Within-District School Lotteries, District Selection, and the Average Partial Effects of School Inputs*

Eleanor Jawon Choi† Hanyang University

Hyungsik Roger Moon‡ USC, USC Dornsife INET and Yonsei University

Geert Ridder§ USC

February 22, 2017

Abstract

This paper proposes an econometric framework to consistently estimate the average partial effects (APE) of school inputs on academic achievement when students are randomly assigned to schools within each school district but endogenously sort across school districts. We illustrate our method by estimating the APE of single-sex schooling and class size on standardized test scores using data from Seoul, Korea in 2008–2009. Our APE estimates are smaller than the estimates from conventional linear regressions with school district fixed effects, which ignore endogenous district choice and thus suffer from selection bias.

JEL Codes: C21, I21

Keywords: average partial effect, within-district school lotteries, selection, treatment effect heterogeneity, education production

*We thank Alan Krueger, David Lee, Nayoung Lee, Heonjae Song, Jungmo Yoon, and seminar participants at the AEA, CUHK, ES NASM, Hanyang-Renmin, KAEA-KEA, KES, KLEA, Sogang, SOLE and SKKU for helpful comments and suggestions; the Korea Education and Research Information Service, the Korean Educational Development Institute, and the Korean Ministry of Education for providing data; and Jae Woo Choi and Sae Won Om for research assistance. Choi's research is supported by the National Research Foundation of Korea Grant from the Korean Ministry of Education (NRF-2015S1A5A8010685). Moon thanks the research award from the USC Lusk Center for Real Estate. All remaining errors are our own.

†College of Economics and Finance, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, Korea. Email: choiej@hanyang.ac.kr.

‡Department of Economics and USC Dornsife INET, University of Southern California, KAP 300, Los Angeles, CA 90089-0253. School of Economics, Yonsei University, Seoul, Korea. Email: moonr@usc.edu.

§Department of Economics, University of Southern California, KAP 300, Los Angeles, CA 90089-0253. Email: ridder@usc.edu. Phone: +1-213-740-7432. Fax: +1-213-740-8543. (corresponding author)

1 Introduction

The effect of school inputs on student academic outcomes has long been an important issue among researchers, policy makers, and parents. When students are free to choose their school, it is difficult to estimate the causal effect of school inputs using observational data on student performance and school characteristics. This is because observed and unobserved student and family attributes influence both the academic achievement and school choice of the students, which could lead to spurious correlation between student performance and school characteristics.[1] Thus, addressing endogenous selection into schools and neighborhoods has been a key econometric issue in studies of school effectiveness. Credible estimates can be obtained when quasi-experimental variation generated from rule-based student assignment is combined with a careful econometric analysis properly taking into account the institutional setting.[2]

This paper estimates causal effects of school inputs on students' academic achievement using quasi-experimental variation from within-district high school lotteries in Seoul, South Korea (henceforth, Korea). We provide an econometric framework to identify the average partial effects (APE) of time-invariant and time-varying school inputs when school districts are endogenously chosen but school assignment is random within each school district. We illustrate our econometric framework by estimating the effects of single-sex schooling and class size, which have been major policy variables in the education production literature.[3]

Students who entered high school in Seoul, Korea before 2010 were randomly assigned to high schools within school districts. Under this assignment rule, students and their parents can only choose the average quality[4] of the schools to which they can potentially be assigned, by choosing a residential neighborhood. Thus, self-selection bias arises from endogenous sorting across school districts but not from endogenous selection into schools within each

[1] For studies on school choice and residential sorting, see, for example, Urquiola (2005); Rothstein (2006); Bayer et al. (2007).

[2] There is a large literature on education production and relevant econometric methods. See Hanushek (2006); Meghir and Rivkin (2011) and references therein.

[3] The class size effect has long been a major topic in education production. Well-known studies include Angrist and Lavy (1999); Krueger (1999); Dearden et al. (2002) to name a few. The effect of single sex education has become an active area of economic research in more recent years. See Hoxby (2000); Lavy and Schlosser (2011); Jackson (2012); Hill (2015) among others. Whitmore (2005) studies the effects of both factors.

[4] Or more generally the distribution of school quality.

school district.

We use data on College Scholastic Ability Test (CSAT) scores and high school characteristics from Seoul in 2008–2009. The population of interest is high school seniors who were randomly assigned to academic high schools and took the CSAT in their senior year. Our analysis sample includes 57,443 male and 52,271 female students in 55 coed, 34 all-girls, and 38 all-boys high schools.

Given the institutional setting, we construct a model of heterogeneous treatment effects that takes into account within-district random assignment and across-district sorting. We specify that potential outcomes—standardized test scores—for each student depend on observed and unobserved school inputs (that are either time varying or time invariant), and unobserved time-invariant district fixed effects, interacted with the student's unobserved productivity (or ability) for these inputs as random coefficients. The random coefficient form of our potential outcomes model allows unobserved individual heterogeneity in the effects of school and district inputs.

In our econometric framework, the random coefficient model of individual-level education production implies a (random coefficient) linear regression of individual academic outcomes on school inputs for each school district. We estimate the APE of observed school inputs in two steps. In the first step, we estimate (district-specific) effects of time-varying and time-invariant school inputs on standardized test scores using data on students and schools from each school district. For this, we first obtain the within estimator of the coefficients on time-varying school inputs (e.g. class size) with school fixed effects to control for time-invariant unobservable school attributes that are potentially confounded with observed school inputs. Next, we estimate the coefficients on time-invariant school inputs (e.g. single-sex education) by regressing the residuals from the first stage within regression on observed time-invariant school characteristics. In the second step, we consistently estimate the APE of school inputs by a weighted average of the district-specific estimates with the fraction of students in each district as the weight.

We also show that a school-level education production function can be derived by aggregating individual potential outcomes. The effects of school inputs in each school's education production function differ by school district but not by school within each school district. The effects of school inputs are potentially correlated with school characteristics, across school districts. This is due to heterogeneous student ability, endogenous sorting into school districts, and random assignment within each school district.

There are several recent studies that exploit quasi-experimental variation from within-district school lotteries in Seoul to estimate school quality effects on students' academic performance. Using various data sets on standardized test scores and secondary school characteristics in Seoul, Park et al. (2013); Lee et al. (2014); Ku and Kwak (2016); Sohn (2016) study the effect of single-sex education, and Hahn et al. (2016) examine the effect of school autonomy.[5] To exploit the institutional fact that students are randomly assigned to different types of schools within school districts, these studies estimate a linear regression model of student outcome on school inputs and school district fixed effects, and thus assume no endogenous sorting across school districts induced by school quality or no individual heterogeneity in school input effects.

In Choi et al. (2014), we suggested how to address endogenous selection into school districts if we use the within-district random assignment. In this study, we extend our earlier work and develop an econometric framework that allows us to estimate the causal APE of school inputs by properly dealing with across-district sorting. We show that students are indeed self-selected into school districts in response to school quality and that the APE of school inputs is estimated with a bias if the self-selection is not properly taken into account.[6]

The APE estimation based on our econometric model finds no evidence of better per-

[5]Park et al. (2013) estimate the effect of single-sex high schools using a data set close to ours. Lee et al. (2014) estimate the effect of school and classroom gender composition at Seoul middle schools. Ku and Kwak (2016) use data on Seoul high schools available for longer periods (1999–2009) but with fewer variables on school characteristics. Besides exploiting within-district random assignment, Ku and Kwak (2016)'s identification strategy mainly relies on the change in types from single-sex to coed for some schools during their analysis period. Sohn (2016) restricts the analysis to public high schools in Seoul to tease out the effect of single-sex schooling not confounded with the public/private status. Hahn et al. (2016) focus on the private school effect as both public and private high schools are subject to the within-district random assignment lotteries in Seoul.

[6]Altonji and Mansfield (2014) suggest that, with rich data on student and family characteristics, school district averages of observed individual characteristics can control for sorting on (un)observables if district choice is independent of school inputs given these averages. However, this does not apply to our case or to Park et al. (2013); Lee et al. (2014); Hahn et al. (2016); Ku and Kwak (2016); Sohn (2016) because individual attributes are mostly unobserved in the test score data currently available in Korea.

formance among students at single-sex schools compared to those at mixed-gender schools, unlike many of the previous studies. Our APE estimates also show that smaller class sizes slightly increase boys' CSAT scores but have no effect on girls.' The district-specific estimates are more heterogeneous for single-sex school effects than for class size effects, which implies a substantial degree of endogenous sorting across school districts based on the likelihood of getting into a single-sex school.

We compare our APE estimates of single-sex schooling and class size effects with estimates from a (homogeneous coefficients) linear regression model with district-year fixed effects (DFE). The DFE model has been commonly estimated in the existing studies to exploit random assignment within school districts, but it is known to yield biased estimates when district selection is endogenous (Choi et al., 2016). Compared to our APE estimates, the estimates from the DFE regressions are less negative or larger. This is due to an upward bias produced by the endogenous district choice that amplifies the positive effect and mutes the negative effect of school inputs.

The remainder of the paper is organized as follows. Section 2 outlines the institutional background of the high school assignment lotteries in Seoul. Section 3 describes the data on CSAT scores and school characteristics. Section 4 presents the potential outcomes model of education production for our institutional setting. Section 5 discusses estimation and statistical inference. Empirical results are given in Section 6. Section 7 concludes.

2 Institutional Background

The education policy in Korea from the early 1970s to the late 2000s mainly focused on equal opportunity in education. In accordance with this emphasis on equality, the Seoul Metropolitan Office of Education (SMOE) started in 1974 the High School Equalization Policy (HSEP). Major goals of the HSEP include providing students with a uniform learning experience and reducing the achievement gap across schools. To achieve these goals, policy makers tried to minimize across-school variation in student ability, teacher quality, curricula and facilities.[7] Although the principles of equal treatment and equal opportunities in primary and secondary education have been maintained until today, during the past several years, the policy focus has shifted from uniformity to diversity in education programs. The HSEP has been substantially reformed since 2010 to broaden students' school choice options, diversify the school curriculum, and encourage competition among schools.

Before the policy reform started in 2010, middle school seniors in Seoul had been randomly assigned to general academic high schools within their school districts. The high school assignment lotteries were designed so that the distribution of student ability—and hence peer quality—was similar across schools within the same school district.[8] In order to accommodate students' preferences for schools, the SMOE abolished the within-district random assignment and introduced a new way of matching students and high schools in 2010. Under the new choice-based system, students have to submit their preference ranking of several high schools and student-school matches are determined by the Boston matching mechanism.[9]

The high school assignment lotteries conducted in the late 2000s before the adoption of the new school choice system covered general academic high schools in ten school districts (districts 1–4 and 6–11) in Seoul. The random assignment system excluded vocational high schools; selective high schools specialized in math and science, foreign languages, arts, or sports; and academic high schools near the city center—mostly in district 5 and some in districts 1, 2, 10 and 11.[10]

[7] For more information on the equalization policy in Korean secondary education and its impacts, see Kang (2007); Kang et al. (2007); Hahn et al. (2008); Kim et al. (2008) among others.

[8] Seoul Metropolitan Office of Education (2008) documents the procedures of the random high school assignment as follows. High school assignments were conducted separately for boys and girls in February, less than a month before the beginning of the school year in early March. 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 school districts. The random assignment was under the restriction that students should attend a school in a commutable sub-area of each school district. Neither the cutoffs of the middle school records nor the boundaries of the commutable areas were ever public information or made available for research purposes. Due to the lack of information on the exact random assignment algorithm, students and their parents were unable to make better predictions than researchers about school assignment outcomes. See Ku and Kwak (2016); Sohn (2016) for further details on the implementation of the computerized lotteries.

[9] See Hwang (2016) for further details on the new choice-based high school assignment that has been implemented in Seoul since 2010.

[10] In 2008, there were 295 high schools in 11 school districts of Seoul—203 academic, 78 vocational, 12 specialized, and two other high schools. Forty of the academic high schools did not participate in the assignment

General academic high schools were subject to the lottery-based assignment system regardless of their resource levels or types. There are various types of high schools—coed, all-girls, and all-boys schools that are either public or private—in each school district. Single-sex schools tend to have a longer history than coeducational schools as high schools started as single-sex schools in the past. Single-sex schools are also more likely to be private because the Korean government has been requiring since 1998 that all newly-opened public schools be coeducational.

The private high schools subject to the random assignment lotteries were similar to charter schools in the US.[11] There was no systematic difference in the curriculum or tuition expenses between public and private high schools in the random assignment system before 2010. As the high school curriculum and the textbooks were developed and regulated by the Ministry of Education, each school had very little discretion in deciding which subjects or material to teach. Both public and private schools were heavily subsidized by the government. The major difference between public and private schools was for the teacher hiring process and the teachers' status. Public school teachers are government employees who have passed the national qualification exam. They serve in a school for three to eight years, and after that, are transferred to another school within the same city.[12] On the contrary, private school teachers are hired by a specific school and usually stay in the same school until they quit or retire.

Non-compliance to the random high school assignment does not appear to be a major issue. Under the random assignment system, students had very little incentive to move to another school district after the initial high school assignment. Once assigned to a high school, students were prohibited to transfer to another school within the same school district. When students (and their family) moved to another school district, their high school assignment

lotteries. The academic high schools outside the random assignment system accepted applications from students in all school districts of Seoul mainly because those schools were located in the business area near the city center, with a small number of residents.

[11]See Chapter 4 of Ministry of Education and KDI School of Public Policy and Management (2013) for more details on government policies and regulations regarding private schools in Korea.

[12]Teachers are not randomly assigned across public high schools. When teachers rotate from one school to another, they usually alternate between schools in wealthy neighborhoods and those in poor neighborhoods. Nevertheless, teachers can reveal their preference for specific schools and schools can also request for specific teachers, which does affect the teacher assignment outcomes across schools. The information on teacher assignment was collected in conversations with public school teachers in Seoul.

was again determined by random lotteries in the new district. Transfers and drop outs after the initial high school assignment were indeed rare (less than 3 percent) and unrelated to school characteristics (Ku and Kwak, 2016; Sohn, 2016). Senior enrollment in each school is on average about 98 percent of the freshman enrollment from 2 years before and the across-school standard deviation of this proportion is less than 4 percentage points in our data.

3 Data and Analysis Sample

We use data on the CSAT scores and high school characteristics provided by the Korean Ministry of Education (MOE). We link individual-level test scores and school-level characteristics by matching school names.

The CSAT is the standardized test used for college admissions in Korea. The test is administered by the Korea Institute of Curriculum and Evaluation and is offered once a year in November. About 600,000 individuals, including high school seniors, high school graduates, and GED holders, take the CSAT every year.[13] The CSAT is a high-stakes exam as the CSAT score is a major factor that determines college admission outcomes and thus potentially affects future earnings.

The CSAT consists of five major sections: Korean, Math, English, Sciences/Social Studies/Vocational Education, and Second Foreign Languages. In this study, the educational outcome is the sum of the Korean and English CSAT scores (hereafter referred to as the CSAT score), which is normalized to have zero mean and unit variance across all observations within each cohort.[14] We do not use the Math score in the main analysis as 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%).[15] Furthermore, students had to select between two types of exams—basic or advanced—for the Math section.[16] It is highly likely that a stu-

[13]In our analysis sample, 93% of high school seniors took the CSAT. As the test taking rate is very high, non-random selection of CSAT takers is less of an issue unlike in the US.

[14]In the raw data, the CSAT scores on Korean, English, and Math are standardized to have a mean of 100 points and a standard deviation of 20 points.

[15]For CSAT applicants who were absent from the exam, missing raw scores are imputed with zeros.

[16]Students who decided to take the advanced Math exam had to further select among calculus, probability and statistics, and discrete mathematics.

dent's choice on whether to take the exam on Math or which type of Math exam to take was made on the basis of the (expected) outcome. We do not analyze scores on Sciences/Social Studies/Vocational Education or Second Foreign Languages for similar reasons. Students had to select between tests for different subjects so that the scores on these different subjects are not directly comparable.

The CSAT score data include the entire population of CSAT takers in Korea but contain no individual characteristics other than gender, whether or not enrolled in high school, and the name of the school currently attending. Data on high school characteristics come from the school information database maintained by the Korea Education and Research Information Service. The database contains information on all primary and secondary schools in Korea, including school type, number of students by gender, class size, number of teachers by gender and employment type, school budget, etc.[17]

To focus on students randomly assigned to academic high schools in Seoul, we restrict the analysis sample to high school seniors in 2008 or 2009 who were attending academic high schools in school districts 3, 4, 6–9. We drop the 2007 or earlier cohorts of high school seniors (who entered high school in 2005 or earlier) as school characteristics in the school information database are available only from 2008. We exclude those admitted under the choice-based system in 2010 or later, as well as those assigned randomly in 2008 or 2009 but exposed to the changes in the high school assignment after the adoption of the new school choice system. We focus on school districts 3, 4, 6–9 where all the general academic high schools admitted students using the high school assignment lotteries. Our analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The analysis sample covers 54% of high school seniors taking the CSAT in Seoul.[18]

Using the analysis sample, Figure 1 plots the within-district distribution of school-year

[17] Information on schools for the most recent four years are publicly available at `http://schoolinfo.go.kr`.

