Discussion Paper Series
No. 1607
October 2016

Good schools or good students?
Evidence on school effects from universal random assignment of students to high schools

Paulo Bastos, Julian Cristia, Beomsoo Kim

The Institute of Economic Research - Korea University
Anam-dong, Sungbuk-ku, Seoul, 136-701, South Korea, Tel: (82-2) 3290-1632, Fax: (82-2) 928-4948

Good schools or good students? Evidence on school effects from universal random assignment of students to high schools*

Paulo Bastos[†] Julian Cristia[‡] Beomsoo Kim[§]

October 2016

Abstract

How much do schools differ in their effectiveness? Answering this question has been complicated by the selection of heterogeneous students into schools, which has made it difficult to distinguish between the influence of school inputs, student selection and peer effects. We exploit universal random assignment of students to high schools in certain areas of South Korea to provide clean estimates of the influence of school inputs. We find statistically significant differences across schools in the effects they have on scores in college entrance exams. However, school effects explain only 0.5% of the variation in learning outcomes in areas where students are randomized to schools. These results suggest that school inputs play a limited role in explaining variation in learning outcomes.

Keywords: Pure school effects; sorting; peer effects; school inputs.

JEL Classification: D44, H75, I21, I23, J16.

*We are grateful to seminar participants at the University of Maryland, University of California-Merced, OECD, and Inter-American Development Bank for very helpful comments and suggestions. Anastasiya Yarygina provided excellent research assistance. Beomsoo Kim gratefully acknowledges financial support from the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014S1A3A2044238). The views expressed in this paper are those of the authors only and should not be attributed to the institutions they are affiliated with. We remain responsible for any errors.

[†]Development Research Group, The World Bank, United States. E-mail: pbastos@worldbank.org

[‡]Research Department, Inter-American Development Bank, United States. E-mail: jcristia@iadb.org

[§]Department of Economics, Korea University, South Korea. E-mail: kimecon@korea.ac.kr

1 Introduction

Do schools differ in the impact they have on academic achievement, and if so how much? Answering these questions is key to inform central elements of education policy. For example, if schools differ markedly in their effectiveness, closing down low-performing schools or disseminating good management practices observed in some schools across the education system could lead to significant improvements in learning.[1] Since the highly influential *Coleman Report* (Coleman et al., 1966), researchers and policymakers have devoted considerable effort to assessing the importance of school inputs, student selection and peer effects in shaping educational outcomes. However, identifying the influence of school inputs has been complicated by the sorting of heterogeneous students into schools (Deming et al., 2014; Ellison and Swanson, 2016).[2]

This paper provides, to our knowledge, the first experimental evidence on how much schools differ in the impact they have on academic achievement. In particular, we assess the role of schools in explaining the variation in learning outcomes once the potential impacts of student heterogeneity and peer effects have been removed. We estimate these pure school effects by exploiting universal random assignment of students to general high schools within school districts in South Korea (henceforth Korea). Random assignment effectively balances student composition across schools within districts, thereby shutting down the potential impacts of student selection and peer effects.

We estimate the variation in pure school effects on student scores in national college entrance exams. We use individual-level administrative data for the period 1995-1997 and focus on a subset of administrative divisions where all students attending general high schools were randomly assigned to schools within districts.[3] We first provide evidence that there are statistically significant differences across schools in the effects they have on learning outcomes. To document the extent of variation in school effectiveness, we estimate

[1]The former is the central hypothesis of the high profile *No Child Left Behind* program. See US Department of Education (2006) for a policy oriented discussion on the importance of improving the quality of high school education.

[2]The set of school inputs leading to potential effects on academic achievement includes learning time, school resources and curriculum, attributes of principals and teachers, organizational practices, school culture, etc.

[3]Korea has sixteen major administrative divisions: eight *Do* [Provinces], six *Gwangyeoksi* [Metropolitan cities], one *Teukbyeolsi* [Special city], and one *Teukbyeoljachi-do* [Special self-governing province]. Metropolitan cities are administrative divisions with population over one million and are not included in provinces. Table A1 in the Appendix lists these administrative divisions and reports the corresponding population estimates in 1997.

the standard deviation of pure school effects. These estimates can be interpreted as the expected increase in learning if students were moved to schools that are one standard deviation above in the school effectiveness distribution. They are analogous to those provided by recent studies documenting the variation in teacher effectiveness (Chetty et al., 2014a,b; Jackson, 2014; Araujo et al., 2016). To facilitate the interpretation of these results, we also estimate the share of total variance in student learning that is explained by schools. To benchmark our estimates, we perform two additional analyses. First, we estimate the standard deviation of school effects in Korean divisions where students were sorted into schools based on an application process that involves student preferences, performance in exams and middle school GPA. Second, we perform a similar analysis for other countries using data from PISA in the year 2000. In these two cases, the variation in school effects can be explained by pure school effects, student sorting and peer effects.

We find that the estimated standard deviation of school effects in randomizing areas in Korea is 0.057. Although this estimate is statistically significantly different from zero, it is about one tenth of that obtained for non-randomizing areas. It is also substantially smaller than those obtained using country data from PISA—which range from 0.251 to 0.835. In randomizing areas, school effects explain only 0.5% of the observed variation in academic achievement. By contrast, school effects account for more than 30% of the variation in test scores in non-randomizing areas. Similarly, in the country-level data from PISA, school effects account for a large share of the variation in student performance.

In addition to the randomness of student assignment within selected areas, Korea's high school system is especially 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 alleviate substantially the concern that the estimated pure school effects might be contaminated by heterogeneity in student composition. Second, the high stakes involved in college entrance exams, taken by 99% of students at the end of general high school, make them a reliable measure of academic achievement. Third, the main results in this paper use data for about 230,000 randomized students. The sheer size of these data allows for the precise estimation of the degree of heterogeneity in school effects, even in specific sub-samples. Finally, there exists significant heterogeneity in school inputs in randomizing areas, which is important for assessing their influence in shaping educational outcomes.

An important consideration in interpreting these findings—and in evaluating their external validity—is the extent to which the degree of heterogeneity in pure school effects is influenced by institutional factors that potentially affect the quality of key school inputs, such as the ability to choose principals and teachers. We examine this issue by comparing the heterogeneity of pure school effects across public and privately-founded schools. In the latter schools, that account for about two thirds of schools in randomizing areas, 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, which account for the other third of schools within randomizing areas, do not have any discretion on the selection of teachers and principals. We find that the standard deviation of pure school effects tends to be larger among privately-founded schools, but still not more than 0.06. Interestingly, using data from the cross-section of countries surveyed in PISA 2000, we also document that schools in Korea have a relatively high degree of autonomy, and display considerable heterogeneity in observable inputs. This evidence suggests that our main results are unlikely to be specific to the Korean context.

