Northwestern | INSTITUTE FOR POLICY RESEARCH

RESEARCH EXCELLENCE • POLICY IMPACT

IPR Working Paper Series

WP-23-45

Good Schools or Good Students? Evidence on School Effects From Universal Random Assignment of Students to High Schools

Paulo Bastos
World Bank

Julian Cristia
Inter-American Development Bank

Beomsoo Kim
Korea University

Ofer Malamud
Northwestern University and IPR

Version: December 19, 2023

Abstract

How much do schools differ in their effectiveness? Recent studies that seek to answer this question exploit random assignment generated by central allocation mechanisms or oversubscribed schools. However, the resulting estimates, while causal, may also reflect peer effects due to differences in student composition across schools. The researchers exploit universal random assignment of students to high schools in certain regions of South Korea to provide estimates of school effects that better reflect the effects of school practices and policies. They find significant effects of schools on scores in high stakes college entrance exams: a 1 standard deviation increase in school quality leads to 0.05–0.08 standard deviations higher average academic achievement in Korean and English. Analogous estimates from areas of South Korea that do not use random assignment, and therefore include the effects of student selection and peer effects, are substantially higher.

The authors are grateful to seminar participants at the University of Maryland, University of California-Merced, OECD, Inter-American Development Bank and at several conferences for very helpful comments and suggestions. Lisseth Escalante, Matias Martinez, and Anastasiya Yarygina provided excellent research assistance. They are grateful to Pedro Carneiro for providing the code for estimating school effects. Beomsoo Kim gratefully acknowledges financial support from a National Research Foundation of Korea Grant funded by the Korean Government (NRF-2018S1A5A2A03028632) and a Korea University Grant (K2312841).

1 Introduction

Measuring differences in school effectiveness has been a central topic in the economics of education (Angrist et al., 2022). Recent studies have used random assignment generated by oversubscribed schools that admit students by lottery to provide causal estimates of school effects in different settings (e.g., Cullen et al., 2006; Abdulkadiroğlu et al., 2011; Deming et al., 2014; Cohodes et al., 2021). More recently, novel methodological approaches have been developed to fully exploit the random order assigned to students with identical match criteria in centralized allocation mechanisms (Abdulkadiroğlu et al., 2017). Yet, in most centralized allocation mechanisms and in settings with oversubscribed schools, not all students are assigned in a randomized fashion. Moreover, student preferences and school priorities also play a substantial role in determining how students are allocated to schools such that student composition typically differs widely across schools in terms of prior academic achievement and socioeconomic status. Hence, school effects estimated using oversubscribed schools or by exploiting the randomness in centralized assignment mechanisms, while causal, may reflect differences in peer quality across schools, in addition to differences that arise from school practices and policies.

Disentangling the sources underlying the variation in school effects is important for policy. As Deming (2014) notes, most papers estimating school effects "cannot unpack the impact of changing school assignment into changes in peer quality, teacher quality, or other important inputs." From a parent's perspective, deciding which school to send their children should be based on causal estimates that include the influences of peer quality because they capture the full effect of schools on children. Indeed, many studies have documented the importance of peer effects in shaping academic outcomes in schools (e.g., Duflo et al., 2011; Busso and Frisancho, 2023, and many others).[1] From a policy perspective, it is important to assess whether school differ in their effectiveness based on inputs and practices that are within their control.[2] For example, if certain schools outperform other schools while keeping peer quality constant, then disseminating their practices could potentially lead to improvements in learning. Similarly, high-stakes decisions involving the reorganization and closure of schools deemed ineffective or the expansion of schools considered highly effective should be based on school effects that do not reflect peer quality which

[1] In this context, the effects of peer quality on student achievement can include direct channels, such as peer learning and class disruptions, as well as indirect channels, such as teacher behavior and parental involvement. (Sacerdote, 2011). We discuss the evidence on peer effects in more detail below.

[2] Raudenbush and Willms (1995) distinguish between school effects that are due to school practices and those due to peer quality and school contexts outside of the control of school administrators and teachers.

are outside of a school's control.

But can we estimate school effects while balancing peer quality across schools? The ideal setting for estimating such school effects would consist of schools in which all students were randomly assigned. This would, on average, balance peer quality across schools so that differences in average school performance would reflect variation in inputs and practices but not differences in student composition related to prior academic achievement and socioeconomic status.

In this paper, we provide estimates of school effects on academic achievement exploiting a natural experiment that resembles this ideal setting. In particular, we focus on a set of administrative divisions in South Korea where all students attending general high schools were randomly assigned to these schools between 1995 and 1997. We use individual-level administrative data to estimate the variation in school effects on student scores in national college entrance exams. To document the extent of variation in school effectiveness, we estimate the standard deviation of school effects. These estimates can be interpreted as the expected increase in learning if students were moved to schools that are one standard deviation higher in the distribution of school effectiveness.

In addition to the randomness of student assignment within selected areas, Korea's high school system is well suited for this analysis. First, it is characterized by virtually universal high school attendance and very low repetition, attrition, or inter-school student movements. These features substantially alleviate the concern that the estimated school effects might be contaminated by endogenous student transfers across schools or differences in high school graduation across schools. Second, the high stakes nature of college entrance exams, taken by 99% of students at the end of general high school, makes them a reliable measure of academic achievement. Third, the main results in this paper use data for more than 230,000 randomized students over three years, representing a large fraction of the total high school population. The sheer size of these data allows for the precise estimation of the variation in school effects, even in smaller sub-samples.

Our estimates indicate that schools vary considerably in their effectiveness. A 1 standard deviation increase in school quality within a district results in 0.05 to 0.08 of a standard deviation higher levels of academic achievement. The estimates are similar for boys and girls, but tend to be higher in the English language (up to 0.10 of a standard deviation) than in Korean (as low as 0.03 of a standard deviation). However, the magnitudes of our estimates are substantially smaller than the 0.20-0.22 of a standard deviation in math achievement documented by Angrist et al. (2017) based on Boston's centralized assignment mechanism. This difference may be due to the relatively muted effects of peer quality in a setting with universal random assignment, although we cannot rule out other differences due to contextual factors.[3] We also find persistence in our estimated school effects, with an average correlation of 0.5 across years, comparable to value-added estimates of school effects in North Carolina (Kane and Staiger, 2002).

Unfortunately, we are not able to estimate traditional value-added models using the college entrance exams because we lack data on baseline test scores. However, to further support the interpretation of our estimates, we perform two additional analyses. First, we estimate the standard deviation of school effects in Korean administrative divisions where students were sorted into schools based on student preferences, performance in exams, and middle school GPA. We find that the standard deviation of academic achievement across schools in non-randomized areas in Korea is almost 10 times larger than in randomized areas, ranging from 0.45 to 0.66 standard deviations. These results are comparable to estimates constructed based on raw average school performance (without controlling for baseline achievement) reported by Angrist et al. (2017) and Deming (2014). Second, we perform a similar analysis using data from the PISA 2000 for Korea and a selection of other countries. The estimates of school effects for Korea using the PISA data, which reflect both randomized and non-randomized districts, are 0.35 standard deviations and on the lower end of the range across countries (0.23 to 0.83 standard deviations).

An important consideration for interpreting our findings and evaluating their external validity is the extent to which the variation of school effects, estimated in the context of balanced peer quality, is influenced by institutional factors and school incentives that potentially affect the allocation of key school inputs. We examine this issue by comparing the variation of school effects between privately-founded schools, which have discretion in selecting principals and teachers, and public schools, which lack such discretion. Interestingly, while we do observe a somewhat larger standard deviation of school effects in privately-founded schools than public schools for girls, this is not the case for boys. Moreover, using data from the cross-section of countries surveyed in PISA 2000, we document that schools in Korea have a relatively high degree of autonomy, and display considerable heterogeneity in observed inputs. This evidence suggests that our main results are likely to apply more broadly.

To our knowledge, this is the first paper to estimate the standard deviation of school effects in a system with *universal* random assignment. As already mentioned, the closest studies to our own are those estimating the effectiveness of oversubscribed schools that admit students

[3] Our estimates are comparable to non-experimental estimates of the standard deviation of school value-added reported by Deming (2014) in Charlotte-Mecklenburg which range from 0.05 to 0.11 across specifications.

using lotteries (e.g., Cullen et al., 2006; Abdulkadiroğlu et al., 2011; Cohodes et al., 2021) and those fully exploiting the randomization embedded in tie-breaking rules in centralized allocation mechanisms (Abdulkadiroğlu et al., 2017). Compared to those from oversubscribed schools, the latter estimates derived from centralized allocation mechanisms are representative of a broader population of schools including those that are not necessarily oversubscribed. However, in all of these studies, the estimates of school effects, while causal, may also reflect peer effects due to differences in student composition across schools.

A related literature examines the validity of value-added methods for estimating the effect of schools on academic achievement by comparing them to those derived from random assignment (Deutsch, 2013; Deming, 2014; Angrist et al., 2017). Evidence from these studies indicate that value-added measures which adjust for lagged achievement are substantially less biased than naïve comparisons of schools' raw test scores. Though we are not able to estimate value-added measures in our context, we do show that the raw test scores differences are vastly different across schools in regions that do not randomly assign students to schools.[4]

There is also an extensive literature on the effectiveness of different school inputs and practices that would be captured by the school effects we seek to estimate, such as class-size (e.g., Krueger, 1999; Angrist and Lavy, 1999), teacher quality (e.g., Rockoff, 2004; Rivkin et al., 2005), the choice of curriculum (e.g., Cortes and Goodman, 2014), and management practices (e.g., Bloom et al., 2015).[5] A complementary literature has examined the role of peer effects in schools. The review by Sacerdote (2011) summarizes the effects of peer background on test scores as "modestly large", while noting the wide range of estimates and the importance of non-linear effects. Moreover, studies that exploit the random assignment of students to generate shocks in peer quality have documented sizeable effects on student achievement. For example, Duflo et al. (2011) show that an increase in baseline peer achievement of one standard deviation increases student achievement by 0.35 standard deviations in primary schools in Kenya. In addition, Busso and Frisancho (2023) analyze the effects of randomly assigning middle school students to tracked and non-tracked classroom in Mexico and infer an important role for peer effects from high achieving students, broadly construed. Furthermore, using quasi-experimental methods, Lavy et al. (2012) and Carrell and Hoekstra (2010) find evidence for substantial negative peer effects from low-achieving students

[4]In addition, some recent studies examine the effect of schools on non-test scores outcomes such as socio-emotional learning, school behavior and educational attainment (Loeb et al., 2018; Jackson et al., 2020).

[5]The link between school practices and school effectiveness is explored by Dobbie and Fryer Jr (2013) who correlate school effects with a variety of different school inputs and practices in charter schools.

in Israel and disruptive students in Florida, respectively.

Finally, some studies have exploited the random assignment of students to high schools in Korea to analyze how different school features affect learning and other outcomes. Park et al. (2013) examine how attending a single-sex school impacts test scores in college entrance exams and college attendance; Park et al. (2018) study how attending these types of schools influences students' interests and major choices; and Hahn et al. (2018) examine how attending privately-founded schools, as opposed to public schools, impacts test scores and college attendance.