[18] About 26% of CSAT takers were from Seoul. The analysis sample includes 36% of the CSAT takers in Seoul. Our sample does not include high school graduates (mostly retakers) comprising 34% of CSAT takers in Seoul. We further drop those attending vocational high schools, selective high schools, or general academic high schools in school districts not fully covered by the high school lotteries, who constituted 30% of CSAT takers in Seoul.

averages of the combined CSAT score. That is, the observations are the school-level average scores in 2008 or 2009 (each school contributes two average scores). The district-specific distribution of the average test score is plotted for boys and girls separately.[19] Each bar in Figure 1 shows the fraction of observations in an interval with width equal to 0.1 standard deviation of the distribution of the average score in each school district (11 intervals). Although in Figure 1 we pool the two years of observations on each school, the distribution looks similar when we plot the distribution for 2008 and 2009 separately. The mean and standard deviation of the average test score are also reported in the top left corner of each bar graph. Similar graphs for Korean, English, and Math CSAT scores are provided in Appendix Figures A1, A2 and A3, respectively. The English score is more dispersed within and across districts compared to the Korean or Math score.

The school-level average test score varies substantially across districts as well as within each district. Based on the district means, the six school districts can be grouped into three categories—high (district 8), middle (districts 4, 6 and 7), and low (districts 3 and 9). The three categories roughly correspond to the income level in the districts—wealthy, middle income, and poor.[20]

Table 1 shows descriptive statistics of student- and school-level variables used in our empirical analysis. In addition to the mean and standard deviation, we also report the proportion of the within-district sum of squares in the total sum of squares.[21] The mean and standard deviation of the CSAT scores are reported at both individual (panel A) and school (panel B) levels. As shown in panels B and C, the school-level average test scores and high school characteristics are heterogeneous within and across school districts. While the within-district variation accounts for 69–99% of the total variation in school characteristics (except for class size), only 44–57% of the total variation in school-level averages of Korean and/or English CSAT scores can be explained by within-district variation. This indicates that a major portion

[19] For mixed-gender schools, average test scores are computed for boys and girls separately, and the gender-specific averages are included in the graphs for each gender.

[20] This can be seen from the 2000–2006 average housing price (in 10,000's of 2009 KRW/m^2): 280.0 in district 3, 229.1 in district 4, 520.4 in district 6, 375.4 in district 7, 676.9 in district 8, and 315.0 in district 9. The housing price statistics are from the database maintained by the Mirae Asset Real Estate, one of the leading real estate investment companies in Korea (`http://www.r114.com`).

[21] For time-varying variables, we compute within-district-year sum of squares.

of the test score variation across schools is likely due to student-level sorting across school districts. The large proportion of within-district variation in school characteristics, on the other hand, implies that it is also important to control for school characteristics to disentangle the APE of a specific school input from many other confounding factors. The mean and standard deviation of the variables by district are shown in Appendix Table A1.

In Appendix A, we check whether high school random assignment was properly implemented within school districts. Using data from the Korean Education Longitudinal Study (KELS), we show that students' characteristics before entering high schools, including parental education, household income, expenditures on private out-of-school education, and 9th-grade standardized test scores, are similar between single-sex and coeducational high schools as well as across high schools with different class sizes, within school districts. We cannot reject the null hypothesis that the baseline characteristics of students are unrelated to school characteristics (single-sex schooling or class size) when we restrict our analysis to students attending high schools subject to the random assignment lotteries. When we repeat this exercise using students attending high schools not subject to the random assignment lotteries, however, we strongly reject the null hypothesis that the distribution of the baseline student characteristics is unrelated to school characteristics. See Appendix A and Table A2 for more details.

4 Potential Outcomes Framework

We construct a potential outcomes model of education production tailored to the institutional setting of our data—random assignment of students into schools within each school district. The potential outcomes framework is also used to derive a school-level education production function in Section 5.

We use the following notation to indicate that our data include a population of students and schools. Let $\mathcal{D} = \{1, ..., N_D\}$ denote the collection of school districts in Seoul where student assignment to high schools is random. Let $\mathcal{I} = \{1, ..., N_I\}$ be the population of high school seniors in these school districts who are randomly assigned to their schools. $\mathcal{S}_d = \{1, ..., N_S(d)\}$ denotes the collection of high schools in school district $d \in \mathcal{D}$, where $N_S(d)$ is the number of high schools in district d. The analysis period is denoted by $\mathcal{T} = \{2008, 2009\}$.

Our econometric framework is based on the *potential outcomes* of student $i \in \mathcal{I}$ attending school $s \in \mathcal{S}_d$ in district $d \in \mathcal{D}$ and year $t \in \mathcal{T}$.[22] We assume a linear random coefficient model of education production:

$$Y_i(s,t,d) = \boldsymbol{\alpha}_i'\mathbf{X}_s + \boldsymbol{\beta}_i'\mathbf{Z}_{s,t} + \gamma_i v_s + \delta_i u_{s,t} + \eta_i c_d. \tag{1}$$

The potential outcome $Y_i(s,t,d)$ is the potential CSAT score of student i if (s)he attends school s (that is in district d) in year t. Observed school inputs are denoted by $\mathbf{X}_s$ and $\mathbf{Z}_{s,t}$, which are a vector of time-invariant and a vector of time-varying school characteristics, respectively. The components v_s and $u_{s,t}$ are time-invariant and time-varying unobserved school inputs, respectively. The variable c_d represents unobserved district characteristics.[23] We allow for educational productivity (or innate ability) to vary across individuals. Heterogeneous effects of school inputs and district characteristics on test scores are represented by a vector of coefficients,

$$\boldsymbol{\theta}_i = (\boldsymbol{\alpha}_i', \boldsymbol{\beta}_i', \gamma_i, \delta_i, \eta_i)'.$$

The potential outcomes model in equation (1) assumes that a student's potential test scores are determined by the interaction between the student's productivity (or ability), $\boldsymbol{\theta}_i$, and school and district characteristics,

$$\mathbf{F}_{s,t,d} = (\mathbf{X}_s', \mathbf{Z}_{s,t}', v_s, u_{s,t}, c_d)'.$$

Thus,

$$Y_i(s,t,d) = \boldsymbol{\theta}_i'\mathbf{F}_{s,t,d}.$$

The potential outcomes model in (1) is motivated by the textbook potential outcomes model defining $Y_{1i} - Y_{0i}$ as the treatment effect for person i when the treatment is binary.[24]

[22] We use 'high school senior' and 'student' interchangeably.

[23] The time effects are not included in equation (1) because each year's CSAT test scores are normalized.

[24] When treatment takes more than two values, the potential outcomes can be expressed by a function $f_i(k)$ for person i and treatment value k as described in Angrist and Pischke (2009). A well-known example is

Note that the textbook potential outcomes model is nonparametric and allows for treatment effect heterogeneity by nature. We assume the linear and interactive functional form in (1) to take into account the multi-dimensional nature of the treatment as each school is a package of various educational inputs. 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 outcomes model can be considered to be a natural extension of the textbook potential outcomes model.[25]

In our data, we observe (Y_i, S_i, T_i, D_i) at the individual level and $(\mathbf{X}'_s, \mathbf{Z}'_{s,t})$ at the school level. S_i denotes the school that student i attends. T_i denotes the senior year of student i, which also represents the birth cohort. D_i is the school district where student i lives and school S_i is located. Note that student i's CSAT score observed in the data set is determined by the realized school assignment, birth cohort, and district choice:

$$Y_i = Y_i\left(S_i, T_i, D_i\right). \tag{2}$$

The sampling scheme of the observed and unobserved variables is described in the following assumptions.

Assumption 1 (Individual Ability) *Individual abilities $\boldsymbol{\theta}_i$ are randomly drawn as $\boldsymbol{\theta}_i \sim$ i.i.d. $(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}})$.*

Assumption 2 (School and District Characteristics) *All of the school- and district-level variables, $\mathbf{F} = \{\mathbf{F}_{s,t,d} : s \in \mathcal{S}_d, t \in \mathcal{T}, d \in \mathcal{D}\}$, are fixed.*

Assumption 3 (District Choice) *The vector of individual district choices $\mathbf{D} = \{D_i : i \in \mathcal{I}\}$ is constant in repeated samples.*

when the treatment is years of education and the outcomes are earnings. In this case, $f_i(m) - f_i(n)$ is the earnings effect of person i getting m instead of n years of education. In a regression model there is a single error at all levels of education which implies that the joint distribution of potential outcomes at all levels of education is identified. Because this joint distribution is not identified we will not use the implications of the one-dimensional error, although we maintain it to facilitate the exposition.

[25] The potential outcomes model (1) suggests that we could potentially identify the joint distribution of potential outcomes at various levels of the inputs. This would deviate from the textbook model where only the marginal distributions of the treated and control outcomes are identified. However this feature of (1) is not used in the identification of the average partial effects.

Assumptions 1, 2, and 3 indicate that in repeated samples the individual coefficients $\boldsymbol{\theta}_i$ are randomly drawn from a population distribution. The observed and unobserved school and district characteristics are fixed, i.e. constant in repeated samples. The same applies to the district choice that is also constant in repeated samples.

Assumption 4 (Random School Assignment within Districts and No Cohort Effects) *By random assignment within districts and if there are no cohort effects we have that*

$$\boldsymbol{\theta}_i \perp (S_i, T_i)|D_i$$

The assumption that the school choice S_i and individual's 'ability' $\boldsymbol{\theta}_i$ are independent within school districts holds because high school assignments are made by a lottery within each school district for each cohort. The assumption that the distribution of $\boldsymbol{\theta}_i$ is the same in the two years does not directly follow from the random assignment. But, given that we are considering cohorts born in two subsequent years, it is reasonable to assume that the random coefficients have the same distribution in 2008 and 2009 within a school district.

In the above assumptions, we do not restrict the relationship between individual's district choice D_i and 'ability' $\boldsymbol{\theta}_i$. In a typical Roy model, that is based on individual and district characteristics, individuals self-select into the school district that gives them the highest expected benefits net of location costs. Therefore the distribution of $\boldsymbol{\theta}_i$ may differ between districts. However Assumption 1 implies that there is no direct effect of district characteristics on $\boldsymbol{\theta}_i$ which is considered to be an individual endowment. We do not (and do not need to) specify an explicit model for school district choice D_i. The following analysis is conditional on the district choices of the students (or their parents) (Assumption 3).

Before starting a discussion on the APE, we introduce notation that will be repeatedly used in the next section. Let $N_I(s,t,d) = \sum_{i \in \mathcal{I}} \mathbb{I}\{S_i = s, T_i = t, D_i = d\}$ denote the number of high school seniors at school s (in district d) in year t.[26] $N_I(s,d) = \sum_{t \in \mathcal{T}} N_I(s,t,d)$ is defined as the number of students at school s (in district d) in all years. Similarly, $N_I(d) =$

[26] $\mathbb{I}\{\cdot\}$ denotes an indicator function that takes value 1 if the statement in the curly brackets is true, and 0 otherwise.

$\sum_{s \in \mathcal{S}_d} N_I(s,d)$ is the number of students in district d in all years. $N_I = \sum_{d \in \mathcal{D}} N_I(d)$ is the total number of students in the sample. Given the notation for the number of individuals, $\hat{w}_d^s = \frac{N_I(s,d)}{N_I(d)}$ and $\hat{w}_{s,d}^t = \frac{N_I(s,t,d)}{N_I(s,d)}$ are consistent estimators of $w_d^s = \mathbb{P}\{S_i = s \mid D_i = d\}$, and $w_{s,d}^t = \mathbb{P}\{T_i = t \mid S_i = s, D_i = d\}$, respectively if $N_I \to \infty$.

5 Estimation of and Inference on Average Partial Effects

The parameters of interest are the APE of observed school inputs $\mathbf{X}_s$ and $\mathbf{Z}_{s,t}$ under district self-selection:

$$\begin{aligned}(\boldsymbol{\alpha}', \boldsymbol{\beta}')' &= \mathbb{E}[(\boldsymbol{\alpha}_i', \boldsymbol{\beta}_i')'] \\ &= \sum_{d \in \mathcal{D}} \mathbb{E}[(\boldsymbol{\alpha}_i', \boldsymbol{\beta}_i')' | D_i = d] \mathbb{P}\{D_i = d\}. \end{aligned} \tag{3}$$

The district choice probability, $w_d = \mathbb{P}\{D_i = d\}$, is estimated by the fraction of individuals choosing school district d, $\frac{N_I(d)}{N_I}$. Under Assumption 3, this estimate of w_d is constant (fixed and nonrandom) in repeated samples.

The average productivity of students in district d is denoted by

$$\boldsymbol{\theta}_d = (\boldsymbol{\alpha}_d', \boldsymbol{\beta}_d', \gamma_d, \delta_d, \eta_d)' = \mathbb{E}[(\boldsymbol{\alpha}_i', \boldsymbol{\beta}_i', \gamma_i, \delta_i, \eta_i)' | D_i = d].$$

We can express the APE under district self-selection defined in (3) as

$$(\boldsymbol{\alpha}', \boldsymbol{\beta}')' = \sum_{d \in \mathcal{D}} (\boldsymbol{\alpha}_d', \boldsymbol{\beta}_d')' w_d. \tag{4}$$

If individual district choices are independent of individual productivity, the average productivity is identical across districts, that is $(\boldsymbol{\alpha}', \boldsymbol{\beta}')' = (\boldsymbol{\alpha}_d', \boldsymbol{\beta}_d')'$. This holds even when productivity levels are heterogeneous across individuals. Thus, by allowing explicitly that the average productivity varies across school districts, we take into account the possibility of district selection being endogenous, i.e. dependent on $\boldsymbol{\theta}_i$.

Assumption 1 states that the individual productivities are i.i.d. This refers to the distribution before self-selection into districts. This assumption is therefore compatible with average productivity being different across school districts after self-selection into districts. Under Assumption 1, the APE under district self-selection is equal to the APE in the population. If the individual productivity depends on the school district, e.g. because of individual variation in the complementarity between school inputs and district characteristics, then the APE defined in (3) depends on the distribution of students over districts. Even in this case it is reasonable to assume that a small change in school inputs does not lead to relocation between school districts and the APE under district self-selection measures the average effect of a marginal change in school inputs on the CSAT score holding district choice constant. In the rest of this paper we assume that the APE under district self-selection is equal to the population APE.

5.1 Identification

To identify the APE $(\boldsymbol{\alpha}', \boldsymbol{\beta}')'$, we need to identify $(\boldsymbol{\alpha}'_d, \boldsymbol{\beta}'_d)'$ for each school district $d \in \mathcal{D}$. This subsection provides sufficient conditions under which the district-specific APE parameters $(\boldsymbol{\alpha}'_d, \boldsymbol{\beta}'_d)'$ are identified.

Based on the potential outcomes in equation (1) and the observed outcome in equation (2), the observed test score of student i attending an assigned school S_i in the district of her choice $D_i = d$ in year T_i is

$$\begin{aligned} Y_i &= \boldsymbol{\alpha}'_i \mathbf{X}_{S_i} + \boldsymbol{\beta}'_i \mathbf{Z}_{S_i,T_i} + \gamma_i v_{S_i} + \delta_i u_{S_i,T_i} + \eta_i c_d \\ &= \boldsymbol{\alpha}'_d \mathbf{X}_{S_i} + \boldsymbol{\beta}'_d \mathbf{Z}_{S_i,T_i} + \gamma_d v_{S_i} + \delta_d u_{S_i,T_i} + \eta_d c_d + \varepsilon_i \end{aligned} \tag{5}$$

where

$$\varepsilon_i = (\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_d)' \mathbf{X}_{S_i} + (\boldsymbol{\beta}_i - \boldsymbol{\beta}_d)' \mathbf{Z}_{S_i,T_i} + (\gamma_i - \gamma_d) v_{S_i} + (\delta_i - \delta_d) u_{S_i,T_i} + (\eta_i - \eta_d) c_d.$$

The independence of $\boldsymbol{\theta}_i$ and (S_i, T_i) conditional on $D_i = d$ in Assumption 4 implies

$$\mathbb{E}[\varepsilon_i | S_i, T_i, D_i = d] = 0. \tag{6}$$

The expectation is the average over the conditional distribution of student productivity $(\boldsymbol{\alpha}_i', \boldsymbol{\beta}_i', \gamma_i, \delta_i, \eta_i)'$ for students in school s, district d and year t.

