacts of strong and weak peers on a student's average score. In other words, on average, complementarity is as large as substitutability for a student's learning.

6.2 Effects of Peers at Different Quantiles

In section 6.1, we suppose that the coefficients of peer variables are common for all levels of student ability. Unlike this assumption, the impacts of peers are likely to vary, depending on the achievement level of the student. Such heterogeneity of peer interactions can be considered in quantile regressions (QR). The potential endogeneity of $\bar{y}^m_{-i}$ is taken into account by instrumental quantile regressions (IVQR).[27]

6.2.1 Effect of the Mean Quality of Peers on Various Quantiles

INSERT TABLE 7 HERE.

Table 7 shows the effects of mean and SD of math scores of peers at five quantiles: 0.1, 0.25, 0.5, 0.75 and 0.9 quantiles. Panel A reports the estimates of ordinary QR that does not control for school effects, while panel B reports those of ordinary QR that does. Panel C presents the estimates of IVQR that controls for school effects. For all estimations, we use a linear term of $\bar{y}^m_{-i}$, since its square terms are rarely significant.

Similar to the findings for the average outcome in Table 4, the QR estimates for $\bar{y}^m_{-i}$ fall at each and every quantile by more than a half when school fixed effects are controlled for. This

[27]In order to consider clustering at the class level in calculating QR and IVQR standard errors, we employ a clustered bootstrap method: we first draw students with replacement from the sample of classrooms; QR and IVQR are performed for each bootstrap sample to calculate the standard deviation of the estimates. The number of bootstrap samples is 50.

suggests that sorting of students across schools takes place for all levels of student achievement. When school effects are taken into account, the performance of weak students at the 0.1 and 0.25 quantiles is strongly related to mean achievement of classroom peers. A one SD greater mean math score of peers is associated with a 0.36 and 0.24 SD higher math score at the 0.1 and 0.25 quantiles, respectively. The math score of the median student is also strongly correlated with mean performance of peers. In addition, the performance of strong students at the 0.75 and 0.9 quantiles is similarly related to mean quality of classroom peers. A one SD greater mean math score of peers is associated with a 0.24 and 0.21 SD higher math score at the 0.75 and 0.9 quantiles, respectively.

Dispersion in peers' quality also shows impacts that substantially vary according to the achievement level of the student. The estimates in panel B imply that a 1-unit greater SD of math scores of peers decreases a student's score at the 0.1 quantile by a 0.73 SD and at the 0.25 quantile by a 0.2 SD. In contrast, a 1-unit greater SD of math scores of peers increases the 0.75 quantile by a 0.28 SD and the 0.9 quantile by a 0.33 SD. That is, when the average peer quality is kept constant, increased heterogeneity of peer quality harms weak students, but benefits strong students.

IVQR estimates of $\bar{y}^m_{-i}$ are presented in panel C. According to Chernozhukov and Hansen [12, 13], IVQR is implemented in the following two steps[28]:

(1) Define $\widehat{\Phi}_i$ is the OLS projection of $\bar{y}^m_{-i}$ on all exogenous variables including IVs.[29] For a given quantile index q, define a grid of values $\{\gamma^h_1,\ h = 1, \cdots, H\}$, and run the ordinary q-quantile regression of $(Y_{ij} - \bar{y}^m_{-i}\gamma^h_1)$ on X_i, $\tilde{y}^m_{-i}$, Z_i, T_{ij}, τ_j and $\widehat{\Phi}_i$ to obtain their estimates.

(2) Choose $\gamma_1(q)$ as the value among $\{\gamma^h_1,\ h = 1, \cdots, H\}$ that makes the estimate for $\widehat{\Phi}_i$ in (1) closest to zero. The estimates of other variables are obtained from the same QR in (1) whereby $\gamma_1(q)$ is produced.

According to Table 7, the IVQR estimates of $\tilde{y}^m_{-i}$ only slightly change from the ordinary QR estimates, but the IVQR estimates of $\bar{y}^m_{-i}$ in panel C show some differences from the ordinary QR estimates in panel B. While the IV 0.1-quantile estimate $\widehat{\gamma}_1$ is similar to its ordinary QR counterpart, the IVQR estimates for γ_1 increases by more than 50 percent at the 0.25 and 0.5

[28]Computer codes for performing IVQR are obtained from the web site "http://www.gsb.uchicago.edu/fac/christian.hansen /research/." We thank Christian Hansen for sharing the computer codes.

[29]For our IVQR estimation, we use both $\bar{y}^s_{-i}$ and $\tilde{y}^s_{-i}$ as IVs for $\bar{y}^m_{-i}$. Employing $\bar{y}^s_{-i}$ alone as an IV does not alter the main results. Such results are available upon request.

quantiles. The estimates suggest that a one SD increase in mean quality of peers enhances the math scores at the 0.25 and 0.5 quantiles by 0.47 and 0.42 SD, respectively. And the estimates are quite precisely estimated. In contrast, the IVQR estimates for γ_1 for the 0.75 and 0.9 quantiles reduce close to zero and even become negative, although not significant. These estimates suggest that mean quality of peers does not affect math achievement of strong students at the 0.75 and 0.9 quantiles.

Nevertheless, our IVQR results are largely in line with ordinary QR results. While strong students above the 0.75 quantile are not affected by mean quality of peers, weak and median students around and below the 0.5 quantile are strongly affected by it. Thus, there is asymmetry of peer interactions across strong and weak students with respect to mean peer quality. As for the effect of peer heterogeneity, when mean peer quality is kept constant, rising heterogeneity of peer quality harms weak students, but benefits strong students. To summarize, weak students are susceptible to peer influence: they are affected by both mean and dispersion of peer quality. In contrast, strong students are less susceptible: they are not affected by mean peer quality, but by dispersion of peer quality alone.

6.2.2 Interpretation Issues

Based on a hypothetical experiment that attempts to measure a change in a student's outcome when she is relocated between mixed and grouped (or tracked) classes, many studies on peer effects suggest that the observed partial association between $\bar{y}_{-i}$ and $\tilde{y}_{-i}$, on the one hand, and y_{ij}, on the other, have direct policy implications for ability mixing and grouping in education (Henderson et al. [31], Glewwe [24], Zimmer and Toma [64], Hanushek et al. [30], Vigdor and Nechyba [62]). Given a pool of students, however, a mixed classroom provides the student with a medium average peer quality and high heterogeneity, while a grouped classroom provides her with either high or low average peer quality and low heterogeneity (see Figure 1[30]).

INSERT FIGURE 1 HERE.

[30]From the figure, when the class format changes from a mixing to a grouping system given the same pool of students, the distribution of a student's peer quality changes from (1) to (2) if she is relocated to a high-track classroom, and from (1) to (3) if relocated to a low-track classroom. Compared with a mixed classroom, the average and dispersion of peer quality is small and the fraction of weak (strong) peers is high (low) in a low-track classroom. Again, compared with a mixed classroom, the average of peer quality is large, its dispersion is small and the fraction of weak (strong) peers is low (high) in a high-track classroom.

Given that the entire distribution of peer quality varies between ability mixed and grouped classrooms, rather than directly supporting ability mixing, the condition that $\gamma_1 > 0$, $\gamma_2 < 0$ and $\gamma_3 > 0$ simply implies that a student benefits from mixing when she moves from a low-track classroom—with low average and low heterogeneity of peer quality—to a mixed classroom—with medium average and high heterogeneity of peer quality. (Recall that γ_1, γ_2 and γ_3 are the coefficients of $\bar{y}_{-i}$, $\bar{y}^2_{-i}$ and $\tilde{y}_{-i}$, respectively, in (1).) If the student switches from a high-track to a mixed classroom, she may experience a loss. Overall effectiveness of mixing and grouping depends on the proportions of those experiencing gains and losses and their relative sizes. In a similar manner, rather than supporting grouping, the condition that $\gamma_1 > 0$, $\gamma_2 > 0$ and $\gamma_3 < 0$ suggests that a student benefits (may suffer) from grouping when she moves from a mixed to a high-track (low-track) classroom. The condition that $\gamma_1 > 0$, $\gamma_2 = 0$ and $\gamma_3 = 0$ does not necessarily imply a zero-sum nature of regrouping students. It is possible that $\gamma_3 < 0$ for weak students and $\gamma_3 > 0$ for strong students drive $\gamma_3 = 0$ at the mean (or at the median). Finally, an interpretation based on either γ_1 and γ_2, or γ_3 alone is even more ambiguous.

Rather than having a direct bearing on ability mixing and grouping in education, we argue that γ_3 reveals the nature of learning that arises from interactions among students—complementarity and substitutability. That is, a positive γ_3 shows substitutability in learning, and a negative γ_3 complementarity. (Recall that the more important the lower (upper) tail of the distribution is in shaping the outcome, the greater the *complementarity* (*substitutability*).) Holding its average fixed, rising (falling) heterogeneity in peer quality means that a student has an increasing (decreasing) degree of contact with both worst-performing and best-performing peers, provided that the distribution of peer quality spreads symmetrically. If a student interacts more closely with best-performing peers than with worst-performing peers—learning is substitutable, γ_3 will be positive. If a student is more affected by worst-performing peers than by best-performing peers—learning is complementary, γ_3 will be negative.