The remainder of the paper proceeds as follows. Section 2 describes the institutional background that gives rise to the natural experiment and the data employed, before providing evidence supporting the validity of the experimental design. Section 3 presents the empirical strategy. Section 4 presents the main results while Section 5 shows that these results are robust to alternative samples and specifications. Section 6 provides a discussion on the external validity of our findings. Section 7 concludes the paper.

2 Experimental design and data

2.1 Institutional background

In Korea, children between the ages of six and fifteen are required to attend school. Compulsory education consists of six years of elementary school, followed by three years of middle school. Students typically attend their local elementary and middle schools, and do not have considerable school choice until the end of compulsory education. After completing middle school, students enter high school, which takes another three years to complete. Although enrollment in high school is not mandatory, over 97% of students from the corresponding population cohort graduated from it in 2005.

High schools are classified as either general, vocational, or selective. General high schools provide advanced general education along with elective courses, which students select on the basis of their intended university studies. Vocational high schools offer the education necessary to enter a specific profession, and are frequently focused on one occupational area, such as agriculture, commerce, or technology. Selective schools provide a more specialized curriculum, have greater autonomy, and select students in a competitive process based on GPAs and interviews. Selective high schools absorbed less than 1% of students entering high schools in randomizing administrative divisions in Korea, while vocational high schools absorbed about one quarter of these students.

The remaining three quarters of students entering high school were randomly assigned to general high schools.[6]

During the relevant period for our analysis, the process for assigning students to high schools had two rounds. In the first round, common to all administrative divisions in the country, interested students applied and were assigned to selective or vocational schools. The second round, that allocated the remaining non-assigned students to general high schools, varied across three groups of administrative divisions. In one group, there was universal random assignment of students to general high schools within school districts. In a second group, students applied to high schools and assignment decisions were made individually by schools largely on the basis of test scores and middle school GPAs. Finally, in a third group, a subset of districts randomized students to high schools, while other districts followed the application and admission procedure.

The geographic variation regarding the assignment of students to general high schools stems from the partial implementation of the "High School Equalization Policy" across administrative divisions. The central feature of this policy was the randomization of students to high schools. It was adopted largely in response to a status quo characterized by fierce competition for elite high schools. In addition to balancing the composition of students across high schools through random student assignment, the equalization policy initially aimed at equalizing the quality of teachers and facilities across schools. Monetary transfers were centralized and balanced across schools, education facilities were upgraded, and teacher training was provided. However, these other components were not successfully implemented, as budgetary constraints made it infeasible to incur the relatively high costs associated with teacher training and facility improvement (Korea Education Development Institute, 1998).

The equalization policy was first implemented in 1974 in Seoul and Busan (the largest metropolitan cities). It was then progressively expanded to include other metropolitan cities, provincial capitals and finally major regional cities.[7] The equalization policy was never implemented in some smaller administrative divisions. Between 1980 and 1995 the system remained essentially stable. For cohorts that started in 1996 and took the national exam in 1998, some limited choice was reintroduced in certain administrative divisions where the equalization policy was implemented. In particular, students identified the two or three schools of their preference. Schools then filled

[6]These proportions were computed using administrative data from CSAT, described below. Data from the 2008 KELS (also described below) reveal that vocational and selective high schools absorbed about 26% and 2% of students entering high school, respectively, whereas general high schools absorbed the remaining 72% of students.

[7]Lee (2012) provides detailed information on this process, while Park (2013) offers a comprehensive description of the equalization policy in Korea.

30 to 40% of slots by random selection among students who showed preference for the school. The remaining slots were then randomized across students residing in the corresponding school district who were not assigned to their preferred school. For this reason, our empirical analysis focuses on the cohorts that took the national exam in 1995-1997, and therefore did not experience this increase in choice.[8]

The high school system was characterized by virtually universal enrollment and very low repetition, attrition, or inter-school student movements. There were different types of general high schools: privately-founded and public schools; single-sex and coeducational schools. High schools also varied in size. All schools operated under similar centralized policies regarding fees and tuition, curriculum, and the qualifications and salary schedules of teachers. Principals controlled daily operations and the allocation of budget and other school resources. All general high schools, regardless of their type, were subject to random student assignment in the corresponding district. Because the majority of general high schools were single-sex, random assignment of students to high schools was performed separately for boys and girls. Students had to accept the randomly assigned school unless they moved to a different school district. If that district was also subject to the equalization policy, these students would be allocated by random assignment. Although it was possible for students not to comply with the random assignment through geographical mobility, evidence suggests that non-compliance was very limited (Park et al., 2013; Park, 2013).

There was substantial variation across public and privately-founded schools regarding personnel matters. In public schools, teachers were government employees who were hired in a centralized fashion based on their performance in a standardized exam. Teachers had to move to a different school within the administrative division every four or five years. Principals in public schools were selected by the regional educational office and could remain in their position at most for two four-year terms. Thus, public schools did not have any discretion regarding the selection of teachers and principals. In contrast, there was substantial discretion regarding personnel decisions in privately-founded schools. The school's board of directors was responsible for the appointment and promotion of principals. Principals in turn had control over the hiring and dismissal of teachers, and the length of their contracts. Thus, privately-founded schools had discretion in the selection and dismissal of principals and teachers.

[8] Until 1993, Korea had a universal college entrance exam named the Student Achievement Test (the *hakruk* exam), but data are not available. The College Scholastic Ability Test started in 1994, but with a slightly different format. Hence, data from this test are available only since 1995.

2.2 Data

We use data from three main sources. First, to assess how schools differ in their impacts on student learning, we use individual-level data from the College Scholastic Ability Test (CSAT). This is a high stakes test required for entry to university and performance has a major impact on their subsequent educational prospects. The CSAT college entrance exams are taken by 99% of students attending general schools at the end of high school. We use CSAT data for the years 1995 to 1997, corresponding to students who were assigned to high schools between 1992 and 1994 before the equalization policy was partially reversed in some areas.[9] For each year, the CSAT data include information on gender, the identity of the high school attended by each student, the name of the administrative division, and the raw scores in each subject.

The structure of the CSAT exam is as follows: Two thirds of the exam are identical across the whole country. This common component assesses proficiency in Korean and English languages, as well as in part of the Mathematics curriculum. The remaining third is choice-based and tests proficiency in Science or Social Science or in the remainder of the Mathematics curriculum (depending on one's curricular focus). Our analysis focuses on test scores in Korean and English, as these scores are comparable across students, schools and districts. We normalize the raw score of each subject to have mean zero and standard deviation one in the full CSAT sample in each year. In addition, we construct a summary measure of academic achievement by averaging the two standardized subject scores, and re-standardized this variable so that it has mean zero and standard deviation one in the full CSAT sample. The CSAT administrative data contain information on the names of the school and the administrative division, but not on school type (single-sex, public) or school district. We merged this information from the annual statistics book of each administrative division.

Second, to examine the validity of the experimental design, we use individual-level data on students in general high schools from the Korean Education Longitudinal Study (KELS). This is an annual longitudinal survey that has been conducted since 2005 by the Korea Educational Development Institute, a government-funded research institute. The first cohort of the KELS consists of 6,908 students in the first year of middle school in 2005. The student and school samples are drawn as a stratified random sample to reflect the national population of seventh graders in middle schools.[10] Students sampled by KELS are administered a series of socio-demographic and

[9] The Korean school year is from March to February. Students take the CSAT test in November of a given year and graduate from high school in February of the following year. The CSAT data are coded by graduation year.

[10] In a first step, 150 schools are selected nationwide in consideration of the regional distribution of schools and

school related questionnaires. In each wave of KELS, student academic performance is measured by achievement tests for three subjects: English, Korean, and Mathematics. We consider a sample of students in the randomizing areas excluding Seoul.[11] Since the KELS data was collected starting in 2005, and the equalization policy started to be reversed in 1996, these data may include some areas where student assignment was not fully random. KELS does not provide information on school districts, and hence we use the middle school as a proxy for the school district.[12]

Finally, to provide a benchmark for the main empirical analysis, we use data from the 2000 round of the Program for International Student Assessment (PISA). PISA is a standardized international assessment coordinated by the Organization for Economic Cooperation and Development that measures academic achievement of 15 year old students in Mathematics, Reading, and Science every three years. We use these data to document how Korea compares to other countries with regard to: (i) the degree of variation in academic achievement across schools; (ii) the degree of heterogeneity in observed school inputs across schools; and (iii) the degree of school autonomy in academic, personnel and budgetary decisions. The school population included in PISA consists of all schools that have at least one 15 year old student attending the school. We use student level weights to generate a representative sample at the national level. For consistency with the analysis using the CSAT and KELS data, we drop observations for vocational schools in constructing the PISA sample. After dropping observations for Canada (which did not provide a representative sample) and for Norway and Poland (that have missing values for measures of school autonomy), we are left with data for 40 countries.

2.3 Sample construction

In this section, we describe how we construct the samples used in the main analysis to estimate the variation in school effects. We start with the individual-level data from CSAT for students taking exams in 1995-1997. We drop students in the CSAT data that were not taking the exam for the first time, as well as those in vocational and selective high schools. The resulting sample contains about 1,150,000 student observations. Additionally, we drop students attending schools with low enrollments of less than 100 students (although we check that our main results are robust

students. In each school, 50 students from the target grade are drawn at random, while all students are drawn if there are fewer than 50 students in the target grade.

[11] As described in more detail in subsection 2.3, we do not include Seoul when using the CSAT data to estimate the variation of school effects among students that were randomly assigned to general high schools. Hence, for consistency, we also drop Seoul when checking for randomization.

[12] Table A1 in the appendix provides summary statistics on these data.

to the inclusion of smaller schools). This restriction excludes small schools from remote areas and ensures a minimum sample size for estimating school effects. Imposing this restriction reduces the sample by 2.7%. Finally, we impose two minor constraints which leaves us with the working sample of 1,083,237 student observations over 1995-1997.[13]

We start by defining the "randomized sample," which we use to estimate school effects arising from differences in school policies and practices. This sample is composed of students that were randomly assigned to general high schools within districts, and for whom we have information on the composition of the school district. To construct this sample, we start by focusing on the 6 out of 17 major administrative divisions that randomized all students to high schools in the final round of the assignment process.[14] We then drop Seoul and Incheon because we do not have sufficient information to specify the universe of schools to which students can be randomly assigned.[15] In contrast, in the metropolitan cities of Busan, Daegu, Gwangju and Daejeon each student was randomized to the set of general high schools included in the corresponding school district, and we are able to determine the exact composition of all school districts in each of these metropolitan cities.

We define a "non-randomized sample" composed of students in the administrative divisions of Gangwon-do and Jeollanam-do that did not randomize any student in the second round of the admission process. The rest of the sample contains students in "mixed divisions," which failed to meet the conditions for inclusion in the randomized sample for varying reasons. Some divisions comprise both urban and rural areas and implemented random assignment in the former areas but not in the latter. For this reason, we restrict the randomized and non-randomized CSAT samples refer to the period 1995-1997. Figure 1 depicts the administrative divisions that compose each of these samples.