Note that equation (5) is a linear regression model for data on students and schools in district d. In the district-specific regression, $\mathbf{X}_{S_i}$ and $\mathbf{Z}_{S_i,T_i}$ are regressors, $\gamma_d v_{S_i} + \delta_d u_{S_i,T_i} + \varepsilon_i$ is a composite regression error, and $\eta_d c_d$ is treated as the intercept.

To identify the regression coefficients $(\boldsymbol{\alpha}_d', \boldsymbol{\beta}_d')'$ in equation (5), we need to impose restrictions on the relation between observed and unobserved school characteristics.

Assumption 5 (Orthogonality Conditions) *For each $d \in \mathcal{D}$, we assume*

$$(i) \sum_{s \in \mathcal{S}_d} w_d^s v_s = 0, \qquad (ii) \sum_{s \in \mathcal{S}_d} w_d^s \mathbf{X}_s v_s = \mathbf{0}.$$

For each $d \in \mathcal{D}$, we assume

$$(iii) \sum_{s \in \mathcal{S}_d} w_d^s \sum_{t \in \mathcal{T}} w_{s,d}^t \left(\mathbf{Z}_{s,t} - \sum_{t \in \mathcal{T}} w_{s,d}^t \mathbf{Z}_{s,t} \right) \left(u_{s,t} - \sum_{t \in \mathcal{T}} w_{s,d}^t u_{s,t} \right) = \mathbf{0}$$

$$(iv) \sum_{s \in \mathcal{S}_d} w_d^s \sum_{t \in \mathcal{T}} w_{s,d}^t \left(\mathbf{X}_s - \sum_{s \in \mathcal{S}_d} w_d^s \mathbf{X}_s \right) u_{s,t} = \mathbf{0}$$

In Assumption 5 (i) and (ii), we assume no correlation between observed and unobserved time-constant school characteristics across schools. In (iii), we assume that the time-varying part of $\mathbf{Z}$ is uncorrelated with the time-varying part of u. The time constant components can be correlated. In (iv), we assume that the time-average of the time-varying unobserved school input is uncorrelated with the time-constant observed school inputs. Our model includes one time-varying and one time-invariant school characteristic of interest—class size and single-sex education, respectively. So for Assumption 5 to hold, we most likely would have to add additional school characteristics as controls, in which case the assumption would be that (*ii*), (*iii*), and (*iv*) hold for the variables of interest, conditional on these control variables. If this assumption does not hold, our estimate of the single-sex effect will be the sum of the APE of single-sex education and its indirect effect through the omitted time-constant school characteristics. For class-size, the estimated effect is the APE and the indirect effect through time-varying characteristics that are correlated with class size.

Under Assumption 5, the following equalities hold:

$$\mathbb{E}[\gamma_d v_{S_i}|D_i = d] = \gamma_d \sum_{s\in\mathcal{S}_d} w_d^s v_s = 0,$$
$$\mathbb{E}[\gamma_d \mathbf{X}_{S_i} v_{S_i}|D_i = d] = \gamma_d \sum_{s\in\mathcal{S}_d} w_d^s \mathbf{X}_s v_s = \mathbf{0},$$

with the expectations equal to average over the distribution of school choice S_i in district d.

Assumption 6 (Rank Conditions) *We assume that the matrices*

$$\sum_{s\in\mathcal{S}_d} w_d^s \sum_{t\in\mathcal{T}} w_{s,d}^t \left(\mathbf{Z}_{s,t} - \sum_{t\in\mathcal{T}} w_{s,d}^t \mathbf{Z}_{s,t}\right)\left(\mathbf{Z}_{s,t} - \sum_{t\in\mathcal{T}} w_{s,d}^t \mathbf{Z}_{s,t}\right)'$$

and

$$\sum_{s\in\mathcal{S}_d} w_d^s \left(\mathbf{X}_{s,d} - \sum_{s\in\mathcal{S}_d} w_d^s \mathbf{X}_{s,d}\right)\left(\mathbf{X}_{s,d} - \sum_{s\in\mathcal{S}_d} w_d^s \mathbf{X}_{s,d}\right)'$$

are positive definite.

Assumption 6 precludes multicollinearity of the observed school inputs.

By equations (5) and (6), we have

$$Y_i - \mathbb{E}[Y_i|S_i, D_i = d] - \boldsymbol{\beta}_d'\left(\mathbf{Z}_{S_i,T_i} - \sum_{t\in\mathcal{T}} w_{S_i,d}^t \mathbf{Z}_{S_i,t}\right) = \boldsymbol{\beta}_d'\left(u_{S_i,T_i} - \sum_{t\in\mathcal{T}} w_{S_i,d}^t u_{S_i,t}\right) + \varepsilon_i.$$

In this expression, $\mathbb{E}[Y_i|S_i, D_i = d]$ is the average outcome in school S_i and district d, i.e.

$$\mathbb{E}[Y_i|S_i, D_i = d] = \sum_{t\in\mathcal{T}} w_{S_i,d}^t \mathbb{E}[Y_i|S_i, T_i = t, D_i = d]$$

Under Assumption 5, the right-hand side error is uncorrelated with

$$\mathbf{Z}_{S_i,T_i} - \sum_{t\in\mathcal{T}} w^t_{S_i,d}\mathbf{Z}_{S_i,t}$$

so that if the inner $\mathbb{E}$ denotes an average over T_i and the outer an average over S_i, we have the moment condition that identifies $\boldsymbol{\beta}_d$.

Identification of $\boldsymbol{\beta}_d$: Under Assumptions 4, 5, and 6, the regression coefficient $\boldsymbol{\beta}_d$ can be identified by the following moment condition,

$$\begin{aligned}
&\mathbb{E}\left[\left\{\begin{array}{c} Y_i - \mathbb{E}[Y_i|S_i, D_i = d] \\ -\mathbf{b}'_d(\mathbf{Z}_{S_i,T_i} - \mathbb{E}[\mathbf{Z}_{S_i,T_i}|S_i, D_i = d]) \end{array}\right\}(\mathbf{Z}_{S_i,T_i} - \mathbb{E}[\mathbf{Z}_{S_i,T_i}|S_i, D_i = d])\middle| D_i = d\right] = \mathbf{0} \\
&\Leftrightarrow \mathbf{b}_d = \boldsymbol{\beta}_d
\end{aligned} \tag{7}$$

where uniqueness follows from Assumption 6.

Identification of $\boldsymbol{\alpha}_d$: Once $\boldsymbol{\beta}_d$ is identified, we can treat the parameter as if it were known. We have

$$Y_i - \mathbb{E}[Y_i|D_i = d] - \boldsymbol{\alpha}'_d\left(\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i}|D_i = d]\right) - \boldsymbol{\beta}'_d\left(\mathbf{Z}_{S_i,T_i} - \mathbb{E}[\mathbf{Z}_{S_i,T_i}|D_i = d]\right) =$$

$$\gamma_d(v_{S_i} - \mathbb{E}[v_{S_i}|D_i = d]) + \delta_d(u_{S_i,T_i} - \mathbb{E}[u_{S_i,T_i}|D_i = d]) + \varepsilon_i.$$

Then, under Assumptions 4 and 5 (*iii*) for the error term on the right-hand side, with the normalization $\mathbb{E}[v_{S_i}|D_i = d] = 0$ by district dummies in equation (5),

$$\mathbb{E}\left[(v_{S_i} - \mathbb{E}[v_{S_i}|D_i = d])\left(\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i}|D_i = d]\right)|D_i = d\right] = 0,$$

and with the normalization $\mathbb{E}[u_{S_i,T_i}|D_i = d] = 0$ by Assumption 5 (*iv*),

$$\mathbb{E}\left[(u_{S_i,T_i} - \mathbb{E}[u_{S_i,T_i}|D_i = d])\left(\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i}|D_i = d]\right)|D_i = d\right] = 0.$$

Therefore by Assumption 6, we can identify the regression coefficients $\boldsymbol{\alpha}_d$ by the following moment condition.

$$\mathbb{E}\left[\left\{\begin{array}{c} Y_i - \mathbb{E}[Y_i|D_i = d] - \boldsymbol{\beta}'_d(\mathbf{Z}_{S_i,T_i} - \mathbb{E}[\mathbf{Z}_{S_i,T_i}|D_i = d]) \\ -\mathbf{a}'_d(\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i}|D_i = d]) \end{array}\right\}(\mathbf{X}_{S_i} - \mathbb{E}[\mathbf{X}_{S_i}|D_i = d])\middle| D_i = d\right] = \mathbf{0}$$
$$\Leftrightarrow \mathbf{a}_d = \boldsymbol{\alpha}_d \tag{8}$$

5.2 Estimation and Statistical Inference

By a sample analog of equation (4), the estimated APE of school inputs is the weighted average of the estimated coefficients on time-invariant and time-varying school characteristics in regression equation (5). That is,

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

To estimate the district-specific effects of school inputs $(\boldsymbol{\alpha}'_d, \boldsymbol{\beta}'_d)'$, we implement a two-step estimation procedure using data on individual test scores and school characteristics from school district d. We first estimate the effect of time-varying school inputs, $\boldsymbol{\beta}_d$, and then estimate the effect of time-invariant school inputs, $\boldsymbol{\alpha}_d$, using the estimates obtained in the first step. $\hat{\boldsymbol{\alpha}}_d$ and $\hat{\boldsymbol{\beta}}_d$ obtained from the two-step estimation are consistent. The asymptotic variances of the school input effect estimators are derived in Appendix B.1 and B.2.

Step 1. Estimation of $\boldsymbol{\beta}_d$: The estimator that solves the sample version of the identifying moment condition in equation (7) is

$$\begin{aligned} \hat{\boldsymbol{\beta}}_d \;=\; & \left[\sum_{s\in\mathcal{S}_d}\sum_{i:S_i=s}\left(\mathbf{Z}_{S_i,T_i} - \frac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)\left(\mathbf{Z}_{S_i,T_i} - \frac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)'\right]^{-1} \\ & \times \sum_{s\in\mathcal{S}_d}\sum_{i:S_i=s}\left(\mathbf{Z}_{S_i,T_i} - \frac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)\left(Y_i - \frac{1}{N_I(s,d)}\sum_{i:S_i=s}Y_i\right). \end{aligned} \tag{9}$$

Note that $\hat{\boldsymbol{\beta}}_d$ is the within estimator obtained from the individual-level ordinary least squares (OLS) regression of Y_i on time-varying school inputs $\mathbf{Z}_{S_i,T_i}$ and school fixed effects using data from school district d.

Step 2. Estimation of $\boldsymbol{\alpha}_d$: The estimator that solves the sample version of the identifying moment condition in equation (8) is

$$\begin{aligned}\hat{\boldsymbol{\alpha}}_d &= \left[\sum_{i:D_i=d}\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)'\right]^{-1} \\ &\quad\times\sum_{i:D_i=d}\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)\left(Y_i-\hat{\boldsymbol{\beta}}_d'\mathbf{Z}_{S_i,T_i}\right). \qquad (10)\end{aligned}$$

Using data from school district d, $\hat{\boldsymbol{\alpha}}_d$ is computed by the individual-level OLS regression of $Y_i-\hat{\boldsymbol{\beta}}_d\mathbf{Z}_{S_i,T_i}$ on $\mathbf{X}_{S_i}$ with a constant term.

The OLS residuals from the first-step estimation are denoted by

$$\hat{\xi}_{i,d}=Y_i-\frac{1}{N_I(S_i,d)}\sum_{j:S_j=S_i}Y_j+\hat{\boldsymbol{\beta}}_d'\left(\mathbf{Z}_{S_i,T_i}-\frac{1}{N_I(S_i,d)}\sum_{k:S_k=S_i}\mathbf{Z}_{S_k,T_k}\right).$$

The OLS residuals from the second-step estimation are denoted by

$$\hat{\zeta}_{i,d} = Y_i-\widehat{\eta_d c_d}-\hat{\boldsymbol{\alpha}}_d'\mathbf{X}_{S_i}-\hat{\boldsymbol{\beta}}_d'\mathbf{Z}_{S_i,T_i},$$

where $\widehat{\eta_d c_d}$ is the estimated constant term.

We estimate the asymptotic variance of $\hat{\boldsymbol{\beta}}_d$ by using White's robust standard errors approach with clustering, to account for heteroskedastic errors in the random coefficient specification and the clustering of the individual outcomes on school and cohort in a school district:

$$\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_d)}=\hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}\hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1}$$

with

$$\hat{\mathbf{A}}_{\boldsymbol{\beta}}=\sum_{s\in\mathcal{S}_d}\hat{w}_d^s\sum_{t\in\mathcal{T}}\hat{w}_{s,d}^t\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)',$$

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}}=\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\left(\hat{w}_d^s\hat{w}_{s,d}^t\right)^2\bar{\xi}_{s,t,d}^2\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)',$$

and the average residual in cluster s, t, d,

$$\bar{\xi}_{s,t,d} = \frac{1}{N_I(s,t,d)} \sum_{i:S_i=s,T_i=t} \hat{\xi}_{i,d}.$$

The variance of $\hat{\boldsymbol{\alpha}}_d$ is estimated by

$$\widehat{\text{Var}(\hat{\boldsymbol{\alpha}}_d)} = \hat{\mathbf{A}}_{\boldsymbol{\alpha}}^{-1}(\hat{\boldsymbol{\Sigma}}_{11} + \hat{\boldsymbol{\Sigma}}_{22} - 2\hat{\boldsymbol{\Sigma}}_{12})\hat{\mathbf{A}}_{\boldsymbol{\alpha}}^{-1}$$

with

$$\begin{aligned}
\hat{\mathbf{A}}_{\boldsymbol{\alpha}} &= \sum_{s\in\mathcal{S}_d} \hat{w}_d^s \left(\mathbf{X}_{s,d} - \sum_{s\in\mathcal{S}_d} \hat{w}_d^s \mathbf{X}_{s,d}\right)\left(\mathbf{X}_{s,d} - \sum_{s\in\mathcal{S}_d} \hat{w}_d^s \mathbf{X}_{s,d}\right)', \\
\hat{\boldsymbol{\Sigma}}_{11} &= \sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} (\hat{w}_d^s \hat{w}_{s,d}^t)^2 \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s'}\right) \bar{\zeta}_{s,t,d}^2 \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s'}\right)', \\
\hat{\boldsymbol{\Sigma}}_{22} &= \hat{\mathbf{B}} \widehat{\text{Var}(\hat{\boldsymbol{\beta}}_d)} \hat{\mathbf{B}}', \\
\hat{\boldsymbol{\Sigma}}_{12} &= \sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} (\hat{w}_d^s \hat{w}_{s,d}^t)^2 \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s'}\right) \bar{\zeta}_{s,t,d}\bar{\xi}_{s,t,d} \left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}} \hat{w}_{s,d}^{t'} \mathbf{Z}_{s,t'}\right)' \hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1}\hat{\mathbf{B}}', \\
\hat{\mathbf{B}} &= \left[\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} \hat{w}_d^s \hat{w}_{s,d}^t \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d} \hat{w}_d^{s'} \mathbf{X}_{s'}\right) \mathbf{Z}_{s,t}'\right],
\end{aligned}$$

and the cluster average of the residuals

$$\bar{\zeta}_{s,t,d} = \frac{1}{N_I(s,t,d)} \sum_{i:S_i=s,T_i=t} \hat{\zeta}_{i,d}.$$

When computing the standard error of $\hat{\boldsymbol{\alpha}}_d$ in the second step, we take into account the sampling variation of $\hat{\boldsymbol{\beta}}_d$ estimated in the first step. Note that omitting $\hat{\boldsymbol{\Sigma}}_{22} - 2\hat{\boldsymbol{\Sigma}}_{12}$ yields White's robust standard errors without considering the sampling variation of $\hat{\boldsymbol{\beta}}_d$.

Recall that the sample analog estimator of the APE of school inputs is

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

where $w_d = \frac{1}{N_I}\sum_{i=1}^{N_I}\mathbb{I}\{D_i = d\}$. Also recall that district choices and thus the fraction of students in each district are fixed once they self-select across school districts. Therefore, the estimated variances of each element in $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ are:

$$\hat{\sigma}^2_{\hat{\alpha}} = \sum_{d\in\mathcal{D}} \mathrm{diag}(\widehat{\mathrm{Var}(\hat{\boldsymbol{\alpha}}_d)})w_d^2, \quad \hat{\sigma}^2_{\hat{\beta}} = \sum_{d\in\mathcal{D}} \mathrm{diag}(\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_d)})w_d^2,$$

where $\mathrm{diag}(\widehat{\mathrm{Var}(\hat{\boldsymbol{\alpha}}_d)})$ and $\mathrm{diag}(\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_d)})$ are diagonal elements of $\widehat{\mathrm{Var}(\hat{\boldsymbol{\alpha}}_d)}$ and $\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_d)}$, respectively.