This paper contributes to several strands of literature. Following the landmark study by Coleman et al. (1966), a large number of studies have examined drivers and effects of school quality, including important contributions by Hanushek (1986), Card and Krueger (1992), Betts (1995), Rockoff (2004) and Rivkin et al. (2005). A growing body of work uses experimental or rule-based designs to assess the effects of school and/or peer quality for students at the margin of admission (Rouse, 1998; Cullen et al., 2006; Hastings et al., 2008; Abdulkadiroglu et al., 2011; Dobbie and Fryer, 2011; Jackson, 2013; Pop-Eleches and Urquiola, 2013; Park et al., 2013). A recent paper by Deming et al. (2014) is perhaps the closest to our own. Exploiting a public school choice lottery in Charlotte-Mecklenburg, they report evidence of a significant overall increase in college attainment among lottery winners who attend their first-choice school. They then show that a variety of non-experimental school value-added measures, which control directly for observed differences in peer quality, are a good predictor of the impacts of school choice. As noted by Deming et al. (2014), an important limitation of this literature is the impossibility of cleanly unpacking the effects of changing school assignment into changes in peer quality, teacher quality, or other inputs that can be directly manipulated by schools. By estimating school value added in areas characterized by universal random assignment of students to general high schools, we are able to circumvent this difficulty and provide experimental evidence on the importance of pure school effects—reflecting the influence of school inputs but not heterogeneity in student composition—for academic achievement.

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, while section 4 presents the results. Section 5 provides a discussion on the external validity of our findings. Section 6 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 may choose to enter high school, which takes another three years to complete. Although enrollment in high school is not mandatory, about 97% of students from the corresponding 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. In turn, selective schools provide a more specialized curriculum, have greater autonomy, and select students in a competitive process that involves the use of information from GPAs and interviews.

The process for assigning students to high schools is composed of two rounds. In the first round, common to all administrative divisions in the country, interested students apply and are assigned to selective or vocational schools. The second round, that involves allocating the remaining non-assigned students to general high schools, varies across three groups of administrative divisions. In a first group, there is universal random assignment of students to general high schools within districts. In a second group, students apply to high schools and assignment decisions are made individually by schools largely on the basis of test scores and middle school GPAs. Finally, in a third group, a subset of districts randomize students to high schools, while other districts follow the application and admission procedure.

During the period of analysis, 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 in these areas were randomly assigned to general high schools.[4]

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. Besides random student assignment, monetary transfers were centralized and balanced across schools, education facilities were upgraded, and some teacher training was provided.[5] The equalization policy was first implemented in 1974 in Seoul and Busan (the two largest cities in the country). Subsequently it was progressively expanded to include metropolitan cities, provincial capitals and finally major regional cities.[6] Between 1980 and 1995 the system remained essentially stable. Because the equalization policy targeted larger administrative divisions, there exist several divisions where the policy was never implemented. Starting in 1996 some limited choice was reintroduced in divisions where the equalization policy was implemented. In particular, students identify the two or three schools of their preference. Schools then fill 30 to 40% of slots from random selection among students who show preference for the school. The remaining slots are randomized across students residing in the corresponding school district who were not assigned to their preferred school.

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

[5]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. However, these other components were not successfully implemented because budgetary constraints made it unfeasible to incur the relatively high costs associated with teacher training and facility improvement (KEDI, 1998).

[6]Lee (2012) provides detailed information on this process in Appendix Table 1. Park (2013) provides a comprehensive description of the equalization policy in Korea.

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

There is substantial variation across public and privately-founded schools regarding personnel matters. In public schools, teachers are government employees who are hired in a centralized fashion based on their performance in a standardized exam. Teachers must move to a different school within the state every four or five years. Principals in public schools are selected by the regional educational office and can remain in their position at most for two four-year terms. Hence public schools do not have any discretion regarding the selection of teachers and principals. In contrast, in privately-founded schools there is substantial discretion at the school level regarding personnel decisions. In particular, the board of directors is responsible for the appointment and promotion of principals. Principals have then command over the hiring and dismissal of teachers, and the length of their contracts. In summary, 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.

2.2 Data

We use data from three main sources. First, to assess how much schools differ in the effects they have on student learning, we use data from the College Scholastic Ability Test (CSAT). Following high school, students who want to continue to university must take this test, which has a major impact on their education prospects. The high stakes involved in these college entrance exams make them a highly reliable measure of academic achievement. The CSAT college entrance exams are taken by about 99% of students attending general schools at the end of high school. The empirical analysis exploits individual-level test scores in CSAT for the years 1995 to 1997.[7] These data therefore contain students assigned to high schools between 1992 and 1994, before the equalization policy started to be partially reversed in some areas.[8] 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 characterized as follows. Two thirds of the exam are identical across the whole country. This homogeneous component assesses proficiency in Korean and English languages, as well as in part of the Mathematics curriculum. By contrast, the remaining third is choice-based and tests proficiency in Science or Social Science and in the remainder of the Mathematics curriculum. We focus our analysis on test scores in Korean and English, as these scores are perfectly 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 addition, we use the simple average of the two standardized subject scores as a summary measure of academic achievement. 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 have merged this information from the annual statistics book of each state.

Second, to examine the validity of the experimental design, we further use individual-level data 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 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.[9] Students sampled by KELS are administered a

[7] Until 1993, Korea had a universal college entrance exam named student achievement test (the *hakruk* exam), but data are not available. CSAT started in 1994, but with a slightly different format. As a result, the CSAT data are available only since 1995.

[8] Students take the CSAT test in November of a given year and graduate from high school in the following year. The CSAT data are coded by the graduation year.

[9] In a first step, 150 schools are selected nationwide in consideration of the regional distribution of schools and students. In each school, 50 students are drawn at random, while all students are drawn if the

series of socio-demographic and school related questionnaires. In each wave of KELS, student academic performance is measured by achievement tests for three subjects: Korean, English and Mathematics. We consider a sample of students in the randomizing areas excluding Seoul.[10] Since the KELS data started to be collected 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 used the middle school as a proxy for the school district. Table A2 in the appendix provides summary statistics on these data.

Finally, to provide an empirical benchmark for the main analysis, we use data from the Program for International Student Assessment (PISA) in the year 2000. PISA is a standardized international assessment coordinated by the Organization for Economic Co-operation and Development that measures literacy of 15 year old students in Reading, Mathematics and Science every three years. We use these data to document how Korea compares with other countries with regard to: (i) the degree of heterogeneity of school effects on test scores; (ii) the degree of heterogeneity in observable 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 CSAT and KELS, in constructing the PISA sample we drop observations for vocational schools. In addition, we drop observations for Canada (that did not provide a representative sample) and for Norway and Poland (that have missing values for measures of school autonomy). After imposing these restrictions, 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, which 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.

school is attended by fewer than 50 students.