Table 1 reports summary statistics for the CSAT data in 1995-1997. Column (1) refers to the full CSAT data, while columns (2) and (3) report statistics for the randomized and non-randomized samples, respectively. The statistics in column (2) reveal that, in the randomized sample, about 71% of students were enrolled in privately-founded schools, 96% of students attended single-sex schools, and the average number of students enrolled in each schools was 540.

[13] First, we drop observations with inconsistent information regarding single-sex status. That is, we drop 83 girls attending all-boys schools. Second, we drop 586 students taking the exam in a city with only two schools.

[14] See Table A2 for a list of Korea's administrative divisions.

[15] In Seoul, students were randomly assigned to high schools for which the commuting time from their homes (using public transportation) was estimated not to exceed 30 minutes (Kim and Kim, 2015). For Incheon, we were unable to obtain the exact composition of school districts

These proportions are larger than in the full sample, and much larger than in the non-randomized sample, where only 42% of students were enrolled in privately-founded schools, about 69% of students attended single- sex schools, and the average number of students enrolled in each school was 374. They also reveal that students in the randomized sample tend to perform considerably better in the CSAT exams than students in the non-randomized sample.

2.4 Validity of the experimental design

The existence of universal random assignment of students to high schools in certain areas of Korea is well documented in the existing literature (e.g., Park et al., 2013, 2018; Hahn et al., 2018). To provide some additional evidence on the validity of the experimental design, we use individual-level data on students in general high schools from the KELS data. We use these data to examine whether middle school test scores and household socio-demographic attributes predict observed characteristics of the high school that students attend. We estimate the following equation:

$$y_{im} = \alpha + \beta X_{im} + \phi_m + \varepsilon_{im} \tag{1}$$

where i indexes the individual student and m the middle school she attended; y_{im} is a different observed high school attribute in each regression (that is, a dummy variable indicating whether the high school the individual is assigned to is privately-founded or single-sex, the total school enrollment and the average class size); X_{im} is a vector of observed attributes of the student or the corresponding household, notably the test scores in middle school, parental education and income, and family size; ϕ_m are fixed effects for the middle school of origin; and ε_{im} is the error term, which is clustered at the middle school level. Given random assignment of students to high schools within districts, we would not expect to observe a systematic association between the attributes of students and the corresponding high schools.

It is important to note that the KELS data have two limitations for the purpose of confirming that students in the randomized sample in CSAT were randomly assigned to general high schools within districts. First, KELS does not identify each administrative division, but contains information on whether students are in a randomized area. Hence we are including all students in randomizing areas. Second, we cannot include district fixed effects because high schools are anonymized in the available KELS data, although we address this issue by adding middle school fixed effects (since students attending the same middle school typically belong to the same high school district).

Table 2 reports the estimation results for randomized and non-randomized samples respectively, separately by girls (panel A) and boys (panel B).[16] The point estimates in column (1) show that students assigned to privately-founded and public schools tend to have similar test scores in middle school, similar levels of parental education and household income, and to be originated in families of similar size. Similarly, columns (2)-(4) show that these student attributes are also unrelated to class size, enrollment, and whether the high school is a single-sex school. The sole exception concerns parental education, which has a positive and weakly significant coefficient for enrollment in the case of girls.

In contrast, the analogous results for the non-randomized sample suggest that student characteristics do predict some of the attributes of the high schools that students end up attending. Most notably, columns (6)-(8) reveal that test scores in middle school are systematically related with school attributes in non-randomized areas, where student admission is determined by entrance exams, middle school test scores, or both. This can also be observed by the higher R-squared values for the non-randomized sample than the randomized sample. Even if this latter sample might contain some contamination, it is reassuring that we do not observe a systematic association between observed attributes of students and the corresponding high schools.

3 Estimation strategy

This section presents the empirical models used to examine whether schools differ in the effects they have on student learning. In line with the literature on classroom and teacher effects (Chetty et al., 2011; Araujo et al., 2016), we assess the role of school effects in shaping academic achievement by estimating an equation of the form:

$$y_{ids} = \gamma_d + \phi_s + \varepsilon_{ids} \tag{2}$$

where y_{ids} is the test score for student i from school district d and school s, γ_d is a district fixed effect, ϕ_s are school effects, and ε_{ids} is the error term. Randomization pools are defined by the interaction of gender, district and year. We will therefore estimate separate regressions by gender and year.

[16] These patterns look similar when we consider an alternative specification in which we regress student and household characteristics on school characteristics, or if we compare average student and household characteristics for different school characteristics (e.g., public versus privately-founded).

The district fixed effects γ_d account for the heterogeneity of students across districts, which might be expected to affect test scores. Because students are randomly assigned to schools within each district, student baseline ability should be orthogonal to ϕ_s. In this framework, ϕ_s are school effects that vary within districts. They reflect school policies and practices such as learning time, school resources and curriculum, attributes of principals and teachers, organizational practices and school culture. With universal random assignment, these school effects should not reflect differences in peer quality because student composition should be balanced across schools. To assess whether these school effects matter for achievement, we estimate (2) using a fixed effects specification for ϕ_s. Following Araujo et al. (2016), we focus on estimating $V(\phi_s)$, where $V(.)$ indicates variance. Note that the variation in school effects within districts could be lower than the variation in school effects across the entire country.

A complication arises because $V(\phi_s)$ overestimates the true variance of the school effects because of sampling error. Again, in line with this literature, we subtract out a term that corrects for over-dispersion due to sampling error in the fixed effects, thereby obtaining the appropriately *shrunken* school effects. For studies examining teacher effects, the adjustment for measurement error tends to reduce the estimates considerably. Given that we have a relatively large number of students per school in our setting, this adjustment only leads to a minor reduction in our estimates (as documented in the next section).

An alternative approach uses random effects to estimate school effects. Assuming that these random effects are normally distributed, we can compute the standard deviation of school effects by averaging the school-level residuals generated from equation (2) following Deming (2014) and others in the teacher value-added literature. In practice, these estimates are very similar to the fixed-effects estimates, but we present both to alleviate any concerns about our estimation method.

Exploiting the orthogonality condition, we can also examine the influence of a vector of observed school attributes, X_s on test scores, by estimating an equation of the form:

$$y_{ids} = \alpha + \gamma_d + \beta X_s + \varepsilon_{ids} \tag{3}$$

where the observed school attributes are measured in the corresponding year of observation and the remaining variables have the meaning defined above. In this case, the estimation does not include school fixed effects.

4 Results

4.1 Main results

We begin by presenting our main estimates of school effects on academic achievement in Korean high school districts where students are randomly assigned into high schools. Table 3 presents these estimates using both the fixed-effects and random-effects methodology described in the previous section, showing separate estimates by gender (Panels A and B), the two compulsory subjects on the high stakes college entrance exams (English and Korean), and for each year of the sample period (1995, 1996 and 1997). Overall, the estimated standard deviation of school effects on academic achievement range from 0.03 to 0.10 standard deviation units. These estimates capture school effects on academic achievement without the confounding effects of student sorting. Moreover, the fact that that all students are randomly assigned implies that these effects do not reflect peer effects due to differences in student composition across schools.

Looking across the different estimates in columns (1) to (3), we observe that the standard deviation of school effects are very similar across the years 1995, 1996, and 1997, respectively. They also appear to be quite similar by gender. On the other hand, the estimates of school effects are always larger in English, where they range from 0.08 to 0.10, than in Korean, where they range from 0.03 to 0.06. This is consistent with the notion that schools have a stronger impact on proficiency in a foreign language, which is not commonly used at home or in everyday life, than on proficiency in the native language. The estimates using random effects models in columns (4) to (6) are very similar to those based on fixed-effects estimates.[17]

It is instructive to compare the magnitude of our causal estimates of school effects to those estimated by Angrist et al. (2017) based on Boston's centralized assignment mechanism. In their Table VI, they show that the standard deviation of school effects for sixth grade math scores range from 0.20 to 0.22, which is substantially higher than our estimates. One explanation for this difference is the relatively muted effects of peer quality in a setting with universal random assignment whereby student composition is balanced across schools. However, we cannot rule out that this difference is due to other contextual factors that vary between Boston and Korea. Interestingly, the magnitudes of our estimates are more comparable to non-experimental estimates reported by Deming (2014) for Charlotte-Mecklenburg.

[17] We also consider estimates from fixed-effects models that do not employ an Empirical Bayes shrinkage procedure, as shown in Appendix Table A3. Given the relatively large number of observations per school, whether or not we apply the Empirical Bayes shrinkage procedure does not make much difference to our estimates.

4.2 Comparison to non-randomized districts

In order to benchmark our main results, we present analogous estimates of the standard deviation of "school effects" in districts where students are not randomly assigned to schools. These are presented in Table 4, which follows the same structure as Table 3. We can observe that the estimates for these non-randomizing districts range from 0.45 to 0.66 standard deviation units, almost an order of magnitude larger than most of the estimates for the randomized sample. Of course, these substantially larger estimates reflect the non-random variation in student characteristics (as well as any resulting peer effects) in addition to the effects of schools on academic achievement.

In terms of heterogeneity, the estimates for the non-randomizing districts are broadly similar across the different years. They are somewhat larger for English than for Korean, mirroring the pattern for the randomized sample but not as strongly. However, there are substantial differences by gender: the estimates for girls are about 0.1 standard deviations larger than for boys, perhaps because the non-random selection into schools is different by gender. Finally, the estimates based on the random-effects model are very similar to those from the fixed-effects model.

The large differences between the estimates for the randomizing and non-randomizing school districts suggests that they might be attributed to the nature of the selection process. However, as shown in Table 1, the randomizing and non-randomizing districts also differ in some dimensions even though funding levels and curriculum are mandated to be the same. Therefore, in a subsequent section 5, we also attempt to adjust for some of these differences explicitly through a city-level comparison.

It is worth noting that our estimates for non-randomized school districts are broadly comparable to those from other studies that are based on raw average school performance, i.e., estimated without controlling for baseline achievement or demographics. For example, Angrist et al. (2017) report that the standard deviation of school effects in an "uncontrolled" regression model is approximately 0.5 using data on math achievement from middle schools in Boston. Similarly, Deming (2014) reports estimates of school effects without the inclusion of any covariates range from 0.40 to 0.43 based on math and English language scores for grades 3-8 in the Charlotte-Mecklenburg school district. This suggests that the patterns of student sorting in Korea are not especially different from those that arise in other contexts.

We also examine the persistence of school effects over time. We follow McCaffrey et al. (2009) and define persistence as the proportion of variability due to persistent effects, corresponding to the cross-year correlation for estimated effects from adjacent years. We calculate the Spearman correlation for school effects by gender for both randomized and non-randomized districts (as shown in Appendix Table A4). The persistence in randomized districts is approximately 0.5, broadly in line with the estimates for value-added scores in Kane and Staiger (2002) for North Carolina. Note that the analogous correlations in non-randomized districts are substantially higher at around 0.9 but comparable to the estimates for level scores in Kane and Staiger (2002).

4.3 Public versus privately-founded schools

An important consideration for interpreting our estimates of school effects is the extent they are influenced by institutional factors that potentially affect the variation in key school inputs, such the ability to select principals and teachers. It is possible that public schools in Korea, characterized by regular rotation of teachers across schools, may be different from those prevailing in areas where there is more sorting of teachers to schools. To provide evidence on this issue, we compare the variation of school effects between public and privately-founded schools. In privately-founded schools, there is a school specific selection process of principals and teachers, who can either be dismissed or stay in the school for long periods of time depending on performance. By contrast, public schools do not have any discretion regarding selection of teachers and principals.

