Using Within-District Random Student Assignments to Estimate the Average Partial Effects of Single-Sex Schooling and School Resources on Academic Achievement

Eleanor Jawon Choi  
Hanyang University

Hyungsik Roger Moon  
USC and Yonsei University

Geert Ridder*  
USC

June 1, 2015

PRELIMINARY. PLEASE DO NOT CITE OR CIRCULATE.

Abstract

This paper estimates the average partial effects (APEs) of single-sex schooling and school resources on academic achievement exploiting the unique student assignment mechanism in Seoul, Korea. Middle school graduates were randomly assigned to high schools within each school district in Seoul until 2009. Exploiting the within-district random student assignments, we derive a school-level education production function by aggregating individual potential outcomes. We propose a two-step method to estimate the school-level education production function and the APEs of school inputs. The coefficient estimates in the education production function considerably differ across school districts, which implies that there may be a substantial degree of endogenous sorting and treatment heterogeneity. The APE estimates show that single-sex schools have a large positive effect on male students’ academic performance whereas the effect is much smaller for female students. The effect of school resources is positive but nearly zero for both male and female students. We also show that the APEs of school inputs can be inconsistently estimated when endogenous district choice and heterogeneous treatment effects are not properly addressed.

JEL Codes: C21, I21

Keywords: average partial effect, within-district random student assignment, selection, heterogeneity, school effect, education production

*Choi: College of Economics and Finance, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, 133-791, Korea. Email: choiej@hanyang.ac.kr. Moon: Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089-0253. School of Economics, Yonsei University, Seoul, Korea. Email: moonr@usc.edu. Ridder: Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089-0253. Email: ridder@usc.edu
1 Introduction

The effect of schools on student academic outcomes has long been an important issue among researchers, policy makers, and parents. It is often difficult to estimate the causal effect of schools from regression analyses using observational data because students and their parents are often allowed to choose schools or neighborhoods in many countries including the US. Addressing endogenous selection into schools or neighborhoods has been a key econometric issue in the literature on school effectiveness. Credible estimates can be obtained when quasi-experimental variation generated from rule-based student assignments are combined with a careful econometric analysis properly taking into account the institutional details.

This paper exploits within-district random student assignments in Seoul, South Korea (hereafter, Korea) to estimate the average partial effects (APEs) of school inputs on students’ academic achievement. We focus on the effects of single-sex education (single-sex versus coeducational schools) and resources (class size) that have been important policy variables in economics of education. There is a large body of literature on the class size effect. Well-known studies include Angrist and Lavy (1999); Krueger (1999); Dearden, Ferri, and Meghir (2002), among others. The effect of single sex education has become an active area of economic research in relatively recent years (Hoxby 2000; Lavy and Schlosser 2011; Jackson 2012). Previous studies looking at the effects of both factors include Whitmore (2005).

On entering high school, students were randomly assigned to schools within each school district in Seoul until 2009. The high school lotteries implemented in Seoul until 2009 had the following unique features. First, by making decisions on residential locations, students and their parents can only choose the average quality of schools to which they can potentially be assigned with non-zero probability. This implies that self-selection bias arises only from endogenous sorting across school districts but not from endogenous selection into schools within each school district. Second, school attributes including resources and teachers are not randomly assigned across schools. This implies that each school within a school district should be considered to be a separate treatment comprising multiple components, including school types, resources, and teacher quality, some of which are observed and others are not.

Studies on school choice and residential sorting include Urquiola (2005); Rothstein (2006); Bayer, Ferreira, and McMillan (2007). Meghir and Rivkin (2011) provides an overview of the literature on education production. Park, Behrman, and Choi (2013) uses data on high school education from Korea that are similar to ours. This is unlike the Tennessee’s Project STAR where not only students but also teachers were randomly assigned to different size classes.
The two important features are often overlooked or not fully taken into account in previous studies exploiting the within-district random student assignments in Seoul to identify school quality effects. For example, regression analyses including school district fixed effects in Park, Behrman, and Choi (2013); Lee, Turner, Woo, and Kim (2014) assume no endogenous sorting of students across school districts as well as no heterogeneity in school input effects. Also, their empirical specifications rely on a strong assumption that unobserved school quality is orthogonal to observed school characteristics. A regression analysis ignoring the key features of the assignment design, i.e. self-selection of students into school districts and the multi-dimensional aspect of the treatment, could result in biased and inconsistent estimates of school quality effects.

Our econometric model specifies individual potential outcomes that depend on observed and unobserved school inputs interacted with heterogeneous individual productivities. We derive a school-level education production function by aggregating individual potential outcomes. Given that our data contain very limited information on individual attributes, this procedure is similar in spirit to the derivation of a market demand function by aggregating over individual choices. The school input effects in the derived school-level education production function differ by district but not by school within each school district. This property is due to endogenous sorting of students into school districts as well as to random student assignment within each district. The random coefficient form of our potential outcome model allows school input effects to be potentially correlated with school characteristics across different school districts.

Under our setting, the APEs of school inputs on student achievement can be computed as a weighted average of the district-specific school input effects using the fraction of students in each district as the weight. Our econometric framework yields a regression equation where we can apply a conventional two-step estimation method on school-level panel data to estimate district-specific effects of time-varying and time-invariant school inputs. We exploit the panel structure to control for time-invariant unobserved factors that are possibly confounded with observed school characteristics. Once we estimate the effects of school inputs for each district, we can obtain a consistent estimate of the APEs by computing the weighted average of the district-specific estimates.

We use data on the College Scholastic Ability Test (CSAT) scores and characteristics of high schools in Seoul. We merge the student-level test scores and the school-level data using school names. The population studied in this paper is high school seniors attending academic high schools within six school districts in Seoul either in 2008 or 2009. We restrict our analysis this way to focus

\footnote{See, for example, Berry, Levinsohn, and Pakes (1995).}
on those whose high school assignment was randomly determined. Our analysis sample includes about 54 percent of CSAT takers who were high school seniors in Seoul over the 2008–2009 period, and covers 55 coed, 34 all-girls, and 38 all-boys high schools.

The empirical results show that single-sex schools and smaller class size lead to higher academic performance among male and female students. The effect of single-sex schools is much larger for boys than for girls. Single-sex schools increase the CSAT scores by about a quarter of a standard deviation for boys but only by 0.04 standard deviations for girls. The effect of class size is positive and statistically significant, but the magnitude of the effect is very small for both male and female students. A decrease in class size by one standard deviation is associated with an increase in the CSAT score by 0.02 standard deviations for boys and by 0.01 standard deviations for girls.

The coefficient estimates of the school-level education production function are substantially different across school districts, which implies that endogenous sorting of students across districts and heterogeneous individual productivity play an important role. We compare our estimates with estimates from a linear regression with school district fixed effects which is the in the previous studies. The results show that the effects of school inputs can be inconsistently estimated when heterogeneous treatment effects and endogenous selection into school districts are not properly taken into account. This aligns with the main findings in Choi, Moon, and Ridder (2015) which focuses on a case of binary treatment in multi-site experiments. This paper extends the econometric framework in Choi, Moon, and Ridder (2015) to a case of multidimensional treatment.

The remainder of the paper is organized as follows. Section 2 outlines the institutional background of lottery-based high school assignment in Seoul. In section 3 we take a closer look at the CSAT and school information data used in this study. The econometric framework is described in sections 4 and 5. Estimation results are given in section 6. Section 7 concludes.

## 2 Institutional Background

The education policy in Korea over the past four decades greatly emphasized equal educational opportunity. In accordance with the policy emphasis, Seoul Metropolitan Office of Education adopted the High School Equalization Policy (HSEP) in 1974. The HSEP was aimed to provide students with uniform learning experience and to close the achievement gap across schools by minimizing across-school variation in student quality, teacher quality, curriculums and facilities. The strong emphasis on equal treatment in education policy has been maintained until 2009. The
policy focus has shifted from uniformity to diversity afterward. Policymakers started to encourage competition among schools in 2010.\(^6\)

Under the HSEP, students were randomly assigned to academic high schools within school districts where they meet residency requirements.\(^7\) The random assignment made the distribution of student ability similar across schools within each district.\(^8\) As results, students within the same district had similar peers on average. The student assignment lottery covered academic high schools in ten school districts, including Districts 1-4, 6-11. High schools excluded from the random assignment were vocational high schools; selective high schools specialized in science, foreign languages, art, or sports; and academic high schools near the city center – mostly in District 5 and some in Districts 1, 2, 10, and 11.\(^9\)

Until choice-based assignment was introduced in 2010, academic high schools were subject to the lottery-based assignment regardless of their type – coed vs. single-sex or public vs. private. Each school district has co-ed, all-girls, and all-boys high schools. Single-sex schools tend to be older and are more likely to be private. This is partly because in the past high schools started as single-sex schools. The government has increased the number of coed schools by requiring since 1998 that all newly-opened public schools are coed.

Unlike in the US and many other countries, private academic high schools were not much different from public academic high schools in terms of educational environment under the HSEP. This is because high school curriculums and textbooks were standardized and regulated by the government. Schools had little discretion over which subjects to teach and which materials to cover in each subject. In addition, tuition costs for public and private high schools were similar as private schools were also heavily subsidized by the government. The biggest difference between public and private schools is on teacher status. Public school teachers are government employees who serve in

---

\(^6\)For more information on the equalization policy in Korean secondary schools and its impacts, see, for example, Kang (2007), Kang, Park, and Lee (2007), Hahn, Sung, and Back (2008), Kim, Lee, and Lee (2008), Lee (2012).

\(^7\)Seoul Metropolitan Office of Education (2008) documents that the random assignment of students was conducted within a commutable sub-area of each school district. Information on the boundaries of those commutable areas are not disclosed to public or for research purposes. Due to the lack of information on the exact random assignment boundary, students and their parents were unable to make any better predictions than researchers about school assignment outcomes.

\(^8\)According to Seoul Metropolitan Office of Education (2008), procedures of the random school assignment was as follows. Students were first classified into top, middle, and bottom groups based on their middle school records. Then, students in each group were randomly assigned to high schools within districts. The cutoffs of middle school records for classifying students into three groups are not available to researchers, students or parents.