[10]As described in more detail in section 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.

The resulting sample contains about 1,150,000 student observations. Additionally, we drop students attending schools with very low enrollment (less than 100 students). This restriction excludes small schools from remote areas and ensures a minimum sample size for estimating school effects. Imposing this restriction reduces the sample in 2.7%. Finally, we impose two additional minor constraints which leaves us with the working sample of 1,083,237 student observations.[11]

We start by defining the "randomized sample", which we use to estimate pure school effects. This sample is composed of students that were randomly assigned to general high schools within districts, and for whom we have information regarding the composition of the school district. To construct this sample, we start by focusing on 6 major administrative divisions that randomized all students to high schools in the final round of the assignment process. 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. 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. 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 further define a "non-randomized sample" that is composed of students in the provinces 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. Seoul and Incheon did randomize students to general high schools in the second round, but for the reasons explained above could not be included in the randomized sample.

In Gwangju, students who were assigned to high school in 1995 and took the CSAT exam in 1998 may have been able to exert preferences, reflecting partial reversal in the

[11] 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, which makes it impossible to estimate variation in school effects in some specifications.

equalization policy.[12] For this reason, 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 on these data. 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, over 70% of students were enrolled in privately-founded schools, more than 96% of students attended single-sex schools, and the majority of students (56%) were enrolled in large schools.[13] These proportions are larger than in the full sample, and considerably larger than in the non-randomized sample. These statistics also reveal that students in the randomized sample tend to perform considerably better in CSAT exams than students in the full sample, and clearly better than those in the non-randomized sample. Table A3 in the appendix reports the number of schools by type for each district in the randomized sample in 1997. It shows that there exists a sufficient number of schools in each district to estimate school effects.

2.4 Validity of the experimental design

To provide evidence on the validity of the experimental design, we use individual-level data from KELS. In particular, we use these data to examine whether middle school test scores and household socio-demographic attributes predict observable characteristics of the high school the student ends up attending. 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 observable 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

[12] We used a news archive to confirm historical changes in either school districts or in the student assignment system. On September 1994, a newspaper reported that Gwangju was considering to introduce student preferences for school assignment in 1995 (implying that students taking CSAT in 1998 would have exerted preferences when entering high school). In 1996, several cities in the randomized sample began to allow for student preferences for high school assignment. For these reasons, our estimation samples do not to include CSAT data from 1998 onward.

[13] Large schools are defined as those with above-median enrollment (i.e. greater than 699 students).

observable 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.

In the context of 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. Table 2 reports the estimation results. The point estimates in column (1) reveal 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. The results in column (2) show that these student attributes are also unrelated to a dummy variable indicating if the school enrolls only students of the same sex or not. The estimates in column (3) show that these student attributes are not a significant predictor of class size. Finally, column (4) reveals that most student attributes also tend to be unrelated to total enrollment in the high school. The sole exception concerns parental education, which has a positive and significant coefficient at the 5% level.[14]

These data have three limitations for the purpose of confirming that students in the randomized sample in CSAT were randomly assigned to general high schools within districts. First, as we noted above, KELS began in 2005 and the equalization policy started to be reversed in 1996. Hence these data may include some areas where student assignment was not fully random. Second, KELS does not identify each state, but contains information on whether students are in a randomized area or in Seoul. Hence we are including all students in randomizing areas (excluding Seoul). Third, we cannot include district fixed effects because schools are anonymized in the available KELS data, although we addressed this issue by adding middle school origin dummies (students attending the same middle school typically belong to the same high school district).

Nevertheless, it is reassuring that, even if this sample might contain some contamination, we do not observe a systematic association between observable attributes of students and the corresponding high schools. Notice further that Table A3 in the appendix reports analogous results for the non-randomized sample, for which we do observe that student

[14]The fact that one out of sixteen coefficients is significant (at the 5% level) would be expected: allowing for a 5% chance of Type I error, and under the strong assumption that results are independent, we would expect to reject the null 0.8 times (16*0.05).

characteristics predict the attributes of the high schools that students end up attending. Most notably, this table reveals 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.

3 Empirical models

This section presents the empirical models used to examine: (i) whether schools differ in the effects they have on student learning; and (ii) how much of the variation in student achievement can be explained by schools. In line with Chetty et al. (2011), we assess the role of pure school effects in shaping academic achievement by estimating an equation of the form:

$$y_{ips} = \alpha + \gamma_p + \phi_s + \varepsilon_{ips} \tag{2}$$

where y_{ips} is the test score for individual i from randomization pool p and school s, γ_p is a randomizing pool fixed effect, ϕ_s are school effects, and ε_{ips} is the error term.[15] Randomization pools are defined by the interaction of gender, district and year. In some specifications, we will also control for a vector of observed school attributes, X_s, which are measured in the corresponding year of observation.

The fixed effects for randomizing pools, γ_p, account for the heterogeneity of students across randomizing pools, which might be expected to affect test scores. Because students are randomly assigned to schools within each randomizing pool, student baseline ability should be orthogonal to ϕ_s. In this framework, ϕ_s are pure school effects, i.e. a bundle of all school level attributes (observable and unobservable) that varied within districts, including, among others, learning time, school resources and curriculum, attributes of principals and teachers, organizational practices and school culture.

To assess whether pure school effects jointly matter for achievement, we first estimate (2) using a fixed effects specification for the school effects, ϕ_s. Under the null hypothesis of no school effects on scores in college entrance exams, the school dummies should not be

[15] Chetty et al. (2011), section V.B., use data from the STAR experiment, in which students and teachers were randomly assigned to classrooms, to explore whether (and how much) the assignment of a student to a certain classroom affected her subsequent educational and economic outcomes. In their setting, students were randomized across classrooms within schools. This is analogous to our setting, in which students were randomized to schools within districts.

statistically significant, as random assignment of students to high schools would remove the potential influence of student composition or peer effects. We test this null hypothesis by performing F-tests for ϕ_s=0. To determine how much pure school effects matter for achievement, we use a random effects model to estimate (2). We then compute the standard deviation of school effects under the assumption that they are normally distributed. To estimate how much of the variation in student academic achievement can be explained by schools, we estimate the intra-cluster correlation as

$$\frac{\sigma_s^2}{\sigma_s^2 + \sigma_w^2} \tag{3}$$

where σ_s^2 is the variance of test scores between clusters (i.e. schools) and σ_w^2 is the variance within schools. In each of these cases, we also estimate specifications where we include a vector of observable attributes of schools.