Table 5 reports the results for public and privately-founded school from 1995 to 1997 for the randomizing districts. For girls, we see that the standard deviation of school effects is indeed larger in privately-founded schools than in public schools: 0.06 versus 0.02 in both 1995 and 1997, with a slight smaller gap in 1996. For boys, however, the evidence is less clear. The standard deviation of school effects is larger in privately-founded schools in 1995 and 1997, but the gap is reversed in 1996. Interestingly, these results suggest that the variation in school effects documented in the main sample is broadly similar to that obtained in settings where there is considerably greater sorting of principals and teachers. This is in contrast to the results for districts which do not randomize students to schools, where the standard deviation of school effects is substantially larger among public schools than privately-founded schools (see Table A5 in the Appendix).

4.4 Effects of observed school attributes

The analysis above suggests that school effects matter for academic achievement. But which factors are able to explain the differences in effectiveness documented across schools? In Table 6 we examine whether observed school attributes are important drivers of academic achievement in college entrance exams. Column (1) estimates the model specified in equation (3) for areas with universal random assignment of students to high schools. For this sample, the results show relatively weak associations between observed high school attributes and performance in college entrance exams (with the exception of enrollment among girls). Column (2) presents point estimates from a similar analysis for districts that do not randomize students to schools. Here, the estimates suggest that girls and boys enrolled in public and large schools tend to perform substantially better in college entrance exams in non-randomized areas. Obviously, in the absence of random assignment, the variation in academic achievement across school types may simply reflect the heterogeneity in student composition across schools. Overall, these results suggest that, once student baseline attributes are balanced across schools, the observed attributes of schools in Korea have a limited impact on academic achievement.

5 City-level comparisons

One important concern when comparing across randomized and non-randomized samples is that they differ in other dimensions beyond the student assignment mechanism. As noted earlier, the equalization policy leading to the random assignment of students to high schools was implemented only in urban areas. Therefore, we conduct an alternative analysis at the city-level where we can adjust for differences in observed attributes across randomized and non-randomized cities. The downside of this approach is that we can only compare average achievement across schools within cities rather than within districts (so our estimates are not completely clean). In addition, we conduct several robustness checks to verify that our estimates are robust to alternative specifications.

The city-level analyses includes three sets of cities: (i) the four administrative divisions/cities used in the baseline analysis where all students attending general high schools were randomly assigned to schools within clearly delineated districts, (ii) other cities that randomized students to schools, including Seoul and Incheon,[18] as well as cities in administrative divisions where only some cities featured student randomization, and (iii) cities that did not randomize students to schools, including those in the two administrative divisions used in the baseline analysis that did not feature student randomization, as well as other cities that did not randomize students in administrative divisions where only some cities featured student randomization. We then restrict attention to cities with at least 4 schools for boys or girls in each respective year. Table A6 in the

[18] As explained in Section 2 we dropped Seoul and Incheon from the baseline analysis because we were unable to obtain the exact composition of school districts within these cities.

appendix lists this initial set of randomized and non-randomized cities included in the city-level analysis.

We merge the CSAT data with city-level characteristics from the 1995 population census using the 1999 School Directory, which contains identifiers for both the high-school and the city.[19] Given potential concerns about the comparability of the randomized and non-randomized samples, we examine how these three sets of cities vary in terms of their 1995 characteristics. Figure 2 plots several city-level characteristics (fractions enrolled in school, under age 18, migrated, married, graduated from high school, and unemployed) against log population for the baseline randomized cities (red), the other randomized cities (black), and the non-randomized cities (grey). While most of the city-level characteristics appear similar across cities, there are striking differences in terms of log population: baseline randomized cities have almost uniformly higher population than non-randomized cities such that their distributions do not overlap at all. It is only cities among the additional randomized cities that have common support with the set of non-randomized cities.

Table A7 in the appendix reports summary statistics on city characteristics for the baseline set of randomized cities, the set of non-randomized cities, as well as a sample of randomized cities within the common support of the non-randomized sample in terms of population (i.e. we restrict the set of randomized cities to those with populations within the range of population of the non-randomized cities). Column (1), (2), and (3) show the mean city-level characteristics; columns (4) and (5) present the "raw" estimated differences between the randomized cities and the non-randomized cities; columns (6) and (7) present analogous results, but controlling for log city population. These results reveal that almost all of the differences between randomized and non-randomized cities are eliminated once we control for city population or restrict to the sample of common support.[20] Thus, in estimating school effects, we compare two sub-samples of randomized cities: the sample of cities within our baseline sample, and the sample of randomized cities which have common support with the non-randomized cities.

To provide a consistent analysis across all randomized and non-randomized cities, we do not include district fixed effects. Rather, we consider the average achievement of all schools within each city. As before, the estimates for randomized cities will represent school effects with student composition balanced across schools, albeit within cities rather than within districts; the

[19]We use CSAT data from 1997 to match city-level characteristics since it is the closest year in our sample to the 1999 school directory. Using other years yields similar results.

[20]The exception is the percentage of married people which remains higher for the non-randomized cities in almost all of the specifications.

estimates for non-randomized cities will reflect differences in student attributes and peer effects as well differences in school effectiveness. This approach makes it possible to provide estimates on the standard deviations of school effects for all cities in a consistent way, and allows us to assess whether the large differences in the standard deviations of school effects between randomized and non-randomized areas documented in Table 3 can be attributed to the system by which students are allocated to high schools.

Figure 3 depicts the standard deviation of school effects for our three sets of cities, in each sub-sample by year and gender. Two clear patterns stand out. First, the standard deviation of school effects varies relatively little across cities featuring random student assignment to high schools.[21] Second, for any given level of population, the standard deviation of school effects tends to be much larger for cities with non-random assignment. Unlike in Figure 2, which displayed how city attributes varied with population, there is a stark difference in the magnitude of the standard deviation of school effects across cities with and without random student assignment whether considering the baseline randomized cities or the other randomized cities.[22]

Table 7 confirms the stark differences in the standard deviation of school effects across cities with and without random assignment.[23] Whether we look at the main randomized sample (corresponding to the districts in the main analysis) or the randomized sample in the common support of log population as the non-randomized sample, the patterns are similar. The standard deviation of school effects range from 0.04 to 0.13 standard deviation units in cities with random assignment, while the estimates range from 0.51 to 0.70 in cities with non-random student assignment. These estimates are similar to those from the main analysis presented in Table 3 in which we included district dummies within cities. Taken together, these findings suggest that differences in the variation of schools effects across cities can be attributed to the system of allocation of students to high schools.

6 External validity

To gain insight into the external validity of our estimates, we compare our baseline results with those obtained from a broad sample of countries that participated in the 2000 round of the

[21] Among cities with random assignment, the standard deviation of schools effects tends to be larger in Seoul.

[22] Among the cities in the non-randomized sample, the city of Iksan has a relatively low standard deviation of school effects. This may be because Iksan was created in 1995 with the merger of Iri city and Iksan gun, and the former city featured random assignment between 1980 and 1990.

[23] These results are based 1997 CSAT data but the patterns are similar when using data from 1995 or 1996

PISA test. We restrict our attention to reading scores in the PISA 2000 data, because not all students took the math and science test.[24] For consistency with the analysis of our CSAT data, we normalize the test scores so that the mean is zero and the standard deviation is one in each country. Of course, in the absence of random assignment to schools, the estimated school effects from PISA reflect differences across schools that also reflect student sorting and peer effects due to differences in student composition.

Table 8 reports results based on student-level data from the PISA 2000 for each country. The standard deviation of school effects varies from 0.23 for girls in Finland and 0.28 for boys in Sweden up to 0.83 in Austria and 0.82 in Hungary for girls and boys respectively. Korea is characterized by low to intermediate levels in the variation of school effects (0.35), above Finland, Sweden, New Zealand, Iceland and Ireland, and just below Spain, Denmark and Australia. Note, however, that the variation in school effects for Korea combines schools in randomizing and non-randomizing areas. Overall, the variation of school effects observed in the cross-section of countries surveyed in PISA 2000 appears to be of a similar order of magnitude to those in non-randomizing districts based on our CSAT data.

We explore whether the variation of school effects in Korea can be representative of other countries by performing two additional analysis using data from PISA 2000. First, we examine how the degree of heterogeneity in observed school inputs in Korea compares with that of other countries. This can shed some light on the extent to which the variation in school effects documented in Korea may be generalized to other contexts. For example, if schools in Korea exhibit smaller heterogeneity in observed inputs than schools in other countries, we might expect greater variation in school effects in other countries. Second, we document the degree of school autonomy with regards to academic, personnel and budgetary decisions in Korea and in other countries. We would expect greater variation in school effects in educational systems for which schools have greater autonomy in key management decisions.

To examine the heterogeneity in observed school inputs, we consider the following indicators: annual hours in school, school size, student-teacher ratio, computer-student ratio, proportion of teachers certified and qualified, additional support to students, private versus public ownership, and single-sex versus coeducational schools.[25] Table 9 reports results from this analysis for each

[24] The math and science tests were taken by 50% of students, respectively. Only 25% of students took tests in all three subjects.

[25] We examine the degree of heterogeneity in each of these school inputs by computing the standard deviation of each indicator across schools for each country, ranking countries by the standard deviation of the indicators, and calculating the corresponding percentiles.

observed school input considered. Column (1) reports statistics for Korea, while the remaining columns present simple averages for countries in Europe, Asia and Oceania, and the Americas. The results reveal that Korea has a relatively high degree of heterogeneity in several school inputs, most notably teacher-student ratio, annual hours in school, and indicators for whether the school is privately-owned or single-sex. By contrast, Korea displays low heterogeneity with regard to additional services to students, as well as in the share of teachers certified or qualified. On average, Korea displays a degree of heterogeneity in observed school inputs that is in line with other countries in Asia and Oceania, higher than Europe and lower than the Americas.

To examine the degree of school autonomy across countries, we follow Hanushek et al. (2014) who compute measures of autonomy along different dimensions at the country-level. In particular, we use data on 6 questions (measuring school autonomy on defining courses, content, textbooks, hiring policies, salaries and budget) to quantify school autonomy along three dimensions: academic, personnel and budgetary. For each country, we compute averages across schools, thereby obtaining country-level measures of autonomy. Table A8 reports the results on these measures for each dimension of autonomy. The results reveal that the degree of school autonomy in Korea is higher than in Europe and the Americas, while below that observed in the rest of Asia and Oceania. While schools in Korea have relatively low autonomy with regard to personnel matters, they have relatively high autonomy over academic and budgetary matters.

Figure 4 summarizes these findings by depicting Korea's relative position in the cross-section of countries with regard to the heterogeneity in observed school inputs and school autonomy. It reveals that the values for Korea are close to average as compared to these 40 countries in terms of both heterogeneity in school inputs and school autonomy. Thus, the evidence presented in this section suggests that the key findings of this paper are likely to apply more broadly across the world.