\(^9\)In 2008, there were 295 high schools in 11 school districts of Seoul, which included 203 academic high schools, 78 vocational high schools, 12 specialized high schools, and two other high schools. Among the academic high schools, 40 were not subject to the assignment lottery and the rest were. The academic high schools outside the random assignment system accepted applications from students living in any school districts of Seoul mainly because those schools were located in business area near the city center with not enough residents.
one school for three to six years and are transferred to another in the same city, whereas private school teachers are hired by a specific school and usually stay there until their retirement.

When students and their families moved to another school district, the students were reassigned randomly to a school in the new district. Thus, school choice of families and student selection of schools had been strictly restricted until the city government reformed the student assignment system by introducing the choice-based assignment in 2010.

3 Data and Descriptive Statistics

3.1 Data

We use data on the CSAT scores and high school characteristics obtained from the Korean Ministry of Education and Korea Education and Research Information Service (KERIS), respectively. We link individual-level test scores and high school characteristics by matching school names.

The CSAT is the standardized test for college admissions in Korea. This test is developed, published, administered, and scored by the Ministry of Education of Korea. The test is offered once a year in November and taken by about 600,000 individuals including high school seniors, high school graduates, and GED holders. The CSAT scores together with high school GPA are the most important factors that determine whether a student is admitted to some college and to which college.

The CSAT scores on Korean are the main outcomes of educational production in this study. We also present results on CSAT English and Math scores in Appendix C. The estimation results for English scores are generally similar to those for Korean scores. The results on Math scores, however, need to be interpreted with caution. The fraction of students who were absent from the exam is more than ten times higher for Math (6.5%) than for Korean (0.03%) or English (0.5%). Furthermore, students had to select between two types – basic and advanced – of exam questions for the Math section. It is highly likely that a student’s choice on whether to take the Math exam or which type of Math exam to take was made endogenously. The CSAT scores on Korean,

---

10 The CSAT score data cover the entire population of CSAT takers in Korea, but contain no individual characteristics other than gender, whether the person was a high school student, and which high school (s)he was attending.
11 The CSAT consists of five major sections: Korean, Math, English, Sciences/Social Studies/Vocational Education, and Second Foreign Languages. We do not analyze scores on Sciences/Social Studies/Vocational Education or Second Foreign Languages because students selected different subordinate subjects in those sections and scores on different subjects were not directly comparable.
12 Advanced Math exam had three subordinates of Calculus, Probability and statistics, and Discrete mathematics. Thus, students who decided to take the advanced Math exam had to further select among the three types of questions.
English, and Math were standardized to have a mean of 100 points and a standard deviation of 20 points.

Data on high school characteristics come from the schoolinfo database maintained by KERIS. The database contains information on all primary and secondary schools in the entire nation, including school types (public vs. private, single-sex vs. co-ed), number of students by gender, class size, number of teachers by gender, number of students on subsidized lunch, development fund spending, etc.\footnote{Information for the most recent four years are publicly available from \url{http://schoolinfo.go.kr}.}

We restrict our analysis to high school seniors in 2008 and 2009, who were randomly assigned to academic high schools within school districts 3, 4, 6-9 of Seoul.\footnote{We focus on 2008 and 2009 since school characteristics are available from 2008 and the HSEP was effectively abolished in 2010.} All academic high schools within each of the six districts participated in the lottery-based assignment. The analysis sample includes 55 coed, 34 all-girls, and 38 all-boys high schools, which are either public or private.\footnote{For the 2008 and 2009 cohorts of seniors, the assignment was conducted in February 2006 and 2007, respectively, as the school year begins in early March and ends in mid-February in Korea.} The analysis sample includes 50,809 students in 2008 and 58,905 in 2009 covering about 54% of CSAT takers who were high school seniors in Seoul.\footnote{About 26% of CSAT takers lived in Seoul during the analysis period. The analysis sample covers about 36% of CSAT takers in Seoul. Our sample does not include high school graduates, who were retakers, consisting of about 34% of the CSAT takers in Seoul. We further drop those attending vocational high schools, selective high schools, or regular high schools not covered by the lottery-based student assignment, consisting about 30% of the high school seniors taking the CSAT in Seoul.}

### 3.2 Summary Statistics

Figure 1 shows the histogram of the school-level average Korean scores by school district. Similar graphs for English and Math CSAT scores are provided in Appendix C. Each histogram in Figure 1 shows the number of schools within each five-point range of school-level average test scores. The two years of data are pooled, so the unit of observation is school-year.\footnote{The distribution looks similar when we plot by year.} As we analyze boys and girls separately, the test score distributions are shown by gender.\footnote{Note that co-ed schools are included both for boys and for girls.} We observe that the average test scores are lower for boys than for girls.

The school average test scores vary substantially across districts as well as within each district. Based on average test scores, the six school districts can be grouped into three: high (District 8), middle (Districts 4, 6, and 7), and low (Districts 3 and 9). The three groups roughly correspond to grouping based on average income: wealthy, middle, and poor. This pattern shows evidence
that students sort endogenously across school districts. Within district dispersions are the largest in District 7. In all districts, English and Math scores are slightly more dispersed across schools compared to Korean scores.

Table 1 reports means and standard deviations of school-level variables that are used in our empirical analysis. We can see that school characteristics are heterogeneous within and across school districts. Thus, it is important to notice that each school within a district is a multi-dimensional treatment.

### 3.3 Random Assignment Check

To test for randomization of high school assignment, we use data from the Korean Education Longitudinal Study (KELS). Note that pretreatment individual characteristics are unavailable from our main data set constructed by merging the students’ CSAT scores and high school characteristics. The KELS provides data on learning experiences of a nationally representative sample of seventh-graders who were first surveyed in 2005 and followed up every year until they become thirty years old. Similar to the National Education Longitudinal Study of 1988 (NELS:1998) in the US, data are collected from students and their parents, teachers and high school principals.

We analyze the KELS sample of students who became 10th graders and entered high school in 2008. By linking the first, third and fourth waves of the KELS, we construct a data set containing baseline characteristics, including parental education, household income, expenditures on private out-of-school education, and 9th-grade standardized test scores, before students entered high school. The treatment variable is whether a student attended a single-sex high school in 2008. We check whether students’ baseline characteristics are balanced between single-sex and co-educational high schools. We conduct this exercise for students attending high schools in areas under the HSEP as the KELS data do not have information on school district or school names. From a logit regression of the treatment indicator on the baseline characteristics, chi-square test of all coefficients equalling zero yields 8.353 and the associated p-value from a chi-squared distribution is 0.499. When we repeat the balance check for those attending high schools not subject to the HSEP, we analyze boys and girls together and allow for different intercepts by including a gender dummy.
HSEP, students’ baseline characteristics are significantly different between the two school types at any conventional level (chi-square test statistic = 23.83, p-value = 0.00457).

4 The Potential Outcome Framework

In this section, we introduce a potential outcome model of education production. The potential outcome model is used to derive a school-level education production function in section 5.

We use the following notations to indicate that our data include not a sample but a population of students and schools. Let \( I = \{1, ..., N_I\} \) be the population of high school senior students in the six school districts of Seoul and years \( T = \{2008, 2009\} \). Let \( D = \{1, ..., N_D\} \) denote the collection of the school districts in Seoul. Let \( S_d = \{1, ..., N_S(d)\} \) denote the collection of high schools in district \( d \).

We consider the following potential outcomes of individual \( i \) at school \( s \in S_d \) in district \( d \in D \) and year \( t \in T \). We assume a linear education production function with random coefficients:

\[
Y_i(s, d, t) = K_s' \alpha_i + L_{s,t} \beta_i + v_s \gamma_i + u_{s,t} \delta_i + c_d \eta_i. \tag{1}
\]

The (potential) outcome, \( Y_i(s, d, t) \), is the (potential) CSAT score of student \( i \) if she attends school \( s \) in district \( d \) and year \( t \). The variables \( K_s \) and \( L_{s,t} \) denote observed school inputs. \( K_s \) is a vector of time-invariant school characteristics and \( L_{s,t} \) is a vector of time-varying school characteristics. The components \( v_s \) and \( u_{s,t} \) are unobserved time-invariant and time-varying school inputs, respectively. The variable \( c_d \) represents unobserved district characteristics. The coefficients

\[
\theta_i = (\alpha_i, \beta_i, \gamma_i, \delta_i, \eta_i)
\]

represent heterogeneous individual responses to school inputs and district characteristics. The potential outcome model in equation (1) assumes that the potential outcome is determined by the interaction of \( \theta_i \), the heterogeneous individual abilities, and

\[
Z_{s,d,t} = (K_s, L_{s,t}, v_s, u_{s,t}, c_d),
\]

the school inputs and the district characteristics.\(^{22}\)

\(^{22}\)We do not include a time effect in (1) because the CSAT test scores are normalized.
The potential outcome model in (1) is motivated by the textbook potential outcome model defining \( Y_{1i} - Y_{0i} \) as the treatment effect for individual \( i \) when the treatment is binary. The textbook potential outcome model is nonparametric and allows for treatment effect heterogeneity. We assume the linear and interactive functional form in (1) to take into account the multi-dimensional feature of the treatment. It is convenient to assume linearity when we analyze the effect of a subset of the multiple treatment components. Other than the functional form assumption, our potential outcome model can be considered a natural extension of the textbook potential outcome model with binary treatment.

In our data, we observe \((Y_i, S_i, D_i, T_i)\) at the individual level and \((K_s, L_{s,t})\) at the school level. The observed outcome, i.e. the CSAT score, of individual \( i \) is
\[
Y_i = Y_i(S_i, D_i, T_i).
\]

\( S_i \) denotes the school that individual \( i \) attends. \( T_i \) denotes the senior year of individual \( i \), which also represents birth cohort. \( D_i \) is the district where individual \( i \) choose to live. \((K_s, L_{s,t})\) are time-invariant and time-varying school characteristics.