Our interpretation of the estimates gives rise to policy implications that are conflicting with the conventional ones. As shown in Table 7, a negative γ_3 is shown for weak students, while a positive γ_3 for strong students. According to conventional interpretations (e.g., Vigdor and Nechyba [62, p.22]), weak students are better off when they are educated in a grouped classroom (most likely a low-track) rather than in a mixed classroom. And strong students benefit from a mixed classroom rather than from a grouped classroom (most likely a high-track). On the contrary, our interpretation indicates that weak students are likely to be *worse off* in a grouped classroom. Under a grouping system, weak students are likely to be assigned to a low-track classroom in which they are exposed to a more frequent contact with other weak and worst-performing peers; hence they suffer negative impacts of grouping.[31] Similarly, *grouping* is likely to enhance the performance of strong students, since they will have more frequent contact with best-performing peers in a high-track classroom; hence they enjoy positive impacts of grouping. The median student showing $\gamma_3 = 0$ seems to display complementarity and substitutability in an offsetting manner.[32] They will gain from the transfer from a mixed classroom to a high-track, but will lose from the movement to a low-track.

As noted by Benabou [5] and Lazear [41], the overall performance of mixing and grouping systems is determined by the degree of complementarity and substitutability in education. Given that the strength of each function varies according to the ability level of the student, it is difficult to determine uniformly one system over the other. A specific design of each system in terms of the mean and dispersion of peers' quality and the fractions of strong and weak students makes a difference in overall performance of each system. Therefore, the ultimate evaluation of the two class formats depends on the detail design of each.

In the Appendix, we attempt a tentative assessment of the overall performance of ability grouping and mixing, employing a simulation based on the raw data and the estimates obtained from quantile regressions. In this evaluation, we try to consider explicitly the difference in entire distribution of peer quality between the two systems (e.g., the fractions of strong and weak students in the classroom). The simulation results suggest that ability mixing may perform better than grouping in terms of mean outcome.[33]

[31] Here, we ignore a possibility that, as proponents of grouping argue, schools adopting grouping align their resources and instruction styles to the ability level of the student. If this is a case, we cannot rule out a possibility that a negative impact of grouping on weak students can be overturned.

[32] Alternatively, $\gamma_3 = 0$ for the median student may also mean that there exist no complementarity and substitutability for her learning. Given that each of them is found to function for other weak and strong students, however, we believe this interpretation makes less sense. We show below that the median student simultaneously and closely interacts with both weak and strong peers in a similar degree.

[33] We wish to emphasize that this evaluation is nothing but tentative. First, as mentioned in footnote (31), once the grouping policy is adopted in a school, the school and teachers will surely align and optimize their resources and instruction styles to take into account the type and ability level of children in each class. Such changes are completely ignored in the simulation. Second, our current estimates of peer variables are obtained using relatively small degrees of variation in peer quality under a mixing system. Thus if the level of peer quality under hypothetical grouping is located outside the boundary currently observed in the data, the current estimates may no longer be valid, and the final outcome of each system may be biased. Third, quantile effects are different from average effects. While the latter can be used to build a counterfactual outcome of the same individual, the former may not be used for the same purpose. In the simulation, we have abused QR estimates as if they were OLS estimates. For an example of the evaluation for ability mixing as opposed to grouping based on an actual

6.2.3 Effect of Strong and Weak Peers on Various Quantiles

INSERT TABLE 8 HERE.

Using the second set of P_i, we confirm the interpretation suggested above and the asymmetric nature of learning for strong and weak students. In Table 8, we show that learning is complementary among weak students at lower quantiles, but substitutable among strong students at higher quantiles. And the complementarity among the weak is found to be greater than substitutability among the strong.

According to the bottom panel, for weak students at the 0.1 and 0.25 quantiles, a 10-percentage-point increase in weak peers is associated with a 0.07 SD lower own math score. However, the increase in strong peers within classroom does not make a significant difference in own outcome of weak students. For strong students at the 0.75 and 0.9 quantiles, a 10-percentage-point increase in strong peers is associated with a 0.04 and 0.05 SD higher own math score, respectively. While the 0.9 quantile is positively affected by weak peers,[34] the 0.75 quantile is not significantly related with the proportion of weak peers. Such findings largely support the second and third view of peer interactions in section 4—the view that the strong inspire the strong's learning alone, not the weak's, and that the weak interrupt the weak's learning, not the strong's.

In addition, the degree of complementarity among weak students is much greater than that of substitutability among strong students. We see a similar trend in Table 7 in which the 0.1-quantile estimate for γ_3 is larger in absolute value than the 0.9-quantile counterpart. This is suggestive of larger detrimental effects of ability grouping on weak students than its beneficial effects on strong students.

The median student is found to interact with both weak and strong peers. And the negative effect of weak peers seems larger than the positive effect of strong peers. (The claim fails to be statistically significant at the conventional levels.) That is, learning appears to be slightly more complementary than substitutable for the median student.

policy change, see Kang et al. [36].

[34] This finding is a bit puzzling. Nevertheless, it is not inexplicable given a possibility that for a strong student weakest peers might represent a challenge and a chance to learn-by-teaching (Winston and Zimmerman [63]).

7 Concluding Remarks

Studies on peer effects are confronted and sometimes plagued by various empirical challenges. Random assignment of peers to individual students is often wanted for consistent estimation. Exploiting near randomization of peers that takes place in middle school of South Korea, we examine whether there exist academic interactions among classroom peers and what form they specifically take.

First, we find that there do exist academic interactions within classroom. A student's achievement (standardized math score) is significantly positively associated with average performance of classroom peers. Employing IV methods, we show that such a relationship is causal: the improvement in peer quality reflected by a one SD increase in mean math score of peers enhances a student's own math score by a 0.30 SD.

Second, using quantile regressions, we also show that the dispersion as well as the average level of peer quality is significantly associated with students of different achievement levels. Achievement of weak students is negatively correlated with the dispersion of peer quality within classroom, while that of strong students is positively correlated. Unlike conventional views that attempt to directly draw a policy implication for ability mixing and grouping from such findings, we suggest that these correlations reveal the nature of learning for different-ability students: weak students display complementarity and strong students substitutability in learning. Contrary to earlier interpretations, this implies that weak students are likely to benefit from ability mixing, while strong students from grouping.

Even if we suggest that ability mixing may benefit weak students and ability grouping strong students, it is often difficult to choose between ability mixing and grouping as the output-maximizing class (or school) organization method. Overall performance of ability mixing and grouping is dependent on many factors: the relative strength of complementarity and substitutability, and the design and student composition of mixed and grouped classrooms. Using a simulation, we attempt a tentative assessment for ability mixing and grouping, and show that mixing may perform better than grouping in terms of mean outcome.

The lack of the pre-determined ability of the student in our data may expose this study to a limitation of omitted variable bias. Nonetheless, we believe that the potential bias is relatively small, since nearly random assignment of peers within middle school of South Korea will minimize a possible correlation between peer variables and the student's ability. Nonetheless, an investigation that includes the pre-determined ability of a student will be a topic for future research.

Acknowledgments

The author is grateful to two anonymous referees for many helpful suggestions. Useful comments have also been received from seminar participants at the National University of Singapore, Korean Educational Development Institute (KEDI) and Korea Research Institute for Vocational Education and Training (KRIVET).

Appendix

A Tentative Assessment of Mixing and Grouping: A Simulation Approach

In the text we argue that an evaluation of the relative effectiveness of ability mixing and grouping is limited when it is based on the relationship between average and dispersion of peer quality, and the student's own outcome alone.

For an evaluation of ability mixing versus grouping, we propose a new estimation model:

$$y_{ij} = \beta_0 + X_i\beta_1 + \sum_{k=1}^{4} F_{ki}\delta_k + Z_i\beta_3 + T_{ij}\beta_4 + \tau_j + u_{ij} \tag{2}$$

where F_1 is the fraction of classroom peers whose math scores are below the 25th percentile of the nationwide math-score distribution; F_2 the fraction of those between the 25th and 40th percentiles; F_3 the fraction of those between the 60th and 85th percentiles; and F_4 the fraction of those above the 85th percentile. The fraction of those between the 40th and 60th percentiles are dropped to avoid collinearity. Using F_k's, we attempt to directly characterize the distribution of peer quality within classroom and its potential shift when class formats are changed between ability mixing and grouping. The smaller the range of F_k, the more closely the distribution of peer quality is approximated, but the less precise the estimates of δ_k's from the OLS and quantile regressions. The OLS and quantile estimates are shown in Appendix Table 1. They generally agree with Table 8.

INSERT Appendix Table 1 HERE.

Given these quantile estimates, we regroup students currently under mixed classrooms to form two tracked (or grouped) classrooms (one is a high-track and the other is a low-track) in order to compare overall performance between the mixing system and hypothetical tracking systems. A current mixed classroom is divided into two tracked classrooms as follows: within each existing classroom, students whose math score is above α quantile ($\alpha \in [0, 1]$) of within-classroom distribution of math score are assigned to the high-track classroom and those below α quantile are assigned to the low-track classroom. By design, the regrouping under α being equal to 0 or 1 reduces to the existing mixed classrooms. We experiment with five different values of α: (1) Tracking I under $\alpha = 0.1$; (2) Tracking II under $\alpha = 0.25$; (3) Tracking III under $\alpha = 0.5$; (4) Tracking IV under $\alpha = 0.75$; (5) Tracking V under $\alpha = 0.9$.