7 Conclusion

Since the highly influential *Coleman Report* (Coleman et al., 1966), researchers and policymakers have devoted considerable effort to assessing the importance of student, school and peer effects in shaping educational outcomes. However, in the absence of universal random assignment of students to schools, it is difficult to cleanly unpack the effects of schools on academic achievement associated with inputs that can be directly manipulated by schools.

In this paper, we exploit universal random assignment of students to high schools within school districts in Korea to circumvent this difficulty. We present quasi-experimental evidence on whether and how much schools effects arising from differences in policies and practices matter for academic achievement. By estimating these school effects for scores in high stakes college entrance exams in areas characterized by universal random assignment of students to high schools, we confirm that schools vary considerably in their effectiveness. A 1 standard deviation increase in school quality within a district results in 0.05 to 0.08 standard deviation higher test scores in language, on average. The estimates are similar for boys and girls, and tend to be higher in English (up to 0.10 standard deviation) than in their native Korean (0.06 standard deviation).

We also show that the standard deviation in school effects within randomizing areas is only one tenth of that observed in comparable areas that allows for student selection. Moreover, the variation in school effects is not systematically larger in privately-founded schools, which have discretion to select principals and teachers. In the cross-section of countries, schools in Korea exhibit a relatively high degree of autonomy and heterogeneity in observed school inputs. Taken together, this evidence suggests that our main results are likely to apply more broadly.

It is important to note that, under universal random assignment, schools may not have the same incentives to perform as well as they would under a system of school choice. However, the educational landscape in Korea has strong reputational and status considerations that encourage schools to perform well. These may be even stronger under universal random assignment if stakeholders attribute high performance to the school rather than to sorting. The interaction between the nature of student assignments and school incentives is an important area for further research.

References

Abdulkadiroğlu, Atila, Joshua D. Angrist, Susan M. Dynarski, Thomas J. Kane, and Parag A. Pathak, "Accountability and Flexibility in Public Schools: Evidence from Boston Charters and Pilots," *Quarterly Journal of Economics*, 2011, *126* (2), 699–748.

_ , _ , Yusuke Narita, and Parag A. Pathak, "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation," *Econometrica*, 2017, *85* (5), 1373–1432.

Angrist, Joshua D and Victor Lavy, "Using Maimonides' rule to estimate the effect of class size on scholastic achievement," *The Quarterly journal of economics*, 1999, *114* (2), 533–575.

Angrist, Joshua D., Peter D. Hull, and Christopher R. Walters, "Methods for measuring school effectiveness*," *National Bureau of Economic Research*, 12 2022, (30803).

_ , _ , Parag A. Pathak, and Christopher R. Walters, "Leveraging Lotteries for School Value-Added: Testing and Estimation*," *The Quarterly Journal of Economics*, 02 2017, *132* (2), 871–919.

Araujo, Caridad M., Pedro Carneiro, Yyannu Cruz-Aguayo, and Norbert Schady, "Teacher Quality and Learning Outcomes in Kindergarten," *Quarterly Journal of Economics*, 2016, *131* (3), 1415–1453.

Bloom, Nicholas, Renata Lemos, Raffaella Sadun, and John Van Reenen, "Does Management Matter in schools?," *The Economic Journal*, 05 2015, *125* (584), 647–674.

Busso, Matías and Verónica Frisancho, "Ability Grouping and Student Performance: Experimental Evidence from Middle Schools in Mexico," January 2023. Inter-American Development Bank.

Carrell, Scott E. and Mark L. Hoekstra, "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids," *American Economic Journal: Applied Economics*, January 2010, *2* (1), 211–28.

Chetty, Raj, John Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Schanzenbach, and Danny Yagan, "How Does Your Kindergarden Classroom Affect Your Earnings? Evidence from Project STAR," *Quarterly Journal of Economics*, 2011, *126* (4), 1593–1660.

Cohodes, Sarah R., Elizabeth M. Setren, and Christopher R. Walters, "Can Successful Schools Replicate? Scaling Up Boston's Charter School Sector," *American Economic Journal: Economic Policy*, February 2021, *13* (1), 138–67.

Coleman, James S., Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York, *Equality of Educational Opportunity*, Washington, D.C., U.S. Government Printing Office, 1966.

Cortes, Kalena E. and Joshua S. Goodman, "Ability-Tracking, Instructional Time, and Better Pedagogy: The Effect of Double-Dose Algebra on Student Achievement," *American Economic Review*, May 2014, *104* (5), 400–405.

Cullen, Julie Berry, Brian A. Jacob, and Steven Levitt, "The Effect of School Choice on Participants: Evidence from Randomized Lotteries," *Econometrica*, 2006, *74* (5), 1191–1230.

Deming, David, "Using School Choice Lotteries to Test Measures of School Effectiveness," *American Economic Review: Papers & Proceedings*, 2014, *104* (5), 406–411.

Deming, David J., Justin S. Hastings, Thomas J. Kane, and Douglas O. Staiger, "School Choice, School Quality, and Postsecondary Attainment," *American Economic Review*, 2014, *104* (3), 991–1013.

Deutsch, Jonah, "Using School Lotteries to Evaluate the Value-Added Model.," *Society for Research on Educational Effectiveness*, 2013.

Dobbie, Will and Roland G Fryer Jr, "Getting beneath the veil of effective schools: Evidence from New York City," *American Economic Journal: Applied Economics*, 2013, *5* (4), 28–60.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, August 2011, *101* (5), 1739–74.

Hahn, Youjin, Liang Choon Wang, and Hee-Seung Yang, "Does greater school autonomy make a difference? Evidence from a randomized natural experiment in South Korea," *Journal of Public Economics*, 2018, *161*, 15–30.

Hanushek, Eric A., Susanne Link, and Ludger Woessmann, "Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA," *Journal of Development Economics*, 2014, *104*, 212–232.

Jackson, C Kirabo, Shanette C Porter, John Q Easton, Alyssa Blanchard, and Sebastián Kiguel, "School effects on socioemotional development, school-based arrests, and educational attainment," *American Economic Review: Insights*, 2020, *2* (4), 491–508.

Kane, Thomas J. and Douglas O. Staiger, "The Promise and Pitfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives*, December 2002, *16* (4), 91–114.

Kim, Gayoung and Beomsoo Kim, "The Impacts of Coeducation in High School on the College Scholastic Ability Test," *KDI Journal of Economic Policy*, 2015, *37*, 65–85.

Korea Education Development Institute, *Study on the History of Modern Korean Education*, KEDI Research Report, RR 98-8, 1998.

Krueger, Alan B, "Experimental estimates of education production functions," *The quarterly journal of economics*, 1999, *114* (2), 497–532.

Lavy, Victor, M. Daniele Paserman, and Analia Schlosser, "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom*," *The Economic Journal*, 2012, *122* (559), 208–237.

Lee, Young S., "Educational Tracking, Residential Sorting and Intergenerational Mobility," 2012. Williams College.

Loeb, Susanna, Michael S Christian, Heather J Hough, Robert H Meyer, Andrew B Rice, and Martin R West, "School Effects on Social-Emotional Learning: Findings from the First Large-Scale Panel Survey of Students. Working Paper.," *Policy Analysis for California Education, PACE*, 2018.

McCaffrey, Daniel F., Tim R. Sass, J. R. Lockwood, and Kata Mihaly, "The Intertemporal Variability of Teacher Effect Estimates," *Education Finance and Policy*, 10 2009, *4* (4), 572–606.

Park, Hyunjoon, *Re-evaluating Education in Japan and Korea: De-mystifying Stereotypes*, Routledge, 2013.

_ , Jere R. Behrman, and Jaesung Choi, "Causal Effects of Single-sex Schools on College Entrance Exams and College Attendance: Random Assignment in Seoul High Schools," *Demography*, 2013, *50* (2), 447–469.

_ , _ , and _ , "Do single-sex schools enhance students' STEM (science, technology, engineering, and mathematics) outcomes?," *Economics of Education Review*, 2018, *62*, 35–47.

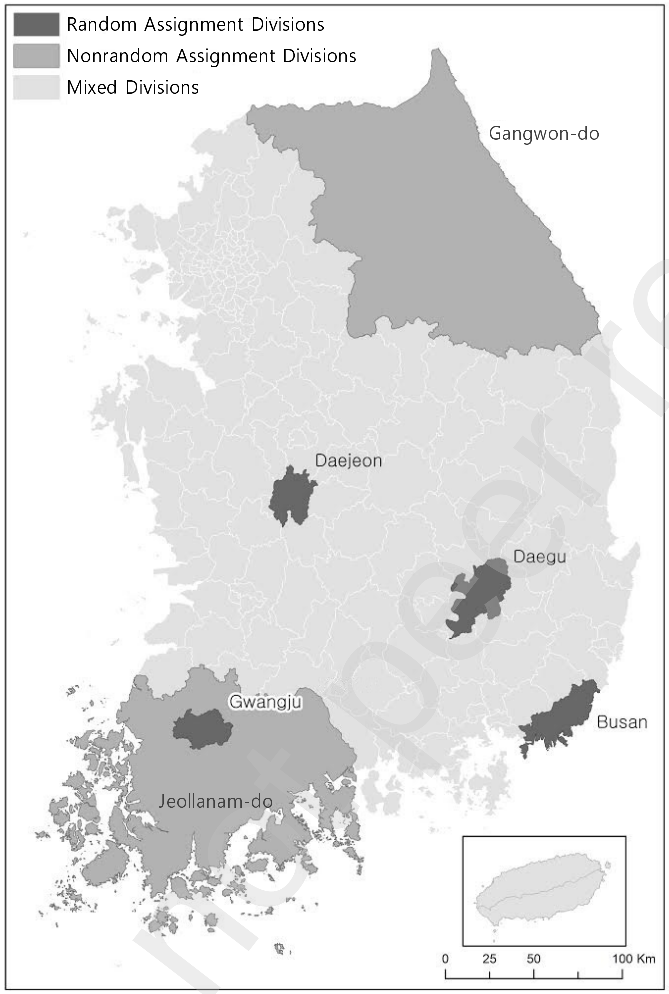
Raudenbush, Stephen W and JDouglas Willms, "The estimation of school effects," *Journal of educational and behavioral statistics*, 1995, *20* (4), 307–335.

Rivkin, Steven, Eric A. Hanushek, and John F. Kain, "Going to a Better School: Effects and Behavioral Responses," *Econometrica*, 2005, *73* (2), 417–458.

Rockoff, Jonah E, "The impact of individual teachers on student achievement: Evidence from panel data," *American economic review*, 2004, *94* (2), 247–252.

