Mixing *versus* Sorting in Schooling: Evidence from the Equalization Policy in South Korea

Taejong Kim
Ju-Ho Lee
Young Lee

September 2003
Working Paper 03-07

Mixing *versus* Sorting in Schooling:

Evidence from the Equalization Policy in South Korea*

by

Taejong Kim
KDI School of Public Policy and Management, Seoul, Korea
+82-2-3299-1085, tjkim@kdischool.ac.kr

Ju-Ho Lee
KDI School of Public Policy and Management, Seoul, Korea
+82-2-3299-1016, jhl@kdischool.ac.kr

Young Lee
Hanyang University, Seoul, Korea
+82-2-2290-1023, younglee@hanyang.ac.kr

September 2003

* The authors are grateful to Ronald Ehrenberg, Jim Poterba, Takao Kato, Debin Ma, Andreas Schleicher, Daiji Kawaguchi and participants at the FASID Hakone conference and the Colgate seminar for their helpful comments and suggestions.

Mixing *versus* Sorting in Schooling: Evidence from the Equalization Policy in South Korea

Abstract

This paper examines the effects of sorting and mixing on academic performance of high school students in South Korea. The Korean government has vigorously promoted mixing for more than three decades, replacing competitive entrance examinations at individual schools by a lottery-based enrollment system. As a result, about half of high schools (grades 10 to 12) as well as all middle schools (grades 7 to 9) are subject to what is locally known as the Equalization Policy (EP), and passively accept students randomly assigned. In contrast, outside the designated EP areas, students are sorted with stratification along ability among schools. This paper employs the difference-in-differences empirical strategy to analyze the newly available data from the Korean National Assessment of Educational Achievement. Two main results emerge. First, sorting raises test scores of students outside the EP areas by roughly 0.3 standard deviations, relative to mixing. Second, more surprisingly, quantile regression results reveal that sorting benefits students across the ability distribution.

JEL classification codes: H4, I0, I2

Key words: public education, sorting, mixing, peer effect, South Korea.

Mixing *versus* Sorting in Schooling: Evidence from the Equalization Policy in South Korea

1. Introduction

Students differ from each other in innate ability, past learning experiences, and other socioeconomic aspects. We may distinguish between two contrasting principles in allocation of disparate students among schools: mixing and sorting. In a mixed system, students learn with heterogeneous peers, while schools themselves look more or less similar to each other. Under sorting, in contrast, student peers are relatively homogeneous in each school, whereas schools are stratified, say along the ability of students. Existing regimes for student allocation may be classified according to the degree of sorting and mixing. Likewise, education reform may bring a system closer toward this or that end of the distribution.

There is a growing literature that gauges the two competing principles in terms of efficiency and equity. In general, mixing students could improve equity by enabling more able students to positively affect peers of lower ability. On the other hand, sorting could raise the overall productivity of the school system by enabling teachers to instruct more efficiently and by encouraging schools to compete to attract students. An often-heard concern over school choice is that choice-induced sorting[1] may damage the low-ability students by cream-skimming high-ability students and leaving low-ability students behind at "forgotten schools" in poorer neighborhoods. On the other hand, advocates of school choice expect a significant boost in the efficiency of schooling that could benefit every student regardless of ability.

Some interesting theoretical models recognize the potential trade-off between equity and efficiency in allocating students among schools. Epple and Romano (1998) delineate an efficient equilibrium of student sorting based on ability and income. Their model predicts that an increase in the extent of sorting through tuition vouchers benefits high-ability students more strongly than low-ability students. Benabou (1996) points out that short-run efficiency gains from segregation (sorting) may be offset in the long run, if integration (mixing) reduces heterogeneity faster

[1] Many school choice programs, such as charter schools and vouchers, target low-income families, and do not necessarily increase the degree of sorting. However, school choice among public schools (Tiebout choice) and for

among community members (students). These concerns about the tradeoff between efficiency and equity will become a moot point, however, if benefits associated with sorting are more or less equally shared by everybody regardless of ability. Indeed, Hoxby (2003a) suggests that school choice could generate large productivity gains that could overwhelm any possible negative allocation effects.

In this paper, we pose and examine two questions: (a) whether and to what extent sorting in student allocation raises efficiency, and (b) how the benefit from sorting, if there is any, is shared among students at various levels of ability. As stressed by Hoxby (2003b) in the context of school choice, theory alone cannot resolve these issues on the effects of school systems on performance because the outcomes typically depend on the values of numerous unknown parameters. Therefore, essential questions should be answered by empirical studies that consider various institutional contexts in specific cases. For the purpose, this paper examines academic performance of high school students in South Korea (Korea hereafter). Korea offers an unusually vivid contrast between sorting and mixing with its national policy experiment, locally known as the Equalization Policy (EP hereafter). .

The Korean government has vigorously promoted mixing for more than three decades, with the EP replacing by a lottery-based enrollment system traditional competitive entrance examinations at individual schools. As a result, about half of high schools (grades 10 to 12) as well as all middle schools (grades 7 to 9) are subject to the EP, and passively accept students randomly assigned. The resulting student mixing in the EP areas is quite thorough, since the area of a typical lottery zone is rather large and the policy encompasses not just public but also private schools. Other nation-wide measures, such as centralized school finance, ban on tracking within school, national curriculum, and uniform teacher policies, have also consolidated student mixing. In contrast, students in non-EP areas are sorted among schools with fairly strict stratification along ability.

This paper utilizes the newly available data from the Korean National Assessment of Educational Achievement (NAEA, hereafter) in 2001. The assessment samples one percent of all

---

college-preparatory private schools is strongly related to student sorting in the US.

students in grades 6, 9, 10, and 11 across the country, and reports test scores on five subjects in Korean, English, Mathematics, Science, and Social Studies. The NAEA test scores are supplemented by survey information on tested students, teachers, and their schools.

Two major findings stand out from our analysis based on the difference-in-differences empirical strategy. First, sorting in the non-EP areas, relative to mixing , significantly raises average test scores by between 0.19 and 0.38 standard deviations in standardized scores, depending on the specification. Second, perhaps more strikingly, quantile regressions reveal that that benefits of sorting accrue to students across the entire range of ability distribution with the possible exception of those in the bottom decile. Our results suggest that a shift from mixing to sorting in school organization through the abolishment of the EP could generate “a rising tide that lifts all boats.”

The remainder of the paper proceeds as follows. Section 2 discusses the relevant institutional backgrounds in Korea with a particular focus on the EP. Section 3 introduces the data and our empirical strategy. Main empirical results will be presented in Section 4, and followed by concluding remarks in Section 5.

2. Institutional Backgrounds

2.1 Equalization Policy (EP)

Major indicators of education in Korea evidence a very rapid educational development, impressive for a country that used to be counted among international basket cases until the early 1960s. Up to the high school level,[2] education has become practically universal with modest tuition.[3] The proportion of high school graduates who advance to either 4-year universities or 2-year technical colleges exceeds 70 percent, among the highest such rates in the world. Average academic achievement of Korean students also appears high in comparison to other OECD

[2] Schooling in Korea starts with 6 years of elementary education, to be followed by 3 years of lower secondary education (middle school) and further 3 years of upper secondary (high school). The first 6 years of education is mandatory. Schooling for middle-schoolers in urban areas is yet to be free of tuition, but will be made so in 2004.
[3] High school tuition is uniformly regulated. The tuition in urban schools is about 700 dollars per year. Rural schools

member countries (U.S. Department of Education, 2001).