**Assumption 1** [Sampling Scheme] All the school and district level variables, \( Z = \{Z_{s,d,t} : s \in S, t \in T, d \in D\} \), are fixed. The individual abilities, \( \theta = (\theta_1', ..., \theta_N')' \), and individuals’ district choices, \( D = (D_1, ..., D_N)' \), are fixed. Given this fixed set up, for each \( i \) with \( D_i = d \), we assume that \((S_i, T_i)\) are randomly drawn from a discrete distribution with probabilities, \[ \mathbb{P}\{S_i = s, T_i = t \mid D_i = d\} = w_{s,t,d}. \]

Assumption 1 describes the sampling scheme of the observed data. We assume school and district characteristics \( Z \), (unobserved) individual characteristics \( \theta \), and individuals’ district choices \( D \) are predetermined as the data used in this study are not from a randomly drawn sample but include a population of schools and students[23]. The random components of the model are \( S_i \) and \( T_i \). The randomness of \( S_i \) accounts that high school assignment is randomly determined within each school district. The randomness of \( T_i \) comes from the fact that birth year can be considered random[24].

Based on the sampling scheme, we introduce non-random objects and random variables that are frequently used in the subsequent analysis. Let \( N_I(s, d, t) = \sum_{i \in I} \mathbb{I}\{S_i = s, D_i = d, T_i = t\} \) denote

[24]Mention about the possible influence of zodiac in Asia?
the number of senior students at school \( s \) in district \( d \) and year \( t \). \( N_I(s,d) = \sum_{t \in T} N_I(s,d,t) \) is defined as the number of senior students at school \( s \) of district \( d \) in all years. Similarly, \( N_I(d) = \sum_{s \in S_d} N_I(s,d) \) is defined as the number of senior students in district \( d \) in all years, and \( N_I = \sum_{d \in D} N_I(d) \) as the total number of students in Seoul over the entire time period of analysis. Under the sampling scheme in Assumption 1, as the individuals’ district choices, \( D = (D_1, ..., D_N)’ \), are fixed, the fraction of students in each district, \( \frac{N_I(d)}{N_I} \), is also fixed and not a random variable. We treat the proportion of students in each school district as the district choice probability that has a fixed value, i.e. \( \frac{N_I(d)}{N_I} = P\{D_i = d\} = w_d \). On the contrary, the fraction of students in each school among those in district \( d \), \( \frac{N_I(s,d)}{N_I(d)} \), and the fraction of students in each cohort within a school, \( \frac{N_I(s,d,t)}{N_I(s,d)} \), are random variables since \((S_i,T_i)\) are randomly drawn conditioning on \( D_i \). We use expressions, \( \hat{w}_s^d = \frac{N_I(s,d)}{N_I(d)} \) and \( \hat{w}_s^t = \frac{N_I(s,d,t)}{N_I(s,d)} \) for the two random variables, as they are consistent estimators of \( P\{S_i = s \mid D_i = d\} = w_s^d \) and \( P\{T_i = t \mid S_i = s, D_i = d\} = w_t^s,d \), respectively.

5 The Average Partial Effects of School Inputs

The parameters of interest are the APE of school inputs, \( K_s \) and \( L_{s,t} \):

\[
\alpha = \mathbb{E}[\alpha_i] \quad \text{and} \quad \beta = \mathbb{E}[\beta_i].
\]

**Assumption 2** For all \( d \in D \), (i) the conditional distribution of \((S_i,T_i)\) given \( D_i = d \) does not depend on \((Z,\theta)\) and (ii) \( S_i \) and \( T_i \) are mutually independent conditioning on \( D_i = d \).

Assumption 2 implies (i) random assignment of students across schools within districts: for all \((s,d,t)\),

\[
\mathbb{E}[(\alpha_i,\beta_i,\gamma_i,\delta_i,\eta_i) \mid S_i = s, D_i = d, T_i = t] = \mathbb{E}[(\alpha_i,\beta_i,\gamma_i,\delta_i,\eta_i) \mid D_i = d, T_i = t]; \quad (3)
\]

and (ii) no cohort effects: for all \((d,t)\),

\[
\mathbb{E}[(\alpha_i,\beta_i,\gamma_i,\delta_i,\eta_i) \mid D_i = d, T_i = t] = \mathbb{E}[(\alpha_i,\beta_i,\gamma_i,\delta_i,\eta_i) \mid D_i = d]. \quad (4)
\]

The average productivity of students in district \( d \) is denoted by

\[
(\alpha_d,\beta_d,\gamma_d,\delta_d,\eta_d) = \mathbb{E}[(\alpha_i,\beta_i,\gamma_i,\delta_i,\eta_i) \mid D_i = d].
\]
If individual district choice is independent of individual productivity, the average productivity should be identical across districts, i.e. $\alpha = \alpha_d$ and $\beta = \beta_d$. In the case we study, however, the average productivity of school inputs may differ by school district because students are likely to be sorted endogenously across districts. In the case we study, however, the average productivity of school inputs may differ by school district because students are likely to select a school district where they can maximize their academic outcomes based on the average and variance of school quality in each district. By allowing the average productivity to vary across districts, we explicitly take into account the possibility that district selection is potentially endogenous.

The APE parameters of interest, $(\alpha, \beta)$, are, by the iterative expectations, expressed as the weighted average of $(\alpha_d, \beta_d)$ as follows:

$$(\alpha, \beta) = \sum_{d \in D} (\alpha_d, \beta_d) \mathbb{P} \{D_i = d\}.$$ 

To estimate $(\alpha, \beta)$, we first estimate $(\alpha_d, \beta_d)$ for each district $d$. Next, we aggregate the estimates $(\hat{\alpha}_d, \hat{\beta}_d)$ weighted by $\mathbb{P} \{D_i = d\}$. Under the sampling scheme in Assumption 1, $\mathbb{P} \{D_i = d\} = \frac{N_I(d)}{N} = w_d$. Therefore,

$$(\hat{\alpha}, \hat{\beta}) = \sum_{d \in D} (\hat{\alpha}_d, \hat{\beta}_d) w_d.$$ 

### 5.1 School-level Education Production Function Implied by the Potential Outcome Model

In this section, we describe the school-level education production function implied by the individual potential outcome in [1] together with the random assignment and no-cohort effect conditions in Assumption 2.

We define $Y_{s,d,t}$ as an aggregate test scores of students at school $s$ in district $d$ and year $t$:

$$Y_{s,d,t} = \frac{\sum_{i \in I} Y_i \mathbb{I} \{S_i = s, D_i = d, T_i = t\}}{\sum_{i \in I} \mathbb{I} \{S_i = s, D_i = d, T_i = t\}}.$$ (5)

Similarly, $\theta_{s,d,t} = (\alpha_{s,d,t}, \beta_{s,d,t}, \gamma_{s,d,t}, \delta_{s,d,t}, \eta_{s,d,t})$ is aggregate productivity of students attending school $s$ in district $d$ and year $t$:

$$\theta_{s,d,t} = \frac{\sum_{i \in I} \theta_i \mathbb{I} \{S_i = s, D_i = d, T_i = t\}}{\sum_{i \in I} \mathbb{I} \{S_i = s, D_i = d, T_i = t\}}.$$
Given the potential outcome (1) and the observed outcome (2), (5) becomes

\[ Y_{s,d,t} = \frac{\sum_{i \in T} Y_i \mathbb{1}\{S_i = s, D_i = d, T_i = t\}}{\sum_{i \in T} \mathbb{1}\{S_i = s, D_i = d, T_i = t\}} \]

\[ = \frac{\sum_{i \in T} Y_i (s, d, t) \mathbb{1}\{S_i = s, D_i = d, T_i = t\}}{\sum_{i \in T} \mathbb{1}\{S_i = s, D_i = d, T_i = t\}} \]

\[ = K_s' \alpha_d + L'_{s,t} \beta_d + v_s \gamma_d + u_{s,t} \delta_d + c_d \eta_d + E_{s,d,t}. \]

Here,

\[ E_{s,d,t} = K'_s (\alpha_{s,d,t} - \alpha_d) + L'_{s,t} (\beta_{s,d,t} - \beta_d) + v_s (\gamma_{s,d,t} - \gamma_d) + u_{s,t} (\delta_{s,d,t} - \delta_d) + c_d (\eta_{s,d,t} - \eta_d) \]

is the aggregation error of individual productivity. The random assignment and no-cohort effect conditions in Assumption 2 implies

\[ \mathbb{E} [E_{s,d,t}|S_i = s, D_i = d, T_i = t] = 0. \]

For notational convenience, we will also use the simplified subscripts as follows:

\[ V_{s,d} = v_s \gamma_d, U_{s,d,t} = u_{s,t} \delta_d, \text{ and } C_d = c_d \eta_d. \]

Then, we can write the aggregate outcome of school \( s \) in district \( d \) and year \( t \) as a function of school inputs, district characteristics, and the error term:

\[ Y_{s,d,t} = K'_{s,d} \alpha_d + L'_{s,d,t} \beta_d + V_{s,d} + U_{s,d,t} + C_d + E_{s,d,t}, \tag{6} \]

which is a school-level education production function.\(^{25}\) Note that we derive the school-level education production function by aggregating the individual outcomes. This procedure is similar in spirit to the derivation of the market demand function as an aggregation over individual choices (Berry, Levinsohn, and Pakes, 1995).

The school-level education production function (6) has a unique feature that the coefficients \( \alpha_d \) and \( \beta_d \) of the observed school inputs are constant across schools within each school district and over time. The conditions indicating the random school assignment within each district and no cohort effects are key for the constant productivity. If students are free to choose their own schools instead

\(^{25}\)Here we use notations \( K_{s,d} \) and \( L_{s,d,t} \) for observed school inputs instead of \( K_s \) and \( L_{s,t} \) to emphasize that school level variables vary across school districts.
of districts, the average productivity of school inputs would become school specific and be correlated with the observed school characteristics \((K_{s,d}, L_{s,d,t})\). Even under self-selection into schools, school input effects can be constant across schools if individual productivity is homogeneous, i.e. \((\alpha_i, \beta_i) = (\alpha, \beta)\). The no heterogeneity assumption is, however, very restrictive and unrealistic. Under self-selection into schools and individual heterogeneity, the school-level education production function becomes a correlated random coefficient model, and identification of the school input coefficients, \((\alpha_s, \beta_s)\), using school level data is challenging. In our setup, however, district specific coefficients, \((\alpha_d, \beta_d)\), are the district averages of \((\alpha_i, \beta_i)\).