5.3 Implications for Education Production Function

In this subsection, we show that a school-level education production function can be derived by aggregating the individual potential outcomes.[27] We also discuss properties of the school-level education production function implied by equations (1), (2) and Assumptions 1–4.

Under Assumption 4, the aggregated student ability in school s and year t is identical across schools and cohorts within each school district:

$$\mathbb{E}[\boldsymbol{\theta}_i|\ S_i = s, T_i = t, D_i = d] = \mathbb{E}[\boldsymbol{\theta}_i|D_i = d] = \boldsymbol{\theta}_d.$$

Given the potential outcome in equation (1) and the observed outcome in equation (2), the average test score in school s, year t, and district d becomes

$$\begin{aligned} \bar{Y}_{s,t,d} &= \frac{\sum_{i\in\mathcal{I}} Y_i \mathbb{I}\{S_i = s, T_i = t, D_i = d\}}{\sum_{i\in\mathcal{I}} \mathbb{I}\{S_i = s, T_i = t, D_i = d\}} \\ &= \frac{\sum_{i\in\mathcal{I}} Y_i(s,t,d) \mathbb{I}\{S_i = s, T_i = t, D_i = d\}}{\sum_{i\in\mathcal{I}} \mathbb{I}\{S_i = s, T_i = t, D_i = d\}} \\ &= \boldsymbol{\alpha}_d'\mathbf{X}_s + \boldsymbol{\beta}_d'\mathbf{Z}_{s,t} + \gamma_d v_s + \delta_d u_{s,t} + \eta_d c_d + \bar{\varepsilon}_{s,t,d}. \end{aligned} \tag{11}$$

Note that equation (11) is a school-level education production function. The school-level education production function derived by aggregating the individual-level outcomes has a

[27] Given that our data contain very limited information on student characteristics or family background, the aggregation procedure is similar in spirit to the derivation of a market demand function by aggregating over individual choices as in Berry et al. (1995).

noteworthy feature that the coefficients on the observed school inputs $(\boldsymbol{\alpha}_d', \boldsymbol{\beta}_d')'$ are constant across schools within each school district and over time due to Assumption 4. Random assignment within each school district is crucial for the within-district constant productivity when effects of school inputs are heterogeneous across individuals.[28]

If students with heterogeneous productivities were free to choose not only school districts but also their own schools, the school-level education production function becomes a correlated random coefficient model. That is, the aggregate productivity in the school-level education production function would be school specific and correlated with the observed school characteristics $(\mathbf{X}_s', \mathbf{Z}_{s,t}')$ even within each school district. In this case, it is difficult to identify the school-specific coefficients on school inputs $(\boldsymbol{\alpha}_s', \boldsymbol{\beta}_s')'$ using school-level data. In our setup under the within-district random assignment, on the contrary, the district-specific coefficients on school inputs $(\boldsymbol{\alpha}_d', \boldsymbol{\beta}_d')'$ are the district-specific averages of individual productivities $(\boldsymbol{\alpha}_i', \boldsymbol{\beta}_i')'$ and can be identified using school-level data.

If individual productivity is homogeneous, i.e. $(\boldsymbol{\alpha}_i', \boldsymbol{\beta}_i')' = (\boldsymbol{\alpha}', \boldsymbol{\beta}')'$, the effects of school inputs would be constant across schools even when students are self-selected into schools. However, no individual heterogeneity is a restrictive assumption.

Given that the school-level education production function takes the school-level panel regression form for each school district, we can implement the two-step estimation procedure using the two-period school-level panel data from each district to estimate the district-specific coefficients on school inputs $(\boldsymbol{\alpha}_d', \boldsymbol{\beta}_d')'$. First, $\boldsymbol{\beta}_d$ can be estimated by the pooled weighted least squares (WLS) regression of $\tilde{Y}_{s,t,d}$ on $\tilde{\mathbf{Z}}_{s,t,d}$ for $s \in \mathcal{S}_d$ and $t \in \mathcal{T}$:

$$\tilde{\boldsymbol{\beta}}_d = \left(\sum_{s \in \mathcal{S}_d} \hat{w}_d^s \sum_{t \in \mathcal{T}} \hat{w}_{s,d}^t \tilde{\mathbf{Z}}_{s,t,d} \tilde{\mathbf{Z}}_{s,t,d}' \right)^{-1} \sum_{s \in \mathcal{S}_d} \hat{w}_d^s \sum_{t \in \mathcal{T}} \hat{w}_{s,d}^t \tilde{\mathbf{Z}}_{s,t,d} \tilde{Y}_{s,t,d}, \tag{12}$$

where

$$\tilde{Y}_{s,t,d} = \bar{Y}_{s,t,d} - \sum_{t \in \mathcal{T}} \hat{w}_{s,d}^t \bar{Y}_{s,t,d}, \qquad \tilde{\mathbf{Z}}_{s,t,d} = \mathbf{Z}_{s,t} - \sum_{t \in \mathcal{T}} \hat{w}_{s,d}^t \mathbf{Z}_{s,t}.$$

[28]The time-invariant productivity is attributed to no cohort effect in Assumption 4.

Now, let $\bar{Y}_{s,\bullet,d} = \sum_{t\in\mathcal{T}} \hat{w}^t_{s,d}\bar{Y}_{s,t,d}$ and $\bar{\mathbf{Z}}_{s,\bullet,d} = \sum_{t\in\mathcal{T}} \hat{w}^t_{s,d}\mathbf{Z}_{s,t}$. Once we estimate $\boldsymbol{\beta}_d$, we can estimate $\boldsymbol{\alpha}_d$ by running the school-level WLS regression of $\bar{Y}_{s,\bullet,d} - \tilde{\boldsymbol{\beta}}'_d\bar{\mathbf{Z}}_{s,\bullet,d}$ on $\mathbf{X}_s$ and a constant term:

$$\begin{aligned}\tilde{\boldsymbol{\alpha}}_d &= \left[\sum_{s\in\mathcal{S}_d}\hat{w}^s_d\left(\mathbf{X}_s - \sum_{s\in\mathcal{S}_d}\hat{w}^s_d\mathbf{X}_s\right)\left(\mathbf{X}_s - \sum_{s\in\mathcal{S}_d}\hat{w}^s_d\mathbf{X}_s\right)'\right]^{-1} \\ &\quad\times\sum_{s\in\mathcal{S}_d}\hat{w}^s_d\left(\mathbf{X}_s - \sum_{s\in\mathcal{S}_d}\hat{w}^s_d\mathbf{X}_s\right)\left(\bar{Y}_{s,\bullet,d} - \tilde{\boldsymbol{\beta}}'_d\bar{\mathbf{Z}}_{s,\bullet,d}\right). \end{aligned} \tag{13}$$

Note that the weights $\hat{w}^s_d = \frac{N_I(s,d)}{N_I(d)}$ and $\hat{w}^t_{s,d} = \frac{N_I(s,t,d)}{N_I(s,d)}$ used in the WLS regressions are constructed based on the number of students.[29]

In Appendix B.3, we show that the two-step estimation procedures using the school-level WLS and the individual-level OLS are equivalent:

$$\hat{\boldsymbol{\beta}}_d = \tilde{\boldsymbol{\beta}}_d, \qquad \hat{\boldsymbol{\alpha}}_d = \tilde{\boldsymbol{\alpha}}_d.$$

6 Empirical Results

6.1 Average Partial Effect Estimates

The main estimation results are presented in Table 2. The estimated effects of single-sex schooling and class size on the CSAT score are reported for boys and girls separately.[30] Recall that the CSAT score refers to the standardized total score on Korean and English with an average of zero and standard deviation of one. Thus the estimates are also in standard deviations of the CSAT score. When we take the weighted average of district-specific estimates to compute the APE, we use the number of CSAT takers in each school district as the weight.[31]

[29] Suppose that we use alternative weights in the WLS estimation: $\hat{w}^s_d = \frac{1}{N_S(d)}$ and $\hat{w}^t_{s,d} = \frac{1}{T}$ that are constructed based on the number of schools instead of the number of students. In this case, we can obtain a within estimator $\breve{\boldsymbol{\beta}}_d$ from the school-level panel regression of $\bar{Y}_{s,t,d}$ on $\mathbf{Z}_{s,t}$ and school fixed effects, and an OLS estimator of $\breve{\boldsymbol{\alpha}}_d$ from the school-level regression of $\frac{1}{T}\sum_{t\in\mathcal{T}}(\bar{Y}_{s,t,d} - \breve{\boldsymbol{\beta}}'_d\mathbf{Z}_{s,t})$ on $\mathbf{X}_s$ and a constant term, using data from school district d. Choi et al. (2014) provides details on this estimation procedure.

[30] We conduct a separate analysis by gender as high school assignment lotteries are separate for male and female students.

[31] The APE estimates change little when we use the number of high school seniors in each district or the number of high school freshman at random assignment in each district as the weight.

Panel A presents our preferred estimates controlling for time-varying and time-invariant school characteristics that are potentially confounded with variables of interest—single-sex school indicator and class size. Time-varying control variables include total enrollment, the number of teachers, the fraction of regular teachers, the fraction of female teachers, and log of annual school spending in thousands of 2009 Korean won (KRW).[32] Time-invariant control variables include a private school indicator, school establishment year, and the interaction between the two variables.[33] The APE estimates change little when more control variables are added to the preferred specification.[34] The estimates with no controls are reported in panel B.

The APE estimates in the first row of panel A show that single-sex schools have nearly zero or insignificant negative effects on academic performance of both male and female students (-0.003 on boys' CSAT scores and -0.041 on girls'). It is particularly important to control for other confounding factors when estimating the effect of a time-invariant school input—single-sex schooling in this study. This is because we rely on the "selection on observables" in Assumption 5 to identify the APE of a time-invariant school input within each school district. Note that the large positive estimates of the single-sex school effect in panel B (0.148 for boys and 0.106 for girls) disappear when the school characteristics listed above are controlled for in the regression.[35][36]

[32] 1,000 KRW is worth approximately 1 USD.

[33] The private status and the age of school can control for unobserved teacher quality, which is known to have a large impact on students' academic achievement (Rivkin et al., 2005). This is because 1) public and private schools have different teacher hiring processes and 2) schools with longer history and stronger alumni network usually try to recruit better teachers to maintain their alumni power. Given that single-sex schools are three times more likely to be private and 23 years older than coed schools, unobserved teacher quality is likely to be systematically different between the two types of schools.

[34] We prefer a parsimonious specification as adding more control variables in the first-step would sacrifice the precision in the second-step estimation.

[35] In Choi et al. (2014), we obtained a much larger APE estimate of single-sex schools especially for boys using a different set of time-varying control variables, including the fraction of students receiving free or reduced price lunch, annual development fund spending per student, and the fraction of female teachers. Appendix Table A3 replicates the estimation results using the specification in Choi et al. (2014). As the specification in Choi et al. (2014) is likely subject to omitted variable bias, we use the updated list of time-varying control variables, including total enrollment, the number of teachers, the fraction of regular teachers, the fraction of female teachers, and log of annual school spending, in this study. After exploring many different sets of control variables, we have found that 1) the APE estimate of single-sex schooling is unstable without school size (represented by total enrollment and number of teachers) being controlled, 2) there are substantial measurement errors in the fraction of students receiving free or reduced price lunch, and 3) annual development fund spending per student is only a limited portion of school spending.

[36] Our finding is consistent with Ku and Kwak (2016)'s differences-in-differences estimates exploiting over-time changes in school types from single-sex to mixed-gender. Their within-school estimates indicate little

Smaller classes appear to slightly increase boys' test scores but have no effect on girls.' For boys, a decrease in class size by one student is associated with an increase in the CSAT scores by 0.014 standard deviations and the effect is statistically significant at the 1% level. For girls, on the other hand, the effect of the class size reduction is less than half in magnitude compared to boys, and not statistically significant at any conventional level. A simple back-of-the-envelope calculation based on the point estimate implies that a class size reduction from the current level to the OECD average (from 35 to 24) can increase the average test score among male high school students by about one seventh of a standard deviation.[37] The APE estimate of the class size effect is very robust to adding different sets of control variables possibly because much of the confounding factors are absorbed by school fixed effects. This can also be seen from the similarity between the estimated APE of class size reported in panels A (with controls) and B (without controls).[38]

The next six rows in panel A of Table 2 show district-specific coefficients on single-sex schooling and class size, including district selection effects. We observe that the estimated effects of single-sex schooling on the CSAT score vary substantially across school districts. For boys, the district-specific estimates range from a large negative effect (-0.727) in district 9 to a large positive effect (0.472) in district 3. The estimated effects on girls' CSAT score are also heterogeneous across districts (ranging from -0.672 to 0.294) but less so than for boys. The district-specific estimates of class size effects, on the other hand, are much more homogeneous across school districts than those of single-sex school effects. The estimated effects for both genders are approximately between -0.02 and 0. When we conduct χ^2 tests on the null hypothesis that the school input effects are uniform across school districts, the null hypothesis is rejected at the 5% significance level for single-sex school effects on both boys and girls, but not rejected at any conventional level of significance for class size effects on both genders. According to anecdotal evidence, the likelihood of getting into a single-sex school is indeed known to be one of major factors affecting students' and their parents' district choice

advantage of single-sex schooling especially in enhancing boys' academic performance.

[37]The OECD average class size is from Table 2.18 in TALIS 2013 Results: An International Perspective on Teaching and Learning (DOI: `http://dx.doi.org/10.1787/9789264196261-table108-en`).

[38]When control variables are omitted, the class size is positively correlated with the CSAT score for girls, but the magnitude of the estimate is small.

in Seoul.

The heterogeneous effects of school inputs imply a substantial degree of endogenous sorting across school districts. To understand the mechanism of sorting, however, we would need more information on individual characteristics from which we could infer how individual preference and productivity interact with school characteristics, which is beyond the scope of this paper.

In Appendix Tables A4–A6, we present estimation results using the CSAT score on each of the three sections—Korean, English, and Math.[39] The results from an analysis of the Korean or English score are qualitatively similar to the main results using the total of Korean and English scores in Table 2. There is more heterogeneity in district-specific effects on the English score than on the Korean score. When using the Math score as an outcome variable, the estimation results for boys are similar to the findings on the Korean or English score. However, we obtain somewhat different results for girls. When estimated without controls, the single-sex schooling effect on girls' Math scores (0.026) is only about a quarter of the effect on the other two subjects (0.106 and 0.093) or on boys' Math scores (0.122). When control variables are added to the regressions, the APE estimate on girls' Math scores doubles from 0.026 to 0.057 (although insignificant), whereas the large positive estimate disappears to zero for girls' Korean scores (from 0.106 to 0.008) or turns negative and insignificant for girls' English scores (from 0.093 to -0.080) and boys' Math scores (from 0.122 to -0.022). The results on the Math score need to be interpreted with caution, because of the endogeneity issue arising on the extensive (whether to take the Math exam or not) and the intensive (which type of Math exam to take) margins of exam taking as described in Section 3.

6.2 Comparison with Estimates from Other Empirical Strategies

In this subsection, we compare the estimated APE of single-sex schooling and class size based on our econometric framework and estimates from a linear regression model with district fixed effects. The district fixed effects model has been commonly used to exploit random assignment within districts but can produce biased estimates when heterogeneous individuals

[39] When the CSAT score on each section is used as an outcome variable, we normalize the test score to have zero mean and unit variance within each section.

sort endogenously across districts as in the case of Seoul's high school assignment.[40]

Since each cohort of students are randomly assigned to schools within each school district, we report results from a specification including district-year fixed effects (DFE) as in Lee et al. (2014); Ku and Kwak (2016); Hahn et al. (2016); Sohn (2016) using multi-year data. When we use district fixed effects instead of district-year fixed effects in the regression, estimation results change little. A linear regression with district-year fixed effects using data from all of the six school districts estimates the following equation:

$$Y_i = \boldsymbol{\alpha}'_{\mathrm{FE}}\mathbf{X}_{S_i} + \boldsymbol{\beta}'_{\mathrm{FE}}\mathbf{Z}_{S_i,T_i} + \sum_{t\in\mathcal{T},d\in\mathcal{D}} \lambda_{t,d}\mathbb{I}\{T_i = t, D_i = d\} + \chi_i.$$

As recently shown in Choi et al. (2016), the fixed effect estimators $(\hat{\boldsymbol{\alpha}}'_{\mathrm{FE}}, \hat{\boldsymbol{\beta}}'_{\mathrm{FE}})'$ can be biased when the district-specific effects $\boldsymbol{\theta}_d = (\boldsymbol{\alpha}'_d, \boldsymbol{\beta}'_d, \gamma_d, \delta_d, \eta_d)'$ (including sorting effects) and the school resources (i.e. treatments) are correlated cross school districts. This happens if individuals make their district choices based on the observed distribution of school resources in each school district. Intuitively, controlling for school districts would introduce selection bias as district choices are in part outcomes of the treatments in this case.