The EP for high school[4] was first introduced in 1974 in Seoul and Pusan, the two largest metropolitan areas in Korea, and has been expanded to cover more and more urban areas. As of 2001, the EP covered all of the 7 metropolitan areas and 11 provincial cities[5], accounting for 51 percent of schools and 65 percent of students out of 1,969 high schools and their 1.91 million students in total.[6]

Replacing traditional entrance examinations at individual high schools, the EP introduced a lottery-based enrollment system. Remarkably, the lottery system decides student assignment not just in public schools but also in private.[7] Student mixing in the EP areas is quite thorough, since the area of a typical enrollment zone is rather large. For instance, Seoul, with its over 10 million inhabitants, is divided into 11 enrollment zones. As discussed below, the government's efforts to create mixed and homogenized schools were bolstered by several other measures.

In contrast, the remaining regions outside the EP regime have retained the traditional student enrollment system, where individual schools select new students based on competitive entrance exams. Well-established rankings exist among high schools, reflecting track records in placement of graduates into elite universities. Entrance exams effectively sort students into schools based on academic achievement in the non-EP areas. The basic idea behind our empirical strategy is to utilize the unusually vivid contrast between the mixed, EP-area schools on the one hand and the sorted, non-EP-area schools on the other, plus the fact that middle school education is universally mixed in the country.

---

charge half the amount. No difference in tuition exists between public and private schools.

[4] The EP for middle schools started in 1969 and soon covered all middle schools across the country. After the completion of the EP for middle schools, the term "EP" is used for general high schools in Korea.

[5] Local government in South Korea breaks down into 7 metropolitan areas (Seoul, Pusan, Taegu, Inchon, Kwangju, Taejon, and Ulsan) and 9 provinces. Metropolitan areas refer to urban centers with population over 1 million. Provinces further break down into 72 provincial cities 91 counties in 2001.

[6] Recent efforts to expand the EP were spurred in part by the strong pro-EP campaign organized by the newly legalized teachers' union. Beginning from March 2002, six additional cities in Kyonggi Province are to be covered by the EP. This is to increase the coverage to 57 percent of general high schools and 74 percent of students.

[7] See more on this in the next subsection.

In our view, it is hard to imagine a better contrast between mixing and sorting than is provided by the EP and non-EP areas in Korea. In that regard, it is important to understand the drastic extent to which the EP bolsters mixing of students across schools and classrooms with an assortment of accompanying measures regulating private schools, tracking among and within schools, school finances, curriculum and teacher policies, and school administration.

As of 2001, 47 percent of high schools are private in the country. Private schools, however, are not in a position to serve as a conduit for school choice in the EP areas. The EP forces them to passively accept students assigned through lottery. In addition, the government has practically taken over school finance even in private schools with its tuition regulation tuition and subsidies.

In secondary education across the world, tracking is prevalent, both among and within school. Korea does offer separate educational tracks at the secondary level for students who would enter the labor market directly and those who would pursue post-secondary education: vocational high schools for the former, accounting for 31 percent of student enrollment in 2001, and general high for the latter. While vocational high schools do conduct entrance exams, whether in EP or non-EP areas, they cannot be expected to promote school choice: Historically, vocational schools enroll students from the lower end of achievement distribution.[8] Our subsequent analysis focuses on general high schools, ignoring vocational high.

Unlike in the US[9], Korean schools admit virtually no tracking within school.[10] Classroom setting in general high schools hardly differs from that in elementary schools. Every student in general high schools takes the same, predetermined sequence of classes without changing classrooms. The degree of heterogeneity in student composition at the school level directly transfers to each and every classroom in the EP areas.

---

[8] In a brief exception to this general pattern, vocational schools attracted a fairly large crop of ambitious students during the 1970s, when the government was energetically promoting heavy and chemical industries.

[9] Argys, Rees and Brewer (1996) estimate the impacts of tracking within school on student performance.

[10] The only exception is for the bifurcation from grade 11 to accommodate students intending to study science and engineering in college and those to study humanities and social sciences. It should be noted that the division is based on the areas of study, not on student ability.

Local finance of public schools could promote school choice through residential sorting.[11] Prior to the introduction of the EP, however, public school finance had already been integrated nationwide. The EP added all private schools, across the country, into the existing system of centralized public school finance. The possibility of residential sorting through Tiebout choice, then, is fairly strictly limited.[12]

Uniform and centralized policies over curriculum and teachers' qualifications have also made Korean schools extremely homogeneous.[13] Instruction is to be provided according to unified national curriculum, based on either designated or certified textbooks and dictated by thick volumes of teachers' guidelines. Teachers in public schools are all public employees of the national government with uniform salary schedules and required qualifications. Private school teachers should also meet the same qualifications required by the government and are also guaranteed with the equivalent salary schedules and other benefits by the government.[14] All these factors, which have made Korean schools very homogeneous, reinforced student mixing through the EP.[15]

2.2 Mixing versus Sorting in Policy Debates

Policy debates on the reform of the EP form an important backdrop for this paper. We

---

[11] There is an ongoing debate over the effects of court-mandated statewide equalization of school finances (Benabou, 1994, and Hoxby, 2001). Hoxby (2000) reports that residential choice facilitated by smaller-sized school districts in the US does promote competition and innovation among schools.

[12] An often-noted exception is the popularity of certain zones within Seoul and other metropolitan areas, famous for the high real estate prices, concentration of private cram schools and professional private tutors. For this and other reasons to be explained shortly, we focus on schools in provincial cities in our empirical analysis.

[13] Although one can easily point to many problems of homogeneous schools associated with the lack of diversity among schools, standardization may have been instrumental in assuaging parental concerns arising from asymmetric information on school quality, especially during early years after introduction of modern education. See Hanushek (2002).

[14] Perhaps the only significant difference is that teachers in public schools are rotated across schools at a 5-6 year interval. Teachers in private schools enjoy probably less job security, but tend to stay longer with the same school.

[15] How may one explain that all the strict regulations were able to take hold in Korea? While referring interested readers to Kim and Lee (2002b) for more detailed discussion on this issue, we briefly note the following observations. First, Korea in the 1970s was the very picture of a developmental state with government in the lead and people largely willing to follow. Second, the EP may have been effective in meeting the rising demands for industrial labor in mid 1970s. Through standardization, the EP helped expand the high school enrollment: from 41 percent in 1975 to 80 percent in 1985.15 Lastly, professed goals of the policy–including alleviation of excessive pressure on young children in middle school and curbing costs for private tutoring – proved popular among concerned parents.

recognize two distinct classroom channels through which sorting and mixing affects student performance: interaction between student peers and interaction between teachers and students. Through the first channel, one may reasonably expect the less able students to benefit from their more able peers in a mixed environment, while the effects on the more able are debatable. The effects on average performance are ambiguous.[16]

Through the second channel, we may think of several reasons why mixing may harm the productivity of schooling regardless of student ability. In a classroom with extreme student mixing, teachers' ability may be more severely strained to tailor their instruction to diverse needs of students. Teachers may also find it more difficult to discipline and motivate students.[17] However, sorting does not necessarily improve the quality of teacher-student interaction, especially in schools left in charge of low-ability students, if morale and motivation of teachers and students are severely reduced, and interest is reduced. Given the complexity of interaction among peers as well as interaction between student and teacher, it is clear that investigation of the effects of sorting and mixing requires an empirical study. Our goal in this paper is to empirically examine how mixing and sorting affects student performance, both in average and at different points of ability distribution.