Previous studies on school quality effect exploiting the within-district random school assignment restrict coefficients on school inputs to be constant across districts and often include district fixed effects to capture district specific unobserved factors. When students are self-selected into school districts, the common coefficients on school inputs across school districts can be justified only when there is no heterogeneity in individual productivity, which is implausible. Some papers in the literature implicitly allow for heterogeneous school input effects by conducting a subgroup analysis although they do not allow school input effects to vary across school districts. A common example is to show that the effect of school inputs on academic performance is different between high and low achievers. This common practice in the empirical literature is self-contradictory as the no heterogeneity assumption underlies the common coefficients on school inputs across districts. Although previous studies believe that the common coefficients represent the average effect of school inputs, it is not true when individuals with heterogeneous productivity are endogenously selected into school districts. In Section 6.2, we show that the school production function can be misspecified when the coefficients on school inputs are restricted to be constant instead of district specific, and discuss how the misspecified model can lead to biased estimates of school input effects.

## 5.2 Estimation of \((\alpha_d, \beta_d)\)

In this section, we explain the two step estimation procedure to measure the effect of school inputs in each school district.

The following notations for weighted averages and mean deviations are used when we describe the estimation procedure. We define a school or district average of a time-varying school-level
variable in district $d$, such as $A_{s,d,t}$, as:

$$\bar{A}_{s,d} = \sum_{t \in T} \hat{w}_{s,d} A_{s,d,t} \text{ and } \bar{A}_{s,d} = \sum_{s \in S_d} \hat{w}_{d} A_{s,d}. $$

Accordingly, a school-level variable deviated from its school or district average can be expressed as:

$$\tilde{A}_{s,d,t} = A_{s,d,t} - \bar{A}_{s,d} \text{ and } \tilde{A}_{s,d} = \bar{A}_{s,d} - \bar{A}_{d}. $$

For a time-invariant school-level variable in district $d$, such as $B_{s,d}$, a district average and a mean deviation from the district average are express as follows:

$$\bar{B}_{s,d} = \sum_{s \in S_d} \hat{w}_{d} B_{s} \text{ and } \bar{B}_{s} = B_{s} - \bar{B}_{d}. $$

To estimate $(\alpha_d, \beta_d)$, the school input effects in district $d$, we implement the following two step estimation method using school-level panel data in each district. We first estimate $\beta_d$, the effect of time-varying school inputs (step 1), and then estimate $\alpha_d$, the effect of time-invariant school inputs based on the estimates from the first step (step 2).

**Step 1 (Estimation of $\beta_d$)** We start from the school-level education production function in (6),

$$Y_{s,d,t} = K_{s,d}' \alpha_d + L_{s,d,t}' \beta_d + V_{s,d} + U_{s,d,t} + C_d + E_{s,d,t}. $$

Note that we have two-period panel data in school level. By taking the weighted average of the school-level education production function over time, we have

$$\bar{Y}_{s,d} = K_{s,d}' \alpha_d + \bar{L}_{s,d} \beta_d + \bar{V}_{s,d} + \bar{U}_{s,d} + C_d + \bar{E}_{s,d}. $$

(7)

Next, subtracting (7) from (6) yields

$$\tilde{Y}_{s,d,t} = \tilde{L}_{s,d,t} \beta_d + \tilde{U}_{s,d,t} + \tilde{E}_{s,d,t}. $$

where $\tilde{Y}_{s,d,t} = Y_{s,d,t} - \bar{Y}_{s,d}$, $\tilde{L}_{s,d,t} = L_{s,d,t} - \bar{L}_{s,d}$, $\tilde{U}_{s,d,t} = U_{s,d,t} - \bar{U}_{s,d}$, and $\tilde{E}_{s,d,t} = E_{s,d,t} - \bar{E}_{s,d}$. 

15
We estimate $\beta_d$ by a pooled weighted least squares (WLS) regression of $\tilde{Y}_{s,d,t}$ on $\tilde{L}_{s,d,t}$:

$$
\hat{\beta}_d = \left( \sum_{s \in S_d} \hat{w}_s^d \sum_{t \in T} \hat{w}_s^t L_{s,d,t} \tilde{L}_{s,d,t}' \right)^{-1} \sum_{s \in S_d} \hat{w}_s^d \sum_{t \in T} \hat{w}_s^t L_{s,d,t} \tilde{Y}_{s,d,t}.
$$

(8)

Step 2 (Estimation of $\alpha_d$) We estimate $\alpha_d$ by a WLS regression of $\bar{Y}_{s,d,\bullet} - \bar{L}_{s,d,\bullet} \hat{\beta}_d$ on $K_{s,d}$ and a constant term:

$$
\hat{\alpha}_d = \left( \sum_{s \in S_d} \hat{w}_s^d \hat{K}_{s,d} \hat{K}_{s,d}' \right)^{-1} \left( \sum_{s \in S_d} \hat{w}_s^d \hat{K}_{s,d} \left( \bar{Y}_{s,d,\bullet} - \bar{L}_{s,d,\bullet} \hat{\beta}_d \right) \right).
$$

(9)

Consistency of the estimator relies on the following large sample approximations of the random variables introduced in section 4. If $N_I(d)$ is large, we have

$$
\hat{w}_{s,d} = \frac{N_I(s,d)}{N_I(d)} \sum_{i \in I} \sum_{S_i = s, D_i = d, T_i = t} \frac{1}{P \{ T_i = t | S_i = s, D_i = d \}} = w_{s,d}^d,
$$

$$
\hat{w}_{s,d} = \frac{N_I(s,d)}{N_I(d)} \sum_{i \in I} \sum_{D_i = d} \frac{1}{P \{ S_i = s | D_i = d \}} = w_{s,d}^d.
$$

Next, the following two assumptions are required for the identification of $\alpha_d$ and $\beta_d$.

**Assumption 3** For all $s \in S_d$, $t \in T$, we assume that (a)

$$
\sum_{s \in S_d} \hat{w}_s^d \sum_{t \in T} \hat{w}_s^t \left( L_{s,d,t} - \sum_{t \in T} \hat{w}_s^t L_{s,d,t} \right) U_{s,d,t} = 0,
$$

and (b)

$$
\sum_{s \in S_d} \hat{w}_s^d \left( K_s - \sum_{s \in S_d} \hat{w}_s^d K_s \right) \left( V_{s,d} + \sum_{t \in T} \hat{w}_s^t U_{s,d,t} \right) = 0.
$$

**Assumption 4** [Rank Conditions] We assume that

$$
\sum_{s \in S_d} \hat{w}_s^d \sum_{t \in T} \hat{w}_s^t \left( L_{s,d,t} - \sum_{t \in T} \hat{w}_s^t L_{s,d,t} \right) \left( L_{s,d,t} - \sum_{t \in T} \hat{w}_s^t L_{s,d,t} \right)' > 0
$$

and

$$
\sum_{s \in S_d} \hat{w}_s^d \left( K_s - \sum_{s \in S_d} \hat{w}_s^d K_s \right) \left( K_s - \sum_{s \in S_d} \hat{w}_s^d K_s \right)' > 0.
$$
Assumption 3 (a) is the standard identification assumption for a within estimator that the time-varying components of school inputs is uncorrelated with the time-varying unobserved school characteristics. This indicates that the time-invariant components of the time-varying school inputs are allowed to be endogenous and can be removed by school fixed effects. Assumption 3 (b) states that time-invariant school inputs are not correlated with unobserved school characteristics. Assumption 4 describes the rank conditions of no perfect multicollinearity.

Under Assumptions 2, 3 and 4, \( (\hat{\alpha}_d, \hat{\beta}_d) \) are an unbiased estimator of \( (\alpha_d, \beta_d) \). We provide empirical evidence in Appendix that Assumption 2 can be justified based on the institutional setting. Assumption 3 (a) would be violated if there is school-specific time trend in the time-varying variable of interest – class size in this study – within each school district. This seems unlikely because the target class size determined by Seoul Metropolitan Office of Education every year is uniform over all schools in each school district. Although students leaving high schools to transfer or to drop out is uncommon in Korea, we control for time-varying school inputs, including fraction of students on subsidized lunch, annual development fund spending per student, and the fraction of female teachers, whose over-time variation can be correlated with that of class size.

We are especially concerned about the violation of Assumption 3 (b). Single-sex indicator, the time-invariant school input of interest in this study, can be correlated with unobserved school characteristics that are time fixed. For example, unobserved teacher quality is a key determinant of school quality and has a large impact on students’ academic achievement. (Rivkin, Hanushek, and Kain 2005) Unobserved teacher quality is likely to be affected by public-private status (through different teacher hiring processes) and age of the school (representing the power of alumni networks), which also influences single-sex status. If we can find a valid instrument that is uncorrelated with unobserved school characteristics and affects students’ academic outcomes only through single-sex status, the two-step estimation procedure employed in this study would be very similar to that described in Hausman and Taylor (1981). Since no valid instrument is available, we take a control function approach instead including private indicator, age of school, and the interaction between the two as covariates.