In the absence of OLS estimates that associate peer quality with own achievement separately for different ability levels of students, we use the quantile estimates to obtain the predicted achievement of each student under the new peer environment. Specifically, given the quantile estimates and a new set of F_k's, T and Z under the hypothetical tracking systems, we replace β and δ in (2) with their quantile estimates, and obtain the predicted achievement of different-ability students. Since we do not have estimates for every quantile, we employ a rule whereby the 0.1 quantile estimates are assigned to those students whose scores are below the 15th percentile of within-school distribution of math score, the 0.25 quantile estimates to those between the 15th and 40th percentiles, the 0.5 quantile estimates to those between the 40th and 60th percentiles, the 0.75 quantile estimates to those between the 60th and 85th percentiles, and the 0.9 quantile estimates to those above the 85th percentile.[35]

Summary statistics of peer variables and the hypothetical achievement of students from five tracking systems are given in Appendix Table 2.

INSERT Appendix Table 2 HERE.

[35]There are two problems with this method. First, as we rely on quantile estimates, we basically generate predicted conditional quantiles of the outcome distribution—not conditional expectations of a student's outcome—under the hypothetical tracking systems. Thus an intra-student comparison between the actual score and the quantile-based predicted score may not be valid. In the absence of valid estimates, we assume that the quantile estimates are such estimates that relate peer quality with own achievement of the student of different ability levels. High-quantile estimates are used for high-ability students and low-quantile estimates for low-ability students. Second, when there is a strong and substantial interaction between students, the quantile estimates of δ_k's may be biased due to $E(F_{kij}\epsilon_{1ij}) \neq 0$. In the simulation, we also suppose that the quantile estimates ($\widehat{\delta}_k$) are consistent and that they show causal relations between y_{ij} and F_k's.

The overall average and standard deviation of $\bar{y}_{-i}$ are larger under the hypothetical tracking than under the existing mixing system. As α increases, the overall average of $\bar{y}_{-i}$ declines and its standard deviation displays an inverted-U shape. Not surprisingly, the average of $\tilde{y}_{-i}$ is smaller under tracking. It is smallest when $\alpha = 0.5$. When we look into each track under tracking, the averages of $\bar{y}_{-i}$ and $\tilde{y}_{-i}$ are lower in low-tracks than in the existing mixing system, while they rise as α increases. The average of $\bar{y}_{-i}$ is higher and that of $\tilde{y}_{-i}$ are lower in high-tracks than in the existing mixing system, while the former rises and the latter falls as α increases.

Overall performance of tracking in terms of the average of hypothetical y_{ij}'s varies under different tracking systems. Given the current estimates and method of dividing students within school, the mixing system usually produces higher average outcomes for all levels of α. Depending on α, the overall average of y_{ij} is greater under mixing than under tracking by 0.01 to 0.09 SD. When α is equal to 0.1 and 0.25, the average gap between mixing and tracking is relatively large (0.04 and 0.09 SD, respectively). The average gap is smaller when α is greater than 0.5.

As α increases, the average gap of low-track students declines as their classroom becomes closer to the mixed classroom. The average achievement gap of high-track students shows an inverted-U shape with the highest gain at $\alpha = 0.75$, as α rises. From these findings, we can infer that when α is low, that is, when the weakest students are concentrated on the low-track, complementarity in learning within the low-track dominates substitutability within the high-track. As α increases, that is, weak students are less concentrated in the low-track, degree of complementarity declines relative to substitutability, while the latter fails to dominate the former in our simulation.

That our simulation shows better overall outcomes under ability mixing than under ability grouping (tracking) does not indicate the universal superiority of ability mixing. As previously emphasized, overall performance of ability mixing and grouping is dependent on relative strength of complementarity and substitutability, and the design and student composition of mixed and grouped classrooms. Difficulties exist in choosing a priori one over another in educating students, unless substantially detailed information becomes available. Moreover, the current comparison between mixing and grouping systems is made in terms of the difference in mean outcome, which is obtained by giving an equal weight to every student. An evaluation becomes even more difficult if we allow the weight to vary by student ability in order to reflect a society's view on the role of able and less able students.

References

[1] D. Aaronson, Using sibling data to estimate the impact of neighborhoods on children's educational outcomes, Journal of Human Resources 33 (1998) 915-946.

[2] J.D. Angrist, K. Lang, Does school integration generate peer effects? Evidence from Boston's Metco program, American Economic Review 94 (2004) 1613-1634.

[3] P. Arcidiacono, S. Nicholson, Peer effects in medical school, Journal of Public Economics 89 (2005) 327-350.

[4] L.M. Argys, D.I. Rees, D.J. Brewer, Detracking America's schools: Equity at zero cost?, Journal of Policy Analysis and Management 15 (1996) 623-645.

[5] R. Benabou, Heterogeneity, stratification, and growth: Macroeconomic implications of community structure and school finance, American Economic Review 86 (1996) 584-609.

[6] J.R. Betts, J.L. Shkolnik, The effects of ability grouping on student achievement and resource allocation in secondary schools, Economics of Education Review 19 (2000) 1-15.

[7] J.R. Betts, A. Zau, Peer groups and academic achievement: Panel evidence from administrative data, Unpublished Manuscript, 2004

[8] M.A. Boozer, S. Cacciola, Inside the 'black box' of Project STAR: Estimation of peer effects using experimental data, Economic Growth Center Discussion Paper 832, Yale University, 2001.

[9] W. A. Brock, S.N. Durlauf, Interactions-based models, in: J.J. Heckman, E. Leamer (Ed.), Handbook of Econometrics, Volume IV, North-Holland: Amsterdam, 2001, pp. 3297-3380.

[10] A.C. Case, L.F. Katz, The company you keep: The effects of family and neighborhood on disadvantaged youths, NBER Working Paper No.3705, 1991.

[11] L.P. Chase-Lansdale, R.A. Gordon, Economic hardship and the development of five- and six-year-olds: Neighborhood and regional perspectives, Child Development 67 (2005) 3338-3367.

[12] V. Chernozhukov, C. Hansen, An IV model of quantile treatment effects, Econometrica 73 (2005) 245-262.

[13] V. Chernozhukov, C. Hansen, Instrumental quantile regression inference for structural and treatment effect models?, Journal of Econometrics (2006) forthcoming.

[14] B.-G. Chung, A Study of the High School Leveling Policy in the Republic of Korea: Its Genesis, Implementation, and Reforms. 197l-1995, Ph. D. thesis. University of Hawaii, 1998.

[15] J.B. Cullen, B.A. Jacob, S. D. Levitt, The impact of school choice on student outcomes: An analysis of the Chicago Public Schools, Journal of Public Economics 89 (2005) 729-760.

[16] J. Currie, A. Yelowitz, Are public housing projects good for kids?, Journal of Public Economics 75 (2000) 99-124.

[17] D. Epple, E. Newlon, R. Romano, Ability tracking, school competition and the distribution of educational benefits, Journal of Public Economics 83 (2002) 1-48.

[18] D. Epple, R. Romano, Competition between private and public schools, vouchers, and peer-group effects, American Economic Review 88 (1998) 33-62.

[19] W.N. Evans, W.E. Oates, R. M. Schwab, Measuring peer group effects: A study of teenage behavior, Journal of Political Economy 100 (1992) 966-991.

[20] D.N. Figlio, M.E. Page, School choice and the distributional effects of ability tracking: Does separation increase equality?, Journal of Urban Economics 51 (2002) 497-514.

[21] A. Gaviria, S. Raphael, School-based peer effects and juvenile behavior, Review of Economics and Statistics 83 (2001) 257-268.

[22] E.L. Glaeser, B. Sacerdote, J. A. Scheinkman, Crime and social interactions, Quarterly Journal of Economics 111 (1996) 507-548.

[23] E.L. Glaeser, J.A. Scheinkman, Measuring social interactions, in: S.N. Durlauf, H.P. Young (Ed.), Social Dynamics, MIT Press: Cambridge, MA, 2001, pp. 83-102.

[24] P.W. Glewwe, Estimating the impact of peer group effects on socioeconomic outcomes: Does the distribution of peer group characteristics matter?, Economics of Education Review 16 (1997) 39-43.

[25] E.J. Gonzalez, T.A. Smith, User Guide for the TIMSS International Database: Primary and Middle School Years, Boston College, MA: TIMSS International Study Center, 1997.

[26] J. Grogger, E. Eide, Changes in college skills and the rise in the college wage premium, Journal of Human Resources 30 (1995) 280-310.

[27] J. Guryan, Desegregation and black dropout rates, American Economic Review 94 (2004) 919-943.

[28] E.A. Hanushek, Teacher characteristics and gains in student achievement: Estimation using micro data, American Economic Review 61 (1971) 280-288.

[29] E.A. Hanushek, Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects, Educational Evaluation and Policy Analysis 21 (1999) 143-163.

[30] E.A. Hanushek, J.F. Kain, J.M. Markman, S.G. Rivkin, Does peer ability affect student achievement?, Journal of Applied Econometrics 18 (2003) 527-544.

[31] V. Henderson, P. Mieszkowski, Y. Sauvageau, Peer group effects and educational production functions, Journal of Public Economics 10 (1978) 97-106.

[32] C.M. Hoxby, Peer effects in the classroom: Learning from gender and race variation. NBER Working Paper No.7867, 2000a.

[33] C.M. Hoxby, Does competition among public schools benefit students and taxpayers?, American Economic Review 90 (2000b) 1209-1238.

[34] C.M. Hoxby, School choice and school competition: Evidence from the United States, Swedish Economic Policy Review 10 (2003) 11-67.

[35] C. Jencks, S. Mayer, The social consequences of growing up in a poor neighborhood, in: L.E. Lynn, M.G.H. McGeary (Ed.), Inner-city poverty in the United States, National Academy Press: Washington, DC., 1990, pp. 111-186.

[36] C. Kang, C. Park, M.-J. Lee, Effects of ability mixing in high school on adulthood earnings: Quasi-experimental evidence from South Korea, Journal of Population Economics forthcoming.