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

$$y_{ips} = \alpha + \gamma_p + \beta X_s + \varepsilon_{ips} \tag{4}$$

where the remaining variables have the meaning defined above. In this case, the estimation does not include fixed or random school effects (in contrast with (2)).

4 Results

4.1 An empirical benchmark

Before turning to the estimation of pure school effects in the randomized sample, it is useful to perform a similar analysis using data for settings characterized by non-random student assignment to schools. We consider two different samples: the non-randomized sample in Korea described earlier, and a sample of countries participating in PISA in the year 2000. In the case of PISA, we analyze scores in Reading because it was the main focus of the 2000 round. Consequently, all students took this test.[16] For consistency with the analysis of CSAT data, we normalize the test scores so that the mean is zero and the standard deviation is one in each country. Since both these samples are characterized by

[16]The Mathematics test was taken by 50% of students, as was the Science test. Only 25% of students took the three tests.

non-random assignment of students to schools, the estimated school effects reflect both heterogeneity in school inputs and in student attributes (including peer effects).

Table 3 presents results for the non-randomized sample in Korea using the models described above. The table reports the p-value of the F-tests on whether all school fixed effects are zero. That is, this test assesses whether there are statistically significant differences across schools in learning outcomes. To provide a measure of the heterogeneity of the school effects, the table also reports the standard deviation for the estimated school effects (based on a model similar to (2)) and the share of the variation in student achievement that is accounted for by schools (based on (3)). Note that in the two provinces that compose the non-randomized sample, students may apply to all high schools in the corresponding province. In practice, however, students are more likely to apply to high schools in more restricted geographic areas (i.e. cities). Hence we consider specifications with gender-province-year fixed effects and gender-city-year fixed effects. In columns (1) and (2), we report results for the full non-randomized sample, while in columns (3) to (6) we report results for each province separately. In all cases, the F-tests reject the null that all school fixed effects are zero. In the complete non-randomized sample, the estimated standard deviation of school effects is 0.542 when adding gender-province-year fixed effects, and 0.519 when adding gender-city-year fixed effects. The results are similar when focusing on the individual provinces composing the non-randomized sample. Finally, the results indicate that between-school heterogeneity accounts for a large share of total variance in academic achievement. This percentage is quite similar across different sub-samples and specifications—it ranges from 31.8% to 36.8% across provinces and specifications.

Table 4 depicts similar findings based on student-level data from PISA 2000 for each country. Inspection of this table reveals that the standard deviation of school effects varies between 0.251 in Iceland and 0.835 in Hungary, with a cross-country unweighted mean of 0.568 and a median of 0.592. Korea is characterized by low to intermediate levels of heterogeneity in school effects (0.348), above Iceland, Finland and Sweden, and just below Denmark, Ireland and New Zealand.[17] The heterogeneity in school effects tends to be particularly high (above 0.6) in less developed countries such as Argentina, Brazil, Chile, Peru, Hungary, Czech Republic and Bulgaria, but also in several high income countries such as Japan, Netherlands, France, Italy, Germany, Belgium, United Kingdom

[17] Note that the heterogeneity in school effects for Korea combines schools in randomizing and non-randomizing areas.

and Australia.

4.2 Main results

In Table 5 we perform a similar exercise using the randomized sample in Korea. Since students were randomly assigned to general high schools, baseline student ability should be orthogonal to school effects. Therefore, in contrast to the samples used in the previous subsection, the baseline observable and unobservable attributes of students are balanced across schools within districts. This feature of the data offers us the possibility of cleanly estimating the heterogeneity in pure school effects on academic achievement. The results in columns (1) of Table 5 reveal that in districts featuring random student assignment, schools differ in the effects they have on academic achievement. The F-test on the joint significance of school effects clearly rejects the null of zero effects. However, these results also reveal that the heterogeneity in school effects within randomizing districts is small when compared to that observed in school districts featuring both student selection and heterogeneity in school inputs: at 0.057 the standard deviation of school effects within randomizing areas is about one tenth of that observed in areas featuring non-random student selection. Moreover, the results indicate that pure school effects account only for about 0.5% of the overall variation in academic achievement. The results in column (2) of this table reveal that these conclusions remain similar when accounting for observable attributes of schools (notably total enrollment, and indicators for whether the school if privately-founded or single-sex).

The heterogeneity in school effects observed in randomized areas is also small when compared to that observed in Table 4 for countries sampled in PISA. In particular, it is less than a quarter of that observed for Finland, and less than one tenth of the average (or median) heterogeneity observed in the cross-section of countries.

4.3 Robustness and heterogeneity of school effects

We proceed by examining the extent to which pure school effects are heterogeneous across different sub-samples. Table 6 reports separate estimates on the standard deviation of school effects for each year of the sample period. The results in column (2) to (4) reveal that the estimates for the standard deviation of school effects are fairly similar across years. Although slightly less precisely estimated, they are also close to that obtained for the full sample (column 1).[18]

Table 7 reports separate estimates by subject and gender. Columns (2) and (3) in this table report estimates of pure school effects for different test scores in each of the subjects measured, respectively Korean and English. The results reveal that the standard deviation of school effects is considerably larger in the foreign language than in the native language: 0.085 versus 0.041. Finally, columns (5) and (6) show that the standard deviation of pure school effects is fairly similar for females and males: 0.050 versus 0.062.

An important consideration in interpreting these estimates is the extent to which the degree of heterogeneity in pure school effects is influenced by institutional factors that potentially affect the variation in key school inputs, such the ability to select principals and teachers. Indeed, a series of recent influential studies for the US suggest that individual teachers have sizable impacts on academic achievement (Chetty et al., 2014ab). Hence it is possible that the context of public schools in Korea, characterized by regular rotation of teachers across schools, may be different to that prevailing in areas in which there is more sorting of teachers to schools. To provide evidence on this issue, we compare the heterogeneity of pure school effects across public and privately-founded schools. In the latter schools, which account for about two thirds of schools in randomizing areas, 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, which account for the other third of schools within randomizing areas, do not have any discretion regarding selection of teachers and principals. Table 8 reports the results for each of these sub-samples. We find that the standard deviation of pure school effects is indeed larger in privately-founded schools than in public schools: 0.060 versus 0.042. Nevertheless, the heterogeneity in school effects among privately-founded schools is still fairly small when compared to that documented for non-randomizing areas in Korea and for the vast majority of countries sampled in PISA (see Table 4).