In the two-step estimation procedure described above, \( (\hat{\alpha}_d, \hat{\beta}_d) \) are defined as the WLS estimators applied to the school-level panel data in district \( d \). An equivalent way of implementing the

\(^{26}\)The number of students assigned to each school equals the target class size multiplied by the number of homeroom classrooms for each cohort in the school.
WLS estimation is to estimate the following regression equation using individual-level data:

\[ Y_i = K'_S \alpha_d + L'_S, T_i \beta_d + \sum_{s \in S_d} \gamma_d v_s I \{ S_i = s \} + \xi_i \]

where \( \xi_i = K'_S (\alpha_i - \alpha_d) + L'_S, T_i (\beta_i - \beta_d) + v_{S_i} (\gamma_i - \gamma_d) + u_{S_i, T_i} \delta_i + c_D \eta_i. \)

Similar to the school-level WLS estimation, the individual-level OLS estimation is also implemented in two-steps to estimate the coefficients on both time-varying and time-invariant school characteristics. First, \( \hat{\beta}_d \) can be estimated from OLS regression of \( Y_i \) on time-varying school inputs, \( L_{S_i, T_i} \), and school indicator variables, \( I \{ S_i = s \}, s \in S_d \). In this case, \( \hat{\beta}_d \) is a within estimator from individual-level regression with school fixed effects. Next, individual-level OLS regression of \( Y_i - L'_S, T_i \hat{\beta}_d \) on \( K_S \) and a constant term yields \( \hat{\alpha}_d \). In Appendix A we show that the OLS estimation using individual-level data with school fixed effects is equivalent to the WLS estimation using school aggregate data when the fraction of individuals, \( \hat{w}_{s,d} = \frac{N_I(s,d)}{N_I(d)} \) and \( \hat{w}_{s,t} = \frac{N_I(s,d,t)}{N_I(s,d)} \), are used as the weights.

If we use alternative weights for the WLS estimation – \( \hat{w}_{s,d} = \frac{1}{N_S(d)} \) and \( \hat{w}_{s,t} = \frac{1}{T} \), constructed based on the number of schools instead of the number of students –, the school-level education production function (6) takes a panel linear regression form using school aggregate data within each district. In this case, we can obtain a within estimator of \( \beta_d \) from a regression of \( Y_{s,d,t} \) on \( L_{s,d,t} \) and school fixed effects, and an OLS estimator of \( \alpha_d \) from a regression of \( \frac{1}{T} \sum_{t \in T} (Y_{s,d,t} - L'_{s,d,t} \hat{\beta}_d) \) on \( K_{s,d} \) and a constant term. Choi, Moon, and Ridder (2014) provides details on the estimation procedure and results from school aggregate data. Comparing results from the two types of weights provides a specification check. We obtain similar estimation results from both types of weights.

6 Estimation Results

6.1 The Average Partial Effect Estimates

Table 2 presents the estimated effect of single-sex education and class size on Korean CSAT scores. The estimates for English and Math scores are reported in Table C1. Regressions also include other time-varying and time-invariant covariates that serve as control variables and are possibly correlated with unobserved school characteristics. Time-varying controls include the fraction of students receiving free or reduced price lunch, annual development fund spending per student, and
the fraction of female teachers. Time-invariant controls include a private school indicator, age of the school in 2008, and the interaction between the two variables. We analyze boys and girls separately.

The estimates in the first six rows are district specific coefficients including the selection effects. We observe that the effect of single-sex education varies substantially across school districts. For boys, the estimated effect ranges from less than one tenth of a standard deviation in District 6 to nearly a half of a standard deviation in District 3 on Korean scores. Similarly, the estimated effect on English scores ranges from no effect in District 6 and 9 to over 0.6 standard deviations in District 3 for boys. For girls, the estimated effect is lower than for boys and even goes negative in District 7. The class size effects are near zero (or insignificant) and generally negative in all districts. The exceptions are the positive effects on English scores for girls in five districts.

The heterogeneous effects imply that endogenous sorting of individuals across districts may play an important role. To understand the mechanism of sorting, we would need more information on individual characteristics from which we could infer how school characteristics interact with individual preference and productivity.

The estimated APE of school inputs are shown in the last row of each table. We use the number of CSAT takers in each district to construct the weighted average. The results on Korean CSAT scores show that single-sex schools and a smaller class size lead to higher academic performance for both boys and girls. The effect of single sex schools is much larger for boys than for girls – about a quarter of a standard deviation for boys and 0.04 standard deviations for girls. The effect of a smaller class size is positive and statistically significant but very small for both boys and girls. A decrease in class size by one standard deviation is associated with an increase in the Korean CSAT scores by 0.02 standard deviations for boys and by 0.01 standard deviations for girls. The APE on English scores is similar to the findings on Korean scores for boys. However, for girls, there is no effect of single sex education and the sign of class size effect is the opposite on English scores.

6.2 Comparison with Estimates from Alternative Estimation Methods

We propose in sections 4 and 5 a novel estimation method that addresses the issue that endogenous sorting of students across school districts and heterogeneous productivity may bias the estimates of school input effects. To gauge the contribution of our estimator, we attempt to assess the bias.

27 The APE estimates change little when we use the number of seniors or the cohort size at random assignment as weights.
in other estimates of school input effects by comparing our estimates to the estimates that ignore endogenous sorting and heterogeneous productivity.

We start by comparing our estimates to the mean-difference of outcomes across schools in treatment and control groups, which we call a naive estimator of the average treatment effect (ATE). Suppose that $K_s$ is a scalar binary treatment variable. In our application, it may be a type of school – single sex versus coeducational, for example. Given that students are randomly assigned to different types of schools within each school district, one might be interested in the effect of a certain type of school. The difference of the average observed outcome $Y_i$ between the two types of schools estimates the following parameter:

$$\alpha^*_d = E[Y_i|K_{S_i} = 1, D_i = d] - E[Y_i|K_{S_i} = 0, D_i = d].$$

When the observed outcome $Y_i$ equals $Y_i(S_i, D_i, T_i)$ in (2) and Assumptions 2 holds, the naive ATE estimator estimates

$$\alpha_d^* = E[Y_i|K_{S_i} = 1, D_i = d] - E[Y_i|K_{S_i} = 0, D_i = d] = \alpha_d + \sum_{s \in S_d, K_s = 1, t \in T} (L_{s,t} \beta_d + v_s \gamma_d + u_{s,t} \delta_d) \frac{N_I(s, d, t)}{\sum_{s \in S_d, K_s = 1} N_I(s, d)} - \sum_{s \in S_d, K_s = 0, t \in T} (L_{s,t} \beta_d + v_s \gamma_d + u_{s,t} \delta_d) \frac{N_I(s, d, t)}{\sum_{s \in S_d, K_s = 0} N_I(s, d)}.$$

From this, we deduce that in general the two parameters $\alpha_d$ and $\alpha_d^*$ are different:

$$\alpha_d \neq \alpha_d^*.$$

The difference between $\alpha_d$ and $\alpha_d^*$ can be see

Table 3 presents the estimates of $\alpha_d$ and $\alpha_d^*$ on Korean CSAT scores. The difference between $\alpha_d$ and $\alpha_d^*$ is due to the difference in the school specific resources and teachers, including the number of student spots, between the two types of schools. This happens because even though students are randomly assigned to schools within each district, the school resources and inputs are not. This implies the treatment includes not only school type but also other observable and unobservable characteristics of the school such as class size, financial status, and alumni reputations.

Next, we compare our estimates to estimates from linear regressions with district fixed effects assuming homogeneous school input coefficients that are often used in the literature. Suppose that
one runs a pooled OLS regression with district fixed effects using data on all schools in Seoul as follows:

\[ Y_{s,d,t} = C_d + K'_{s,d}\alpha + L'_{s,d,t}\beta + \chi_{s,d,t}, \] (10)

where

\[ \chi_{s,d,t} = U_{s,d,t} + V_{s,d} + K'_{s,d} (\alpha_d - \alpha) + L'_{s,d,t} (\beta_d - \beta). \]

The fixed effect estimates of \( \alpha \) and \( \beta \) would be biased under our framework. The bias arises from the omitted school fixed effects that can be correlated with school inputs \((K_{s,d}, L_{s,d,t})\). Another source of the bias is endogenous district choice. When students’ district choice \( D_i = d \) is related with the variation of school inputs in district \( d \), such as \( \sum_{s \in S_d} (K_{s,d} - \bar{K}_{\cdot,d}) (K_{s,d} - \bar{K}_{\cdot,d})' \) and \( \sum_{s \in S_d} (L_{s,d} - \bar{L}_{\cdot,d}) (L_{s,d} - \bar{L}_{\cdot,d})' \), the fixed effect estimates can be biased.

Table 4 shows the substantial difference between our estimates and estimates from the specification with district fixed effects and homogeneous school input coefficients. When endogenous sorting and heterogeneous productivity are ignored, the effect of single-sex education is substantially underestimated (for boys) or the estimated effect becomes marginally significant (for girls). The district fixed effect specification is unable to capture the negative effect of larger class size.

7 Conclusion

This paper estimates the average partial effects (APEs) of single-sex schooling and school resources exploiting within-district random assignment of students into high schools in Seoul, Korea. We derive a school-level education production function by aggregating individuals’ potential outcomes. We propose a two-step method to estimate the school-level education production function and the APEs of school inputs. The estimation results show that the coefficient estimates of the school-level education production function substantially differ across districts due to self-selection of students into school districts and individual heterogeneity in productivity. We find a substantial positive effect of single sex schools on academic achievement among male students whereas the effect is much smaller for female students. The effect of smaller class size is positive but very small for both male and female students. We also show that the APEs can be inconsistently estimated if endogenous district choice and heterogeneous ability are ignored.
References


Figure 1: School-level Average Korean CSAT Scores by District

A. Boys

<table>
<thead>
<tr>
<th>District</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Nambu</td>
<td>91.9</td>
<td>3.7</td>
</tr>
<tr>
<td>4. Bukbu</td>
<td>97.6</td>
<td>3.3</td>
</tr>
<tr>
<td>6. Gangdong</td>
<td>98.2</td>
<td>2.1</td>
</tr>
<tr>
<td>7. Gangseo</td>
<td>98.4</td>
<td>4.7</td>
</tr>
<tr>
<td>8. Gangnam</td>
<td>102.8</td>
<td>3.1</td>
</tr>
<tr>
<td>9. Dongjak</td>
<td>94.2</td>
<td>2.4</td>
</tr>
</tbody>
</table>