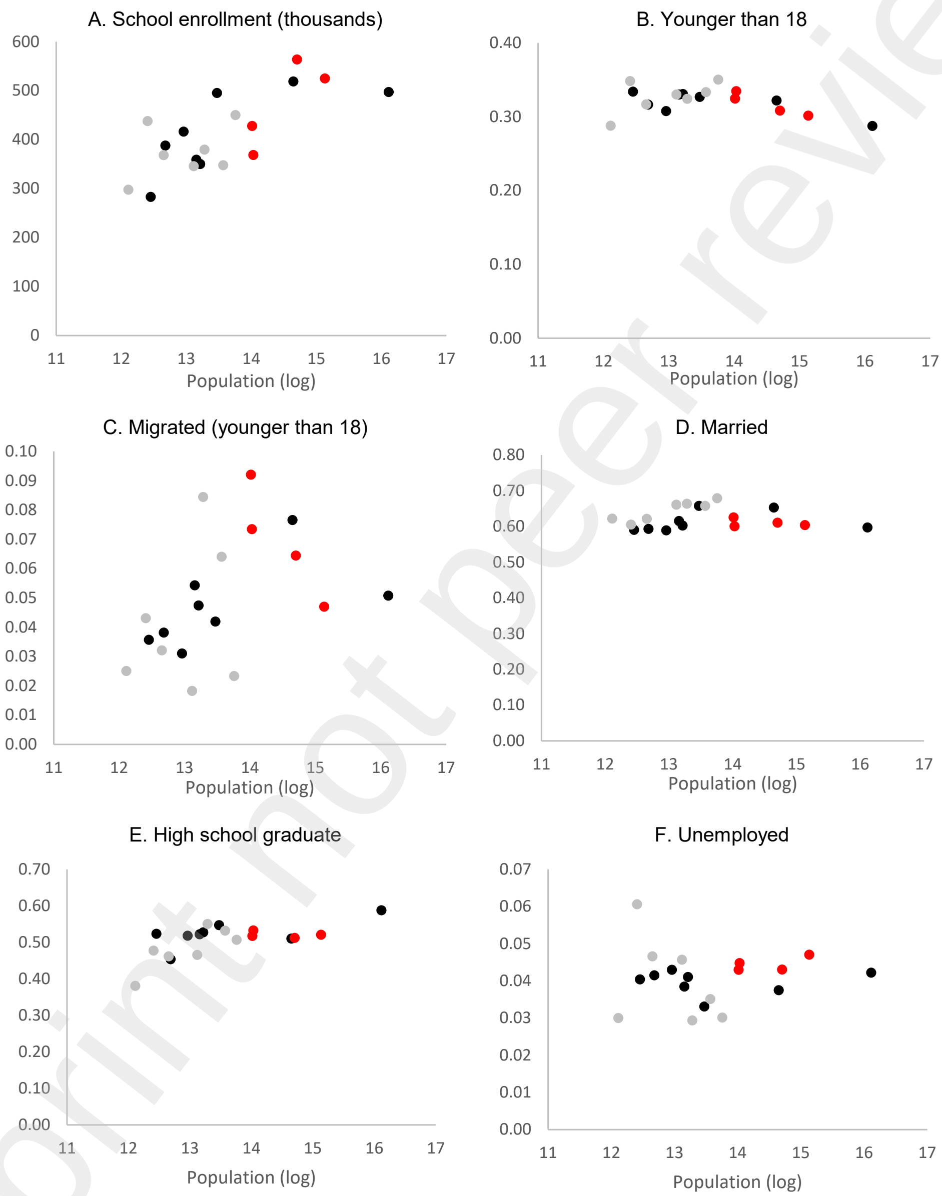
Sacerdote, Bruce, "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds.*, Vol. 3 of *Handbook of the Economics of Education*, Elsevier, 2011, pp. 249–277.

Figure 1: Randomizing and non-randomizing administrative divisions



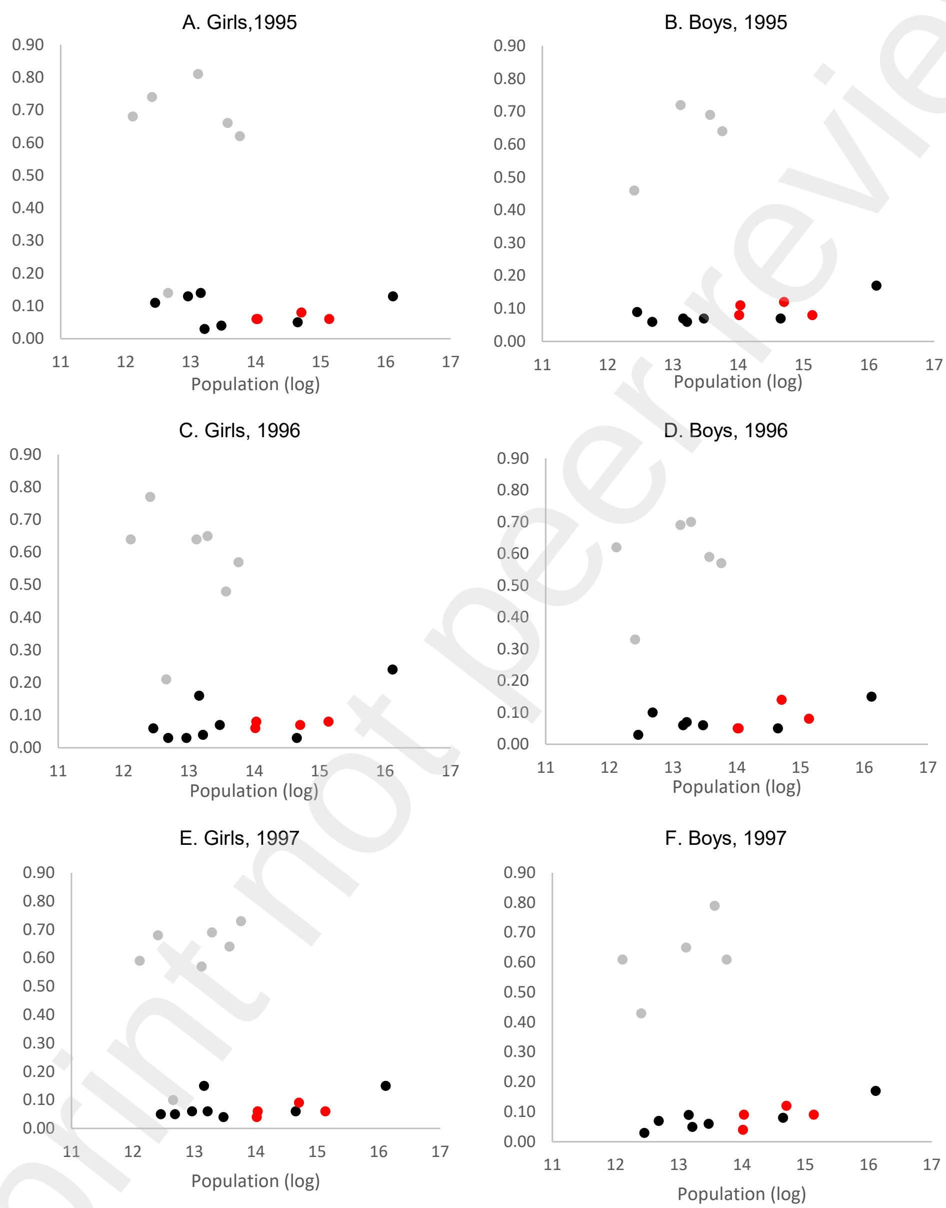
Notes: The main random assignment sample is composed of the metropolitan cities of Daejeon, Daegu, Busan and Gwangju. The main nonrandom assignment sample is composed of the administrative divisions of Gangwon-do and Jeollanam-do.

Figure 2: Average characteristics of cities



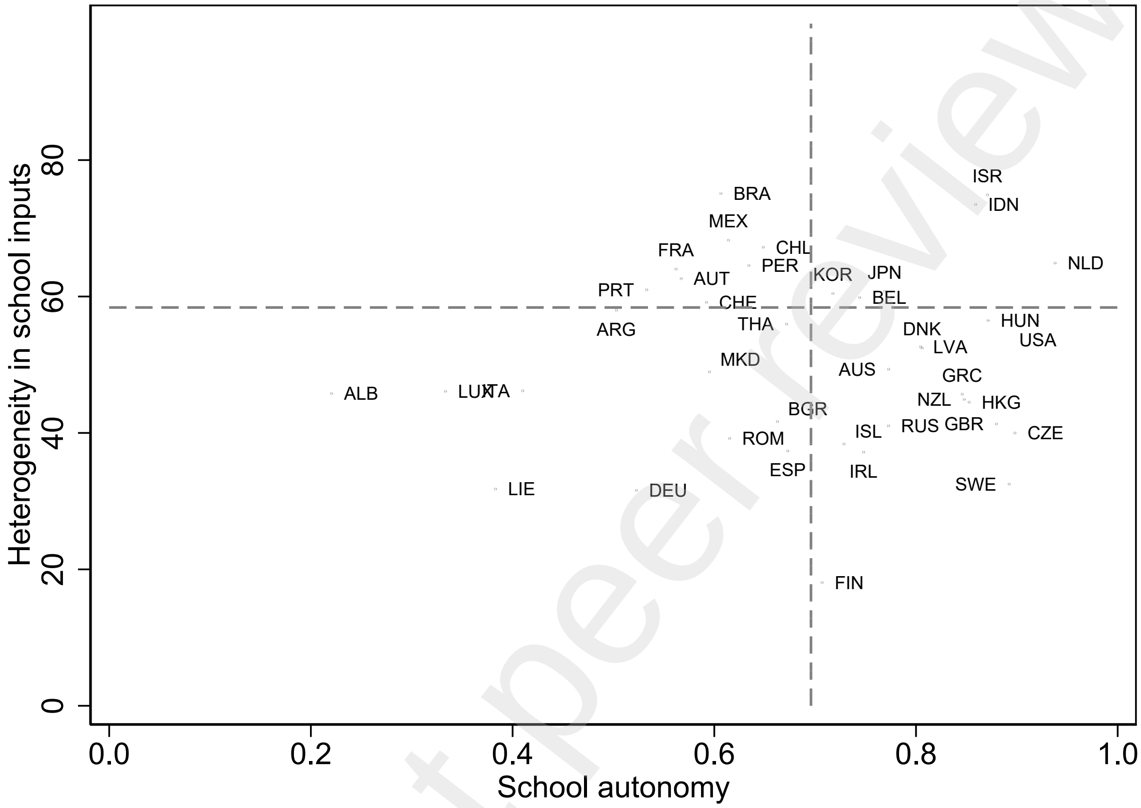
Notes: City-level population and other characteristics were obtained from the 1995 population census. Each point represents one city. Only cities with at least 4 high schools attended by girls in 1997 are included in the analysis. Red dots correspond to cities in the random assignment administrative divisions (those used in the main analysis in Tables 3, 5 and 7). Black dots correspond to other randomized cities. Grey dots correspond to non-randomized cities.

Figure 3: Standard deviations of school effects at the city-level



Notes: This figure presents standard deviation of school effects at the city level based on the 1995-1997 CSAT. City-level population was obtained merging the CSAT data with the 1995 population census using the 1999 School Directory (that contains high school identifier and city identifier). Each point represents one city. Only cities with at least 4 high schools attended by girls in 1997 are included in the analysis. Red dots correspond to cities in random assignment administrative states (these cities were used in the main analysis presented in Tables 3, 5 and 7). Black dots correspond to other randomized cities. Grey dots correspond to non-randomized cities. Standard deviations of school effects are computed following the methodology described in Section 3.