While this paper is motivated by the ongoing policy debates in Korea, we believe that there are at least three reasons why our results should be of interest to researchers in economics of education elsewhere in the world.

First, our findings may have an important implication for the school choice literature. School choice is generally expected to promote competition among schools. Yet, advocates of choice often turn defensive when pointed out that the accompanying re-sorting may hurt students

[16] Positive peer effects are likely to extend to activities outside the classroom. Sacerdote (2000) reports that peer effects among randomly matched college roommates are strong in such areas as joining clubs, but rather weak in more narrowly academic aspects.
[17] According to a recent report from the OECD (2002), student satisfaction with teachers is the lowest in Korea among OECD countries and lower than in some other developing countries such as Brazil and Russia sampled in the study. The proportion of students who answered yes to the statement "students do not listen to what the teacher says" was also the highest in Korea. While cultural preconditioning may affect the way students answer these questions in different countries, these results might mirror the real difficulties faced by teachers in charge of an extremely diverse student body.

with low ability. Below in the paper, we report our finding that sorting raises performance of students throughout the range of ability distribution. Therefore, ours is evidence that increased sorting among schools does not necessarily harm the students of lower ability. It is worth mentioning that the EP and the accompanying policies deprive of Korean school both incentives and means for competition. Our estimates for the impact of sorting are thus mainly driven by the "student composition effects", not "competition effects".

Second, the contrast between EP and non-EP areas in Korea represents a case of "regime change", as opposed to marginal changes in the composition of student peers, the focus of the growing literature on peer effects in the US. Hoxby (2000) and Hanushek, Kahn, Markman and Rivkin (2003) consider changes in the peer composition that arise from small changes across grades and cohorts in demographics and prior achievement. Angrist and Lang (2002) examine the peer effects using the data from Metco, a Boston-area desegregation program that sends mostly black students out of the Boston public school district to attend schools in more affluent suburban districts. The results from these studies may not be a proper guide, if an envisioned reform entails a much larger compositional change. Unlike marginal changes, regime changes in peer group composition may alter the nature of instruction in the classroom. Levels of expectations, efforts, motivation of teachers, as well as teaching methods could completely change between two opposite regimes of homogeneous peers (sorting) and heterogeneous peers (mixing) even in cases the same teachers teach with the same curriculum.

Third, in spite of the dominance of East Asian countries in the top rankings of test scores in international tests of student achievement in mathematics and science (Hanushek and Luque, 2003), research on the link between student performance and the educational institutions in East Asia has yet to be carried out with the same depth and width devoted to the other part of the world. As suggested by Hanushek (2002), studies on the organization, funding, and institutions of different countries and their impacts on student performance could offer a useful clue to unresolved questions that have not been fully addressed so far. Given the ubiquitous regulations and interventions over education in East Asia, it seems a highly pertinent question whether the high performance comes about because of the regulations or in spite of them.

3. Data and Empirical Strategy

This section introduces data and empirical strategy, setting the stage before addressing the essential empirical questions: Does sorting improve the overall performance, and if so, to what extent? What are the consequences for students at the lower end of ability distribution?

3.1 The NAEA 2001

The data analyzed in the next section come from the Korean National Assessment of Educational Achievement 2001. The core of the NAEA consists of individual test scores on five subjects, Korean, English, math, science, and social studies.[18] The tests were administered on a 1 percent nationally representative sample of students in grades 6, 9, 10, and 11.[19] As high schools in Korea instruct students in grades from 10 to 12, the paper utilizes mainly the portion of data pertaining to grades 10 and 11. It should be noted that our data is not longitudinal. However, the availability of data on two cohorts, grades 10 and 11, allows us to implement 'difference-in-differences' estimation. Tests for grade 10 are designed to assess attainment of curriculum goals for the preceding three grades in middle school; tests for grade 11 assess what students learned in grade 10.

Supplementing the individual test scores are surveys of students and their teachers responsible for instruction of the test subjects. Teacher surveys provide essential information on school characteristics and teacher characteristics. School characteristics include location, being public or private, being coed or gender-segregated, whether the region is subject to equalization, size of the school, and size of the class tested. Teacher characteristics include age, education, teaching experience, and workload.

While student surveys do provide some bare essentials on family characteristics, such as

[18] The tests are not of high stake because they are not linked to any policy measures affecting individual school administrators, teachers, and students. Results regarding individual students or schools are released neither to the test participants and nor to the general public.
[19] Sampling for the test was conducted in two stages. The first stage selected classes through stratified sampling

parental education and whether the mother has a job outside home, some of the usual family background variables, such as income, occupation, and age of parents, are not provided. Students do report, however, their use of time after school, including hours spent on self-study and hours spent to receive private tutoring. The information on private tuition is potentially valuable in the Korean context, since it is a prevalent practice among high school students.[20] Kim and Lee (2002a) show that the demand for private tuition is in part a parental response to the EP equalization policy and that the incidence of private tutoring is higher in the EP areas after controlling for income and other family traits. Their result suggests that, to isolate the effects of sorting and mixing, participation in private tutoring should be controlled for.

In addition to general high schools, Korea also has vocational high schools and a small number of special-purpose schools designed to nurture talented students in such areas as arts, sports, science, and languages. We drop vocational as well as special-purpose schools from the sample, since the two school classes are fundamentally distinguished from the more usual general schools in terms of purposes and contents of instruction and student selection mechanism.

The data are further trimmed to drop metropolitan areas and rural counties. It may be recalled that all metropolitan cities are subjected to the EP, while all provincial counties are exempted. Of the 72 cities in the provinces, 10 are mixed under the EP regime with the remaining 62 under the sorting regime. The trimming means that our analysis contrasts provincial cities in the mixed EP areas against provincial cities in the sorted non-EP areas. There are three reasons for this trimming.

First, some high-income districts in metropolitan areas are known to be a popular choice for residence among parents with school-age children. These districts are famous for the concentration of elite cram schools and renowned professional tutors. Including Seoul and other metropolitan areas may thus complicate our analysis by confounding effects of sorting and mixing with effects from residential choice. Second, rural counties may have so few schools to begin with, that the extent of sorting is severely limited even with entrance exams in individual

based on region and class size. Then the second stage selected individual examinees based on cluster sampling.

[20] In our data, about 50 percent of students in grade 11 are shown to receive at least some private tutoring. The

schools. In other words, most schools in rural areas are essentially mixed, even though rural counties are nominally under the non-EP sorting regime. Third, provincial cities under the alternative arrangements are expected to more closely comparable than is the case between the whole mixed areas *vs.* the whole sorted areas across the entire country. Indeed, our preliminary examination showed that provincial cities, mixed and sorted, are almost indistinguishable in terms of parental education and after-school learning activities. In contrast, students in the mixed areas as a whole have significantly better-educated parents and also spend significantly more time in receiving private tuition than their counterparts in the nation's sorted areas.

From the initial number of observations of 15,054, these exclusions leave us with 3,024 individual students. Of them, 1,560 are in grade 10 and 1,464 in grade 11. Among the grade 10 students, 317 reside in EP cities and 1,243 in non-EP cities. Among the grade 11 students, the figures are 276 and 1,188, respectively.