B. Girls

<table>
<thead>
<tr>
<th>District</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Nambu</td>
<td>97.9</td>
<td>3.7</td>
</tr>
<tr>
<td>4. Bukbu</td>
<td>102.8</td>
<td>3.2</td>
</tr>
<tr>
<td>6. Gangdong</td>
<td>103.0</td>
<td>2.4</td>
</tr>
<tr>
<td>7. Gangseo</td>
<td>103.0</td>
<td>4.3</td>
</tr>
<tr>
<td>8. Gangnam</td>
<td>106.8</td>
<td>3.1</td>
</tr>
<tr>
<td>9. Dongjak</td>
<td>100.2</td>
<td>2.7</td>
</tr>
</tbody>
</table>

Notes. Units are school×year.
<table>
<thead>
<tr>
<th>A. Boys</th>
<th>District 3</th>
<th>District 4</th>
<th>District 6</th>
<th>District 7</th>
<th>District 8</th>
<th>District 9</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single-sex schools (%)</td>
<td>28.6</td>
<td>35.3</td>
<td>35.7</td>
<td>52.9</td>
<td>47.4</td>
<td>41.7</td>
<td>40.9</td>
</tr>
<tr>
<td>Private schools (%)</td>
<td>21.4</td>
<td>41.2</td>
<td>57.1</td>
<td>70.6</td>
<td>52.6</td>
<td>50.0</td>
<td>49.5</td>
</tr>
<tr>
<td>Class size</td>
<td>35.9 [2.4]</td>
<td>35.0 [1.9]</td>
<td>36.3 [2.7]</td>
<td>35.1 [2.6]</td>
<td>34.3 [2.5]</td>
<td>34.7 [2.2]</td>
<td>35.2</td>
</tr>
<tr>
<td>Female teachers (%)</td>
<td>49.7 [18.0]</td>
<td>44.6 [13.3]</td>
<td>37.5 [17.0]</td>
<td>31.2 [19.4]</td>
<td>39.1 [19.4]</td>
<td>38.8 [23.0]</td>
<td>40.0</td>
</tr>
<tr>
<td>Female senior enrollment</td>
<td>297.6 [114.9]</td>
<td>316.7 [175.9]</td>
<td>417.2 [143.0]</td>
<td>347.4 [156.8]</td>
<td>344.7 [153.5]</td>
<td>282.8 [140.9]</td>
<td>335.9</td>
</tr>
<tr>
<td>Number of male CSAT takers</td>
<td>267.6 [101.3]</td>
<td>297.2 [170.1]</td>
<td>380.0 [130.3]</td>
<td>326.0 [152.2]</td>
<td>314.2 [138.4]</td>
<td>257.6 [130.7]</td>
<td>308.8</td>
</tr>
<tr>
<td>Number of high schools</td>
<td>14 17 14 17 19 12 93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Girls</td>
<td>District 3</td>
<td>District 4</td>
<td>District 6</td>
<td>District 7</td>
<td>District 8</td>
<td>District 9</td>
<td>All</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>-------</td>
</tr>
<tr>
<td>Single-sex schools (%)</td>
<td>28.6</td>
<td>35.3</td>
<td>40</td>
<td>46.7</td>
<td>41.2</td>
<td>36.4</td>
<td>38.2</td>
</tr>
<tr>
<td>Private schools (%)</td>
<td>14.3</td>
<td>47.1</td>
<td>53.3</td>
<td>60.0</td>
<td>52.9</td>
<td>45.5</td>
<td>46.1</td>
</tr>
<tr>
<td>Class size</td>
<td>36.1 [16.0]</td>
<td>35.1 [17.2]</td>
<td>36.5 [30.1]</td>
<td>35.6 [23.6]</td>
<td>34.8 [30.4]</td>
<td>34.8 [27.0]</td>
<td>35.5</td>
</tr>
<tr>
<td>Female teachers (%)</td>
<td>60.6 [7.7]</td>
<td>53.3 [5.6]</td>
<td>47.0 [15.7]</td>
<td>53.0 [8.8]</td>
<td>53.4 [12.2]</td>
<td>52.7 [16.5]</td>
<td>53.3</td>
</tr>
<tr>
<td>Female senior enrollment</td>
<td>255.6 [114.1]</td>
<td>290.1 [168.0]</td>
<td>361.4 [185.5]</td>
<td>346.5 [186.6]</td>
<td>334.3 [158.8]</td>
<td>258.4 [91.6]</td>
<td>312.4</td>
</tr>
<tr>
<td>Number of female CSAT takers</td>
<td>236.9 [108.6]</td>
<td>286.4 [165.6]</td>
<td>339.6 [177.2]</td>
<td>330.5 [180.3]</td>
<td>307.6 [145.3]</td>
<td>242.6 [88.3]</td>
<td>293.7</td>
</tr>
<tr>
<td>Number of high schools</td>
<td>14</td>
<td>17</td>
<td>15</td>
<td>15</td>
<td>17</td>
<td>11</td>
<td>89</td>
</tr>
</tbody>
</table>

Notes. All variables are for the school level. Standard deviations in brackets. 1000 KRW is worth approximately 1 USD.
Table 2: School Input Effects on Korean CSAT Scores

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single-sex</td>
<td>Class size</td>
</tr>
<tr>
<td>d = 3</td>
<td>9.12        (0.86)***</td>
<td>-0.45         (0.37)</td>
</tr>
<tr>
<td>d = 4</td>
<td>3.58        (0.69)***</td>
<td>-0.18         (0.09)*</td>
</tr>
<tr>
<td>d = 6</td>
<td>1.30        (0.53)**</td>
<td>-0.32         (0.09)***</td>
</tr>
<tr>
<td>d = 7</td>
<td>7.74        (0.68)***</td>
<td>0.13          (0.18)</td>
</tr>
<tr>
<td>d = 8</td>
<td>4.18        (0.60)***</td>
<td>-0.17         (0.12)</td>
</tr>
<tr>
<td>d = 9</td>
<td>2.09        (0.87)**</td>
<td>-0.13         (0.23)</td>
</tr>
<tr>
<td>APE</td>
<td>4.65        (0.28)***</td>
<td>-0.17         (0.07)**</td>
</tr>
</tbody>
</table>

Notes. *, **, and *** indicate significance at 10 percent, 5 percent, and 1 percent level, respectively. Robust standard errors in parentheses. Standard errors clustered in school level for coefficients on time-varying regressors. Time-varying control variables include the fraction of students receiving free or reduced price lunch, annual development fund spending per student, and the fraction of female teachers. Time-invariant control variables include a private indicator, age of the school in 2008, and the interaction between the two.
<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th></th>
<th>Girls</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha_d$</td>
<td>$\alpha^*_d$</td>
<td>$\alpha_d$</td>
<td>$\alpha^*_d$</td>
</tr>
<tr>
<td>$d = 3$</td>
<td>9.12 (0.86)***</td>
<td>0.92 (0.48)*</td>
<td>3.88 (0.69)***</td>
<td>1.60 (0.45)***</td>
</tr>
<tr>
<td>$d = 4$</td>
<td>3.58 (0.69)***</td>
<td>3.22 (0.40)***</td>
<td>5.33 (0.87)***</td>
<td>3.11 (0.37)***</td>
</tr>
<tr>
<td>$d = 6$</td>
<td>1.30 (0.53)**</td>
<td>0.73 (0.40)*</td>
<td>-0.59 (0.44)</td>
<td>0.42 (0.38)</td>
</tr>
<tr>
<td>$d = 7$</td>
<td>7.74 (0.68)***</td>
<td>3.63 (0.44)***</td>
<td>-4.68 (0.51)***</td>
<td>0.65 (0.40)*</td>
</tr>
<tr>
<td>$d = 8$</td>
<td>4.18 (0.60)***</td>
<td>3.21 (0.40)***</td>
<td>0.69 (0.57)</td>
<td>3.51 (0.36)***</td>
</tr>
<tr>
<td>$d = 9$</td>
<td>2.09 (0.87)**</td>
<td>3.45 (0.54)***</td>
<td>1.81 (0.83)**</td>
<td>3.08 (0.50)***</td>
</tr>
</tbody>
</table>

Notes. *, **, and *** indicate significance at 10 percent, 5 percent, and 1 percent level, respectively. Robust standard errors in parentheses.
Table 4: Comparison with District Fixed Effect Estimates

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th>Girls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single-sex</td>
<td>Class size</td>
</tr>
<tr>
<td>Our approach</td>
<td>4.65 (0.28)**</td>
<td>-0.17 (0.07)**</td>
</tr>
<tr>
<td>Homogeneity</td>
<td>0.76 (0.68)</td>
<td>0.06 (0.13)</td>
</tr>
</tbody>
</table>

Notes. *, **, and *** indicate significance at 10 percent, 5 percent, and 1 percent level, respectively. Robust standard errors in parentheses.
A Appendix

A.1 Equivalence of \( \hat{\beta}_d \) with the Individual Regression Coefficient Estimators

First we show that \( \hat{\beta}_d \) is the OLS estimator of \( \beta_d \) in the individual level regression with school fixed effects (or school dummies) within district \( d \),

\[
Y_i = K_{S_i} \alpha_d + L'_{S_i,T_i} \beta_d + \sum_{s \in S_d} \gamma_d \nu_s \mathbb{I} \{ S_i = s \} + \xi_i
\]  

(11)

where

\[
\xi_i = K'_{S_i} (\alpha_i - \alpha_d) + L'_{S_i,T_i} (\beta_i - \beta_d) + \nu_s (\gamma_i - \gamma_d) + c_D_i \eta_i + u_{S_i,T_i} \delta_i.
\]