Figure 4: Heterogeneity of school inputs and school autonomy across countries



Notes: This figure reports measures of (i) the degree of heterogeneity in observed school inputs across schools and (ii) measures of school autonomy for 40 countries based on data from PISA 2000. The dashed lines identify the average for each dimension. For consistency with the analysis using CSAT and KELS, the estimation sample from PISA does not include observations for vocational schools.

Table 1: Summary statistics (CSAT data)

	All	Randomized sample	Non-randomized sample
	(1)	(2)	(3)
A. Girls			
School characteristics			
Privately-founded	0.62	0.71	0.42
Single sex	0.80	0.96	0.69
Enrollment	505.41	540.40	374.40
Mean test scores			
Average	0.08	0.41	-0.14
Korean	0.09	0.37	-0.14
English	0.06	0.39	-0.12
N (students)	482,281	100,957	33,860
B. Boys			
School characteristics			
Privately-founded	0.65	0.70	0.58
Single sex	0.79	0.97	0.74
Enrollment	511.89	527.97	377.04
Mean test scores			
Average	-0.07	0.17	-0.26
Korean	-0.08	0.14	-0.27
English	-0.05	0.17	-0.21
N students	600,956	132,539	38,513

Notes: The sample includes students that graduated from high school in 1996 and took the CSAT in that year. Students attending selective and vocational schools are excluded. Schools with less than 100 students are dropped. Randomized sample includes students in the administrative divisions that satisfy these conditions: (i) students are randomized to schools in all school districts; (ii) there are no distance-based rules that introduce exemptions to the randomization (i.e. Seoul is dropped); (iii) we were able to identify the sample of schools included in each school district (i.e. Incheon is dropped). The randomized sample includes these administrative divisions: Busan, Daegu, Gwangju, and Daejeon. The non-randomized sample includes the administrative divisions in which there is no school district in which students are randomized to schools. This sample includes these administrative divisions: Gangwon-do and Jeollanam-do. Average test scores are computed using scores in English and Korean.

Table 2: Randomization tests (KELS data)

	Randomized sample				Non-randomized sample			
	Private	Single-sex	Class size	Enrollment	Private	Single-sex	Class size	Enrollment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Girls								
Scores in middle school	-0.03	-0.01	0.07	0.00	-0.00	0.05	0.47*	0.06**
	(0.03)	(0.02)	(0.15)	(0.01)	(0.04)	(0.04)	(0.23)	(0.02)
Parents' education	-0.00	0.01	0.10	0.01*	0.01	-0.01	-0.01	-0.00
	(0.01)	(0.01)	(0.06)	(0.01)	(0.01)	(0.01)	(0.06)	(0.01)
Log household income	0.00	0.02	-0.00	0.00	-0.03*	-0.01	0.00	0.00
	(0.01)	(0.02)	(0.08)	(0.01)	(0.01)	(0.02)	(0.05)	(0.01)
Family size	0.03	0.04	0.15	0.01	0.01	-0.00	0.28	0.01
	(0.04)	(0.04)	(0.21)	(0.02)	(0.02)	(0.02)	(0.26)	(0.02)
R-square	0.47	0.45	0.45	0.51	0.50	0.65	0.89	0.84
N students	499	499	499	499	498	498	498	498
B. Boys								
Scores in middle school	0.02	0.04	0.15	0.01	0.04	0.07**	0.66**	0.06**
	(0.03)	(0.03)	(0.16)	(0.01)	(0.03)	(0.03)	(0.17)	(0.01)
Parents' education	-0.01	0.00	0.07	0.00	0.02	-0.01	0.04	0.01**
	(0.01)	(0.01)	(0.07)	(0.00)	(0.01)	(0.01)	(0.03)	(0.00)
Log household income	0.04	0.00	-0.15	-0.01	-0.03*	0.02	0.17	0.00
	(0.04)	(0.03)	(0.24)	(0.02)	(0.01)	(0.01)	(0.14)	(0.01)
Family size	0.04	0.02	0.14	0.01	-0.01	0.02	0.20	0.03*
	(0.05)	(0.02)	(0.16)	(0.01)	(0.03)	(0.02)	(0.14)	(0.01)
R-square	0.22	0.50	0.54	0.57	0.49	0.59	0.87	0.87
N students	481	481	481	481	596	596	596	596

Notes: The sample consists of individual-level data on test scores from KELS, which collected information from 7th graders from 150 middle schools in 2005 and followed them over time. High school characteristics (privately-founded, single-sex, class size and enrollment) were extracted from wave 4 collected in 2008. Class size is calculated for the whole school. Enrollment is also for whole school and is divided by 1000. Data on test scores in middle-school, log household income and family size come from wave 3 (collected in 2007). Data on parental education is only available in wave 1 (collected in 2005). KELS do not provide information on school district. Middle-school is used as a proxy for the school district. Standard errors are clustered by middle-school. ** significant at 1% level; * significant at 5% level.

Table 3: Standard deviation of school effects, Randomized sample

	FE-Empirical Bayes			Random Effects		
	1995	1996	1997	1995	1996	1997
	(1)	(2)	(3)	(4)	(5)	(6)
A. Girls						
Average	0.06	0.07	0.06	0.06	0.07	0.05
English	0.10	0.10	0.09	0.09	0.10	0.09
Korean	0.04	0.06	0.03	0.03	0.06	0.03
N students	33,711	33,175	34,071	33,711	33,175	34,071
N schools	67	68	69	67	68	69
N districts	9	9	9	9	9	9
B. Boys						
Average	0.08	0.07	0.07	0.08	0.07	0.07
English	0.09	0.08	0.10	0.09	0.08	0.10
Korean	0.06	0.06	0.05	0.06	0.05	0.05
N students	44,999	43,814	43,726	44,999	43,814	43,726
N schools	90	90	91	90	90	91
N districts	9	9	9	9	9	9

Notes: Estimates are based on individual-level data on test scores in CSAT exams for high-school graduates in the randomized sample described in Table 1. Column titles indicate the year used for the analysis. Average corresponds to the mean of English and Korean test scores. Standard deviations of school effects are computed following the methodology described in section 3.

Table 4: Standard deviation of school effects, Non-randomized sample

	FE-Empirical Bayes sample			Random Effects sample		
	1995	1996	1997	1995	1996	1997
	(1)	(2)	(3)	(4)	(5)	(6)
A. Girls						
Average	0.66	0.63	0.62	0.64	0.58	0.60
English	0.64	0.61	0.60	0.61	0.57	0.58
Korean	0.61	0.55	0.56	0.59	0.51	0.53
N students	10,774	11,454	11,632	10,774	11,454	11,632
N schools	45	47	48	45	47	48
N districts	9	10	10	9	10	10
B. Boys						
Average	0.53	0.52	0.55	0.51	0.51	0.52
English	0.49	0.52	0.52	0.48	0.50	0.50
Korean	0.50	0.46	0.50	0.48	0.45	0.48
N students	11,800	13,387	13,326	11,800	13,387	13,326
N schools	54	56	56	54	56	56
N districts	9	10	10	9	10	10

Notes: Estimates are based on individual-level data on test scores in CSAT exams for high-school graduates in the non-randomized sample described in Table 1. Column titles indicate the year used for the analysis. Average corresponds to the mean of English and Korean test scores. Standard deviations of school effects are computed following the methodology described in section 3.

Table 5: Standard deviation of school effects by school type, Randomized sample

	Privately-founded			Public		
	1995	1996	1997	1995	1996	1997
	(1)	(2)	(3)	(4)	(5)	(6)
A. Girls						
Average	0.06	0.07	0.06	0.02	0.05	0.02
English	0.09	0.11	0.10	0.04	0.07	0.03
Korean	0.04	0.04	0.06	0.01	0.04	0.01
N students	23,821	23,636	24,232	9,890	9,539	9,839
N schools	46	47	48	21	21	21
N districts	9	9	9	9	9	9
B. Boys						
Average	0.08	0.06	0.08	0.06	0.08	0.06
English	0.10	0.08	0.11	0.07	0.09	0.09
Korean	0.03	0.07	0.05	0.05	0.07	0.05
N students	31,709	30,679	30,679	13,290	13,135	13,185
N schools	61	61	61	29	29	29
N districts	9	9	9	9	9	9

Notes: Estimates are based on individual-level data on test scores in CSAT exams for high-school graduates in the randomized sample described in Table 1. Column titles indicate the year used for the analysis. Average corresponds to the mean of English and Korean test scores. Standard deviations of school effects are computed using Fixed-Effects models with Empirical Bayes adjustment following the methodology described in section 3.

Table 6: Associations of school characteristics and academic achievement, 1996

	Randomized sample (1)	Non-randomized sample (2)
A. Girls		
Privately-founded	-0.01	-0.62***
	(0.02)	(0.18)
Single-sex	-0.06*	0.36
	(0.03)	(0.22)
Enrollment	0.32**	2.20***
	(0.13)	(0.71)
N students	33,175	11,454
R-square	0.007	0.259
B. Boys		
Privately-founded	-0.01	-0.63***
	(0.02)	(0.14)
Single-sex	0.02	0.04
	(0.04)	(0.13)
Enrollment	-0.06	2.97***
	(0.11)	(0.49)
N students	43,814	13,387
R-square	0.019	0.216

Notes: Estimates are obtained from regressions of individual-level average test scores in CSAT exams for high-school graduates in 1996 on school characteristics. The randomized and non-randomized samples are described in Table 1. Test scores are averages of English and Korean. Regressions control for school district fixed-effects. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table 7: Standard deviation of school effects by subject (city level analysis, main analysis sample, 1997)

	Randomized sample (Main)	Randomized (Common Support)	Non-randomized sample
	(1)	(2)	(3)
A. Girls			
Average	0.06	0.07	0.57
English	0.04	0.06	0.51
Korean	0.11	0.10	0.64
N students	33,344	12,799	14,504
N schools	66	32	40
B. Boys			
Average	0.06	0.07	0.57
English	0.06	0.04	0.67
Korean	0.13	0.09	0.70
N students	43,160	12,377	17,683
N schools	89	33	48

Notes: This table presents standard deviation of school effects at the city level for students participating in the 1997 CSAT. Only cities with at least four high schools are included in the analysis. Standard deviations of school effects are computed following the methodology described in section 3. Average corresponds to the mean of English and Korean test scores.

Table 8: **Standard deviation of school effects, PISA 2000**

	Girls	Boys		Girls	Boys
Finland	0.23	0.33	Israel	0.56	0.65
Sweden	0.27	0.28	Portugal	0.56	0.59
New Zealand	0.31	0.40	Hong Kong	0.57	0.67
Iceland	0.32	0.29	Luxembourg	0.58	0.60
Ireland	0.34	0.42	Brazil	0.60	0.67
Korea	0.35	0.38	Argentina	0.60	0.60
Spain	0.37	0.47	Romania	0.61	0.60
Denmark	0.37	0.41	United Kingdom	0.61	0.72
Australia	0.40	0.43	France	0.63	0.71
Latvia	0.46	0.52	Netherlands	0.63	0.65
Greece	0.46	0.66	Czech Republic	0.64	0.67
United States	0.47	0.56	Belgium	0.64	0.72
Thailand	0.49	0.55	Chile	0.65	0.68
Albania	0.50	0.60	Italy	0.65	0.72
Russia	0.50	0.60	Mexico	0.67	0.68
Liechtenstein	0.53	0.68	Peru	0.69	0.69
Japan	0.53	0.66	Bulgaria	0.69	0.79
Macedonia	0.54	0.61	Hungary	0.71	0.82
Indonesia	0.55	0.52	Germany	0.71	0.68
Switzerland	0.55	0.58	Austria	0.83	0.78

Notes: This table reports the standard deviation of school effects for 40 countries. School effects are estimated using individual-level data on test scores in reading from PISA in year 2000. For consistency with the analysis using CSAT and KELS, the estimation sample from PISA does not include observations for vocational schools. Standard deviations of school effects are computed following the methodology described in section 3.