Table 3 presents the estimated effects of single-sex schooling and class size from our approach as well as from the DFE regressions. As shown in panel A with controls, the DFE model tends to overestimate the effect of single-sex schools. Compared to the negative APE estimates from our approach, the estimates from the DFE regressions are opposite in sign (positive) and larger in magnitude for both boys (-0.003 vs. 0.012) and girls (-0.041 vs. 0.061). The estimates from both methods are not significantly different from zero though. For class size effects, estimates from the DFE regressions are closer to zero (less negative) or slightly positive compared to our APE estimates although the difference is not large.[41]

It makes sense that our approach addressing the district selection problem yields smaller or more negative estimates of school input effects compared to the DFE regression. Endoge-

[40] See Choi et al. (2016) for discussions on inconsistency of ATE estimates in a linear regression with site fixed effects under endogenous site selection.

[41] As reported in panel B without controls, the two approaches happen to yield very similar estimates on the single-sex school effect, whereas estimates from the DFE regressions are substantially larger than our estimates on the class size effect.

nous migration across school districts can amplify the positive effects and arbitrage away the negative effects of school inputs. This would produce an upward bias in estimates from empirical strategies not dealing with the endogenous selection problem, and the bias would be larger for a school input that induces more sorting—school coed type in our example.

We also report estimates from pooled OLS regressions without district or district-year fixed effects, which would be biased as neither within-district random assignment nor endogenous sorting across districts is taken into account. The OLS estimates of single-sex and class size effects are indeed quite different from our APE estimates or the DFE estimates. Also, there does not seem to be a systematic pattern in the sign or magnitude of the bias.

7 Conclusion

In this study, we propose an econometric framework to estimate the average partial effects (APE) of school inputs on students' academic performance exploiting within-district high school lotteries in Seoul, Korea. Our estimation method is implemented using data on CSAT scores and high school characteristics from Seoul in 2008–2009. Identification relies on the fact that self-selection arises across school districts but not into schools within each district. We focus on the effect of single-sex schooling and class size to illustrate the APE estimation of time-invariant and time-varying school inputs.

In our potential outcomes model, each student's academic outcomes are determined by observed and unobserved school inputs interacted with the student's productivity. From the random coefficient model of individual-level education production, we derive a district-specific linear regression equation of individual academic outcomes on school inputs. The district-specific effects of school inputs are estimated in two steps within each district: the coefficients on time-varying school inputs are estimated first and the coefficients on time-invariant school inputs are estimated in the next step based on the first-step estimates. The APE estimates are computed by taking the weighted average of the district-specific effect estimates.

Our empirical analysis finds that single-sex schools have zero effect on boys' CSAT scores and an insignificant negative effect on girls.' Also, smaller class sizes have a small positive effect on boys' test scores but no effect on girls.' There is a substantial degree of across-district heterogeneity especially in the effect of single-sex schools, which confirms that endogenous sorting does occur in school district choice. We also show that estimates from a conventional linear regression with district(-year) fixed effects, which ignore endogenous sorting across school districts, are subject to upward selection bias. To understand the mechanism of sorting, however, we would need to know how individual preference and productivity interact with school characteristics, which we leave for future research.

References

Altonji JG, Mansfield RK. 2014. Group-average observables as controls for sorting on unobservables when estimating group treatment effects: the case of school and neighborhood effects. Working Paper 20781, National Bureau of Economic Research. DOI: 10.3386/w20781.

Angrist JD, Lavy V. 1999. Using maimonides' rule to estimate the effect of class size on scholastic achievement. *Quarterly Journal of Economics* **114**: 533–575. DOI: 10.1162/003355399556061.

Angrist JD, Pischke JS. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Bastos P, Cristia J, Kim B. 2016. Good schools or good students? Evidence on school effects from universal random assignment of students to high schools. Working Paper: `link`.

Bayer P, Ferreira F, McMillan R. 2007. A unified framework for measuring preferences for schools and neighborhoods. *Journal of Political Economy* **115**: 588–638. DOI: 10.1086/522381.

Berry S, Levinsohn J, Pakes A. 1995. Automobile prices in market equilibrium. *Econometrica* **63**: 841–890. DOI: 10.2307/2171802.

Choi EJ, Moon HR, Ridder G. 2014. Estimation of an education production function under random assignment with selection. *American Economic Review: Papers and Proceedings* **104**: 206–211. DOI: 10.1257/aer.104.5.206.

Choi EJ, Moon HR, Ridder G. 2016. The (l)ate and fixed effects in multi-site experiments with endogenous site selection. Working Paper.

Dearden L, Ferri J, Meghir C. 2002. The effect of school quality on educational attainment and wages. *Review of Economics and Statistics* **84**: 1–20. DOI: 10.1162/003465302317331883.

Hahn S, Sung T, Baek J. 2008. Mixing versus sorting: Entering top universities. *Economics Letters* **100**: 43–46. DOI: 10.1016/j.econlet.2007.10.028.

Hahn Y, Wang CL, Yang HS. 2016. Does greater school autonomy make a difference? Evidence from a randomized natural experiment in South Korea. Working Paper: `link`.

Hanushek EA. 2006. School resources. In Hanushek E, Welch F (eds.) *Handbook of the Economics of Education*, volume 2, chapter 14. Elsevier, 865–908. DOI: 10.1016/S1574-0692(06)02014-9.

Hill AJ. 2015. The girl next door: The effect of opposite gender friends on high school achievement. *American Economic Journal: Applied Economics* **7**: 147–177. DOI: 10.1257/app.20140030.

Hoxby C. 2000. Peer effects in the classroom: Learning from gender and race variation. Working Paper 7867, National Bureau of Economic Research. DOI: 10.3386/w7867.

Hwang SI. 2016. A robust redesign of high school match. Working Paper: `link`.

Jackson CK. 2012. Single-sex schools, student achievement, and course selection: Evidence from rule-based student assignments in Trinidad and Tobago. *Journal of Public Economics* **96**: 173–187. DOI:10.1016/j.jpubeco.2011.09.002.

Kang C. 2007. Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics* **61**: 458–495. DOI: 10.1016/j.jue.2006.07.006.

Kang C, Park C, Lee MJ. 2007. Effects of ability mixing in high school on adulthood earnings: quasiexperimental evidence from South Korea. *Journal of Population Economics* **20**: 269–297. DOI: 10.1007/s00148-006-0090-y.

Kim T, Lee JH, Lee Y. 2008. Mixing versus sorting in schooling: Evidence from the equalization policy in South Korea. *Economics of Education Review* **27**: 697–711. DOI: 10.1016/j.econedurev.2006.07.014.

Krueger AB. 1999. Experimental estimates of education production functions. *Quarterly Journal of Economics* **114**: 497–532. DOI: 10.1162/003355399556052.

Ku H, Kwak DW. 2016. Together or separate: Disentangling the effects of single-sex schooling from the effects of single-sex schools. Working Paper: `link`.

Lavy V, Schlosser A. 2011. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics* **3**: 1–33. DOI: 10.1257/app.3.2.1.

Lee S, Turner LJ, Woo S, Kim K. 2014. All or nothing? The impact of school and classroom gender composition on effort and academic achievement. Working Paper 20722, National Bureau of Economic Research. DOI: 10.3386/w20722.

Meghir C, Rivkin S. 2011. Econometric methods for research in education. In Eric A Hanushek SM, Woessmann L (eds.) *Handbook of the Economics of Education*, volume 3, chapter 1. Elsevier, 1–87. DOI: 10.1016/B978-0-444-53429-3.00001-6.

Ministry of Education, KDI School of Public Policy and Management. 2013. 2012 Modularization of Koreas development experience: Role of private schools in Korea's educational development. `link`.

Park H, Behrman JR, Choi J. 2013. Causal effects of single-sex schools on college entrance exams and college attendance: Random assignment in Seoul high schools. *Demography* **50**: 447–469. DOI: 10.1007/s13524-012-0157-1.

Rivkin SG, Hanushek EA, Kain JF. 2005. Teachers, schools, and academic achievement. *Econometrica* **73**: 417–458. DOI: 10.1111/j.1468-0262.2005.00584.x.

Rothstein JM. 2006. Good principals or good peers? Parental valuation of school characteristics, tiebout equilibrium, and the incentive effects of competition among jurisdictions. *American Economic Review* **96**: 1333–1350. DOI: 10.1257/aer.96.4.1333.

Seoul Metropolitan Office of Education. 2008. High school student assignment plan (in Korean). `link`.

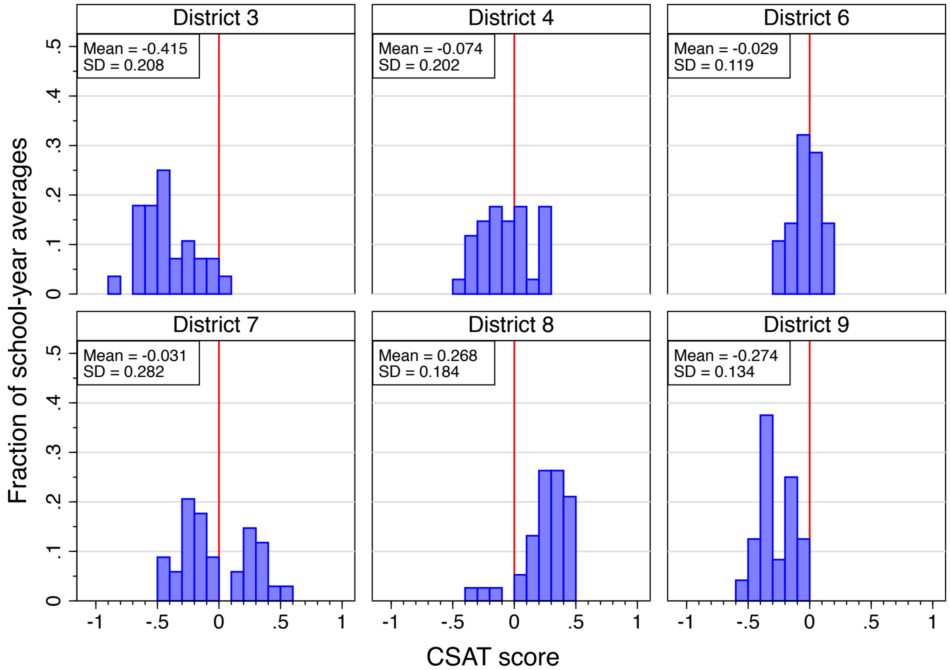
Sohn H. 2016. Mean and distributional impact of single-sex high schools on students' cognitive achievement, major choice, and test-taking behavior: Evidence from a random assignment policy in Seoul, Korea. *Economics of Education Review* **52**: 155–175. DOI: 10.1016/j.econedurev.2016.02.007.

Urquiola M. 2005. Does school choice lead to sorting? Evidence from tiebout variation. *American Economic Review* **95**: 1310–1326. DOI: 10.1257/0002828054825484.

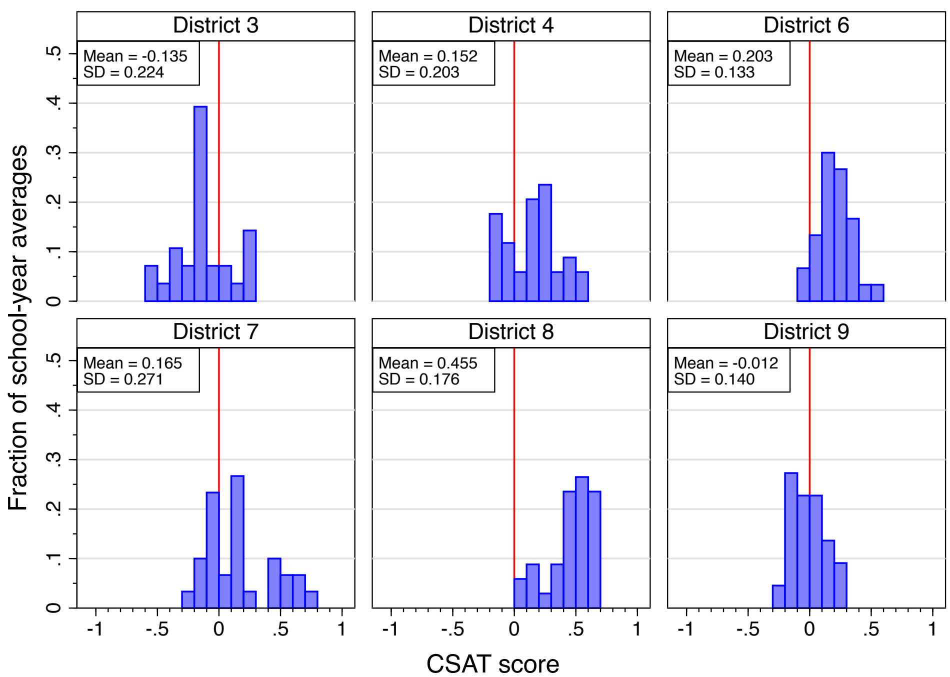
Whitmore D. 2005. Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review: Papers and Proceedings* **95**: 199–203. DOI: 10.1257/000282805774670158.

Figure 1: The School-Year Average CSAT Score by School District

A. Boys



B. Girls



Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The unit of observation is school-year as the CSAT score is averaged at the school-year level. The CSAT score refers to the total of Korean and English CSAT scores and is standardized to have zero mean and unit variance across all test takers in a given year. For CSAT applicants who were absent for the exam, missing raw scores on Korean or English are imputed with zeros. The mean and standard deviation in the top left box in each bar graph are computed using school-year average test scores from each school district.

Table 1: Descriptive Statistics

	Boys			Girls		
	Mean [SD]	% within-dist. variation[a]	Obs.[b]	Mean [SD]	% within-dist. variation[a]	Obs.[b]
A. Student-level CSAT scores[c]						
Combined (Korean and English)	-0.023 [1.001]	95.1	57,443	0.205 [0.894]	95.5	52,271
Korean	-0.082 [1.015]	97.0	57,443	0.166 [0.898]	97.5	52,271
English	0.032 [0.976]	94.3	57,443	0.214 [0.890]	94.4	52,271
Math	0.092 [0.913]	98.1	57,443	-0.032 [1.096]	99.4	52,271
B. School-level average CSAT scores[c]						
Combined (Korean and English)	-0.066 [0.294]	44.2	186	0.155 [0.270]	51.9	178
Korean	-0.119 [0.240]	47.5	186	0.126 [0.211]	57.1	178
English	-0.012 [0.309]	42.8	186	0.162 [0.293]	49.1	178
Math	0.064 [0.184]	57.5	186	-0.052 [0.186]	79.6	178
C. School-level characteristics						
Single-sex school	0.409 [0.494]	97.2	93	0.382 [0.489]	98.7	89
Private school	0.495 [0.503]	91.1	93	0.461 [0.501]	91.6	89
School establishment year	1977.0 [25.0]	88.9	93	1978.3 [25.9]	85.6	89
Class size	35.2 [2.5]	42.1	186	35.5 [2.6]	40.6	178
Total enrollment	1,422.8 [279.2]	76.1	186	1,398.7 [288.6]	68.6	178
Number of teachers	83.6 [13.5]	79.5	186	82.8 [15.0]	78.9	178
Fraction of female teachers	0.400 [0.190]	90.5	186	0.533 [0.120]	88.5	178
Fraction of regular teachers	0.895 [0.067]	92.0	186	0.892 [0.065]	89.8	178
Annual school spending (in 1000's of 2009 KRW)	2,146,877 [1,052,292]	88.4	186	2,014,230 [893,447]	87.1	178

Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The CSAT scores are standardized to have zero mean and unit variance. 1,000 KRW is worth approximately 1 USD. Standard deviations in brackets.

[a] For time-invariant variables, the proportion of the within-district sum of squares in the total sum of squares is reported. For time-varying variables, the proportion of the within-district-year sum of squares in the total sum of squares is reported.

[b] The unit of observation is individual for student-level variables, school for time-invariant school-level variables, and school-year for time-varying school-level variables.

[c] For CSAT applicants who were absent for the exam, missing raw scores are imputed with zeros.