The equivalence of \( \hat{\beta}_d \) follows since

\[
\left( \frac{1}{N_f(d)} \sum_{s \in S_d} \sum_{i:S_i = s} \left( L_{S_i,d,T_i} - \frac{1}{N_f(s,d)} \sum_{i:S_i = s} L_{S_i,d,T_i} \right) \left( L_{S_i,d,T_i} - \frac{1}{N_f(s,d)} \sum_{i:S_i = s} L_{S_i,d,T_i} \right) \right)^{-1}
\]

\[
\times \frac{1}{N_f(d)} \sum_{s \in S_d} \sum_{i:S_i = s} \left( L_{S_i,d,T_i} \right) \left( Y_i - \frac{1}{N_f(s,d)} \sum_{i:S_i = s} Y_i \right)
\]

\[
= \left( \frac{1}{N_f(d)} \sum_{s \in S_d} \sum_{i:S_i = s} \left( L_{S_i,d,T_i} - \frac{1}{N_f(s,d)} \sum_{i:S_i = s} L_{S_i,d,T_i} \right) \left( L_{S_i,d,T_i} - \frac{1}{N_f(s,d)} \sum_{i:S_i = s} L_{S_i,d,T_i} \right) \right)^{-1}
\]

\[
\times \frac{1}{N_f(d)} \sum_{s \in S_d} \sum_{i:S_i = s} \left( L_{S_i,d,T_i} \right) \left( Y_i - \frac{1}{N_f(s,d)} \sum_{i:S_i = s} Y_i \right)
\]

\[
= \left( \frac{1}{N_f(d)} \sum_{s \in S_d} \sum_{i:T \in T} \left( L_{S_i,d,t} - \frac{1}{N_f(s,d)} \sum_{i:T \in T} L_{S_i,d,t} \right) \left( L_{S_i,d,t} - \frac{1}{N_f(s,d)} \sum_{i:T \in T} L_{S_i,d,t} \right) \right)^{-1}
\]

\[
\times \frac{1}{N_f(d)} \sum_{s \in S_d} \sum_{i:T \in T} \left( L_{S_i,d,t} \right) \left( Y_i - \frac{1}{N_f(s,d)} \sum_{i:T \in T} Y_i \right)
\]

\[
= \left( \sum_{s \in S_d} \frac{N_f(s,d)}{N_f(d)} \frac{N_f(s,d)}{N_f(s,d)} L_{S_i,d,t} \right) \left( \sum_{s \in S_d} \frac{N_f(s,d)}{N_f(d)} \frac{N_f(s,d)}{N_f(s,d)} L_{S_i,d,t} \right)^{-1}
\]

\[
\times \sum_{t \in T} \frac{N_f(s,d)}{N_f(s,d)} \frac{N_f(s,d)}{N_f(s,d)} L_{S_i,d,t} \sum_{t \in T} \frac{N_f(s,d)}{N_f(s,d)} \frac{N_f(s,d)}{N_f(s,d)} L_{S_i,d,t}
\]

\[
= \hat{\beta}_d.
\]

Next, we show that \( \hat{\alpha}_d \) is the OLS estimator of \( \alpha_d \) in the individual level regression of \( Y_i - L'_{S_i,T_i} \hat{\beta}_d \) on \( K_{S_i} \) and a constant term) district by district:

\[
Y_i - L'_{S_i,T_i} \hat{\beta}_d = C_d + K_{S_i} \alpha_d + \zeta_i,
\]

(13)

where \( \hat{\beta}_d \) is from (12) and

\[
\zeta_i = K'_{S_i} (\alpha_i - \alpha_d) + L'_{S_i,T_i} (\beta_i - \beta_d) + c_D_i (\eta_i - \eta_d) + \nu_s (\gamma_i - \gamma_d) + u_{S_i,T_i} \delta_i.
\]

30
To see this, the OLS estimator in (13) is
\[
\begin{bmatrix}
\sum_{i:D_i=d} \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i=d} K_{S_i} \right) \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i=d} K_{S_i} \right)'
\end{bmatrix}^{-1}
\times
\sum_{i:D_i=d} \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i=d} K_{S_i} \right) \left( Y_i - L' \hat{\beta}_d \right).
\]

Since
\[
\frac{1}{N_I(d)} \sum_{i:D_i=d} K_{S_i} = \frac{1}{N_I(d)} \sum_{s \in S_d} \sum_{t \in T} \sum_{i \in I} K_s \mathbb{1} \{S_i = s, D_i = d, T_i = t\}
\]
\[
= \sum_{s \in S_d} \left( \frac{1}{N_I(d)} \sum_{t \in T} \sum_{i \in I} \mathbb{1} \{S_i = s, D_i = d, T_i = t\} \right) K_s
\]
\[
= \sum_{s \in S_d} \frac{N_I(s, d)}{N_I(d)} K_s = \sum_{s \in S_d} \hat{w}_d^s K_s,
\]
we have
\[
\frac{1}{N_I(d)} \sum_{i:D_i=d} \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i=d} K_{S_i} \right) \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i=d} K_{S_i} \right)'
\]
\[
= \frac{1}{N_I(d)} \sum_{s \in S_d} \sum_{t \in T} \sum_{i \in I} \mathbb{1} \{S_i = s, D_i = d, T_i = t\}
\]
\[
\times \left( K_s - \sum_{s \in S_d} \hat{w}_d^s K_s \right) \left( K_s - \sum_{s \in S_d} \hat{w}_d^s K_s \right)'
\]
\[
= \sum_{s \in S_d} \left( K_s - \sum_{s \in S_d} \hat{w}_d^s K_s \right) \left( K_s - \sum_{s \in S_d} \hat{w}_d^s K_s \right)'
\]
\[
= \sum_{s \in S_d} \frac{N_I(s, d)}{N_I(d)} \left( K_d - \sum_{s \in S_d} \hat{w}_d^s K_s \right) \left( K_s - \sum_{s \in S_d} \hat{w}_d^s K_s \right)'
\]
\[
= \sum_{s \in S_d} \hat{w}_d^s K_{s,d} \tilde{K}_{s,d}'.
\]
Next, note that since the school aggregate test score is
\[ Y_{s,d,t} = \sum_{i \in I} \frac{1}{N_I(d)} \sum_{i \in I} K_{s_i} (Y_i - \hat{L}_{S_i,T_i} \hat{\beta}_d) \]
we have
\[ \frac{1}{N_I(d)} \sum_{i:D_i = d} \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i = d} K_{S_i} \right) \left( Y_i - \hat{L}_{S_i,T_i} \hat{\beta}_d \right) \]
\[ = \frac{1}{N_I(d)} \sum_{s \in S_d} \sum_{t \in T} \sum_{i \in I} \mathbb{I} \{ S_i = s, D_i = d, T_i = t \} \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i = d} K_{S_i} \right) \left( Y_i - \hat{L}_{S_i,T_i} \hat{\beta}_d \right) \]
\[ = \sum_{s \in S_d} \left( K_s - \sum_{s \in S_d} \bar{w}_s^d K_s \right) \left( \frac{1}{N_I(d)} \sum_{t \in T} \sum_{i \in I} \mathbb{I} \{ S_i = s, D_i = d, T_i = t \} \left( Y_i - \hat{L}_{S_i,T_i} \hat{\beta}_d \right) \right) \]
\[ = \sum_{s \in S_d} \left( K_s - \sum_{s \in S_d} \bar{w}_s^d K_s \right) \left( \frac{N_I(s,d)}{N_I(d)} \sum_{t \in T} \frac{N_I(s,d,t)}{N_I(s,d)} \left( Y_{s,d,t} - \hat{L}_{s,d,t} \hat{\beta}_d \right) \right) \]
\[ = \sum_{s \in S_d} \left( K_s - \sum_{s \in S_d} \bar{w}_s^d K_s \right) \bar{w}_s^d \left( \bar{Y}_{s,d,t} - \bar{L}_{s,d,t} \hat{\beta}_d \right) \].

This shows that the individual OLS estimator of \( Y_i - \hat{L}_{S_i,T_i} \hat{\beta}_d \) on \( K_{S_i} \) (with a constant term) with \( \hat{\beta}_d \) in \([12]\) is equivalent to \( \hat{\alpha}_d \) :

\[ \hat{\alpha}_d = \left[ \sum_{i:D_i = d} \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i = d} K_{S_i} \right) \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i = d} K_{S_i} \right) \right]^{-1} \]
\[ \times \sum_{i:D_i = d} \left( K_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i = d} K_{S_i} \right) \left( Y_i - \hat{L}_{S_i,T_i} \hat{\beta}_d \right) \],
as required. \( \blacksquare \)
B Online Appendix: Not for Publication

B.1 Derivation of $\alpha_d^*$

We show ...

The average outcome among each type of schools can be written

$$E[Y_i|K_{S_i} = 1, D_i = d, T_i = t] = E[\alpha_i|K_{S_i} = 1, D_i = d, T_i = t]$$

$$+ E[L'_{S_i,T_i}\beta_i + v_s\gamma_i + u_{S_i,T_i}\delta_i + c\eta_i|K_{S_i} = 1, D_i = d, T_i = t]$$

(14)

and

$$E[Y_i|K_{S_i} = 0, D_i = d, T_i = t] = E[L'_{S_i,T_i}\beta_i + v_s\gamma_i + u_{S_i,T_i}\delta_i + c\eta_i|K_{S_i} = 0, D_i = d, T_i = t].$$

(15)