3.2 Empirical strategy

To identify the effects of sorting, relative to mixing, on scholastic performance, our chief empirical strategy is to employ a difference-in-differences (or DD) estimator. The DD estimator is often used to estimate a causal treatment effect when the outcome of interest is observed for both treatment and comparison groups before and after the treatment in question. Angrist and Krueger's chapter (1999) in the Handbook of Labor Economics provides an excellent exposition on the estimator together with its potential pitfalls.[21]

In the context of the paper, the treatment is exposure to the sorting regime. We take students under the EP as a comparison group. Test scores for grade 10 examinees are taken to index scholastic attainment levels, for both the treatment and the comparison groups, attained before they get subjected to the alternative regimes. This interpretation is justified, as grade 10 exams are designed to test grasp of materials taught during the preceding three years in middle

---

modal hours per day spent on this activity is two hours among the participants.

[21] Recent implementations include Card (1990) and Eissa (2001). Card (1990) uses the DD method to estimate the effect of a large scale Cuban immigration on the Miami labor market. Eissa (2001) uses the same method to measure the effect of the 1986 US tax reform on female labor supply. The method is also known as the "non-equivalent

school.[22] One may be concerned that test instruments are different between the two grades. Partly to address this problem, and also to facilitate interpretation of coefficient estimates, we will analyze standardized test scores, individual score less national average divided by standard deviation.

Intuitively, the DD strategy uses the evolution of observed outcomes in the comparison group to answer the counterfactual question: What would have happened to the treatment group if it hadn't been subject to the treatment?

Let Y_{0i} be student *i*'s scholastic performance under mixing, and let Y_{1i} be *i*'s performance under sorting. The average performance of students in area *c* in grade *g* under mixing is $E[Y_{0i} \mid c, g]$, and that under sorting $E[Y_{1i} \mid c, g]$. Let $c=M$ represent areas under mixing with the EP, and $c = S$ areas under sorting; $g = 10$ grade 10, and $g = 11$ grade 11. The parameter of interest is the treatment effect δ, or the effect of sorting, in areas $c = S$:[23]

$$\delta = E[Y_{1i} \mid S, 11] - E[Y_{0i} \mid S, 11].$$

The problem is that we observe only the sample moment corresponding to the first term on the right hand side, but not the second. The key identifying assumption in the DD strategy is that the counterfactual evolution of performance in areas $c = S$ is the same as that in areas $c = M$:

$$E[Y_{0i} \mid S, 11] - E[Y_{0i} \mid S, 10] = E[Y_{0i} \mid M, 11] - E[Y_{0i} \mid M, 10].$$

or

$$E[Y_{0i} \mid S, 11] = E[Y_{0i} \mid S, 10] + \{E[Y_{0i} \mid M, 11] - E[Y_{0i} \mid M, 10]\}.$$

control group pretest-posttest design" in psychology. See Campbell (1969).

[22] Middle schools are equalized across the country, *i.e.,* middle-school students learn among mixed peers regardless of location. Equalization for middle schools was introduced in 1969 across the country, and has been maintained since without much controversy, unlike its high-school counterpart.

[23] In our terminology, living in area $c = S$ does not necessarily mean that the student is exposed to sorting. The timing and the test coverage of the NAEA exams mean that it is only those students with $c = S$ AND $g = 11$ that has

The assumption implies that the treatment effect δ may be estimated by double differencing sample moments as follows:

$$\delta = \{E[Y_{1i} \mid S, 11] - E[Y_{0i} \mid S, 10]\} - \{E[Y_{0i} \mid M, 11] - E[Y_{0i} \mid M, 10]\}.$$

All the sample moments corresponding to the four terms on the right hand side above are available. Conveniently, the DD estimate can be computed in a regression of stacked individual data using the following equation:

$$Y_i = \beta_c + \gamma_g + \delta M_i + \varepsilon_i$$

where β_c denotes the area effect, and γ_g the grade effect, $E[\varepsilon_I \mid c, g] = 0$, and M_i is a dummy variable that equals 1 if i was exposed to sorting by living in area S in grade 11. Thus the regressors are dummies for areas ($c \in \{S, M\}$) and grades ($g \in \{S, M\}$), and M_i, with the constant term suppressed.

Similarly, a regression-adjusted version of the DD estimator adds a vector of covariates X_i to the equation above:

$$Y_i = X'_i\beta_0 + \beta_c + \gamma_g + \delta M_i + \varepsilon_i$$

Both raw and regression-adjusted DD estimates will be reported in the next section.

Table 1 presents summary statistics for test scores in grades 10 (top panel) and 11 (lower panel). Table 2 does the same for explanatory variables.

In Table 1, top six rows in each panel list raw test scores-out of the full score of 100 points- in five subjects: Korean, English, math, science and social studies, together with the average score across the five subjects. The following six rows show standardized test scores, using the average score and standard deviation from the national distribution. The first two columns, or columns (1)-(2), pertain to mixed, EP area schools; columns (3)-(4) sorted, non-EP

---

been subjected to sorting.

area schools; columns (5)-(6) report for reference mean and standard deviation for the entire sample, including metropolitan and rural areas. The table shows that, even between provincial cities, the performance gap between the mixed and the sorted areas is remarkable. Upon entering high school, the difference in the average score is as large as over 9 points in grade 10, or more than 0.6 standard deviations apart, even though it narrows down in grade 11. This suggests that the comparability between the mixed provincial cities and the sorted provincial cities goes only so far. Our main analysis will address this problem with an extensive list of family, school, and teacher covariates.[24]

4. Results

4.1 Illustration

The top panel (A) in Table 3 provides an intuitive illustration of the DD strategy.

In the mixed cities, students' average score goes down by 0.26 standard deviations between grade 10 and 11. In contrast, students in the sorted cities suffer a smaller reduction worth 0.07 standard deviations. The difference between the grade differences, or 0.19 standard deviations, is considered as measuring the sorting effect. It is important to recall that the NAEA tested students in grade 10 about what they learned in middle school, which is mixed without an exception across the country. The DD strategy makes the identifying assumption that students in the sorted cities (non-EP) would have experienced the same drop of 0.26 standard deviations in their score as their counterparts in the mixed cities (EP) did, if they had been subjected to mixing (EP) as well. The assumption allows us to estimate the effect of sorting to be the difference in differences, or 0.191=-0.072-(-0.263).

[24] The higher initial performance of students in mixed cities may reflect the government's strategic choice of target cities in the implementation of the EP. The government first started with Seoul and Pusan, and then expanded the EP to cover the other metropolitan cities, and finally to cover the remaining traditional centers of provincial education with elite high schools. We consider this to be another potential source of downward bias in our positive estimates of the sorting impact.

The lower panel (B) conducts an informal test of the identifying assumption for the DD strategy, taking advantage of the scores available for the lower-grade students (grades 6 and 9). While the first and the second columns are titled "mixed" and "sorted" respectively, it should be noted that children in grades up to 9, in elementary and middle schools, are instructed in mixed schools and classrooms across the nation, whether in the EP or non-EP areas. If provincial cities in EP and non-EP areas are comparable, the double differencing should yield a number close to zero.

The actual result in panel B resoundingly rejects the claim that, in the absence of sorting, score evolutions for non-EP area students would have been identical to those for EP area students. The first column shows that the average score improves between two grades in mixed areas; the second shows that it goes down over time in non-EP areas. Given that up to grade 9, all schools are mixed, the result suggests that EP cities hold an important innate advantage over non-EP cities, perhaps reflecting stronger preferences for education among parents and availability of better stimuli.[25] [26] If true, the DD estimate found in the panel A is likely to understate the true effect of sorting. Regression controls may help address this problem, to the extent that the advantage of EP cities is accounted for by observed variables. Later we will present some evidence for this optimism.