The analysis above suggests that pure school effects matter for academic achievement. But it also reveals that heterogeneity in school inputs explains a relatively small proportion of the overall variation in learning outcomes. In Table 9 we examine whether observable school attributes in Korea are an important driver of the observed variation in

[18]Notice that these estimates are essentially the standard deviation of average test scores across schools. Therefore, they are a non-linear function of these school average scores, which implies that estimates for the full sample might potentially range outside those observed for sub-samples.

learning, on average. To provide a benchmark, column (1) presents point estimates for the non-randomized sample, based on the model specified in equation (4). These estimates suggest that students enrolled in public single-sex schools tend to perform relatively better in college entrance exams. However, in the absence of random assignment, this heterogeneity in achievement across school types might simply reflect heterogeneity in student composition across schools. In column (2) we perform a similar analysis, but focusing now on areas featuring universal random assignment of students to high schools. In this case, the results do not show any systematic association between observed high school attributes and performance in college entrance exams. This evidence suggests that once student baseline attributes are balanced across schools, none of the observed attributes of schools in Korea has a systematic impact on academic achievement.

5 External validity

The results reported above suggest that pure school effects are a statistically significant driver of academic achievement. However, they also suggest that the heterogeneity in school effects within randomizing areas is about one tenth of that observed in areas featuring non-random student assignment, which is in turn close to the average heterogeneity of school effects observed in the cross-section of countries surveyed in PISA 2000. Interestingly, these findings also hold among privately-founded high schools in Korea, which have greater freedom to choose key school inputs, notably principals and teachers. Taken together, these results suggest that the variation in inputs that can be directly manipulated by schools plays a limited role in explaining differences in mean school performance, especially when compared with the role played by student heterogeneity (including peer effects).

To gain further insights on the external validity of our results, we first inspect the degree of heterogeneity in observable school inputs across randomized and non-randomized areas using data from KELS. Apart from the marked differences in the process of student assignment, schools across Korea operate in a similar institutional environment with regard to the choice of inputs. Table 10 reports the mean and standard deviation of several observable school attributes in randomizing and non-randomizing areas using data from KELS. The statistics in this table reveal that schools in randomizing areas tend to be larger, are more likely to be privately-founded and single-sex. The ratio of teachers and computers per student and the share of teachers with education majors tend to be smaller among schools in randomizing areas, though these differences are not large. The degree of heterogeneity in school attributes across schools does not appear to be systematically higher in non-randomizing areas. Besides having a higher share of privately-founded schools, which have greater discretion over the selection of principals and teachers, schools in randomizing areas display greater dispersion in the proportion of teachers that have education majors. By contrast, the degree of heterogeneity in school size, teacher-student ratio and computer-student ratio tends to be considerably larger in the non-randomized areas.

To explore whether the variation of school effects in Korea may be representative of the expected variation in other countries we perform two additional analysis using data from PISA 2000. First, we examine how the degree of heterogeneity in observable school inputs in Korea compares with that of other countries. Information about the dispersion of observable school inputs 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 variation in observable inputs than schools in other countries, we may expect larger heterogeneity in school effects in other countries. Second, we document the degree of school autonomy in academic, personnel and budgetary decisions in Korea and in other countries. We would expect larger variation in school effects in educational systems in which schools have greater autonomy in key management decisions.

To examine heterogeneity in observable 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.[19] To understand how we examine the degree of heterogeneity in each of these school inputs across countries, consider for example the number of hours that students spend at school. First, we start with a data set where the unit of observation is a school in a country, which contains a variable indicating the number of annual hours that students spend at school. Second, we compute the standard deviation of this indicator for each country. Third, we rank countries by the standard deviation of this indicator and compute percentiles for the over-

[19] As noted above, Korea has only public and privately-founded high schools. In PISA 2000, privately-founded schools are classified as private.

all distribution. This analysis indicates that Korea is at the 80th percentile in terms of variation in the number of annual hours that students spend at school.

Table 11 reports results from this analysis for each observable 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, computer-student ratio, 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 observable 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. (2013) 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 12 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 2 summarizes these findings by depicting the relative position of Korea in the cross-section of countries with regard to the heterogeneity in observable school inputs and school autonomy. It reveals that schools in Korea are characterized by relatively high measures along both these dimensions. Among the forty countries studied, only five countries display both higher heterogeneity in school inputs and greater school autonomy. Taken together, the evidence presented in this section suggests that the key findings of this paper are likely to apply more broadly across the world.

6 Conclusion

Since the landmark study by 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 has been difficult to cleanly unpack the effects of changing school assignment into changes in peer quality, teacher quality, or other inputs that can be directly manipulated by schools.

In this paper, we have exploited universal random assignment of students to high schools within school districts in Korea to circumvent this difficulty. We have presented experimental evidence on whether and how much pure schools effects matter for academic achievement. By estimating school value added for scores in high stakes college entrance exams in areas characterized by universal random assignment of students to high schools, we were able to confirm that pure school effects are a statistically significant determinant of academic achievement. However, we have also shown that the standard deviation in school effects within randomizing areas is only one tenth of that observed in areas featuring both student selection and heterogeneity in school inputs. Moreover, we have shown that the heterogeneity in pure school effects is only moderately 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.

References

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

Araujo, Caridad, Pedro Carneiro, Yyannu Cruz, and Norbert Schady (2016). "Teacher Quality and Learning Outcomes in Kindergarden." *Quarterly Journal of Economics*, forthcoming.

Betts, Julian R. (1995). “Does School Quality Matter? Evidence from the National Longitudinal Survey of Youth.” *Review of Economics and Statistics*, 77(2): 231-250.

Card, David, and Alan B. Krueger (1992). “Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States.” *Journal of Political Economy*, 100(1): 1-40.

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

Chetty, Raj, John Friedman, and Jonah Rockoff (2014a). "Measuring the Impact of Teachers I: Evaluating Bias in Teacher Value-added Estimates.” *American Economic Review*, 104(9): 2593-2632.

Chetty, Raj, John Friedman, and Jonah Rockoff (2014b). "Measuring the Impact of Teachers II: The Long-term Impacts of Teachers.” *American Economic Review*, 104(9): 2593-2632.

Coleman, J., E. Campbell, C. Hobson, J. McPartland, A. Mood, F. Weinfeld, and R. York (1966). *Equality of Educational Opportunity*, Washington, D.C., U.S. Government Printing Office.

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

Deming, David J., Justine S. Hastings, Thomas J. Kane, and Douglas O. Staiger (2014). “School Choice, School Quality, and Postsecondary Attainment.” *American Economic Review*, 104(3): 991-1013.

Dobbie, Will, and Roland G. Fryer, Jr. (2011). “Are High-Quality Schools Enough to Increase Achievement among the Poor? Evidence from the Harlem Childrens Zone.” *American Economic Journal: Applied Economics*, 3(3): 158-187.