[37] J.E. Kaufman, J.E. Rosenbaum, The education and employment of low-income black youth in white suburbs, Educational Evaluation and Policy Analysis 14 (1992) 229-240.

[38] W. Kremer, D.M. Levy, Peer effects and alcohol use among college students. NBER Working Paper No.9876, 2003

[39] T. Leventhal, J. Brooks-Gunn, The neighborhoods they live in: The effects of neighborhood residence on child and adolescent outcomes, Psychological Bulletin 126 (2000) 309-337.

[40] T. Leventhal, J. Brooks-Gunn, A randomized study of neighborhood effects on low-income children's educational outcomes, Developmental Psychology 40 (2004) 488-507.

[41] E.P. Lazear, Educational production, Quarterly Journal of Economics 116 (2001) 777-803.

[42] L. Lefgren, Educational peer effects and the Chicago public schools, Journal of Urban Economics 56 (2004) 169-191.

[43] J. Ludwig, G.J. Duncan, P. Hirschfield, Urban poverty and juvenile crime: Evidence from a randomized housing mobility experiment, Quarterly Journal of Economics 116 (2001) 655-679.

[44] C. Manski, Identification of endogenous social effects: The reflection problem, Review of Economic Studies 60 (1993) 531-542.

[45] C. Manski, Economic analysis of social interactions, Journal of Economic Perspectives 14 (2000) 114-136.

[46] R. Marlow-Ferguson, C. Lopez, World Education Encyclopedia: A Survey of Educational Systems Worldwide (2nd edition). Detroit, MI: Thomson Gale, 2001.

[47] D. Marmaros, B. Sacerdote, Peer and social networks in job search, European Economic Review 46 (2002) 870-879.

[48] D. Marmaros, B. Sacerdote, How do friendships form?, Quarterly Journal of Economics 121 (2006) 79-119.

[49] M.O. Martin, D.L. Kelly, TWSS technical report Volume I: Design and development. Technical report, International Association for the Evaluation of Educational Achievement, Boston College, Chestnut Hill, MA., 1996.

[50] P.J. McEwan, Peer effects on student achievement: Evidence from Chile, Economics of Education Review 22 (2003) 131-141.

[51] R.A. Moffitt, Policy intcrvcntions, low-level equilibria, and social interactions, in: S.N. Durlauf, H.P. Young (Ed.), Social Dynamics, MIT Press: Cambridge, MA., 2001, pp. 45-82.

[52] R.J. Mumane, J.B. Willett, F. Levy, The growing importance of cognitive skills in wage determination, Review of Economics and Statistics 77 (1995) 251-266.

[53] OECD (Organization for Economic Co-operation and Development), Reviews of National Policies for Education: Korea. Paris: OECD Publications, 1998.

[54] S.G. Rivkin, E.A. Hanushek, J.F. Kain, Teachers, schools, and academic achievement, Econometrica 73 (2005) 417-458.

[55] D. Robertson, J. Symons, Do peer groups matter? Peer group versus schooling effects on academic attainment, Economica 70 (2003) 31-53.

[56] J.E. Rockoff, The impact of individual teachers on student achievement: Evidence from panel data, American Economic Review 94 (2004) 247-252.

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

[58] T.R. Stinebrickner, R. Stinebrickner, What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds, Unpublished Manuscript, 2004.

[59] J.H. Stock, J.H. Wright, M. Yogo, A survey of weak instruments and weak identification in generalized method of moments, Journal of Business and Economic Statistics 20 (2002) 518-529.

[60] A.A. Summers, B.L. Wolfe, Do schools make a difference?, American Economic Review 67 (1977) 639-652.

[61] V. Vandenberghe, Evaluating the magnitude and the stakes of peer effects analysing science and math achievement across OECD, Applied Economics 34 (2002) 1283-1290.

[62] J. Vigdor, T. Nechyba, Peer effects in elementary school: Learning from "apparent" random assignment, Unpublished Manuscript, 2004.

[63] G.C. Winston, D.J. Zimmerman, Peer effects in higher education, NBER Working Paper No.9501, 2003.

[64] R.W. Zimmer, E. Toma, Peer effects in private and public schools across countries, Journal of Policy Analysis and Management 19 (2000) 75-92.

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

Figure 1: Distribution of Peer Quality

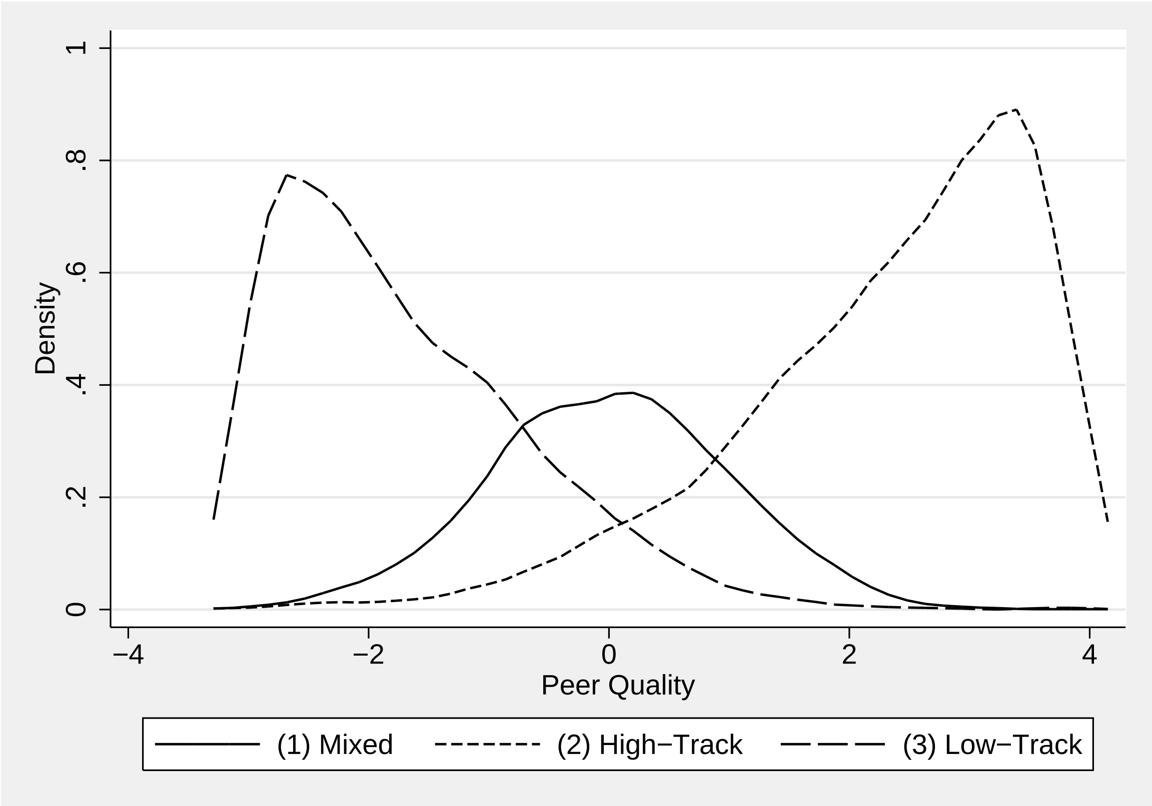

Table 1: Descriptive Statistics

	Unweighted		Weighted	
Individual Characteristics	Mean	S.D.	Mean	S.D.
Mathematics Test Score	600.1	(106.7)	601.2	(106.5)
Standardized Score	0.07	(0.99)	0.08	(0.99)
Age	13.61	(0.60)	13.61	(0.60)
Male	0.56	(0.50)	0.56	(0.50)
Both Parents Present	0.87	(0.33)	0.87	(0.33)
Mother's Education				
0-6 Years	0.10	(0.30)	0.10	(0.29)
7-9 Years	0.23	(0.42)	0.22	(0.42)
10-12 Years	0.40	(0.49)	0.40	(0.49)
13-14 Years	0.06	(0.23)	0.06	(0.23)
15+ Years	0.13	(0.34)	0.13	(0.34)
Not Reported	0.09	(0.28)	0.09	(0.28)
Father's Education				
0-6 Years	0.06	(0.24)	0.06	(0.23)
7-9 Years	0.15	(0.35)	0.14	(0.35)
10-12 Years	0.38	(0.48)	0.38	(0.48)
13-14 Years	0.09	(0.29)	0.09	(0.29)
15+ Years	0.25	(0.43)	0.25	(0.43)
Number of Books at Home				
0-10	0.08	(0.27)	0.07	(0.26)
11-25	0.09	(0.28)	0.09	(0.28)
26-100	0.33	(0.47)	0.33	(0.47)
101-200	0.25	(0.43)	0.25	(0.43)
201+	0.25	(0.43)	0.25	(0.44)
Computer at Home	0.40	(0.49)	0.41	(0.49)
Grade 8	0.50	(0.50)	0.50	(0.50)
Teacher Characteristics				
Male	0.50	(0.50)	0.49	(0.50)
Post-graduate Degree	0.10	(0.30)	0.10	(0.30)
Years of Experience	12.41	(9.36)	12.32	(9.39)
Hours of Teaching Activities	8.34	(4.37)	8.37	(4.34)
Frequent Teacher Meeting	0.44	(0.50)	0.44	(0.50)
Gender-matching	0.60	(0.49)	0.60	(0.49)
Class/School Characteristics				
Class Size	19.45	(0.85)	55.81	(6.53)
Boys-Only Class	0.37	(0.48)	0.37	(0.48)
Girls-Only Class	0.34	(0.47)	0.34	(0.47)
Coed Class	0.30	(0.46)	0.29	(0.45)
City Center Location	0.64	(0.48)	0.65	(0.48)
Peer Group Characteristics				
Mean Math Score	0.07	(0.31)	0.08	(0.30)
SD of Math Score	0.95	(0.16)	0.93	(0.16)
Mean Science Score	0.04	(0.29)	0.05	(0.28)
SD of Science Score	0.97	(0.17)	0.95	(0.16)
Proportion below 25th pt	0.25	(0.12)	0.25	(0.12)
Proportion above 75th pt	0.25	(0.12)	0.25	(0.12)
Proportion of Male	0.56	(0.46)	0.56	(0.46)
Father's Education	12.38	(1.28)	12.43	(1.26)
Mother's Education	11.29	(1.26)	11.33	(1.26)
Books over 200	0.50	(0.16)	0.51	(0.16)
Computer at Home	0.40	(0.16)	0.41	(0.15)
Number of Students	4,813		13,598	
Number of Schools	124		124	
Number of Classes	248		248	

Note: The weight variable is Wihtin-Class Student Weight, which is equal to the integer part of (WGTFAC3*WGTADJ3+0.5).