Table 9: Heterogeneity in observed inputs across schools in PISA 2000

	Korea	Europe	Asia-Oceania	Americas
	(1)	(2)	(3)	(4)
Annual hours in school	80.0	39.1	62.8	88.3
School size	62.5	42.8	52.8	85.8
Teacher-student ratio	77.5	54.6	43.4	47.1
Computer-student ratio	72.5	51.8	55.0	43.8
Private school	100.0	41.9	70.3	69.7
Coed school	95.0	40.2	83.1	56.7
Additional services to students	51.3	41.0	59.9	82.5
Teachers certified (%)	2.6	54.1	39.5	56.1
Teachers with qualifications (%)	2.6	48.4	58.0	54.3
Average	60.4	46.0	58.3	64.9

Notes: This table reports measures of heterogeneity in observed school inputs in Korea and in different regions of the world based on school-level data for 40 countries from PISA in the year 2000. Heterogeneity measures in columns (2)-(4) correspond to the simple average of the heterogeneity measures for the corresponding countries.

Table A1: Summary statistics, KELS data set

	Randomized sample (1)	Non-randomized sample (2)
A. Girls		
Parents' education	13.22	12.76
Log household income	12.71	12.71
Number of siblings	2.21	2.34
Test scores in middle-school	0.42	0.22
Single-sex	0.71	0.36
Privately-founded	0.49	0.38
Enrollment	1.35	1.00
Class size	37.09	34.46
N (obs.)	499	498
B. Boys		
Parents' education	13.42	12.76
Log household income	12.83	12.73
Number of siblings	2.11	2.21
Test scores in middle-school	0.07	-0.02
Single-sex	0.72	0.32
Privately-founded	0.54	0.38
Enrollment	1.22	0.95
Class size	35.68	33.68
N students	481	596

Notes: Sample consists of individual-level data from KELS, which collected information from 7th graders from 150 middle schools in 2005 and followed them over time. High school characteristics like privately-founded, single-sex, class size and enrollment were extracted from wave 4 collected in 2008. Class size is calculated for the whole school. Enrollment is also for whole school and divided by 1000. Data on test scores in middle school, log household income and family size come from wave 3 (collected in 2007). Data on parental education is only available in wave 1 (collected in 2005).

Table A2: Population of major administrative divisions, 1997

Name	Administrative division status	Population
Seoul	special city	10,389,057
Busan	metropolitan city	3,865,114
Daegu	metropolitan city	2,501,928
Incheon	metropolitan city	2,460,906
Gwangju	metropolitan city	1,326,478
Daejeon	metropolitan city	1,323, 009
Ulsan	metropolitan city	1,013,070
Gyeonggi-do	province	8,514,716
Gyeongsangnam-do	province	3,058,479
Gyeongsangbuk-do	province	2,811,586
Jeollanam-do	province	2,166,247
Jeollabuk-do	province	2,007,379
Chungcheongnam-do	province	1,822,543
Gangwon-do	province	1,540,307
Chungcheongbuk-do	province	1,475,448
Jeju	special self-governing province	528,360

Notes: This table lists the administrative divisions of Korea and their population estimates in 1997. Population estimates are from the National Statistical Office of the Republic of Korea.

Table A3: Standard deviation of school effects, FE-unshrunken

	Randomized			Non-randomized		
	1995	1996	1997	1995	1996	1997
	(1)	(2)	(3)	(4)	(5)	(6)
A. Girls						
Average	0.06	0.08	0.06	0.66	0.65	0.63
English	0.10	0.11	0.10	0.64	0.63	0.61
Korean	0.05	0.07	0.05	0.62	0.58	0.57
N students	33,711	33,175	34,071	10,774	11,454	11,632
N schools	67	68	69	45	47	48
N districts	9	9	9	9	10	10
B. Boys						
Average	0.09	0.08	0.09	0.54	0.53	0.55
English	0.10	0.09	0.11	0.50	0.52	0.53
Korean	0.08	0.07	0.06	0.52	0.47	0.51
N students	44,999	43,814	43,726	11,800	13,387	13,326
N schools	90	90	91	54	56	56
N districts	9	9	9	9	10	10

Notes: Estimates are based on individual-level data on test scores in CSAT exams for high-school graduates in randomized and non-randomized samples described in Table 1. Column titles indicate the year used for the analysis. Average corresponds to the mean of English and Korean test scores. Standard deviations of school effects are computed following the fixed-effects (FE) methodology described in section 3, without using Empirical Bayes shrinkage adjustment.

Table A4: Persistence of school effects over time

	Randomized sample (1)	Non-randomized sample (2)
A. Girls		
Corr(1995,1996)	0.53	0.97
Corr(1996,1997)	0.50	0.94
Corr(1995,1997)	0.49	0.92
N students	100,957	33,860
N schools	67	45
N districts	9	9
B. Boys		
Corr(1995,1996)	0.46	0.90
Corr(1996,1997)	0.55	0.92
Corr(1995,1997)	0.56	0.85
N students	132,539	38,513
N schools	90	54
N districts	9	9

Notes: Persistence is defined as the proportion of variability due to persistent effects, corresponding to the cross-year correlation for estimated effects from adjacent years.

Table A5: Standard deviation of school effects by school type, Non-randomized sample

	Privately-founded			Public		
	1995	1996	1997	1995	1996	1997
	(1)	(2)	(3)	(4)	(5)	(6)
A. Girls						
Average	0.50	0.43	0.36	0.71	0.66	0.68
English	0.47	0.42	0.34	0.68	0.63	0.65
Korean	0.48	0.39	0.34	0.48	0.39	0.34
N students	4,515	4,711	4,880	6,259	6,743	6,752
N schools	18	18	18	27	29	30
N districts	7	7	7	9	9	10
B. Boys						
Average	0.50	0.41	0.41	0.47	0.55	0.56
English	0.45	0.41	0.38	0.47	0.55	0.56
Korean	0.50	0.37	0.40	0.50	0.37	0.40
N students	6,930	7,799	7,826	4,870	5,588	5,500
N schools	25	26	26	29	30	30
N districts	7	8	8	9	9	9

Notes: Estimates are based on individual-level data on test scores in CSAT exams for high-school graduates in the randomized sample described in Table 1. Column titles indicate the year used for the analysis. Average corresponds to the mean of English and Korean test scores. Standard deviations of school effects are computed using Fixed-Effects models with Empirical Bayes adjustments following the methodology described in section 3.

Table A6: Cities included in the city-level analysis

City	Major administrative division
A. Randomized	
Seoul	Seoul
Incheon	Incheon
Busan^{B}	Busan
Daegu^{B}	Daegu
Gwangju^{B}	Gwangju
Daejeon^{B}	Daejeon
Cheongju^{cs}	Daejeon
Suwon^{cs}	Gyeonggi-do
Jeonju^{cs}	Jeollabuk-do
Masan^{cs}	Gyeongsangnam-do
Jinju^{cs}	Gyeongsangnam-do
B. Non-randomized	
Ulsan	Gyeongsangbuk-do
Bucheon	Gyeonggi-do
Anyang	Gyeonggi-do
Pohang	Gyeongsangbuk-do
Iksan	Jeollanam-do
Gyeongju	Gyeongsangbuk-do
Yeongju	Gyeongsangbuk-do
Mokpo	Jeollanam-do
Gunsan	Jeollabuk-do
Andong	Gyeongsangbuk-do

Notes: This table lists the set of randomized and non-randomized cities included in the city-level analysis.

^{B}: Cities in baseline analysis.

^{cs}: Cities in common support analysis.

Table A7: City characteristics at the city level

	Randomized baseline (1)	Randomized common support (2)	Non-randomized (3)	Raw diff. baseline (4)	Raw diff. CS (5)	Adj. diff. baseline (6)	Adj. diff. CS (7)
Population (millions)	2.16	0.50	0.42	1.74**	0.09	0.46	-0.06
				(0.38)	(0.13)	(0.46)	(0.04)
School enrollment (thousands)	0.47	0.40	0.36	0.11*	0.04	0.00	0.02
				(0.04)	(0.03)	(0.05)	(0.03)
% Population younger 18	0.32	0.32	0.32	0.00	0.00	-0.03	0.00
				(0.01)	(0.01)	(0.02)	(0.01)
% Migrated (younger 18)	0.07	0.04	0.04	0.03**	0.01	0.02	0.00
				(0.01)	(0.01)	(0.02)	(0.01)
% High school graduates	0.52	0.51	0.46	0.06*	0.05	-0.05	0.02
				(0.03)	(0.03)	(0.03)	(0.02)
% Married	0.61	0.61	0.64	-0.03*	-0.03	-0.07***	-0.04***
				(0.01)	(0.01)	(0.02)	(0.01)
% Unemployed	0.04	0.04	0.04	0.00	0.00	0.01	0.00
				(0.01)	(0.01)	(0.01)	(0.01)
N cities	4	5	10	14	15	14	15

Notes: This table presents statistics on city characteristics and standard deviation of school effects at the city level. City-level characteristics were obtained merging the CSAT data with the 1995 population census using the 1999 School Directory (that contains high school identifier and city identifier). Only cities with at least four high schools attended by boys are included in the analysis. Standard deviations of school effects are computed following the methodology described in Section 3. Test scores are averages of English and Korean. Column (1) presents means for cities in random assignment administrative divisions (the cities used in the main analysis presented in Tables 3, 5 and 7). Column (2) presents means for all randomized cities. Column (3) presents means for non-randomized cities. Columns (4) to (7) present coefficients and standard errors from OLS regressions. Columns (4) presents estimated differences between the randomized cities included in the main analysis (column (1)) and the non-randomized cities (column (3)). Columns (5) presents estimated differences between all randomized cities (column (2)) and the non-randomized cities (column (3)). Results in column (6) and (7) present results analogous to columns (4) and (5) but controlling for log city population. Standard errors are reported in parentheses. *** significant at 1% level; ** significant at 5% level; * significant at 10% level.

Table A8: Measures of school autonomy

	Korea	Europe	Asia-Oceania	Americas
	(1)	(2)	(3)	(4)
Academic	0.97	0.69	0.95	0.82
Personnel	0.25	0.41	0.48	0.43
Budget	0.93	0.90	0.95	0.72
Average	0.72	0.67	0.80	0.65

Notes: This table reports measures of school autonomy in Korea and across different regions of the world based on school-level data for 40 countries from PISA in the year 2000. Autonomy measures in columns (2)-(4) are simple averages of autonomy measures in the corresponding countries.