Under the random assignment of students across schools within district, the first term after the equal sign in (14) becomes

$$E[\alpha_i|K_{S_i} = 1, D_i = d, T_i = t]$$

$$= \sum_{s \in S_d, K_s = 1} E[\alpha_i|K_{S_i} = 1, S_i = s, D_i = d, T_i = t] E[S_i = s|K_{S_i} = 1, D_i = d, T_i = t]$$

$$= \sum_{s \in S_d, K_s = 1} E[\alpha_i|S_i = s, D_i = d, T_i = t] E[S_i = s|K_{S_i} = 1, D_i = d, T_i = t]$$

$$= \alpha_d \sum_{s \in S_d, K_s = 1} E[S_i = s|K_{S_i} = 1, D_i = d, T_i = t]$$

$$= \alpha_d.$$

Also, the random assignment condition leads to the following expression for the second term after the equal sign in (14):

$$E[L'_{S_i,T_i}\beta_i + v_s\gamma_i + u_{S_i,T_i}\delta_i + c\eta_i|K_{S_i} = 1, D_i = d, T_i = t]$$

$$= \sum_{s \in S_d, K_s = 1} \left\{ E\left[ L'_{S_i,T_i}\beta_i + v_s\gamma_i + u_{S_i,T_i}\delta_i + c\eta_i|K_{S_i} = 1, S_i = s, D_i = d, T_i = t \right] \right\}$$

$$\times E[S_i = s|K_{S_i} = 1, D_i = d, T_i = t]$$

$$= \sum_{s \in S_d, K_s = 1} \left\{ E\left[ L'_{S_i,T_i}\beta_i + v_s\gamma_i + u_{S_i,T_i}\delta_i + c\eta_i|S_i = s, D_i = d, T_i = t \right] \right\}$$

$$\times E[S_i = s|K_{S_i} = 1, D_i = d, T_i = t]$$

$$= \sum_{s \in S_d, K_s = 1} \left\{ E[L'_{s,t}\beta_d + v_s\gamma_d + u_{s,t}\delta_d + c\eta_d]|S_i = s, D_i = d, T_i = t \right\} \times E[S_i = s|K_{S_i} = 1, D_i = d, T_i = t]$$

$$= \sum_{s \in S_d, K_s = 1} \left\{ E[L'_{s,t}\beta_d + v_s\gamma_d + u_{s,t}\delta_d + c\eta_d] \times E[S_i = s|K_{S_i} = 1, D_i = d, T_i = t] \right\}.$$
Therefore, we have

\[
\alpha_d^* = \alpha_d + \sum_{s \in S_d, \ k_s = 1} \left( \frac{1}{N_S} \sum_{d \in D} \sum_{s \in S_d} \sum_{t \in T} (L_{s,d,t} - \bar{L}_{\bullet,d,t})^2 \right)^{-1} \left( \frac{1}{N_D N_S T} \sum_{d \in D} \sum_{s \in S_d} \sum_{t \in T} (L_{s,d,t} - \bar{L}_{\bullet,d,t}) (Y_{s,d,t} - \bar{Y}_{d,t}) \right)
\]

\[
= \beta + \left( \frac{1}{N_D N_S T} \sum_{d \in D} \sum_{s \in S_d} \sum_{t \in T} (L_{s,d,t} - \bar{L}_{\bullet,d,t})^2 \right)^{-1} \left( \frac{1}{N_D N_S T} \sum_{d \in D} \sum_{s \in S_d} \sum_{t \in T} (L_{s,d,t} - \bar{L}_{\bullet,d,t}) \chi_{s,d,t} \right).
\]

Note that

\[
\frac{1}{N_D N_S T} \sum_{d \in D} \sum_{s \in S_d} \sum_{t \in T} (L_{s,d,t} - \bar{L}_{\bullet,d,t}) \chi_{s,d,t} = \frac{1}{N_D N_S T} \sum_{d \in D} \sum_{s \in S_d} \sum_{t \in T} (L_{s,d,t} - \bar{L}_{\bullet,d,t}) U_{s,d,t} + \frac{1}{N_D} \sum_{d \in D} \frac{1}{N_S} \sum_{s \in S_d} V_{s,d} \left( \frac{1}{T} \sum_{t \in T} (L_{s,d,t} - \bar{L}_{\bullet,d,t}) \right)
\]

\[
+ \frac{1}{N_D} \sum_{d \in D} (\beta_d - \beta) \frac{1}{N_S} \sum_{s \in S_d} \frac{1}{T} \sum_{t \in T} (L_{s,d,t} - \bar{L}_{\bullet,d,t})^2.
\]

In this case, since the omitted school fixed effects \( V_{s,d} \) and the time average of the school inputs \( \frac{1}{T} \sum_{t \in T} (L_{s,d,t} - \bar{L}_{\bullet,d,t}) \) can be correlated. When the district choice is endogenous, the district specific sorting effects \( (\beta_d - \beta) \) and the variation of the school input within the district can also be correlated. In this case, the district fixed effect estimator \( \bar{\beta} \) is a biased estimator of \( \beta \).
C Appendix Tables and Figures: Not for Publication

Figure C1: School-level Average English CSAT Scores by District

A. Boys

<table>
<thead>
<tr>
<th>District</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nambu</td>
<td>90.9</td>
<td>4.4</td>
</tr>
<tr>
<td>Bukbu</td>
<td>98.7</td>
<td>4.7</td>
</tr>
<tr>
<td>Gangdong</td>
<td>100.1</td>
<td>2.8</td>
</tr>
<tr>
<td>Gangseo</td>
<td>99.7</td>
<td>6.6</td>
</tr>
<tr>
<td>Gangnam</td>
<td>107.3</td>
<td>4.2</td>
</tr>
<tr>
<td>Dongjak</td>
<td>94.3</td>
<td>3.0</td>
</tr>
</tbody>
</table>

B. Girls

<table>
<thead>
<tr>
<th>District</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nambu</td>
<td>95.8</td>
<td>5.2</td>
</tr>
<tr>
<td>Bukbu</td>
<td>102.5</td>
<td>4.7</td>
</tr>
<tr>
<td>Gangdong</td>
<td>104.1</td>
<td>3.2</td>
</tr>
<tr>
<td>Gangseo</td>
<td>103.0</td>
<td>6.6</td>
</tr>
<tr>
<td>Gangnam</td>
<td>110.6</td>
<td>4.1</td>
</tr>
<tr>
<td>Dongjak</td>
<td>98.7</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Notes. Units are school×year.
Figure C2: School-level Average Math CSAT Scores by District

A. Boys

3. Nambu

Mean = 93.8
SD = 3.3

4. Bukbu

Mean = 98.7
SD = 4.7

6. Gangdong

Mean = 100.1
SD = 2.8

7. Gangseo

Mean = 99.7
SD = 6.6

8. Gangnam

Mean = 107.3
SD = 4.2

9. Dongjak

Mean = 94.3
SD = 3.0

B. Girls

3. Nambu

Mean = 94.6
SD = 3.0

4. Bukbu

Mean = 100.0
SD = 3.9

6. Gangdong

Mean = 100.3
SD = 5.7

7. Gangseo

Mean = 107.0
SD = 4.3

8. Gangnam

Mean = 107.3
SD = 3.2

9. Dongjak

Mean = 97.5
SD = 2.8

Notes. Units are school × year.
Table C1: School Input Effects on English and Math CSAT Scores

<table>
<thead>
<tr>
<th></th>
<th>Boys</th>
<th></th>
<th>Girls</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Single-sex Class size</td>
<td></td>
<td>Single-sex Class size</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. English</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d = 3$</td>
<td>12.17 (0.84)*** -0.56 (0.39)</td>
<td></td>
<td>4.97 (0.72)*** -0.20 (0.29)</td>
<td></td>
</tr>
<tr>
<td>$d = 4$</td>
<td>5.09 (0.68)*** -0.39 (0.15)**</td>
<td></td>
<td>3.95 (0.90)*** 0.14 (0.08)*</td>
<td></td>
</tr>
<tr>
<td>$d = 6$</td>
<td>0.36 (0.53) -0.30 (0.09)***</td>
<td></td>
<td>-0.85 (0.46)** 0.31 (0.13)**</td>
<td></td>
</tr>
<tr>
<td>$d = 7$</td>
<td>5.67 (0.67)*** 0.06 (0.15)</td>
<td></td>
<td>-5.12 (0.52)*** 0.29 (0.11)**</td>
<td></td>
</tr>
<tr>
<td>$d = 8$</td>
<td>5.54 (0.61)*** -0.34 (0.10)***</td>
<td></td>
<td>-0.83 (0.59) 0.50 (0.23)**</td>
<td></td>
</tr>
<tr>
<td>$d = 9$</td>
<td>-0.35 (0.83) -0.40 (0.24)</td>
<td></td>
<td>1.26 (0.82) 0.41 (0.16)**</td>
<td></td>
</tr>
<tr>
<td>APE</td>
<td>4.76 (0.28)*** -0.30 (0.07)***</td>
<td></td>
<td>0.19 (0.27) 0.26 (0.07)***</td>
<td></td>
</tr>
<tr>
<td>B. Math</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$d = 3$</td>
<td>4.97 (0.81)*** -0.37 (0.25)</td>
<td></td>
<td>3.51 (0.69)*** -0.10 (0.18)</td>
<td></td>
</tr>
<tr>
<td>$d = 4$</td>
<td>5.61 (0.66)*** -0.17 (0.18)</td>
<td></td>
<td>6.42 (0.92)*** 0.17 (0.15)</td>
<td></td>
</tr>
<tr>
<td>$d = 6$</td>
<td>0.65 (0.54) -0.40 (0.13)*</td>
<td></td>
<td>-1.75 (0.50)*** 0.19 (0.11)</td>
<td></td>
</tr>
<tr>
<td>$d = 7$</td>
<td>11.48 (0.65)*** 0.25 (0.25)</td>
<td></td>
<td>-6.81 (0.58)*** 0.25 (0.13)*</td>
<td></td>
</tr>
<tr>
<td>$d = 8$</td>
<td>8.65 (0.66)*** -0.38 (0.16)**</td>
<td></td>
<td>4.19 (0.68)*** 0.15 (0.16)</td>
<td></td>
</tr>
<tr>
<td>$d = 9$</td>
<td>3.92 (0.88)*** -0.56 (0.36)</td>
<td></td>
<td>0.31 (0.84) 0.41 (0.22)*</td>
<td></td>
</tr>
<tr>
<td>APE</td>
<td>6.19 (0.28)*** -0.24 (0.09)**</td>
<td></td>
<td>0.88 (0.29)*** 0.17 (0.06)**</td>
<td></td>
</tr>
</tbody>
</table>

Notes. *, **, and *** indicate significance at 10 percent, 5 percent, and 1 percent level, respectively. Robust standard errors in parentheses. Standard errors clustered in school level for coefficients on time-varying regressors. Time-varying control variables include the fraction of students receiving free or reduced price lunch, annual development fund spending per student, and the fraction of female teachers. Time-invariant control variables include a private indicator, age of the school in 2008, and the interaction between the two.