Table 2: Evidence of Randomization: Regression of Peer Variables on Student and Teacher Characteristics

	Dependent Variables			
	Mean of Math Score	SD of Math Score	Proportion Below 25th pt	Proportion Above 75th pt
Individual Characteristics				
Age	-0.002 (0.006)	-0.001 (0.006)	0.001 (0.003)	-0.003 (0.003)
Male	0.023 (0.034)	0.014 (0.023)	-0.014 (0.013)	-0.004 (0.015)
Both Parents Present	0.015 (0.006)*	0.000 (0.005)	-0.005 (0.003)*	0.003 (0.003)
Mother's Education				
7-9 Years	-0.004 (0.008)	-0.006 (0.006)	-0.001 (0.004)	0.000 (0.003)
10-12 Years	0.005 (0.009)	-0.007 (0.007)	-0.006 (0.004)	0.001 (0.004)
13-14 Years	0.011 (0.014)	0.003 (0.012)	-0.002 (0.006)	0.010 (0.008)
15+ Years	-0.001 (0.013)	-0.012 (0.009)	-0.008 (0.006)	0.002 (0.005)
Not Reported	0.002 (0.011)	0.002 (0.008)	-0.002 (0.005)	0.004 (0.004)
Father's Education				
7-9 Years	0.000 (0.007)	0.006 (0.006)	0.004 (0.004)	0.003 (0.003)
10-12 Years	-0.009 (0.007)	0.002 (0.006)	0.007 (0.004)*	-0.001 (0.003)
13-14 Years	0.000 (0.010)	-0.011 (0.010)	0.005 (0.005)	0.002 (0.005)
15+ Years	-0.014 (0.010)	0.006 (0.006)	0.013 (0.004)**	-0.004 (0.005)
Number of Books				
11-25	-0.007 (0.016)	0.001 (0.007)	0.000 (0.007)	-0.006 (0.005)
26-100	-0.010 (0.012)	-0.003 (0.006)	0.003 (0.005)	-0.007 (0.004)
101-200	0.001 (0.012)	0.002 (0.008)	0.001 (0.005)	0.002 (0.005)
201+	0.004 (0.012)	-0.002 (0.007)	-0.002 (0.005)	-0.001 (0.005)
Computer at Home	-0.002 (0.004)	0.000 (0.003)	0.002 (0.002)	0.000 (0.002)
Grade 8	0.005 (0.029)	-0.001 (0.021)	-0.021 (0.012)	-0.009 (0.013)
Teacher Characteristics				
Male	-0.031 (0.052)	0.011 (0.028)	0.002 (0.022)	-0.017 (0.023)
Post-graduate Degree	0.012 (0.069)	-0.015 (0.048)	-0.010 (0.027)	0.021 (0.032)
Years of Experience	-0.004 (0.002)	0.001 (0.002)	0.002 (0.001)*	-0.002 (0.001)
Hours of Teaching Activities	-0.006 (0.006)	0.000 (0.004)	0.003 (0.003)	-0.003 (0.002)
Frequent Teacher Meeting	-0.001 (0.044)	-0.054 (0.031)	-0.022 (0.020)	-0.010 (0.021)
Gender-matching	0.042 (0.042)	0.010 (0.022)	-0.014 (0.018)	-0.001 (0.018)
Intercept	0.544 (0.245)*	1.304 (0.204)**	0.261 (0.114)*	0.455 (0.103)**
School Fixed Effects	Yes	Yes	Yes	Yes
Number of Students	13,305	13,305	13,305	13,305
R-Square	0.753	0.487	0.686	0.690
F-test: joint sig of student and teacher variables (p-value)	0.085	0.278	0.026	0.034

Note: The regressions are weighted by Within-Class Student Weight, which is equal to the interger part of (WGTFAC3*WGTADJ3+0.5). Robust and cluster-adjusted standard errors are in parentheses.
* and ** indicate the estimate is significant at the 0.05 and 0.01 levels, respectively.

Table 3: Proportion of Within-Class/Within-School Variance of Mathematics Score and Prevalence of Grouping by Country

	Variance Decomposition		
Country	Within-Classroom Variance	Within-School Variance	Same Course of Math
South Korea	0.954	0.954	1.000
Japan	0.888	0.888	1.000
Thailand	0.664	0.699	0.774
Singapore	0.559	0.658	0.205
Hong Kong	0.535	0.573	1.000
Denmark	0.898	0.915	0.971
U.K	0.750	0.753	0.371
Switzerland	0.593	0.606	0.673
Netherlands	0.538	0.592	0.395
Germany	0.531	0.520	0.795
U.S.A	0.525	0.690	0.143

Source: Third International Mathematics and Science Study, 1995
Note: Total Student Weight (TOTWGT) and non-rural sample are used for calculation.

Table 4: OLS Estimation Results: Mean and SD of Peers' Scores and Own Average Math Scores

	OLS: No School Fixed Effects			OLS: School Fixed Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Peer Group Characteristics						
Mean Math Score	0.811 (0.020)**	0.550 (0.028)**	0.653 (0.032)**	0.344 (0.058)**	0.267 (0.055)**	0.258 (0.059)**
Mean Math Score^2	0.044 (0.041)	-0.051 (0.059)	-0.006 (0.049)	-0.062 (0.108)	-0.066 (0.094)	-0.041 (0.094)
SD of Math Score	-0.010 (0.026)	0.047 (0.051)	0.012 (0.043)	0.004 (0.063)	0.028 (0.071)	0.034 (0.073)
Proportion of Male			-0.271 (0.075)**			-0.228 (0.102)*
Father's Education			-0.046 (0.012)**			-0.043 (0.021)*
Mother's Education			0.024 (0.011)*			0.026 (0.023)
Books over 200			-0.201 (0.067)**			-0.014 (0.112)
Computer at Home			-0.150 (0.059)*			-0.137 (0.101)
Individual Characteristics						
Age		-0.045 (0.039)	-0.040 (0.040)		-0.040 (0.040)	-0.036 (0.041)
Male		0.139 (0.044)**	0.277 (0.071)**		0.230 (0.056)**	0.280 (0.070)**
Both Parents Present		0.096 (0.042)*	0.099 (0.042)*		0.098 (0.042)*	0.107 (0.043)*
Mother's Education						
7-9 Years		-0.022 (0.052)	-0.023 (0.057)		-0.021 (0.053)	-0.026 (0.058)
10-12 Years		0.048 (0.056)	0.038 (0.061)		0.056 (0.058)	0.040 (0.062)
13-14 Years		-0.195 (0.082)*	-0.215 (0.088)*		-0.198 (0.084)*	-0.216 (0.089)*
15+ Years		0.041 (0.067)	0.054 (0.075)		0.059 (0.069)	0.052 (0.076)
Not Reported		-0.146 (0.063)*	-0.140 (0.065)*		-0.139 (0.065)*	-0.141 (0.066)*
Father's Education						
7-9 Years		0.053 (0.054)	0.054 (0.061)		0.055 (0.054)	0.055 (0.062)
10-12 Years		0.135 (0.054)*	0.157 (0.065)*		0.144 (0.054)**	0.157 (0.065)*
13-14 Years		0.074 (0.068)	0.105 (0.078)		0.089 (0.069)	0.107 (0.079)
15+ Years		0.310 (0.062)**	0.350 (0.073)**		0.328 (0.064)**	0.347 (0.073)**
Number of Books						
11-25		0.174 (0.068)*	0.143 (0.070)*		0.172 (0.070)*	0.145 (0.072)*
26-100		0.515 (0.055)**	0.501 (0.056)**		0.521 (0.056)**	0.508 (0.058)**
101-200		0.722 (0.060)**	0.711 (0.062)**		0.733 (0.061)**	0.713 (0.063)**
201+		0.801 (0.058)**	0.791 (0.059)**		0.818 (0.059)**	0.798 (0.060)**
Computer at Home		0.168 (0.031)**	0.165 (0.032)**		0.175 (0.032)**	0.163 (0.032)**
Grade 8		0.072 (0.038)	0.065 (0.039)		0.072 (0.040)	0.071 (0.042)
Teacher Characteristics						
Male		-0.011 (0.017)	-0.008 (0.013)		-0.031 (0.027)	-0.023 (0.026)
Post-graduate Degree		0.014 (0.024)	0.005 (0.016)		0.021 (0.037)	0.015 (0.038)
Years of Experience		-0.002 (0.001)**	-0.001 (0.001)		-0.003 (0.001)**	-0.003 (0.001)*
Hours of Teaching Activities		0.001 (0.002)	0.000 (0.001)		-0.001 (0.003)	-0.003 (0.003)
Frequent Teacher Meeting		0.007 (0.015)	0.002 (0.011)		0.005 (0.023)	0.011 (0.023)
Gender-matching		0.012 (0.017)	0.017 (0.015)		0.030 (0.030)	0.029 (0.030)
Intercept	0.021 (0.024)	-0.309 (0.520)	-0.118 (0.526)	0.337 (0.109)**	-0.306 (0.560)	-0.129 (0.615)
Class/School Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Number of Students	13,598	13,305	12,964	13,598	13,305	12,964
R-Square	0.064	0.184	0.183	0.072	0.194	0.190

Note: The regressions are weighted by Within-Class Student Weight, which is equal to the integer part of (WGTFAC3*WGTADJ3+0.5).
Robust and cluster-adjusted standard errors are in parentheses. * and ** indicate the estimate is significant at the 0.05 and 0.01 levels, respectively.