Figure 1 illustrates the evolution of standardized average scores across grades in provincial mixed cities (EP) and sorted cities (non-EP). Presumably reflecting area effects, the scores in the mixed cities exhibit an upward trend, and those in the sorted a gradual downward trend, that is, until the 10-th grade. During the first year in high school, when the students get exposed to alternative regimes, we observe a sharp change in the two areas' relative positions in the national distribution.

4.2 Sorting and average performance

---

[25] Once again, this may reflect the strategic choice of target cities for the EP on the part of the government.

[26] Similar results hold for DD between grades 9 and 10. These are however more complicated to interpret, since a large change takes place in terms of student composition between grades 9 and 10, reflecting the beginning of the

Table 4 implements the DD strategy in a regression framework, and reports our first key results. The central results are reported in columns (3)-(6). In particular, column (3) reproduces the DD calculation reported in a tabular form in panel A, Table 3. The DD estimate is given by the coefficient estimate for the interaction term between the sorted (non-EP) area dummy and the upper grade, or grade 11, dummy. Naturally, the regression DD estimate is identical to the one reported in Table 3.

Columns (4)-(6) implement regression-adjusted versions of DD. They control for a cumulatively increasing set of covariates, first for school characteristics, then for student characteristics, and finally for after-school time allocation by students. They all point at a larger impact of sorting than is suggested by the result from the raw DD. It is reassuring that all the estimates from regression-adjusted DD estimations are roughly comparable.

Columns (1)-(2) report grade-by-grade regression of test scores. The dummy for the sorted areas (non-EP) has a significant negative estimated coefficient for grade 10 (column 2), but the deficit largely disappears by grade 11 (column 1). This change in estimated coefficients reflects the comparable effect as captured by the DD methods.

Column (7) reproduces the raw DD result for grades 6 and 9 in Table 3-B. Employing regression controls in column (8) significantly reduces the magnitude of the treatment dummy coefficient to a level that is not significantly different from zero at conventional levels of significance. With regression controls, the result in column (8) informally demonstrates that there is some merit in our identifying assumption that in the absence of sorting, evolution of scores in the EP and non-EP areas would have been comparable. The result also suggests that regression-adjusted DD estimates, as reported in columns (4)-(6) should be preferred to the raw DD estimate in column (3).[27]

Estimated signs of regression controls in columns (4)-(6) mostly conform to our prior

---

vocational school track in high school.

[27] Regression controls in the model (8) do not include dummies for private school, girls-only school, and boys-only school. Private schools are rare at the primary level in Korea, and our small-city sample does not include any. As well, all primary schools in Korea are co-ed.

expectations. Teacher experience is a significant booster for student performance. Parental education is positively associated with children's performance. The significantly large values of private school dummy coefficients are somewhat surprising, given the virtual nationalization of private schools and the inability of students and parents to choose them at will.[28] Perhaps another surprise is that private tutoring shows a negligible impact, while self-study hours are a strong predictor of performance. We do note, however, that private tutoring shows a stronger impact at an earlier stage, as shown in column (8).

Other findings from Table 4 may be of interest to readers, especially those who are interested in the changing impacts of environmental variables on educational outcome as children grow. The dummy variable for girls has a significant positive coefficient estimate in the lower grades (column 8); the gender effect gets reversed in later years (column 6). This finding is consistent with the often-heard view that schools "shortchange" girls.[29] With the exception of mother's education, family environment variables, such as father's education, mother absent, and mother housewife, are shown to have stronger impacts on performance when children are younger. Again compare coefficient estimates for the relevant variables in columns 6 and 8. The stronger impact of mother's education on children's performance in high school may have to do with the famously intense involvement of mothers in children's preparation for college entrance exams, a peculiar phenomenon in Korea. Finally, while private tutoring is shown to have some influence on performance of young children up to grade 9, this effect practically disappears by the high school phase. In contrast, hours of self-study get much more important as a factor in the determination of student performance in high schools.

Table 5 presents both raw and regression-adjusted DD estimates for individual subjects. The sorting impact is particularly pronounced for English and math, and somewhat less so for science and social studies. The estimated impact on test scores in Korean is insignificant. The results suggest that the nature of a subject may matter in the relationship between effective instruction and sorting/mixing arrangements.

---

[28] The only significant difference between public and private schools is the mandatory teacher rotation between schools in the public school system.

The estimated effects of sorting are statistically significant. But how important are they for practical purposes? Whether about 0.3 standard deviations in the national distribution of scores are a large change may depend on the eyes of the evaluator.

One way to answer the question is to refer to standardized scores at different percentile points. For instance, in mixed areas in grade 11, the 90 percentile score is 1.23, 80 percentile is 0.95, and 70 percentile 0.74. These figures suggest that the sorting advantage roughly amounts to a 10 percentile head start in the national distribution of scores. Alternatively, the estimated impact of sorting may be compared against alternative ways to raise scholastic performance, as implied by coefficient estimates for regression controls in columns (4)-(6) in Table 4. From column (6), for instance, we may infer that to buy a similar increase in average performance would need to raise average teachers' experience by 3 years in all the tested subjects, which might be quite costly. To buy a similar increase through increase in parental education would be significantly more costly. While these naive calculations should not be taken too seriously, they suggest that the sorting effect is practically important.

In addition, there are some reasons to believe that our DD estimates understate the true impact of sorting. First of all, we have to recall that our estimates are derived from the comparison of grade 10 and grade 11 students. The full impact of sorting through three full years of exposure to the sorting treatment in the duration of high school education may be substantially larger. Second, we have repeatedly noted that the government systematically selected traditional centers of education in the expansion of the EP. Thus, mixed (EP) cities may hold a more than their fair share of traditional elite schools. Preference for education among parents and students may also be stronger there.

One may be concerned that a Tiebout-style migration may bias upward the estimates of the sorting effect. Naturally, parents who value their children's education more highly would move to an area where schooling is known to be more effective, other things being equal. However, it seems that the potential bias due to migration is likely to be insignificant. Once a child starts attending high school, parents tend to get more reluctant to move, not to disrupt the

---

[29] See the AAUW report “How Schools Shortchange Girls” (1995), for instance.

child's preparation for college entrance exams. The national regulations of schools also mean that the diversity among schools is fairly limited to begin with. Finally, if student migration does bias our estimates, it is likely to be in the downward direction, because better students are more likely to be attracted to the traditional center of education in the EP areas.

4.3 Sorting and distributional effects

Table 6 reports quantile regression results to examine the distributional effects of sorting on test scores. Quantile regression is a straightforward generalization of median regression. Median regression is also a 50-percentile quantile regression and chooses coefficient estimates such that half of residuals are negative and the other half positive. A 90-percentile quantile regression chooses coefficients so that 90 percent of residuals are negative, and the remaining 10 percent positive. Quantile regressions at different percentiles are similarly interpreted. Roughly speaking, the coefficient estimates in a 90-percentile regression capture the effects of the corresponding variables near the 90 percentile of test scores. All the columns in Table 6 take on the same specification as in column (6), Table 4. Again, the variable of key interest is the sorting treatment dummy, or the interaction term between grade 11 and sorted area dummies.

Strikingly, the results in Table 6 suggest that sorting benefits students regardless of their ability level. Furthermore, the impacts are quite comparable at different percentile points, ranging from 0.25 to 0.38 standard deviations. This implies that abolition of the EP, or a regime change from mixing to sorting, could well be the proverbial rising tide that lifts all boats, a Pareto improvement.