Table 2: The Effect of School Inputs on the CSAT Score

	Boys				Girls			
	Single-sex		**Class size**		**Single-sex**		**Class size**	
A. With controls								
All districts								
APE	-0.003	(0.084)	-0.014	(0.003)***	-0.041	(0.068)	-0.006	(0.002)**
By district								
District 3	0.472	(0.171)***	0.000	(0.009)	0.041	(0.123)	0.002	(0.006)
District 4	0.068	(0.294)	-0.016	(0.008)*	0.294	(0.137)**	-0.009	(0.004)**
District 6	-0.004	(0.049)	-0.022	(0.006)***	-0.077	(0.096)	0.000	(0.005)
District 7	-0.115	(0.218)	-0.020	(0.006)***	-0.672	(0.279)**	-0.016	(0.006)***
District 8	0.116	(0.111)	-0.007	(0.005)	0.162	(0.080)**	-0.004	(0.006)
District 9	-0.727	(0.361)**	-0.017	(0.009)*	0.086	(0.168)	-0.004	(0.008)
Testing school input effects identical across districts								
χ^2-statistic	12.61	[0.027]	7.50	[0.186]	13.48	[0.019]	7.53	[0.184]
B. No controls								
All districts								
APE	0.148	(0.027)***	-0.010	(0.002)***	0.106	(0.029)***	0.005	(0.001)***
By district								
District 3	0.068	(0.078)	-0.004	(0.007)	0.085	(0.077)	0.003	(0.005)
District 4	0.214	(0.061)***	-0.014	(0.005)***	0.200	(0.052)***	0.010	(0.003)***
District 6	0.034	(0.045)	-0.015	(0.002)***	0.018	(0.053)	0.004	(0.002)*
District 7	0.190	(0.090)**	0.001	(0.004)	0.029	(0.109)	0.008	(0.003)***
District 8	0.186	(0.047)***	-0.008	(0.005)*	0.146	(0.044)***	0.011	(0.003)***
District 9	0.189	(0.045)***	-0.024	(0.006)***	0.192	(0.058)***	-0.013	(0.005)***
Testing school input effects identical across districts								
χ^2-statistic	10.43	[0.064]	18.71	[0.002]	8.76	[0.119]	21.75	[0.001]

Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The CSAT scores refers to the total of Korean and English CSAT scores and is standardized to have zero mean and unit variance. For CSAT applicants who were absent for the exam, missing raw scores on Korean or English are imputed with zeros. Time-varying control variables include total enrollment, the number of teachers, the fraction of regular teachers, the fraction of female teachers, and log annual school spending in thousands of 2009 KRW. Time-invariant control variables include a private indicator, school establishment year, and the interaction between the two. Robust standard errors clustered at the school-year level in parentheses. Section 5.2 provides more details on standard error computation. p-values associated with χ^2-statistics in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Comparison with Other Empirical Strategies

	Boys			Girls		
	Our approach	**DFE**	**OLS**	**Our approach**	**DFE**	**OLS**
A. With controls						
Single-sex	-0.003	0.012	-0.070	-0.041	0.061	0.041
	(0.084)	(0.047)	(0.053)	(0.068)	(0.052)	(0.055)
Class size	-0.014	-0.008	-0.033	-0.006	0.003	-0.019
	(0.003)***	(0.011)	(0.009)***	(0.002)**	(0.010)	(0.008)**
B. No controls						
Single-sex	0.148	0.148	0.196	0.106	0.099	0.145
	(0.027)***	(0.027)***	(0.041)***	(0.029)***	(0.029)***	(0.040)***
Class size	-0.010	0.009	-0.009	0.005	0.031	0.013
	(0.002)***	(0.009)	(0.009)	(0.001)***	(0.008)***	(0.008)*

Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The CSAT score refers to the total of Korean and English CSAT scores and is standardized to have zero mean and unit variance. For CSAT applicants who were absent for the exam, missing raw scores on Korean or English are imputed with zeros. Time-varying control variables include total enrollment, the number of teachers, the fraction of regular teachers, the fraction of female teachers, and log annual school spending in thousands of 2009 KRW. Time-invariant control variables include a private indicator, school establishment year, and the interaction between the two. Robust standard errors clustered at the school-year level in parentheses. See Section 5.2 for more details on standard error computation for our approach. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Online Appendix (Not for Publication)

A Validity of the Within-District Random Assignment Design

To verify if high school assignment was random within school districts before 2010, we use data from the Korean Education Longitudinal Study (KELS). It is not feasible to test the random assignment design using our main data set as it does not contain variables on pretreatment individual characteristics. The KELS provides data on learning experiences of a nationally representative sample of 6,908 seventh-graders who were first surveyed in 2005 and are followed up every year. The data are collected from students, parents, teachers and school principals, similar to the National Education Longitudinal Studies in the US.

We analyze the KELS sample of students who entered high schools as 10th graders in 2008. Given that the high school random assignment lotteries had been effective until 2009, the KELS data can be used to verify whether the random assignment system worked properly. By linking the first, third and fourth waves of the KELS, we construct a data set containing students' baseline characteristics, including parental education, household income, expenditures on private out-of-school education, and the 9th-grade standardized test score, before they are assigned to high schools.[42]

To check whether the baseline characteristics are balanced between single-sex and coeducational high schools as well as across high schools with different class sizes within school districts, we test the null hypothesis that coefficients on baseline characteristics are all equal to zero in the following regression model:

$$Q_i = \pi_0 + \boldsymbol{\pi}_1'\mathbf{W}_i + \sum_m \phi_m \mathbb{I}\{M_i = m\} + e_i. \tag{14}$$

Here, Q_i is a treatment variable—single-sex status or average class size of the high school attended by student i. $\mathbf{W}_i$ is a vector of student i's baseline characteristics. M_i indicates which middle school student i attended.

By including middle school fixed effects in the regression (14), we compare students who had attended the same middle school but were assigned to different high schools. In Korea, middle school assignment is random within school districts (or within middle school attendance zones that are usually smaller than school districts).[43] Thus students from the same middle school can be thought of as from the same school district and randomly assigned to general academic high schools within the school district, as long as their school districts are under the High School Equalization Policy (HSEP) and their high schools are subject to the random assignment lotteries.[44] This institutional setting helps overcome the limitation of the KELS data in which school district identifiers or school names are not provided. Note also that, due to randomness in the middle school assignment process, selection bias from endogenous sorting is not a major issue in the middle school fixed effects regression unlike in the regression analysis including school district fixed effects.

Table A2 reports estimated coefficients on students' baseline characteristics and their joint significance in the balance check regression (14). To increase the power of the test, we pool observations in all regions without restricting the sample only to Seoul. In panel A, we first do the estimation and the balance check only with those attending "random high schools," which are general aca-

[42]The parents' educational attainment is from the first wave of KELS. All the other baseline characteristics are from the third wave when the students were in the ninth grade. The fourth wave of the survey provides information on high schools attended by the KELS sample of students.

[43]See Kang (2007); Lee et al. (2014) for more details on the middle school assignment process in Korea.

[44]We thank Beomsoo Kim for sharing the idea of comparing students within middle schools, which also appears in his recent work (Bastos et al., 2016).

demic high schools subject to the random assignment lotteries in regions under the HSEP. For boys, coefficients on students' baseline characteristics are jointly insignificant at the 5 (10) percent level when the dependent variable is single-sex (class size). For girls, the baseline characteristics cannot explain the variation in both of the treatment variables (with p-values > 0.48).

In panel B, we repeat this exercise for students attending high schools not subject to the random assignment lotteries. The "non-random high schools" are either schools in the non-HSEP regions or selective or vocational high schools in the HSEP regions. In contrast to the results in panel A, the null hypothesis of zero coefficients on pre-treatment characteristics can be rejected at any conventional level of significance in all cases by treatment variables and gender (with p-values $<$ 0.006).

Note that the baseline characteristics may be weakly correlated with high school characteristics across students from the same middle school, even when we restrict to those attending high schools subject to the random assignment lotteries. Suppose that some students have moved to another school district after graduating from middle school but before entering high school, and then randomly assigned to a high school in their new school districts. The higher the fraction of the movers, the stronger the correlation between student attributes and high school characteristics. This is probably why the different between single-sex and mixed-gender high schools in male students' baseline characteristics is statistically significant at the 10 percent level (although insignificant at the 5 percent level). This is consistent with the anecdotal evidence in Korea that parents prefer single-sex schools especially when their child is male.

B Proofs

B.1 Variance of $\hat{\boldsymbol{\beta}}_d$

We derive the variance of the within estimator in Section 5.2. In this derivation, we have to consider two complications. First, the data are clustered in school-cohort groups and all students in a cluster share the same unobservable time-varying school input $u_{s,t}$. Second, the individual error terms ε_i are heteroskedastic, as they are derived from a random coefficient regression model. This model implies that the variance of the error varies in clusters, i.e. it is $\sigma^2_{s,d,t}$. The number of students in each school district d, i.e. $N_I(d)$, is assumed to become large. However, the number of schools in each district and the number of cohorts is fixed.

Because for all $s \in \mathcal{S}_d$

$$\sum_{i:S_i=s}\left(\mathbf{Z}_{S_i,T_i} - \frac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right) = 0,$$

we have for the within estimator

$$\begin{aligned}\hat{\boldsymbol{\beta}}_d - \boldsymbol{\beta}_d \quad = \quad & \left[\tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{i:S_i=s}\left(\mathbf{Z}_{S_i,T_i} - \tfrac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)\left(\mathbf{Z}_{S_i,T_i} - \tfrac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)'\right]^{-1} \\ & \times \tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{i:S_i=s}\left(\mathbf{Z}_{S_i,T_i} - \tfrac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)\left(\delta_d u_{S_i,T_i} + \varepsilon_i\right),\end{aligned}$$

where $N_I(s,d)$ denotes the number of students in school s that is in district d.

We consider the second factor. Because

$$\sum_{i:S_i=s}\left(\mathbf{Z}_{S_i,T_i}-\frac{1}{N_I\left(s,d\right)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)\left(\delta_d u_{S_i,T_i}+\varepsilon_i\right)=$$

$$\sum_{t\in\mathcal{T}}N_I(s,t,d)\left(\mathbf{Z}_{s,t}-\frac{1}{N_I\left(s,d\right)}\sum_{t'\in\mathcal{T}}N_I(s,t',d)\mathbf{Z}_{s,t'}\right)\left(\delta_d u_{s,t}+\frac{1}{N_I(s,t,d)}\sum_{i:S_i=s,T_i=t}\varepsilon_i\right),$$

the second factor is

$$\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\frac{N_I(s,t,d)}{N_I\left(d\right)}\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}\frac{N_I(s,t',d)}{N_I\left(s,d\right)}\mathbf{Z}_{s,t'}\right)\left(\delta_d u_{s,t}+\frac{1}{N_I(s,t,d)}\sum_{i:S_i=s,T_i=t}\varepsilon_i\right)=$$

$$\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\hat{w}_d^s\hat{w}_{s,d}^t\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\delta_d u_{s,t}+\frac{1}{N_I(s,t,d)}\sum_{i:S_i=s,T_i=t}\varepsilon_i\right).$$

Note that if $N_I(s,t,d)\to\infty$ for all s,t, then

$$\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\hat{w}_d^s\hat{w}_{s,d}^t\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\delta_d u_{s,t}+\frac{1}{N_I(s,t,d)}\sum_{i:S_i=s,T_i=t}\varepsilon_i\right)\xrightarrow{p}$$

$$\sum_{s\in\mathcal{S}_d}w_d^s\sum_{t\in\mathcal{T}}w_{s,d}^t\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}w_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\delta_d u_{s,t}=0$$

by Assumption 5 (iii), so that $\hat{\boldsymbol{\beta}}_d$ is weakly consistent for $\boldsymbol{\beta}_d$.

Because the statistics, $\hat{w}_d^s$ and $\hat{w}_{s,d}^t$, are ancillary for $\boldsymbol{\beta}_d$ and independent of ε_i, we condition on them, i.e. we treat them as fixed constants. If we define

$$\hat{\mathbf{A}}_{\boldsymbol{\beta}}=\sum_{s\in\mathcal{S}_d}\hat{w}_d^s\sum_{t\in\mathcal{T}}\hat{w}_{s,d}^t\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)'$$

with the probability limit $\mathbf{A}_{\boldsymbol{\beta}}$, then in large samples

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_d)=\mathbf{A}_{\boldsymbol{\beta}}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{\beta}}\mathbf{A}_{\boldsymbol{\beta}}^{-1}$$

with

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}}=\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\left(w_d^s w_{s,d}^t\right)^2\left(\delta_d^2u_{s,t}^2+\frac{\sigma_{s,t,d}^2}{N_I(s,t,d)}\right)\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}w_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}w_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)'.$$

For the cross-product terms

$$\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}w_d^s w_{s,d}^t\left(\mathbf{Z}_{s,t}-\sum_{t'\in\mathcal{T}}w_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\delta_d u_{s,t}+\frac{1}{N_I(s,t,d)}\sum_{i:S_i=s,T_i=t}\varepsilon_i\right)\cdot\left(\delta_d u_{p,q}+\frac{1}{N_I(p,q,d)}\sum_{i:S_i=p,T_i=q}\varepsilon_i\right)\xrightarrow{p}0$$

by Assumption 5 (iii) if $p \neq s, q \neq t$. For the other cross-product terms, we assume

$$\sum_{s \in \mathcal{S}_d} \sum_{t \in \mathcal{T}} (w_d^s)^2 w_{s,d}^t w_{s,d}^q \left(\mathbf{Z}_{s,t} - \sum_{t' \in \mathcal{T}} w_{s,d}^{t'} \mathbf{Z}_{s,t'} \right) \delta_d^2 u_{s,t} u_{s,q} \left(\mathbf{Z}_{s,q} - \sum_{t' \in \mathcal{T}} w_{s,d}^{t'} \mathbf{Z}_{s,t'} \right)' = 0$$

and

$$\sum_{s \in \mathcal{S}_d} \sum_{t \in \mathcal{T}} w_d^s w_{s,d}^t w_{p,d}^t \left(\mathbf{Z}_{s,t} - \sum_{t' \in \mathcal{T}} w_{s,d}^{t'} \mathbf{Z}_{s,t'} \right) \delta_d^2 u_{s,t} u_{p,t} \left(\mathbf{Z}_{p,t} - \sum_{t' \in \mathcal{T}} w_{p,d}^{t'} \mathbf{Z}_{p,t'} \right)' = 0.$$

These assumptions amount to assuming that the unobserved time-varying inputs for a school are not correlated over time, nor are they correlated among schools in the same time-period.

To estimate the variance of the within estimator, we define the residuals

$$\hat{\xi}_{i,d} = Y_i - \frac{1}{N_I(S_i, d)} \sum_{j: S_j = S_i} Y_j + \hat{\boldsymbol{\beta}}_d' \left(\mathbf{Z}_{S_i, T_i} - \frac{1}{N_I(S_i, d)} \sum_{k: S_k = S_i} \mathbf{Z}_{S_k, T_k} \right).$$

Define the average residual in cluster s, t, d by

$$\overline{\xi}_{s,t,d} = \frac{1}{N_I(s,t,d)} \sum_{i: S_i = s, T_i = t} \hat{\xi}_{i,d}.$$

A consistent estimator of the large-sample variance is

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_d) = \hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} \hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1}$$

with

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}} = \sum_{s \in \mathcal{S}_d} \sum_{t \in \mathcal{T}} \left(\hat{w}_d^s \hat{w}_{s,d}^t \right)^2 \overline{\xi}_{s,t,d}^2 \left(\mathbf{Z}_{s,t} - \sum_{t' \in \mathcal{T}} \hat{w}_{s,d}^{t'} \mathbf{Z}_{s,t'} \right) \left(\mathbf{Z}_{s,t} - \sum_{t' \in \mathcal{T}} \hat{w}_{s,d}^{t'} \mathbf{Z}_{s,t'} \right)'.$$

B.2 Variance of $\hat{\boldsymbol{\alpha}}_d$

As shown earlier, $\hat{\boldsymbol{\alpha}}_d$ is the OLS estimator of $\boldsymbol{\alpha}_d$ in the individual-level regression of $Y_i - \hat{\boldsymbol{\beta}}_d' \mathbf{Z}_{S_i, T_i}$ on $\mathbf{X}_{S_i}$ and a constant term using data from school district d:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_d \quad = \quad & \left[\sum_{i: D_i = d} \left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)} \sum_{i: D_i = d} \mathbf{X}_{S_i} \right) \left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)} \sum_{i: D_i = d} \mathbf{X}_{S_i} \right)' \right]^{-1} \\ & \times \sum_{i: D_i = d} \left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)} \sum_{i: D_i = d} \mathbf{X}_{S_i} \right) \left(Y_i - \hat{\boldsymbol{\beta}}_d' \mathbf{Z}_{S_i, T_i} \right). \end{aligned}$$

Since

$$Y_i - \hat{\boldsymbol{\beta}}_d' \mathbf{Z}_{S_i, T_i} = \boldsymbol{\alpha}_d' \mathbf{X}_{S_i} - \left(\hat{\boldsymbol{\beta}}_d - \boldsymbol{\beta}_d \right)' \mathbf{Z}_{S_i, T_i} + c_d \eta_d + v_{S_i} \gamma_d + u_{S_i, T_i} \delta_d + \varepsilon_i,$$

we have

$$\hat{\boldsymbol{\alpha}}_d - \boldsymbol{\alpha}_d$$
$$= \left[\frac{1}{N_I(d)}\sum_{i:D_i=d}\left(\mathbf{X}_{S_i} - \frac{1}{N_I\,(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)\left(\mathbf{X}_{S_i} - \frac{1}{N_I\,(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)'\right]^{-1}$$
$$\times\frac{1}{N_I(d)}\sum_{i:D_i=d}\left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)\left(v_{S_i}\gamma_d + u_{S_i,T_i}\delta_d + \varepsilon_i - (\hat{\boldsymbol{\beta}}_d - \boldsymbol{\beta}_d)'\mathbf{Z}_{S_i,T_i}\right).$$

The matrix in the first factor is the inverse of

$$\hat{\mathbf{A}}_{\boldsymbol{\alpha}} = \sum_{s\in\mathcal{S}_d}\hat{w}_d^s\left(\mathbf{X}_{s,d} - \sum_{s\in\mathcal{S}_d}\hat{w}_d^s\mathbf{X}_{s,d}\right)\left(\mathbf{X}_{s,d} - \sum_{s\in\mathcal{S}_d}\hat{w}_d^s\mathbf{X}_{s,d}\right)'$$

with an obvious probability limit $\mathbf{A}_{\boldsymbol{\alpha}}$.