Table 5: OLS and IV Estimation Results: Mean and SD of Peers' Scores and Own Average Math Scores

	Reduced from models (OLS)					Structural Model (2SLS)	
Dependent variable:	Own Math Score			Mean Math Score of Peers		Own Math Score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Peer Group Characteristics							
Mean Math Score	0.254 (0.059)**	0.224 (0.074)**	0.226 (0.074)**			0.307 (0.076)**	0.304 (0.077)**
Mean Science Score		0.053 (0.063)	0.054 (0.063)	0.639 (0.049)**	0.636 (0.049)**		
SD of Math Score	0.039 (0.073)	0.051 (0.073)	0.047 (0.074)	0.095 (0.067)	0.101 (0.068)	0.043 (0.068)	0.043 (0.068)
SD of Science Score			0.045 (0.084)		-0.072 (0.092)		
Proportion of Male	-0.230 (0.102)*	-0.239 (0.103)*	-0.246 (0.103)*	-0.054 (0.072)	-0.043 (0.074)	-0.235 (0.098)*	-0.234 (0.098)*
Father's Education	-0.043 (0.021)*	-0.041 (0.021)*	-0.042 (0.021)*	0.009 (0.023)	0.010 (0.023)	-0.042 (0.019)*	-0.042 (0.020)*
Mother's Education	0.026 (0.022)	0.026 (0.022)	0.027 (0.022)	0.011 (0.024)	0.010 (0.024)	0.025 (0.021)	0.025 (0.021)
Books over 200	-0.016 (0.111)	-0.013 (0.112)	-0.018 (0.112)	0.430 (0.106)**	0.437 (0.108)**	-0.049 (0.108)	-0.046 (0.109)
Computer at Home	-0.138 (0.100)	-0.138 (0.101)	-0.136 (0.101)	0.060 (0.098)	0.057 (0.098)	-0.143 (0.094)	-0.142 (0.094)
Individual Characteristics							
Age	-0.036 (0.041)	-0.036 (0.041)	-0.036 (0.041)	0.006 (0.004)	0.006 (0.004)	-0.036 (0.041)	-0.036 (0.041)
Male	0.280 (0.070)**	0.280 (0.070)**	0.279 (0.070)**	0.004 (0.006)	0.004 (0.006)	0.279 (0.070)**	0.279 (0.070)**
Both Parents Present	0.107 (0.043)*	0.107 (0.043)*	0.107 (0.043)*	0.010 (0.004)**	0.010 (0.004)**	0.106 (0.043)*	0.107 (0.043)*
Mother's Education							
7-9 Years	-0.025 (0.058)	-0.026 (0.058)	-0.026 (0.058)	-0.003 (0.004)	-0.003 (0.004)	-0.025 (0.058)	-0.025 (0.058)
10-12 Years	0.041 (0.062)	0.041 (0.062)	0.041 (0.062)	0.002 (0.005)	0.002 (0.005)	0.040 (0.062)	0.040 (0.062)
13-14 Years	-0.217 (0.089)*	-0.218 (0.089)*	-0.218 (0.089)*	-0.002 (0.007)	-0.002 (0.007)	-0.217 (0.088)*	-0.217 (0.088)*
15+ Years	0.052 (0.076)	0.052 (0.076)	0.052 (0.076)	0.001 (0.006)	0.001 (0.006)	0.052 (0.075)	0.052 (0.075)
Not Reported	-0.141 (0.066)*	-0.141 (0.066)*	-0.141 (0.066)*	0.010 (0.005)	0.010 (0.005)	-0.142 (0.066)*	-0.142 (0.066)*
Father's Education							
7-9 Years	0.055 (0.062)	0.056 (0.062)	0.055 (0.062)	0.005 (0.005)	0.006 (0.005)	0.055 (0.062)	0.055 (0.062)
10-12 Years	0.156 (0.065)*	0.157 (0.065)*	0.157 (0.065)*	0.002 (0.004)	0.003 (0.005)	0.157 (0.065)*	0.157 (0.065)*
13-14 Years	0.106 (0.079)	0.107 (0.079)	0.107 (0.079)	0.006 (0.006)	0.006 (0.006)	0.107 (0.078)	0.107 (0.078)
15+ Years	0.346 (0.073)**	0.347 (0.073)**	0.347 (0.073)**	-0.001 (0.005)	-0.001 (0.005)	0.347 (0.073)**	0.347 (0.073)**
Number of Books							
11-25	0.145 (0.072)*	0.144 (0.072)*	0.145 (0.072)*	-0.013 (0.008)	-0.013 (0.008)	0.145 (0.071)*	0.145 (0.071)*
26-100	0.507 (0.058)**	0.507 (0.058)**	0.507 (0.058)**	-0.014 (0.006)*	-0.014 (0.006)*	0.508 (0.057)**	0.508 (0.057)**
101-200	0.713 (0.063)**	0.713 (0.063)**	0.713 (0.063)**	-0.009 (0.006)	-0.009 (0.006)	0.714 (0.062)**	0.714 (0.062)**
201+	0.798 (0.060)**	0.797 (0.060)**	0.797 (0.060)**	-0.005 (0.007)	-0.005 (0.007)	0.798 (0.059)**	0.798 (0.059)**
Computer at Home	0.163 (0.032)**	0.163 (0.032)**	0.163 (0.032)**	-0.001 (0.002)	-0.001 (0.002)	0.163 (0.032)**	0.163 (0.032)**
Grade 8	0.071 (0.042)	0.072 (0.042)	0.071 (0.042)	0.025 (0.017)	0.026 (0.017)	0.070 (0.041)	0.070 (0.041)
Teacher Characteristics							
Male	-0.025 (0.026)	-0.027 (0.026)	-0.026 (0.026)	-0.036 (0.026)	-0.036 (0.026)	-0.024 (0.024)	-0.024 (0.024)
Post-graduate Degree	0.015 (0.038)	0.015 (0.038)	0.015 (0.037)	0.006 (0.039)	0.006 (0.039)	0.015 (0.035)	0.015 (0.035)
Years of Experience	-0.003 (0.001)**	-0.003 (0.001)**	-0.003 (0.001)**	-0.005 (0.001)**	-0.005 (0.001)**	-0.003 (0.001)**	-0.003 (0.001)**
Hours of Teaching Activities	-0.003 (0.003)	-0.003 (0.003)	-0.003 (0.003)	-0.005 (0.003)	-0.005 (0.003)	-0.003 (0.003)	-0.003 (0.003)
Frequent Teacher Meeting	0.010 (0.023)	0.013 (0.024)	0.013 (0.024)	0.049 (0.026)	0.049 (0.026)	0.008 (0.021)	0.009 (0.021)
Gender-matching	0.030 (0.030)	0.029 (0.030)	0.030 (0.030)	0.020 (0.021)	0.018 (0.021)	0.028 (0.029)	0.028 (0.029)
Intercept	-0.152 (0.607)	-0.190 (0.614)	-0.232 (0.611)	-0.353 (0.340)	-0.285 (0.336)	-0.161 (0.595)	-0.160 (0.596)
School FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Students	12,964	12,964	12,964	12,964	12,964	12,964	12,964
R-Square	0.190	0.190	0.190	0.862	0.863		

Note: The regressions are weighted by Within-Class Student Weight, which is equal to the integer part of (WGTFAC3*WGTADJ3+0.5).
Robust and cluster-adjusted standard errors are in parentheses. * and ** indicate the estimate is significant at the 0.05 and 0.01 levels, respectively.