5. Concluding remarks

Societies with a heterogeneous population have to wrestle with how best to organize communities. Segregated communities are expected to better reduce frictions and conflicts, compared to integrated communities. On the other hand, segregated communities may confirm and even raise the level of personal heterogeneities across the society. (Benabou, 1996)

A similar problem arises in the education sector. Since students and families differ from each other in their abilities and backgrounds, the society must decide at what level of integration (mixing) or segregation (sorting) it is to organize its schools.

The high-school equalization policy (EP) in Korea provides a rare opportunity to compare the relative performance of the alternative regimes, mixing and sorting. In addition to the use of lottery for randomized mixing of heterogeneous students, the EP systematically suppresses room for school diversity and room for choice through private schools, local school finance, tracking, and curriculum innovation. While subjected to the same set of regulations, schools in areas exempted from the EP use entrance exams to sort students. The resulting contrast between mixed and sorted schools allows us to isolate student composition effects from competition effects due to choice. The policy experiment also allows us a chance to examine the effects of a regime change in school organization, rather than marginal changes in the composition of students.

Utilizing a difference-in-differences empirical strategy, the paper examined test scores of high schools students from the newly available Korean National Assessment of Educational Achievement. Two major findings emerged.

First, evidence suggests a rather substantial gain in average performance due to sorting, corresponding to 0.3 standard deviations in the national distribution of scores. There are some good reasons to believe that the estimate understates the full true impact of sorting. Second, sorting seems to benefit students across the distribution of ability.

We see three policy implications from these findings. First, the extreme form of student mixing practiced under the school equalization policy in Korea holds back progress of both high- and low-ability students. If complete revocation of the policy is not feasible, measures should be sought to introduce flexibility into the system. Allowing tracking within school should be a sensible move, since it would make it easier for schools to address diverse needs of students.

Second, our findings may have some implications bearing on the school choice debate in the US. While advocates stress beneficial consequences from choice, the concern still seems to linger on that re-sorting of students in the form of so-called cream-skimming may hurt less able students. Our results suggest that this does not necessarily have to be the case. Caution is warranted with this interpretation, however, since the contrast we analyzed in the paper is between alternative regimes, not small changes at the margin in student composition.

Third, we consider the results as a contribution in the attempts to answer the question: has the high student performance in East Asian countries come about because of government regulations or in spite of them? The paper identified one instance where straightjacket regulations do harm progress of students.

References

Angrist, Joshua D., and Alan B. Krueger (1999), "Empirical Strategies in Labor Economics," in Handbook of Labor Economics, Vol. 3A, editors Orley C. Ashenfelter and David Card, North-Holland.

Angrist, Joshua D., and Kevin Lang (2002), "How Important Are Classroom Peer Effects? Evidence from Boston's Metro Program," NBER Working paper 9263.

Argys, L.M., D.I. Rees and D.J. Brewer (1996), "Detracking America's Schools: Equity at Zero Cost," Journal of Policy Analysis and Management, vol. 15, no. 4, pp. 623-645.

American Association of University Women Educational Foundation, *"How Schools Shortchange Girls: The AAUW Report",* Marlowe & Company, New York, 1995.

Benabou, Roland (1994), "Education, Income Distribution, and Growth: The Local Connection," NBER Working Paper 4798.

Benabou, Roland (1996), "Heterogeneity, Stratification and Growth: Macroeconomic Implications of Community Structure and School Finance," American Economic Review, vol. 86, no. 3, pp. 584-609.

Buchinsky, Moshe (1994), "Changes in the US Wage Structure 1963-1987: Application of Quantile Regression," Econometrica, vol. 62, pp. 405-458.

Campbell, Donald T (1969), "Reforms as Experiments," American Psychologist XXIV: 409-429.

Card, David E. (1990), "The Impact of the Mariel Boatlift on the Miami Labor Market," Industrial and Labor Relations Review, 43: 245-257.

Eissa, Nada (2001), "Taxation and Labor Supply of Married Women: the 1986 Tax Reform Act as a Natural Experiment," NBER working paper 5023.

Epple, D., and R.E. Romano (1998), "Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects," American Economic Review, vol. 88, no. 1, pp. 33-62.

Godo, Yoshihisa (2001), "Accumulation of Education in Korea's Economic Development: 1910-2000", mimeograph, Meiji Gaukuin University.

Hanushek, Eric A. (2002), "Publicly provided education," in Handbook of Public Economics, Vol. 4, edited by A. J. Auerbach and M. Feldstein. Elsevier.

Hanushek, E.A. and J.A. Luque (2003). "Efficiency and Equity in Schools Around the World," Economics of Education Review, vol. 22, no. 4.

Hanushek, E.A., J.F. Kain, J.M. Markman and S.G. Rivkin (forthcoming), "Does Peer Ability Affect Student Achievement?" Journal of Applied Econometrics

Hoxby, Caroline Minter (2000), "Peer Effects in the Classroom: Learning from Gender and Race Variation," NBER Working Paper 7867.

Hoxby, Caroline Minter (2001), "All Schools Finance Equalizations Are Not Created Equal?" Quarterly Journal of Economics, vol. 116, no. 4, pp. 1189-1231.

Hoxby, Caroline Minter (2003a), "School Choice and School Productivity (or Could School Choice Be a Tide That Lifts All Boats?)," in The Economics of School Choice, edited by C. Hoxby, university of Chicago Press.

Hoxby, Caroline Minter (2003b), "School Choice and School Competition: Evidence from the United States," unpublished manuscript.

Kim, Sunwoong, and Ju-Ho Lee (2002a), "Private Tutoring and Demand for Education in South Korea", KDI School Working Paper.

Kim, Sunwoong, and Ju-Ho Lee (2002b), "The Secondary School Equalization Policy in South Korea," KDI School Working Paper.

Sacerdote, Bruce (2000), "Peer Effects with Random Assignment: Results for Dartmouth Roommates," NBER Working paper 7469.

US Department of Education,

Zimmer R.W., and E.F. Toma (2000), "Peer Effects in Private and Public Schools Across Countries," Journal of Policy Analysis and Management, vol. 19, no. 1, pp. 75-92.

Table 1. Summary Statistics for Test Scores

	Mixed		sorted		all	
	(1)	(2)	(3)	(4)	(5)	(6)
	mean	St. dev	mean	st. dev	mean	st. dev
10th Graders	*N=317*		*N=1243*		*N=4315*	
test score – All (0-100)	62.57	11.19	53.15	16.24	52.33	15.52
– Korean (0-100)	67.35	11.92	61.70	15.44	60.42	15.73
– English (0-100)	60.38	17.05	50.17	19.98	49.64	20.25
– Mathematics (0-100)	59.58	16.24	47.86	21.09	47.55	20.19
– Science (0-100)	65.49	15.31	55.40	19.02	54.28	17.73
– Social Studies (0-100)	60.04	12.62	50.62	16.85	49.74	16.06
stand'd test score – All	0.66	0.72	0.05	1.05	0.00	1.00
– Korean	0.44	0.76	0.08	0.98	0.00	1.00
– English	0.53	0.84	0.03	0.99	0.00	1.00
– Mathematics	0.60	0.80	0.02	1.04	0.00	1.00
– Science	0.63	0.86	0.06	1.07	0.00	1.00
– Social Studies	0.64	0.79	0.05	1.05	0.00	1.00
11th Graders	*N=276*		*N=1188*		*N=4381*	
test score – All (0-100)	61.42	10.36	54.81	17.75	55.11	15.90
– Korean (0-100)	71.30	11.73	63.80	18.30	64.72	16.73
– English (0-100)	51.61	13.63	47.51	20.31	47.48	19.10
– Mathematics (0-100)	47.67	18.12	41.74	21.76	42.06	20.60
– Science (0-100)	71.41	11.62	64.06	18.87	64.31	17.39
– Social Studies (0-100)	65.11	12.91	56.96	19.72	56.98	17.56
stand'd test score – All	0.40	0.65	-0.02	1.12	0.00	1.00
– Korean	0.39	0.70	-0.06	1.09	0.00	1.00
– English	0.22	0.71	0.00	1.06	0.00	1.00
– Mathematics	0.27	0.88	-0.02	1.06	0.00	1.00
– Science	0.41	0.67	-0.01	1.08	0.00	1.00
– Social Studies	0.46	0.74	0.00	1.12	0.00	1.00