Ellison, Glenn, and Ashley Swanson (2016). “Do Schools Matter for High Math Achievement? Evidence from the American Mathematics Competitions”, *American Economic Review*, forthcoming.

Hanushek, Eric A. (1986). "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, 24(3): 1141-1177.

Hanushek, Eric A., Susanne Link, and Ludger Woessmann (2013). "Does School Autonomy Make Sense Everywhere? Panel Estimates from PISA." *Journal of Development Economics*, 104: 212-232.

Hastings, Justine S., and Jeffrey M. Weinstein (2008). "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics*, 123(4): 1373-1414.

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

Jackson, C. Kirabo (2013). "Can higher-achieving peers explain the benefits to attending selective schools? Evidence from Trinidad and Tobago." *Journal of Public Economics*, 108(C): 63-77.

Jackson, C. Kirabo (2014). "Teacher Quality at the High School Level: The Importance of Accounting for Tracks." *Journal of Labor Economics* 32(4): 645-684.

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

Kim, Ga-young, and Beomsoo Kim (2015). "The Impacts of Coeducation in High School on the College Scholastic Ability Test." *KDI Journal of Economic Policy*, 37: 65-85.

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

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

Pop-Eleches, Cristian, and Miguel Urquiola (2013). "Going to a Better School: Effects and Behavioral Responses." *American Economic Review*, 103(4): 1289-1324.

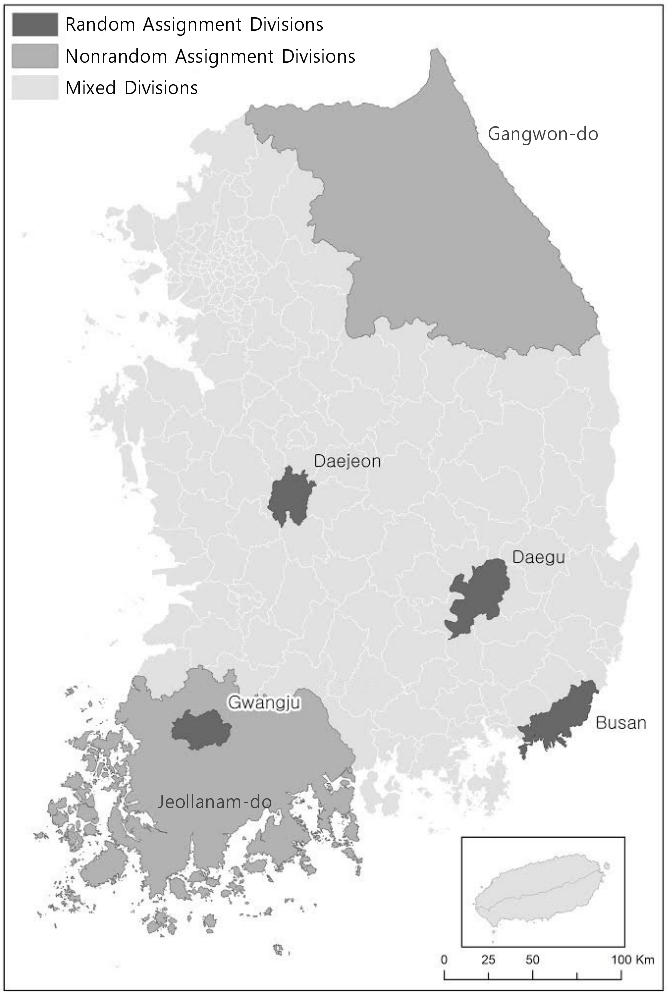
Rivkin, Steven G., Eric A. Hanushek, and John F. Kain (2005). "Teachers, Schools, and Academic Achievement." *Econometrica*, 73(2): 417-58.

Rockoff, Jonah E. (2004). “The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data.” *American Economic Review*, 94: 247-252.

Rouse, Cecilia Elena (1998). “Private School Vouchers and Student Achievement: An Evaluation of the Milwaukee Parental Choice Program.” *Quarterly Journal of Economics*, 113(2): 553-602.

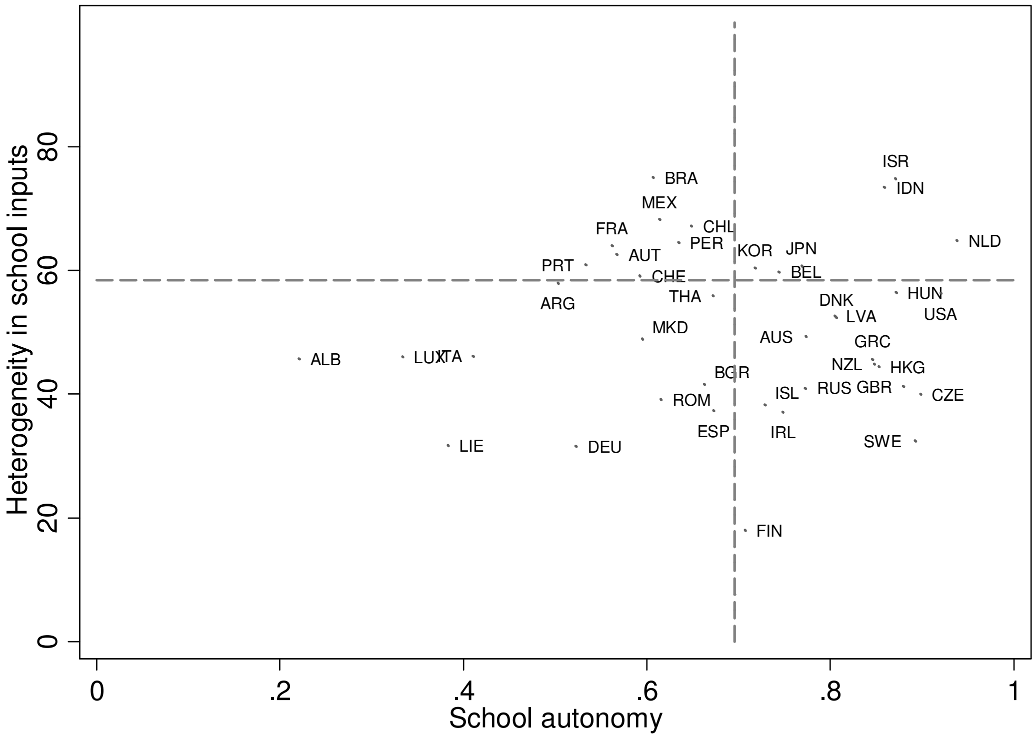
US Department of Education (2006). “Action Plan for Higher Education: Improving Accessibility, Affordability, and Accountability.” Washington, DC: US Department of Education.