Table 6: OLS Estimation Results: Fractions of Weak and Strong Peers and Own Average Math Scores

	OLS: No School Fixed Effects			OLS: School Fixed Effects		
	(1)	(2)	(3)	(4)	(5)	(6)
Peer Group Characteristics						
Proportion below 25th pt	-1.159 (0.075)**	-0.741 (0.107)**	-0.854 (0.107)**	-0.369 (0.137)**	-0.303 (0.137)*	-0.261 (0.140)
Proportion above 75th pt	1.112 (0.082)**	0.680 (0.096)**	0.793 (0.094)**	0.408 (0.117)**	0.295 (0.121)*	0.263 (0.128)*
Proportion of Male			-0.294 (0.083)**			-0.217 (0.108)*
Father's Education			-0.034 (0.016)*			-0.035 (0.024)
Mother's Education			0.019 (0.016)			0.018 (0.025)
Books over 200			-0.143 (0.094)			0.019 (0.121)
Computer at Home			-0.113 (0.074)			-0.113 (0.108)
Individual Characteristics						
Age		-0.045 (0.039)	-0.038 (0.040)		-0.039 (0.040)	-0.036 (0.041)
Male		0.134 (0.047)**	0.280 (0.071)**		0.233 (0.057)**	0.281 (0.070)**
Both Parents Present		0.098 (0.042)*	0.100 (0.043)*		0.100 (0.042)*	0.108 (0.043)*
Mother's Education						
7-9 Years		-0.021 (0.053)	-0.023 (0.058)		-0.023 (0.053)	-0.026 (0.058)
10-12 Years		0.050 (0.057)	0.040 (0.062)		0.055 (0.058)	0.041 (0.062)
13-14 Years		-0.197 (0.083)*	-0.216 (0.088)*		-0.199 (0.084)*	-0.216 (0.089)*
15+ Years		0.041 (0.067)	0.053 (0.075)		0.056 (0.069)	0.050 (0.076)
Not Reported		-0.145 (0.063)*	-0.143 (0.065)*		-0.141 (0.065)*	-0.142 (0.066)*
Father's Education						
7-9 Years		0.052 (0.053)	0.050 (0.061)		0.055 (0.054)	0.055 (0.062)
10-12 Years		0.137 (0.054)*	0.154 (0.065)*		0.144 (0.054)**	0.155 (0.065)*
13-14 Years		0.080 (0.068)	0.104 (0.079)		0.089 (0.069)	0.106 (0.079)
15+ Years		0.319 (0.062)**	0.349 (0.073)**		0.329 (0.064)**	0.346 (0.073)**
Number of Books						
11-25		0.166 (0.069)*	0.134 (0.071)		0.172 (0.070)*	0.145 (0.072)*
26-100		0.514 (0.055)**	0.498 (0.057)**		0.521 (0.056)**	0.507 (0.058)**
101-200		0.721 (0.061)**	0.705 (0.063)**		0.732 (0.061)**	0.712 (0.063)**
201+		0.803 (0.059)**	0.788 (0.060)**		0.819 (0.060)**	0.798 (0.060)**
Computer at Home		0.169 (0.031)**	0.165 (0.032)**		0.175 (0.032)**	0.163 (0.032)**
Grade 8		0.064 (0.039)	0.058 (0.041)		0.070 (0.041)	0.070 (0.043)
Teacher Characteristics						
Male		-0.012 (0.021)	-0.008 (0.019)		-0.035 (0.028)	-0.026 (0.028)
Post-graduate Degree		0.004 (0.028)	-0.005 (0.023)		0.015 (0.041)	0.009 (0.043)
Years of Experience		-0.003 (0.001)*	-0.001 (0.001)		-0.003 (0.001)*	-0.003 (0.001)*
Hours of Teaching Activities		0.000 (0.002)	-0.001 (0.002)		0.000 (0.003)	-0.002 (0.003)
Frequent Teacher Meeting		0.002 (0.019)	-0.002 (0.017)		-0.002 (0.025)	0.004 (0.026)
Gender-matching		0.019 (0.020)	0.023 (0.019)		0.040 (0.030)	0.037 (0.031)
Intercept	0.087 (0.032)**	-0.231 (0.530)	-0.189 (0.537)	0.387 (0.097)**	-0.207 (0.554)	-0.055 (0.617)
Class/School Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Number of Students	13,598	13,305	12,964	13,598	13,305	12,964
R-Square	0.056	0.180	0.177	0.071	0.194	0.190

Note: The regressions are weighted by Within-Class Student Weight, which is equal to the integer part of (WGTFAC3*WGTADJ3+0.5).
Robust and cluster-adjusted standard errors are in parentheses. * and ** indicate the estimate is significant at the 0.05 and 0.01 levels, respectively.

Table 7: Quantile Estimation Results: Mean and SD of Peers' Scores and Quantiles of Math Score

Quantiles:	Average	0.10	0.25	0.50	0.75	0.90
	A. Ordinary Quantile Regression					
Peer Characteristics						
Mean Math Score	0.652 (0.032)**	0.739 (0.060)**	0.633 (0.034)**	0.625 (0.033)**	0.675 (0.029)**	0.642 (0.061)**
SD of Math Score	0.013 (0.044)	-0.865 (0.080)**	-0.425 (0.058)**	0.097 (0.052)	0.449 (0.053)**	0.617 (0.091)**
Proportion of Male	-0.270 (0.074)**	-0.256 (0.111)*	-0.193 (0.078)*	-0.218 (0.062)**	-0.416 (0.099)**	-0.237 (0.103)*
Father's Education	-0.046 (0.012)**	-0.050 (0.022)*	-0.046 (0.017)**	-0.061 (0.019)**	-0.074 (0.019)**	-0.014 (0.024)
Mother's Education	0.024 (0.011)*	0.017 (0.021)	0.035 (0.015)*	0.024 (0.016)	0.054 (0.017)**	-0.003 (0.024)
Books over 200	-0.201 (0.067)**	-0.307 (0.101)**	-0.321 (0.082)**	-0.136 (0.071)	-0.052 (0.093)	-0.039 (0.113)
Computer at Home	-0.151 (0.059)*	-0.067 (0.113)	-0.228 (0.087)**	-0.001 (0.082)	-0.226 (0.085)**	-0.307 (0.116)**
School Fixed Effects	No	No	No	No	No	No
	B. Ordinary Quantile Regression					
Peer Characteristics						
Mean Math Score	0.254 (0.059)**	0.362 (0.208)	0.238 (0.071)**	0.275 (0.073)**	0.238 (0.090)**	0.209 (0.119)
SD of Math Score	0.039 (0.073)	-0.727 (0.155)**	-0.204 (0.081)*	0.084 (0.091)	0.279 (0.074)**	0.327 (0.127)*
Proportion of Male	-0.230 (0.102)*	-0.319 (0.279)	-0.270 (0.120)*	-0.051 (0.104)	-0.347 (0.124)**	-0.402 (0.145)**
Father's Education	-0.043 (0.021)*	-0.028 (0.070)	-0.042 (0.038)	-0.047 (0.034)	-0.072 (0.029)*	0.048 (0.045)
Mother's Education	0.026 (0.022)	-0.024 (0.117)	0.051 (0.037)	-0.007 (0.034)	0.062 (0.032)	-0.038 (0.049)
Books over 200	-0.016 (0.111)	-0.276 (0.184)	-0.364 (0.146)*	-0.010 (0.127)	0.172 (0.173)	0.021 (0.198)
Computer at Home	-0.138 (0.100)	-0.198 (0.218)	0.048 (0.168)	0.196 (0.126)	-0.106 (0.154)	-0.236 (0.196)
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
	C. IV Quantile Regression					
Peer Characteristics						
Mean Math Score	0.307 (0.076)**	0.369 (0.128)**	0.473 (0.115)**	0.423 (0.116)**	0.031 (0.143)	-0.057 (0.173)
SD of Math Score	0.043 (0.068)	-0.725 (0.112)**	-0.223 (0.085)**	0.094 (0.087)	0.248 (0.081)**	0.369 (0.126)**
Proportion of Male	-0.235 (0.098)*	-0.331 (0.184)	-0.361 (0.109)**	-0.048 (0.110)	-0.305 (0.132)*	-0.483 (0.163)**
Father's Education	-0.042 (0.019)*	-0.029 (0.037)	-0.059 (0.036)	-0.048 (0.034)	-0.063 (0.029)*	0.019 (0.048)
Mother's Education	0.025 (0.021)	-0.019 (0.037)	0.060 (0.035)	-0.006 (0.033)	0.052 (0.033)	-0.020 (0.048)
Books over 200	-0.049 (0.108)	-0.287 (0.185)	-0.463 (0.159)**	-0.060 (0.144)	0.420 (0.191)*	0.299 (0.209)
Computer at Home	-0.143 (0.094)	-0.188 (0.209)	-0.158 (0.162)	0.134 (0.118)	-0.211 (0.160)	-0.262 (0.207)
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Students	12,964	12,964	12,964	12,964	12,964	12,964

Note: The regressions are weighted by Within-Class Student Weight. Standard errors are in parentheses. * and ** indicate the estimate is significant at the 0.05 and 0.01 levels, respectively.