Table 2. Summary Statistics for Independent Variables

	Mixed		sorted		all	
	(1)	(2)	(3)	(4)	(5)	(6)
	mean	st. dev	mean	st. dev	mean	st. dev
10th Graders	*N=317*		*N=1243*		*N=1560*	
female student (0, 1)	0.35	0.48	0.56	0.50	0.46	0.50
mother's education (6-18)	11.98	2.53	11.47	2.49	11.67	2.62
father's education (6-18)	13.37	2.73	12.65	2.83	12.83	2.90
mother absent (0, 1)	0.01	0.10	0.01	0.09	0.01	0.12
mother housewife (0, 1)	0.62	0.49	0.58	0.49	0.60	0.49
private tutoring hours (0, 5)	1.10	1.45	1.17	1.60	1.50	1.70
self study hours (0, 5)	1.23	1.20	1.13	1.15	1.10	1.09
teacher avg age (yrs)	36.57	6.15	35.87	3.58	37.44	4.74
teacher avg experience(yrs)	10.39	5.93	10.13	3.50	11.60	4.58
school size (no. of classes)	28.30	2.81	26.22	6.82	28.51	10.56
private school (0, 1)	0.65	0.48	0.42	0.49	0.53	0.50
girl-only school (0, 1)	0.23	0.42	0.25	0.43	0.21	0.41
boy-only school (0, 1)	0.54	0.50	0.24	0.43	0.35	0.48
11th Graders	*N=276*		*N=1188*		*N=1464*	
Female student (0, 1)	0.39	0.49	0.40	0.49	0.48	0.50
mother's education (6-18)	11.94	2.16	11.15	2.26	11.47	2.72
father's education (6-18)	13.08	2.66	12.17	3.08	12.58	3.03
mother absent (0, 1)	0.01	0.08	0.01	0.10	0.02	0.13
mother housewife (0, 1)	0.53	0.50	0.61	0.49	0.60	0.49
Private tutoring hours (0, 5)	0.87	1.33	0.90	1.35	1.26	1.56
self study hours (0, 5)	1.29	1.09	1.43	1.39	1.26	1.21
teacher avg age (yrs)	37.32	3.64	38.13	3.59	38.77	4.22
teacher avg experience(yrs)	11.67	2.80	12.03	3.37	12.61	4.10
School size (no. of classes)	30.52	3.43	24.85	7.73	28.74	10.89
private school (0, 1)	0.67	0.47	0.48	0.50	0.56	0.50
girl-only school (0, 1)	0.39	0.49	0.16	0.37	0.31	0.46
boy-only school (0, 1)	0.61	0.49	0.37	0.48	0.36	0.48

Table 3. Difference in Difference of Test Score

Panel A. 10th graders and 11th graders

	Mixed	sorted	sorted-mixed
10th Graders	0.660	0.053	-0.607
	(0.721)	(1.046)	(0.062)**
11th Graders	0.397	-0.019	-0.416
	(0.701)	(1.117)	(0.070)**
11th Graders – 10th Graders	-0.263	-0.072	0.191
	(0.057)**	(0.044)	(0.093)*

Panel B. 6th graders and 9th graders

	Mixed	sorted	sorted-mixed
6th Graders	-0.065	0.140	0.205
	(0.964)	(0.983)	(0.054)**
9th Graders	0.165	0.038	-0.127
	(0.978)	(0.981)	(0.054)*
9th Graders – 6th Graders	0.230	-0.102	-0.333
	(0.064)**	(0.040)**	(0.076)**

Standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Table 4. Regression Results, Standardized Score for All Subject Averages

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
sample graders	11^{th}	10^{th}		11^{th} and 10^{th}			6^{th} and 9^{th}	
sorted vs. mixed								
dummy treatment group			0.191	0.384	0.361	0.304	-0.333	-0.149
(sorted & upper grade)			(0.093)*	(0.089)**	(0.086)**	(0.082)**	(0.076)**	(0.078)+
dummy for upper grade			-0.263	-0.394	-0.358	-0.362	0.230	0.046
			(0.084)**	(0.080)**	(0.077)**	(0.074)**	(0.065)**	(0.067)
dummy sorted	-0.068	-0.230	-0.607	-0.394	-0.332	-0.339	0.205	0.143
	(0.073)	(0.056)**	(0.064)**	(0.062)**	(0.060)**	(0.058)**	(0.054)**	(0.056)*
school characteristics								
average teacher age	-0.034	-0.206		-0.140	-0.119	-0.103		0.012
	(0.021)	(0.020)**		(0.015)**	(0.015)**	(0.014)**		(0.008)
average teacher exp.	0.059	0.202		0.132	0.116	0.101		0.002
	(0.021)**	(0.021)**		(0.015)**	(0.015)**	(0.014)**		(0.008)
school size	0.038	0.213		0.087	0.092	0.101		0.016
	(0.021)+	(0.024)**		(0.017)**	(0.016)**	(0.015)**		(0.006)**
square of school size	0.000	-0.003		-0.001	-0.001	-0.001		0.000
	(0.000)	(0.000)**		(0.000)**	(0.000)**	(0.000)**		(0.000)**
private school	0.015	0.432		0.301	0.277	0.231		
	(0.061)	(0.054)**		(0.043)**	(0.041)**	(0.040)**		
girls-only school	0.372	-0.285		-0.112	-0.005	0.000		
	(0.084)**	(0.061)**		(0.048)*	(0.052)	(0.050)		
boys-only school	0.078	0.294		0.313	0.230	0.157		
	(0.074)	(0.069)**		(0.043)**	(0.052)**	(0.049)**		
individual characteristics								
dummy for girls	-0.213	-0.083			-0.128	-0.130		0.080
	(0.078)**	(0.063)			(0.053)*	(0.050)**		(0.032)*
mother's education	0.047	0.040			0.046	0.040		0.023
	(0.013)**	(0.011)**			(0.009)**	(0.009)**		(0.008)**
father's education	0.033	0.036			0.053	0.039		0.067
	(0.011)**	(0.010)**			(0.008)**	(0.008)**		(0.008)**
mother absent	-0.069	-0.078			-0.056	-0.046		-0.363
	(0.236)	(0.229)			(0.176)	(0.168)		(0.119)**
mother housewife	0.006	0.047			0.026	0.026		0.106
	(0.049)	(0.042)			(0.034)	(0.033)		(0.033)**
private tutoring hours	0.011	0.012				0.018		0.086
	(0.018)	(0.014)				(0.011)		(0.010)**
self study hours	0.264	0.180				0.228		0.033
	(0.019)**	(0.018)**				(0.013)**		(0.016)*
Observations	1464	1560	3024	3024	3024	3024	3355	3355
Adjusted R-squared	0.278	0.360	0.043	0.166	0.220	0.291	0.005	0.113

Standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%

Note: Constants are included in the regressions, but not reported in the table.