The second factor is the difference of

$$\frac{1}{N_I(d)}\sum_{i:D_i=d}\left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)}\sum_{j:D_j=d}\mathbf{X}_{S_j}\right)\left(v_{S_i}\gamma_d + u_{S_i,T_i}\delta_d + \varepsilon_i\right) \tag{15}$$

and

$$\frac{1}{N_I(d)}\sum_{i:D_i=d}\left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)}\sum_{j:D_j=d}\mathbf{X}_{S_j}\right)'\mathbf{Z}'_{S_i,T_i}(\hat{\boldsymbol{\beta}}_d - \boldsymbol{\beta}_d). \tag{16}$$

We rewrite (15) and (16) as

$$\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\hat{w}_d^s\hat{w}_{s,d}^t\left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d}\hat{w}_d^{s'}\mathbf{X}_{s'}\right)\left(v_s\gamma_d + u_{s,t}\delta_d + \overline{\varepsilon}_{s,t,d}\right)$$

and

$$\left[\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\hat{w}_d^s\hat{w}_{s,d}^t\left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d}\hat{w}_d^{s'}\mathbf{X}_{s'}\right)\mathbf{Z}'_{s,t}\right]$$
$$\times\left[\sum_{s\in\mathcal{S}_d}\hat{w}_d^s\sum_{t\in\mathcal{T}}\hat{w}_{s,d}^t\left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)'\right]^{-1}$$
$$\times\left[\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\hat{w}_d^s\hat{w}_{s,d}^t\left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\delta_d u_{s,t} + \overline{\varepsilon}_{s,t,d}\right)\right]$$

with

$$\overline{\varepsilon}_{s,t,d} = \frac{1}{N_I(s,t,d)}\sum_{i:S_i=s,T_i=t}\varepsilon_i.$$

Recall that

$$\hat{\mathbf{A}}_{\boldsymbol{\beta}} = \sum_{s\in\mathcal{S}_d} \hat{w}_d^s \sum_{t\in\mathcal{T}} \hat{w}_{s,d}^t \left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}} \hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}} \hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)'.$$

And, define

$$\hat{\mathbf{B}} = \left[\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} \hat{w}_d^s \hat{w}_{s,d}^t \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d} \hat{w}_d^{s'}\mathbf{X}_{s'}\right)\mathbf{Z}_{s,t}'\right].$$

The variances and the covariance of (15) and (16) before pre- and post-multiplying $\hat{\mathbf{A}}_\alpha^{-1}$ are

$$\begin{aligned}
\boldsymbol{\Sigma}_{11} &= \sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} \left(\hat{w}_d^s\hat{w}_{s,d}^t\right)^2 \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d}\hat{w}_d^{s'}\mathbf{X}_{s'}\right)(v_s\gamma_d + u_{s,t}\delta_d + \bar{\varepsilon}_{s,t,d})^2 \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d}\hat{w}_d^{s'}\mathbf{X}_{s'}\right)', \\
\boldsymbol{\Sigma}_{22} &= \hat{\mathbf{B}}\hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1} \sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} \left(\hat{w}_d^s\hat{w}_{s,d}^t\right)^2 \left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)(u_{s,t}\delta_d + \bar{\varepsilon}_{s,t,d})^2 \left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)' \hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1}\hat{\mathbf{B}}', \\
\boldsymbol{\Sigma}_{12} &= \sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} \left(\hat{w}_d^s\hat{w}_{s,d}^t\right)^2 \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d}\hat{w}_d^{s'}\mathbf{X}_{s'}\right)(v_s\gamma_d + u_{s,t}\delta_d + \bar{\varepsilon}_{s,t,d})(u_{s,t}\delta_d + \bar{\varepsilon}_{s,t,d}) \left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)' \\
&\quad \times \hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1}\hat{\mathbf{B}}'.
\end{aligned}$$

The variance of $\hat{\boldsymbol{\alpha}}_d$ is then

$$\mathrm{Var}(\hat{\boldsymbol{\alpha}}_d) = \mathbf{A}_{\boldsymbol{\alpha}}^{-1}(\boldsymbol{\Sigma}_{11} + \boldsymbol{\Sigma}_{22} - 2\boldsymbol{\Sigma}_{12})\mathbf{A}_{\boldsymbol{\alpha}}^{-1}.$$

To estimate this variance we use the residuals

$$\hat{\zeta}_{i,d} = Y_i - \widehat{\eta_d c_d} - \hat{\boldsymbol{\beta}}_d'\mathbf{Z}_{S_i,T_i} - \hat{\boldsymbol{\alpha}}_d'\mathbf{X}_{S_i}$$

and their cluster average

$$\bar{\zeta}_{s,t,d} = \frac{1}{N_I(s,t,d)} \sum_{i:S_i=s,T_i=t} \hat{\zeta}_{i,d}.$$

The estimate of the variance is obtained from $\hat{\mathbf{A}}_{\boldsymbol{\alpha}}$ and

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{11} &= \sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} \left(\hat{w}_d^s\hat{w}_{s,d}^t\right)^2 \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d}\hat{w}_d^{s'}\mathbf{X}_{s'}\right)\bar{\zeta}_{s,t,d}^2 \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d}\hat{w}_d^{s'}\mathbf{X}_{s'}\right)', \\
\hat{\boldsymbol{\Sigma}}_{22} &= \hat{\mathbf{B}}\hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1} \sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} \left(\hat{w}_d^s\hat{w}_{s,d}^t\right)^2 \left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)\bar{\xi}_{s,t,d}^2 \left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)' \hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1}\hat{\mathbf{B}}' \\
&= \hat{\mathbf{B}}\widehat{\mathrm{Var}(\hat{\boldsymbol{\beta}}_d)}\hat{\mathbf{B}}', \\
\hat{\boldsymbol{\Sigma}}_{12} &= \sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}} \left(\hat{w}_d^s\hat{w}_{s,d}^t\right)^2 \left(\mathbf{X}_s - \sum_{s'\in\mathcal{S}_d}\hat{w}_d^{s'}\mathbf{X}_{s'}\right)\bar{\zeta}_{s,t,d}\bar{\xi}_{s,t,d} \left(\mathbf{Z}_{s,t} - \sum_{t'\in\mathcal{T}}\hat{w}_{s,d}^{t'}\mathbf{Z}_{s,t'}\right)' \hat{\mathbf{A}}_{\boldsymbol{\beta}}^{-1}\hat{\mathbf{B}}',
\end{aligned}$$

with $\bar{\xi}_{s,t,d}$ defined in Section B.1.

B.3 Equivalence of the School-Level WLS and the Individual-Level OLS 2-Step Estimation Procedures

First, we show the equivalence of $\hat{\boldsymbol{\beta}}_d$ in equation (9) from the individual-level regression with school fixed effects, and $\tilde{\boldsymbol{\beta}}_d$ in equation (12) from the school-level WLS regression.

Recall that

$$\tilde{Y}_{s,t,d} = \bar{Y}_{s,t,d} - \sum_{t\in\mathcal{T}} \hat{w}^t_{s,d}\bar{Y}_{s,t,d}, \qquad \tilde{\mathbf{Z}}_{s,t,d} = \mathbf{Z}_{s,t} - \sum_{t\in\mathcal{T}} \hat{w}^t_{s,d}\mathbf{Z}_{s,t},$$
$$\bar{Y}_{s,\bullet,d} = \sum_{t\in\mathcal{T}} \hat{w}^t_{s,d}\bar{Y}_{s,t,d}, \qquad \bar{\mathbf{Z}}_{s,\bullet,d} = \sum_{t\in\mathcal{T}} \hat{w}^t_{s,d}\mathbf{Z}_{s,t},$$

with the school-year average test score $\bar{Y}_{s,t,d} = \frac{\sum_{i\in\mathcal{I}} Y_i(s,t,d)\mathbb{I}\{S_i=s,T_i=t,D_i=d\}}{\sum_{i\in\mathcal{I}} \mathbb{I}\{S_i=s,T_i=t,D_i=d\}}$ as in equation (11).

Then,

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_d &= \left[\tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{i:S_i=s}\left(\mathbf{Z}_{S_i,T_i} - \tfrac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)\left(\mathbf{Z}_{S_i,T_i} - \tfrac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)'\right]^{-1} \\
&\qquad \times\tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{i:S_i=s}\left(\mathbf{Z}_{S_i,T_i} - \tfrac{1}{N_I(s,d)}\sum_{i:S_i=s}\mathbf{Z}_{S_i,T_i}\right)\left(Y_i - \tfrac{1}{N_I(s,d)}\sum_{i:S_i=s}Y_i\right) \\
&= \left[\tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\sum_{i:S_i=s,T_i=t}\left\{\begin{array}{c}\left(\mathbf{Z}_{S_i,T_i} - \tfrac{1}{N_I(s,d)}\sum_{t\in\mathcal{T}}\sum_{i:S_i=s,T_i=t}\mathbf{Z}_{S_i,T_i}\right)\\ \times\left(\mathbf{Z}_{S_i,T_i} - \tfrac{1}{N_I(s,d)}\sum_{t\in\mathcal{T}}\sum_{i:S_i=s,T_i=t}\mathbf{Z}_{S_i,T_i}\right)'\end{array}\right\}\right]^{-1} \\
&\qquad \times\tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\sum_{i:S_i=s,T_i=t}\left\{\begin{array}{c}\left(\mathbf{Z}_{S_i,T_i} - \tfrac{1}{N_I(s,d)}\sum_{t\in\mathcal{T}}\sum_{i:S_i=s,T_i=t}\mathbf{Z}_{S_i,T_i}\right)\\ \times\left(Y_i - \tfrac{1}{N_I(s,d)}\sum_{t\in\mathcal{T}}\sum_{i:S_i=s,T_i=t}Y_i\right)\end{array}\right\} \\
&= \left[\tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\sum_{i:S_i=s,T_i=t}\left\{\left(\mathbf{Z}_{S_i,T_i} - \sum_{t\in\mathcal{T}}\tfrac{N_I(s,t,d)}{N_I(s,d)}\mathbf{Z}_{s,t}\right)\left(\mathbf{Z}_{S_i,T_i} - \sum_{t\in\mathcal{T}}\tfrac{N_I(s,t,d)}{N_I(s,d)}\mathbf{Z}_{s,t}\right)'\right\}\right]^{-1} \\
&\qquad \times\tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\sum_{i:S_i=s,T_i=t}\left\{\left(\mathbf{Z}_{S_i,T_i} - \sum_{t\in\mathcal{T}}\tfrac{N_I(s,t,d)}{N_I(s,d)}\mathbf{Z}_{s,t}\right)\left(Y_i - \sum_{t\in\mathcal{T}}\tfrac{N_I(s,t,d)}{N_I(s,d)}\bar{Y}_{s,t,d}\right)\right\} \\
&= \left(\tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}\left(\mathbf{Z}_{s,t} - \bar{\mathbf{Z}}_{s,\bullet,d}\right)\left(\mathbf{Z}_{s,t} - \bar{\mathbf{Z}}_{s,\bullet,d}\right)'\mathbb{I}\left\{S_i=s,T_i=t,D_i=d\right\}\right)^{-1} \\
&\qquad \times\tfrac{1}{N_I(d)}\textstyle\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}\left(\mathbf{Z}_{s,t} - \bar{\mathbf{Z}}_{s,\bullet,d}\right)\left(Y_i - \bar{Y}_{s,\bullet,d}\right)\mathbb{I}\left\{S_i=s,T_i=t,D_i=d\right\} \\
&= \left(\sum_{s\in\mathcal{S}_d}\frac{N_I(s,d)}{N_I(d)}\sum_{t\in\mathcal{T}}\frac{N_I(s,t,d)}{N_I(s,d)}\tilde{\mathbf{Z}}_{s,t,d}\tilde{\mathbf{Z}}'_{s,t,d}\right)^{-1}\sum_{s\in\mathcal{S}_d}\frac{N_I(s,d)}{N_I(d)}\sum_{t\in\mathcal{T}}\frac{N_I(s,t,d)}{N_I(s,d)}\tilde{\mathbf{Z}}_{s,t,d}\tilde{Y}_{s,t,d} \\
&= \left(\sum_{s\in\mathcal{S}_d}\hat{w}^s_d\sum_{t\in\mathcal{T}}\hat{w}^t_{s,d}\tilde{\mathbf{Z}}_{s,t,d}\tilde{\mathbf{Z}}'_{s,t,d}\right)^{-1}\sum_{s\in\mathcal{S}_d}\hat{w}^s_d\sum_{t\in\mathcal{T}}\hat{w}^t_{s,d}\tilde{\mathbf{Z}}_{s,t,d}\tilde{Y}_{s,t,d} \\
&= \tilde{\boldsymbol{\beta}}_d.
\end{aligned}$$

Next, we show that $\hat{\boldsymbol{\alpha}}_d$ in equation (10) and $\tilde{\boldsymbol{\alpha}}_d$ in equation (13) are equivalent. To see this, the OLS estimator from the individual-level regression of $Y_i - \hat{\boldsymbol{\beta}}'_d\mathbf{Z}_{S_i,T_i}$ on $\mathbf{X}_{S_i}$ (with a constant term) is

$$
\begin{aligned}
\hat{\boldsymbol{\alpha}}_d &= \left[\sum_{i:D_i=d}\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)'\right]^{-1} \\
&\quad \times \sum_{i:D_i=d}\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)\left(Y_i-\hat{\boldsymbol{\beta}}_d'\mathbf{Z}_{S_i,T_i}\right).
\end{aligned}
$$

Since

$$
\begin{aligned}
\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i} &= \frac{1}{N_I(d)}\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}\mathbf{X}_s\mathbb{I}\{S_i=s,T_i=t,D_i=d\} \\
&= \sum_{s\in\mathcal{S}_d}\left(\frac{1}{N_I(d)}\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}\mathbb{I}\{S_i=s,T_i=t,D_i=d\}\right)\mathbf{X}_s \\
&= \sum_{s\in\mathcal{S}_d}\frac{N_I(s,d)}{N_I(d)}\mathbf{X}_s \\
&= \sum_{s\in\mathcal{S}_d}\hat{w}_d^s\mathbf{X}_s,
\end{aligned}
$$

we have

$$
\begin{aligned}
&\frac{1}{N_I(d)}\sum_{i:D_i=d}\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)' \\
=\ & \frac{1}{N_I(d)}\sum_{s\in\mathcal{S}_d}\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}\mathbb{I}\{S_i=s,T_i=t,D_i=d\} \\
& \times\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)\left(\mathbf{X}_{S_i}-\frac{1}{N_I(d)}\sum_{i:D_i=d}\mathbf{X}_{S_i}\right)' \\
=\ & \sum_{s\in\mathcal{S}_d}\left(\mathbf{X}_s-\sum_{s\in\mathcal{S}_d}\hat{w}_d^s\mathbf{X}_s\right)\left(\mathbf{X}_s-\sum_{s\in\mathcal{S}_d}\hat{w}_d^s\mathbf{X}_s\right)'\frac{1}{N_I(d)}\sum_{t\in\mathcal{T}}\sum_{i\in\mathcal{I}}\mathbb{I}\{S_i=s,T_i=t,D_i=d\} \\
=\ & \sum_{s\in\mathcal{S}_d}\frac{N_I(s,d)}{N_I(d)}\left(\mathbf{X}_d-\sum_{s\in\mathcal{S}_d}\hat{w}_d^s\mathbf{X}_s\right)\left(\mathbf{X}_s-\sum_{s\in\mathcal{S}_d}\hat{w}_d^s\mathbf{X}_s\right)'.
\end{aligned}
$$