Figure 1: Randomizing and non-randomizing administrative divisions



Notes: The CSAT data for 1995-1997 were divided into three samples, depending on: (i) the rules of student assignment to general high schools; and (ii) the availability of precise information on the definition of school districts. In the random assignment sample (composed of the metropolitan cities of Daejeon, Daegu, Busan and Gwangju), following a first round of admission to selective or vocational schools, the remaining students were randomly assigned (separately by gender) to general high schools within districts. In the nonrandom assignment sample (composed of the provinces of Gangwon-do and Jeollanam-do), student admission in the second round was determined by exams. To enter the random assignment sample, administrative divisions had to meet two conditions. First, the existence of information on the exact composition of the school district. Second, the observation of universal random assignment of students to general high schools within school districts. Divisions identified as mixed failed to meet these conditions, but for varying reasons. Some have universal random assignment, but the available information does not make it possible to identify the exact definition of school districts. Others comprise both urban and rural areas and implemented random assignment in the former areas but not in the latter. Finally, Seoul imposed a limit of 30 minutes of commuting time between the students' residency and the assigned school (using public transportation), thus failing to meet the condition of universal random assignment within precisely-defined school districts.

Figure 2: Heterogeneity of school inputs and school autonomy across countries



Notes: Figure reports measures of (i) the degree of heterogeneity in observable school inputs across schools and (ii) measures of school autonomy for 40 countries based on data from PISA 2000. 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

	All	Randomized sample	Non-randomized sample
	(1)	(2)	(3)
A. Share of observations by year			
1995	0.329	0.337	0.312
1996	0.334	0.330	0.343
1997	0.338	0.333	0.345
B. Share of observations by school type			
Privately-founded	0.636	0.705	0.507
Single sex	0.796	0.963	0.722
Large	0.499	0.562	0.069
C. Mean test scores (standard deviations)			
Average	0.000	0.252	-0.189
	(0.931)	(0.822)	(0.943)
Korean	0.000	0.235	-0.210
	(1.000)	(0.882)	(1.021)
English	0.000	0.269	-0.167
	(1.000)	(0.910)	(1.991)
N (obs.)	1,083,237	233,496	72,373

Notes: Sample consists of individual-level data on test scores in CSAT exams in Korean and English for high school graduates in 1995, 1996 and 1997. Large schools are defined as those with above-median enrollment (i.e. greater than 699 students).

Table 2: Randomization tests

Dependent variable:	Privately-founded	Single-sex	Class size	Enrollment
	(1)	(2)	(3)	(4)
Test scores in middle-school	-0.000	0.018	0.116	0.010
	(0.021)	(0.018)	(0.111)	(0.010)
Parents' education	-0.003	0.006	0.080	0.007*
	(0.008)	(0.008)	(0.043)	(0.003)
Log household income	0.012	0.014	-0.043	-0.003
	(0.014)	(0.014)	(0.083)	(0.006)
Family size	0.034	0.028	0.142	0.008
	(0.029)	(0.021)	(0.131)	(0.011)
R-square	0.348	0.469	0.532	0.568
N (obs.)	980	980	980	980

Notes: 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 and multiplied by 1000. 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: Estimates of school effects, non-randomized sample

	Pooled		Gangwon-do		Jeollanam-do	
	(1)	(2)	(3)	(4)	(5)	(6)
p-value of F-test on school effects	0.000	0.000	0.000	0.000	0.000	0.000
Standard deviation of school effects	0.542**	0.519**	0.565**	0.543**	0.525**	0.501**
	(0.043)	(0.041)	(0.069)	(0.066)	(0.054)	(0.052)
Between-school variance of test scores	35.1%	33.3%	36.8%	35.1%	33.7%	31.8%
Gender-province-year fixed effects	Y	N	N	N	N	N
Gender-year fixed effects	N	N	Y	N	Y	N
Gender-city-year fixed effects	N	Y	N	Y	N	Y
N (obs.)	72,373	72,373	28,761	28,761	43,612	43,612

Notes: Estimates are based on individual-level data on test scores in CSAT exams in Korean and English for high school graduates in 1995-1997 in the non-randomized sample described in Table 1. Standard errors are below each estimate in parentheses. ** significant at 1% level; * significant at 5% level.

Table 4: Standard deviation of school effects, PISA 2000

Iceland	0.251	Israel	0.603
Finland	0.256	Japan	0.607
Sweden	0.264	Argentina	0.609
Korea	0.348	Romania	0.620
Denmark	0.357	Brazil	0.630
Ireland	0.360	Hong Kong	0.631
New Zealand	0.363	Greece	0.643
Spain	0.419	Netherlands	0.657
Australia	0.428	Czech Republic	0.670
Latvia	0.496	France	0.682
Thailand	0.507	Mexico	0.684
United States	0.507	Italy	0.687
Indonesia	0.539	Germany	0.694
Albania	0.562	Chile	0.697
Portugal	0.562	Belgium	0.702
Russia	0.564	United Kingdom	0.711
Luxembourg	0.572	Bulgaria	0.731
Switzerland	0.573	Peru	0.732
Macedonia	0.574	Austria	0.813
Liechtenstein	0.581	Hungary	0.835

Notes: 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.

Table 5: Estimates of school effects, randomized sample

	(1)	(2)
p-value of F-test on school effects	0.000	0.000
Standard deviation of school effects	0.057**	0.057**
	(0.004)	(0.004)
Variation in test scores explained by schools	0.5%	0.5%
School covariates	N	Y
N (obs.)	233,496	233,496

Notes: Estimates are based on individual-level data on test scores in CSAT exams in Korean and English for high school graduates in 1995-1997 in the randomized sample described in Table 1. All regressions include fixed effects by randomizing pool (defined by the interaction between gender, district and year). Regressions in column (2) additionally include school covariates. Standard errors are below each estimate in parentheses. ** significant at 1% level; * significant at 5% level.

Table 6: Estimates of school effects by year, randomized sample

	Pooled (1)	1995 (2)	1996 (3)	1997 (4)
p-value of F test on school effects	0.000	0.000	0.000	0.000
Standard deviation of school effects	0.057**	0.064**	0.063**	0.062**
	(0.004)	(0.005)	(0.005)	(0.005)
Variation in test scores explained by schools	0.5%	0.7%	0.6%	0.6%
N (obs.)	233,496	78,710	76,989	77,797

Notes: Estimates are based on individual-level data on test scores in CSAT exams in Korean and English for high school graduates in the randomized sample described in Table 1. All regressions include fixed effects by randomizing pool (defined by the interaction between gender, district and year). Standard errors are below each estimate in parentheses. ** significant at 1% level; * significant at 5% level.