Table 5. Difference in Difference Estimates, Individual Subjects, Grades 10 and 11

	All		English		Math	
	(1)	(2)	(3)	(4)	(5)	(6)
dummy treatment group (sorted & upper grade)	0.191 (0.093)*	0.304 (0.082)**	0.289 (0.090)**	0.366 (0.081)**	0.292 (0.093)**	0.326 (0.083)**
dummy grade 11	-0.263 (0.084)**	-0.362 (0.074)**	-0.314 (0.081)**	-0.365 (0.073)**	-0.323 (0.083)**	-0.368 (0.075)**
dummy sorted	-0.607 (0.064)**	-0.339 (0.058)**	-0.504 (0.062)**	-0.260 (0.057)**	-0.580 (0.064)**	-0.336 (0.058)**
Other control variables included? [a]	No	Yes	No	Yes	No	Yes
Observations	3024	3024	3024	3024	3024	3024
Adjusted R-squared	0.043	0.291	0.026	0.252	0.033	0.266

	Science		Social Studies		Korean	
	(7)	(8)	(9)	(10)	(11)	(12)
dummy treatment group (sorted & upper grade)	0.146 (0.094)	0.243 (0.087)**	0.123 (0.095)	0.228 (0.087)**	-0.089 (0.090)	0.108 (0.085)
dummy grade 11	-0.224 (0.085)**	-0.319 (0.078)**	-0.178 (0.085)*	-0.282 (0.078)**	-0.047 (0.081)	-0.199 (0.076)**
dummy sorted	-0.569 (0.065)**	-0.343 (0.061)**	-0.587 (0.065)**	-0.321 (0.061)**	-0.359 (0.062)**	-0.189 (0.059)**
Other control variables included? [a]	No	Yes	No	Yes	No	Yes
Observations	3024	3024	3024	3024	3024	3024
Adjusted R-squared	0.039	0.228	0.041	0.233	0.029	0.193

Standard errors in parentheses + significant at 10%; * significant at 5%; ** significant at 1%

Note: [a] Other control variables included are all the variables in categories "school characteristics" and "individual characteristics" in Table 4.

Table 6. Distributional Effects of Sorting, Quantile Regression Results, Standardized Score for All Subject Averages, 10^{th} and 11^{th} Graders, n=3024

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
percentile	10%	20%	30%	40%	50%	60%	70%	80%	90%
sorted vs. mixed									
Dummy treatment group	0.254	0.342	0.253	0.279	0.295	0.378	0.360	0.298	0.273
	(0.131)+	(0.130)**	(0.102)*	(0.107)**	(0.099)**	(0.098)**	(0.113)**	(0.105)**	(0.122)*
Dummy for upper	-0.331	-0.404	-0.340	-0.330	-0.369	-0.352	-0.330	-0.331	-0.321
	(0.120)**	(0.118)**	(0.092)**	(0.096)**	(0.089)**	(0.088)**	(0.102)**	(0.095)**	(0.112)**
Dummy sorted	-0.645	-0.526	-0.400	-0.367	-0.285	-0.305	-0.253	-0.198	-0.090
	(0.091)**	(0.090)**	(0.071)**	(0.075)**	(0.069)**	(0.069)**	(0.080)**	(0.074)**	(0.086)
School characteristics									
average teacher age	-0.174	-0.148	-0.123	-0.101	-0.096	-0.084	-0.073	-0.066	-0.036
	(0.022)**	(0.023)**	(0.018)**	(0.018)**	(0.017)**	(0.017)**	(0.019)**	(0.018)**	(0.021)+
average teacher	0.184	0.149	0.121	0.097	0.091	0.073	0.059	0.061	0.042
	(0.024)**	(0.024)**	(0.018)**	(0.019)**	(0.017)**	(0.017)**	(0.019)**	(0.018)**	(0.022)+
School size	0.087	0.133	0.131	0.128	0.130	0.113	0.132	0.100	0.102
	(0.028)**	(0.027)**	(0.020)**	(0.021)**	(0.018)**	(0.018)**	(0.020)**	(0.018)**	(0.020)**
Square of school size	-0.001	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002	-0.002
	(0.001)	(0.001)**	(0.000)**	(0.000)**	(0.000)**	(0.000)**	(0.000)**	(0.000)**	(0.000)**
Private school	0.328	0.291	0.285	0.235	0.242	0.230	0.192	0.154	0.098
	(0.067)**	(0.064)**	(0.049)**	(0.052)**	(0.048)**	(0.047)**	(0.054)**	(0.051)**	(0.061)
girls-only school	0.045	-0.109	-0.027	-0.025	0.003	-0.010	-0.001	0.012	-0.018
	(0.079)	(0.078)	(0.061)	(0.064)	(0.060)	(0.059)	(0.069)	(0.063)	(0.073)
boys-only school	0.267	0.241	0.212	0.204	0.131	0.100	0.081	0.026	-0.052
	(0.075)**	(0.076)**	(0.060)**	(0.064)**	(0.059)*	(0.059)+	(0.069)	(0.065)	(0.077)
individual characteristics									
Gender of students	0.120	0.087	0.022	-0.014	-0.170	-0.261	-0.335	-0.371	-0.353
	(0.082)	(0.080)	(0.062)	(0.065)	(0.060)**	(0.060)**	(0.070)**	(0.065)**	(0.077)**
Mother's education	0.045	0.047	0.036	0.033	0.043	0.039	0.047	0.048	0.036
	(0.015)**	(0.014)**	(0.011)**	(0.012)**	(0.011)**	(0.010)**	(0.012)**	(0.011)**	(0.013)**
father's education	-0.007	0.010	0.021	0.034	0.036	0.044	0.043	0.050	0.062
	(0.012)	(0.012)	(0.010)*	(0.010)**	(0.009)**	(0.009)**	(0.011)**	(0.010)**	(0.012)**
Mother absent	0.167	0.169	0.037	0.021	-0.148	-0.239	-0.124	-0.137	-0.367
	(0.237)	(0.257)	(0.205)	(0.210)	(0.198)	(0.192)	(0.229)	(0.209)	(0.221)+
Mother housewife	-0.018	-0.007	-0.010	0.023	0.025	-0.007	0.016	0.027	0.042
	(0.053)	(0.052)	(0.041)	(0.043)	(0.039)	(0.039)	(0.045)	(0.042)	(0.050)
Private tutoring hours	0.054	0.034	0.022	0.017	0.020	0.012	0.018	0.019	-0.004
	(0.019)**	(0.018)+	(0.014)	(0.015)	(0.014)	(0.013)	(0.015)	(0.014)	(0.017)
self study hours	0.192	0.219	0.223	0.249	0.253	0.249	0.230	0.208	0.174
	(0.023)**	(0.022)**	(0.017)**	(0.017)**	(0.016)**	(0.016)**	(0.018)**	(0.017)**	(0.019)**

Standard errors in parentheses. + significant at 10%; * significant at 5%; ** significant at 1%
Note: Constants are included in the regressions, but not reported in the table.

Figure 1. Average Score by Grade and Region