Table 8: Quantile Estimation Results: Fractions of Weak and Strong Peers and Quantiles of Math Score

Quantiles:	Average	0.10	0.25	0.50	0.75	0.90
Peer Characteristics						
Proportion below 25th pt	-0.854 (0.107)**	-1.677 (0.096)**	-1.347 (0.107)**	-0.900 (0.107)**	-0.363 (0.107)**	-0.050 (0.107)
Proportion above 75th pt	0.793 (0.094)**	0.182 (0.109)	0.402 (0.094)**	0.795 (0.094)**	1.333 (0.094)**	1.282 (0.094)**
Proportion of Male	-0.294 (0.083)**	-0.354 (0.177)*	-0.232 (0.083)**	-0.255 (0.083)**	-0.227 (0.083)**	-0.284 (0.083)**
Father's Education	-0.034 (0.016)*	-0.040 (0.034)	-0.047 (0.016)**	-0.049 (0.016)**	-0.026 (0.016)	0.032 (0.016)
Mother's Education	0.019 (0.016)	0.019 (0.040)	0.034 (0.016)*	0.012 (0.016)	0.007 (0.016)	-0.032 (0.016)*
Books over 200	-0.143 (0.094)	-0.130 (0.178)	-0.225 (0.094)*	-0.068 (0.094)	-0.012 (0.094)	-0.053 (0.094)
Computer at Home	-0.113 (0.074)	0.076 (0.203)	-0.059 (0.074)	-0.024 (0.074)	-0.243 (0.074)**	-0.294 (0.074)**
School Fixed Effects	No	No	No	No	No	No
Peer Characteristics						
Proportion below 25th pt	-0.261 (0.140)	-0.658 (0.236)**	-0.721 (0.182)**	-0.538 (0.125)**	-0.121 (0.200)	0.475 (0.233)*
Proportion above 75th pt	0.263 (0.128)*	0.117 (0.227)	-0.091 (0.183)	0.283 (0.135)*	0.435 (0.181)*	0.461 (0.218)*
Proportion of Male	-0.217 (0.108)*	-0.484 (0.198)*	-0.380 (0.120)**	-0.065 (0.113)	-0.278 (0.120)*	-0.298 (0.182)
Father's Education	-0.035 (0.024)	-0.064 (0.038)	-0.032 (0.038)	-0.057 (0.030)	-0.053 (0.039)	0.035 (0.045)
Mother's Education	0.018 (0.025)	0.008 (0.050)	0.042 (0.041)	-0.005 (0.028)	0.038 (0.041)	-0.043 (0.052)
Books over 200	0.019 (0.121)	-0.174 (0.204)	-0.309 (0.158)	0.004 (0.114)	0.253 (0.191)	0.303 (0.199)
Computer at Home	-0.113 (0.108)	-0.144 (0.216)	0.041 (0.168)	0.239 (0.141)	-0.114 (0.169)	-0.281 (0.201)
School Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Number of Students	12,964	12,964	12,964	12,964	12,964	12,964

Note: The regressions are weighted by Within-Class Student Weight. Standard errors are in parentheses. * and ** indicate the estimate is significant at the 0.05 and 0.01 levels, respectively.

Appendix Table 1: Quantile Estimation Results: Fractions of Different-Quality Peers, and Average and Quantiles of Math Score

Quantiles:	Average	0.10	0.25	0.50	0.75	0.90
Quantile Regression: School FE						
Peer Group Characteristics						
Proportion below 25th pt	-0.161 (0.222)	-0.984 (0.306)**	-1.056 (0.250)**	-0.629 (0.225)**	-0.171 (0.243)	0.865 (0.289)**
Proportion: 25th and 40th pt	-0.520 (0.214)*	-0.887 (0.289)**	-0.776 (0.312)*	-0.624 (0.241)**	-0.606 (0.291)*	-0.074 (0.325)
Proportion: 60th and 75th pt	0.370 (0.268)	-0.223 (0.342)	-0.175 (0.296)	0.081 (0.232)	0.178 (0.277)	1.033 (0.424)*
Proportion above 75th pt	0.186 (0.225)	-0.368 (0.275)	-0.466 (0.247)	0.022 (0.236)	0.264 (0.254)	0.817 (0.267)**
Proportion of Male	-0.259 (0.136)	-0.446 (0.218)*	-0.481 (0.168)**	-0.139 (0.121)	-0.342 (0.154)*	-0.344 (0.206)
Father's Education	-0.060 (0.030)*	-0.090 (0.057)	-0.047 (0.040)	-0.051 (0.035)	-0.073 (0.032)*	0.005 (0.045)
Mother's Education	0.047 (0.032)	0.041 (0.067)	0.077 (0.043)	-0.009 (0.036)	0.053 (0.036)	-0.026 (0.050)
Books over 200	0.047 (0.175)	-0.166 (0.200)	-0.312 (0.155)*	0.108 (0.132)	0.266 (0.200)	0.230 (0.188)
Computer at Home	-0.266 (0.148)	-0.221 (0.197)	-0.045 (0.211)	0.086 (0.176)	-0.221 (0.203)	-0.562 (0.256)*
Class Size	0.003 (0.003)	0.007 (0.005)	0.003 (0.004)	-0.003 (0.003)	-0.002 (0.005)	0.004 (0.006)
Number of Students	12,964	12,964	12,964	12,964	12,964	12,964

Note: The regressions are weighted by Within-Class Student Weight. Standard errors are in parentheses. * and ** indicate the estimate is significant at the 0.05 and 0.01 levels, respectively.

Appendix Table 2: Overall Achievement under Mixing and Hypothetical Tracking System

Education System :	Mixing				Tracking I ($\alpha = 0.1$)				Tracking II ($\alpha = 0.25$)			
Variables	Average	SD	Min	Max	Average	SD	Min	Max	Average	SD	Min	Max
Peer Variable	(N=12,964)				(N=12,908)				(N=12,964)			
Mean Math Score	0.083	0.306	-0.842	1.055	0.096	0.607	-3.233	1.154	0.090	0.757	-2.566	1.479
SD of Math Score	0.936	0.157	0.433	1.578	0.745	0.222	0.000	1.359	0.616	0.183	0.097	1.161
Individual Math Score	0.093	0.984	-3.731	3.491	0.059	1.086	-4.435	3.422	0.007	1.178	-4.345	3.416
Change in Math Score					-0.042	0.215	-1.230	0.191	-0.086	0.306	-0.992	0.296
Proportion of Decrease					0.281	0.450	0.000	1.000	0.383	0.486	0.000	1.000
					Low-Track under Tracking I				Low-Track under Tracking II			
Peer Variable					(N=1,153)				(N=3,156)			
Mean Math Score					-1.542	0.469	-3.233	0.126	-1.108	0.405	-2.566	0.488
SD of Math Score					0.253	0.216	0.000	1.263	0.449	0.188	0.097	1.161
Individual Math Score					-2.231	0.579	-4.435	0.208	-1.682	0.679	-4.345	0.528
Change in Math Score					-0.686	0.213	-1.230	0.191	-0.577	0.165	-0.992	0.093
Proportion of Decrease					0.995	0.072	0.000	1.000	0.993	0.085	0.000	1.000
					High-Track under Tracking I				High-Track under Tracking II			
Peer Variable					(N=11,755)				(N=9,808)			
Mean Math Score					0.256	0.308	-0.688	1.154	0.475	0.308	-0.438	1.479
SD of Math Score					0.794	0.153	0.329	1.359	0.670	0.145	0.271	1.047
Individual Math Score					0.284	0.835	-2.534	3.422	0.551	0.687	-1.659	3.416
Change in Math Score					0.021	0.042	-0.126	0.119	0.072	0.111	-0.255	0.296
Proportion of Decrease					0.211	0.408	0.000	1.000	0.187	0.390	0.000	1.000

(Continued to the next page)

Appendix Table 2: Overall Achievement under Mixing and Hypothetical Tracking System (Continued)

Education System :	Tracking III ($\alpha = 0.5$)				Tracking IV ($\alpha = 0.75$)				Tracking V ($\alpha = 0.9$)			
Variables	Average	SD	Min	Max	Average	SD	Min	Max	Average	SD	Min	Max
Peer Variable	(N=12,964)				(N=12,964)				(N=12,960)			
Mean Math Score	0.090	0.807	-1.757	1.854	0.088	0.745	-1.302	2.593	0.084	0.612	-1.018	3.173
SD of Math Score	0.557	0.162	0.156	1.297	0.626	0.209	0.084	1.461	0.748	0.239	0.000	1.494
Individual Math Score	0.057	1.175	-4.287	3.282	0.088	1.074	-3.931	3.300	0.076	0.998	-3.796	3.144
Change in Math Score	-0.035	0.277	-0.637	0.490	-0.005	0.165	-0.582	0.620	-0.017	0.085	-0.623	0.796
Proportion of Decrease	0.560	0.496	0.000	1.000	0.735	0.441	0.000	1.000	0.789	0.408	0.000	1.000
	Low-Track under Tracking III				Low-Track under Tracking IV				Low-Track under Tracking V			
Peer Variable	(N=6,391)				(N=9,624)				(N=11,625)			
Mean Math Score	-0.655	0.349	-1.757	0.708	-0.305	0.325	-1.302	0.885	-0.092	0.313	-1.018	0.979
SD of Math Score	0.584	0.163	0.156	1.297	0.706	0.155	0.282	1.461	0.809	0.152	0.341	1.494
Individual Math Score	-0.915	0.786	-4.287	1.181	-0.371	0.813	-3.931	1.535	-0.104	0.876	-3.796	2.357
Change in Math Score	-0.265	0.142	-0.637	0.179	-0.072	0.058	-0.270	0.079	-0.022	0.025	-0.113	0.051
Proportion of Decrease	0.974	0.159	0.000	1.000	0.916	0.278	0.000	1.000	0.824	0.381	0.000	1.000
	High-Track under Tracking III				High-Track under Tracking IV				High-Track under Tracking V			
Peer Variable	(N=6,573)				(N=3,340)				(N=1,335)			
Mean Math Score	0.815	0.318	-0.108	1.854	1.222	0.348	0.324	2.593	1.610	0.424	0.581	3.173
SD of Math Score	0.530	0.156	0.167	0.937	0.396	0.172	0.084	0.955	0.212	0.186	0.000	0.942
Individual Math Score	1.003	0.554	-0.737	3.282	1.408	0.472	-0.019	3.300	1.642	0.511	0.478	3.144
Change in Math Score	0.188	0.176	-0.458	0.490	0.189	0.212	-0.582	0.620	0.032	0.251	-0.623	0.796
Proportion of Decrease	0.157	0.364	0.000	1.000	0.213	0.410	0.000	1.000	0.488	0.500	0.000	1.000