Table 7: Estimates of school effects by subject and gender, randomized sample

	Pooled (1)	Korean (2)	English (3)	Females (4)	Males (5)
p-value of F test on school effects	0.000	0.000	0.000	0.000	0.000
Standard deviation of school effects	0.057**	0.041**	0.085**	0.050**	0.062**
	(0.004)	(0.003)	(0.005)	(0.005)	(0.005)
Variation in test scores explained by schools	0.5%	0.2%	0.9%	0.5%	0.5%
N (obs.)	233,496	233,496	233,496	100,957	132,539

Notes: Estimates are based on individual-level data on test scores in CSAT exams in Korean and English for high school graduates in the randomized sample described in Table 1. All regressions include fixed effects by randomizing pool (defined by the interaction between gender, district and year). Standard errors are below each estimate in parentheses. ** significant at 1% level; * significant at 5% level.

Table 8: Estimates of school effects by school type, randomized sample

	Pooled (1)	Privately-founded (2)	Public (3)
p-value of F test on school effects	0.000	0.000	0.000
Standard deviation of school effects	0.057**	0.060**	0.042**
	(0.004)	(0.005)	(0.006)
Variation in test scores explained by schools	0.5%	0.6%	0.3%
N (obs.)	233,496	164,618	68,878

Notes: Estimates are based on individual-level data on test scores in CSAT exams in Korean and English for high school graduates in the randomized sample described in Table 1. All regressions include fixed effects by randomizing pool (defined by the interaction between gender, district and year). Standard errors are below each estimate in parentheses. ** significant at 1% level; * significant at 5% level.

Table 9: Effects of school characteristics on learning

	Non-randomized sample (1)	Randomized sample (2)
Privately-founded	-0.485**	0.008
	(0.126)	(0.011)
Single-sex	0.488**	-0.018
	(0.140)	(0.022)
Large	0.289	-0.008
	(0.172)	(0.014)
R-square	0.177	0.033
N (obs.)	72,373	233,496

Notes: Estimates are based on individual-level data on test scores in CSAT exams in Korean and English for high school graduates in the non-randomized and randomized samples described in Table 1. All regressions include fixed effects by randomizing pool (defined by the interaction between gender, district and year). Standard errors are below each estimate in parentheses. ** significant at 1% level; * significant at 5% level.

Table 10: School inputs in randomizing and non-randomizing areas

	Non-randomizing		Randomizing	
	Mean	St. Dev.	Mean	St. Dev.
	(1)	(2)	(3)	(4)
Enrollment	974	0.486	1289	0.265
Average teacher-student ratio	0.066	0.024	0.059	0.006
Computer-student ratio	0.106	0.073	0.084	0.039
Privately-founded	0.378	0.485	0.511	0.500
Single-sex	0.342	0.475	0.712	0.453
Share of teachers with education majors	0.632	0.201	0.566	0.216
N (obs.)	1094		980	
N (schools)	199		189	

Notes: Sample consists of school-level data from KELS. High school characteristics (privately-founded, single-sex, class size and enrollment) were extracted from wave 4 collected in 2008. Enrollment is for whole school and multiplied by 1,000. Average teacher-student ratio is total number of teachers divided by enrollment. Randomized areas are identical to those defined in Table 2. ** significant at 1% level; * significant at 5% level.

Table 11: Heterogeneity in school inputs across schools in PISA 2000

	Korea	Europe	Asia and 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
Proportion teachers certified	2.6	54.1	39.5	56.1
Proportion teachers with qualifications	2.6	48.4	58.0	54.3
Average	60.4	46.0	58.3	64.9

Notes: Table reports measures of heterogeneity in observable 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 12: Measures of school autonomy

	Korea	Europe	Asia and Oceania	Americas
	(1)	(2)	(3)	(4)
Academic	0.972	0.689	0.953	0.817
Personnel	0.250	0.412	0.478	0.428
Budget	0.930	0.896	0.954	0.718
Average	0.718	0.666	0.795	0.654

Notes: 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) correspond to the simple average of autonomy measures in the corresponding countries.

Table A1: 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
Gangwon-do	province	1,540,307
Chungcheongbuk-do	province	1,475,448
Chungcheongnam-do	province	1,822,543
Jeollabuk-do	province	2,007,379
Jeollanam-do	province	2,166,247
Gyeongsangbuk-do	province	2,811,586
Gyeongsangnam-do	province	3,058,479
Jeju	special self-governing province	528,360

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

Table A2: Summary statistics, KELS data set

	Randomized sample (1)	Non-randomized sample (2)
Parents' education	13.317	12.760
Log household income	12.770	12.722
Number of siblings	2.162	2.273
Test scores in middle-school	0.248	0.086
Single-sex	0.712	0.342
Privately-founded	0.511	0.378
Enrollment	1.289	0.974
Class size	36.400	34.036
N (obs.)	980	1094

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 and multiplied by 1000. 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). KELS do not provide information on school district. Middle school is used to proxy for the school district.

Table A3: Number of schools by district and type, randomized sample, 1997

District	All (1)	Privately-founded (2)	Single-sex (3)	Large (4)
Busan 1	13	8	13	8
Busan 2	15	9	15	8
Busan 3	15	10	15	5
Busan 4	19	10	18	8
Daegu 1	11	9	10	7
Daegu 2	12	10	12	10
Daegu 3	12	9	12	6
Gwangju	32	27	30	1
Daejeon	24	16	21	3
Total	153	108	146	56

Notes: Table reports number of schools by type in samples for male and female students from the CSAT administrative records in 1997.

Table A4: Predicting high school characteristics using pre-determined student variables in non-randomizing areas

	Privately-founded	Single-sex	Enrollment	Class size
	(1)	(2)	(3)	(4)
Test scores in middle school	0.022	0.063**	0.066**	0.580**
	(0.026)	(0.023)	(0.011)	(0.136)
Parents' education	0.015*	-0.007	0.006	0.023
	(0.006)	(0.006)	(0.003)	(0.031)
Log household income	-0.031**	0.005	0.004	0.083
	(0.010)	(0.013)	(0.006)	(0.071)
Family size	0.000	0.009	0.017	0.244
	(0.018)	(0.014)	(0.011)	(0.146)
R-square	0.492	0.619	0.858	0.878
N (obs.)	1,094	1,094	1,094	1,094

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 (privately-founded, single-sex, class size and enrollment) were extracted from wave 4 collected in 2008. Class size is calculated for the whole school and multiplied by 1000. 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 to proxy for the school district. Standard errors are clustered by middle school.