Also, we have

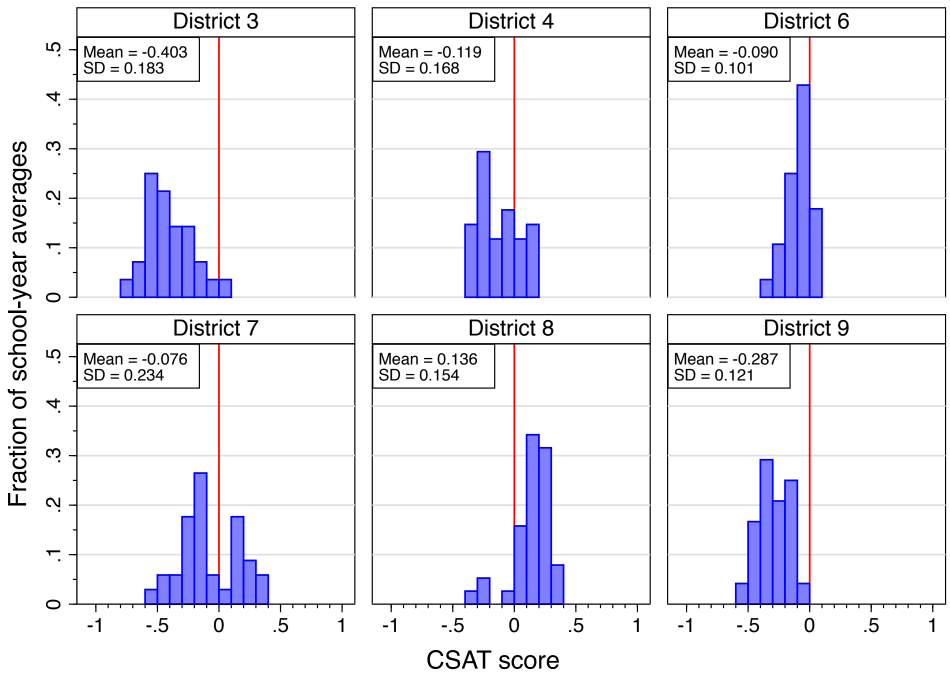
$$
\begin{aligned}
& \frac{1}{N_I(d)} \sum_{i:D_i=d} \left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i=d} \mathbf{X}_{S_i} \right) \left(Y_i - \hat{\boldsymbol{\beta}}_d' \mathbf{Z}_{S_i,T_i} \right) \\
= \; & \frac{1}{N_I(d)} \sum_{s \in \mathcal{S}_d} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} \mathbb{I}\{S_i = s, T_i = t, D_i = d\} \left(\mathbf{X}_{S_i} - \frac{1}{N_I(d)} \sum_{i:D_i=d} \mathbf{X}_{S_i} \right) \left(Y_i - \hat{\boldsymbol{\beta}}_d' \mathbf{Z}_{S_i,T_i} \right) \\
= \; & \sum_{s \in \mathcal{S}_d} \left(\mathbf{X}_s - \sum_{s \in \mathcal{S}_d} \hat{w}_d^s \mathbf{X}_s \right) \frac{1}{N_I(d)} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{I}} \mathbb{I}\{S_i = s, T_i = t, D_i = d\} \left(Y_i - \hat{\boldsymbol{\beta}}_d' \mathbf{Z}_{S_i,T_i} \right) \\
= \; & \sum_{s \in \mathcal{S}_d} \left(\mathbf{X}_s - \sum_{s \in \mathcal{S}_d} \hat{w}_d^s \mathbf{X}_s \right) \frac{N_I(s,d)}{N_I(d)} \sum_{t \in \mathcal{T}} \frac{N_I(s,t,d)}{N_I(s,d)} \left(\bar{Y}_{s,t,d} - \hat{\boldsymbol{\beta}}_d' \mathbf{Z}_{s,t} \right) \\
= \; & \sum_{s \in \mathcal{S}_d} \left(\mathbf{X}_s - \sum_{s \in \mathcal{S}_d} \hat{w}_d^s \mathbf{X}_s \right) \hat{w}_d^s \left(\bar{Y}_{s,\bullet,d} - \hat{\boldsymbol{\beta}}_d' \bar{\mathbf{Z}}_{s,\bullet,d} \right).
\end{aligned}
$$

Thus, we have

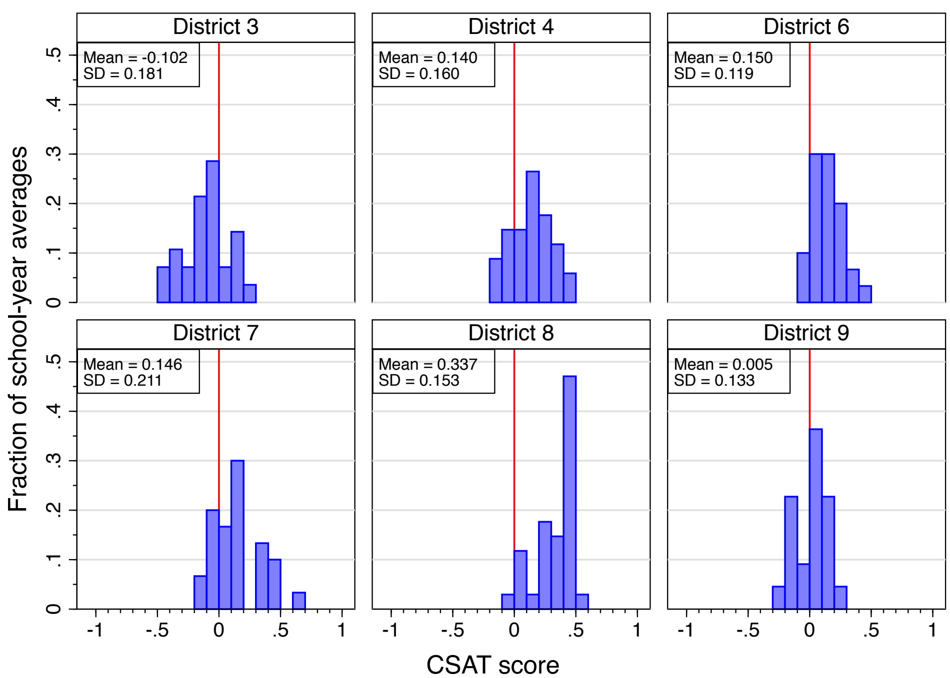
$$
\hat{\boldsymbol{\alpha}}_d = \left[\sum_{s \in \mathcal{S}_d} \hat{w}_d^s \tilde{\mathbf{X}}_{s,d} \tilde{\mathbf{X}}_{s,d}' \right]^{-1} \sum_{s \in \mathcal{S}_d} \hat{w}_d^s \tilde{\mathbf{X}}_{s,d} \left(\bar{Y}_{s,\bullet} - \tilde{\boldsymbol{\beta}}_d' \bar{\mathbf{Z}}_{s,\bullet,d} \right) = \tilde{\boldsymbol{\alpha}}_d.
$$

Figure A1: The School-Year Average Korean CSAT Score by School District

A. Boys



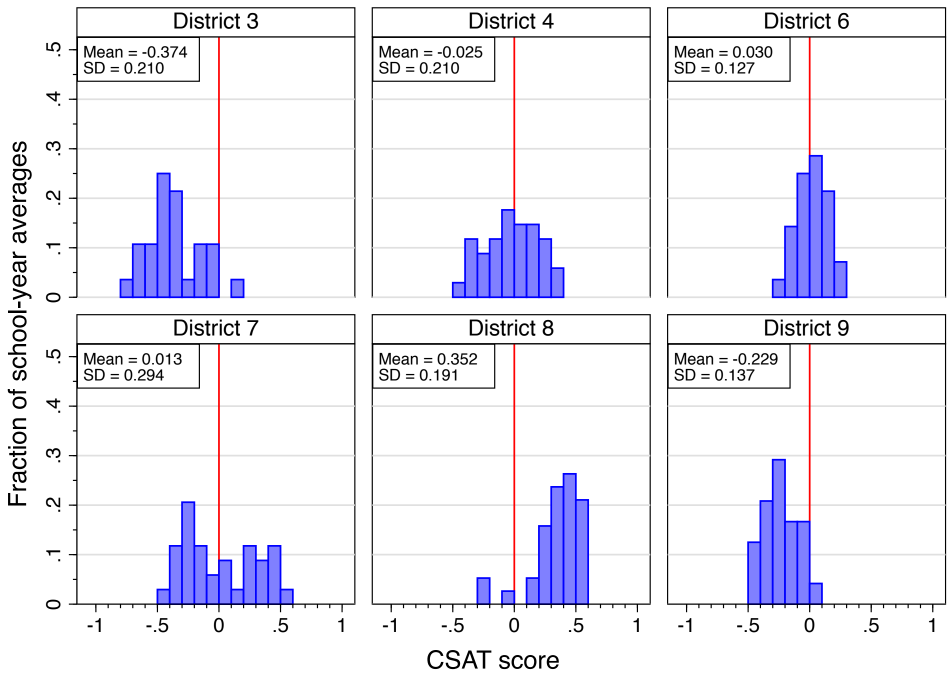
B. Girls



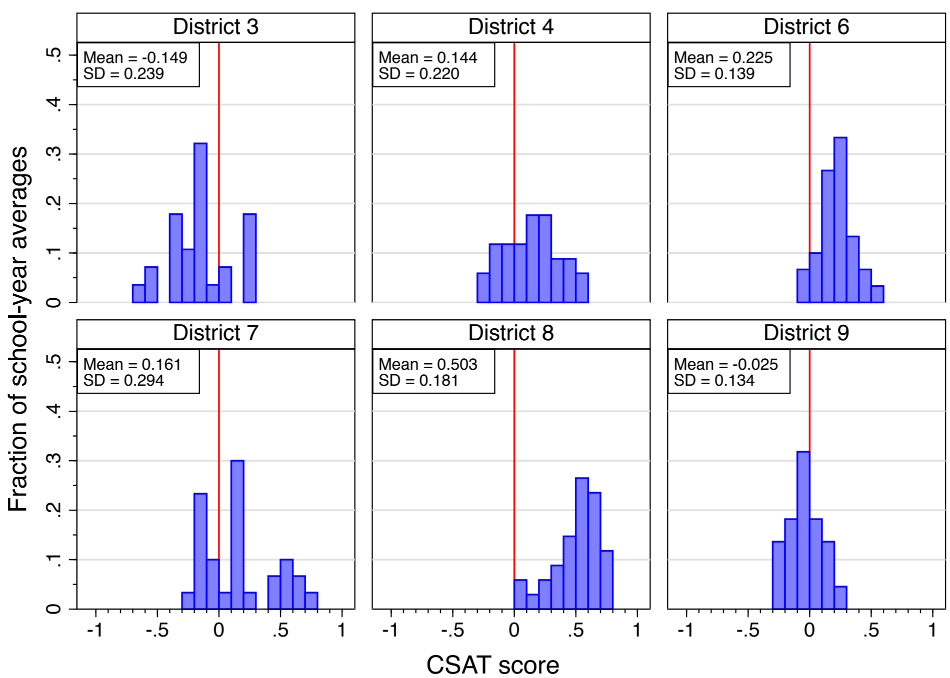
Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The unit of observation is school-year as the Korean CSAT score is averaged at the school-year level. The Korean CSAT score is standardized to have zero mean and unit variance across all test takers in a given year. For CSAT applicants who were absent for the exam, missing raw scores are imputed with zeros. The mean and standard deviation in the top left box in each bar graph are computed using school-year average test scores from each school district.

Figure A2: The School-Year Average English CSAT Score by School District

A. Boys



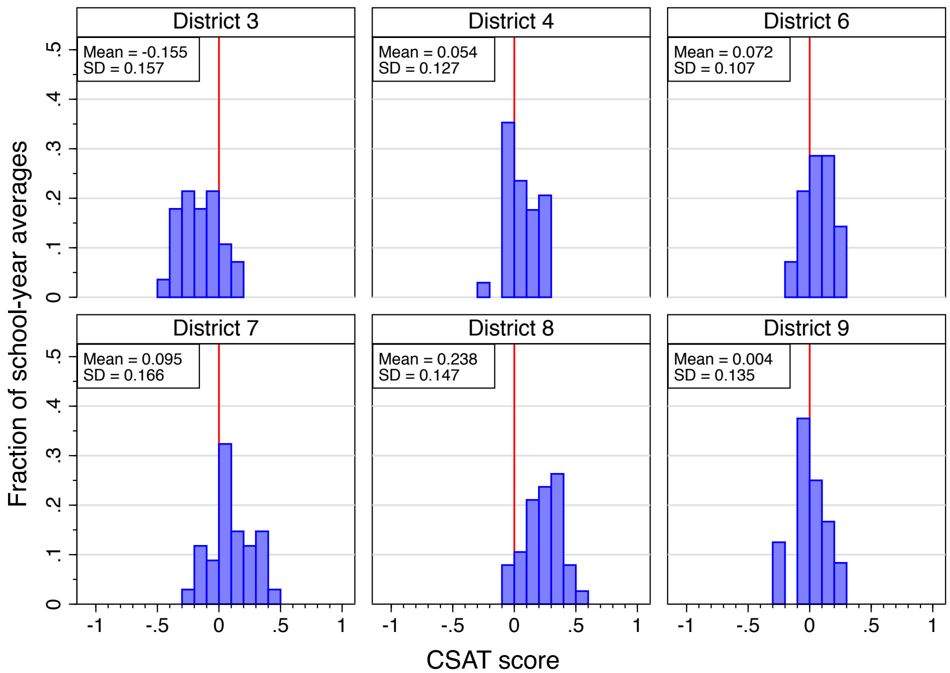
B. Girls



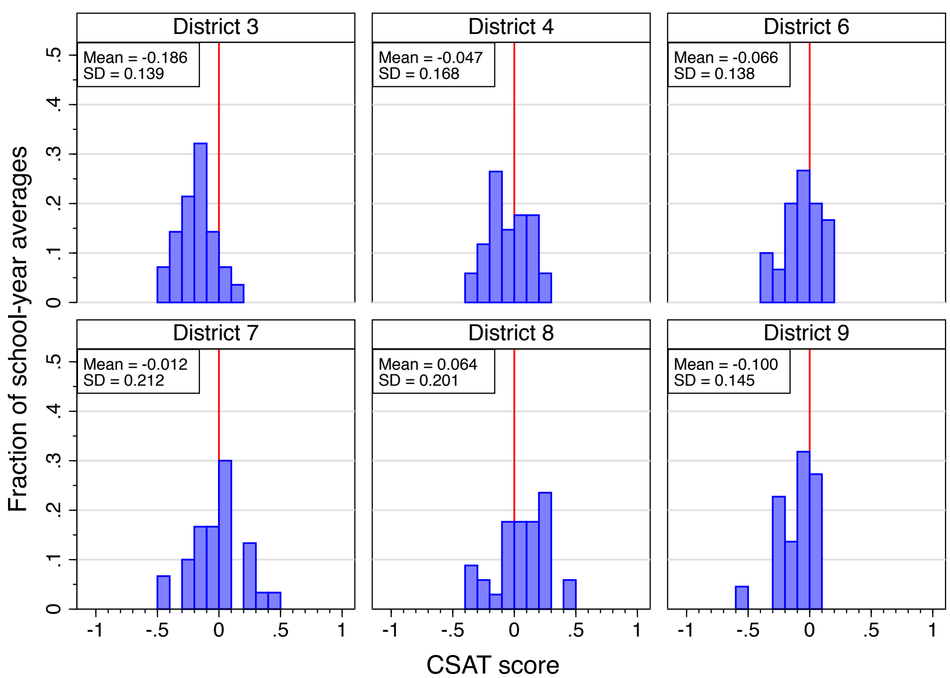
Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The unit of observation is school-year as the English CSAT score is averaged at the school-year level. The English CSAT score is standardized to have zero mean and unit variance across all test takers in a given year. For CSAT applicants who were absent for the exam, missing raw scores are imputed with zeros. The mean and standard deviation in the top left box in each bar graph are computed using school-year average test scores from each school district.

Figure A3: The School-Year Average Math CSAT Score by School District

A. Boys



B. Girls



Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The unit of observation is school-year as the Math CSAT score is averaged at the school-year level. The Math CSAT score is standardized to have zero mean and unit variance across all test takers in a given year. For CSAT applicants who were absent for the exam, missing raw scores are imputed with zeros. The mean and standard deviation in the top left box in each bar graph are computed using school-year average test scores from each school district.

Table A1: School-level Characteristics by School District

	District 3	**District 4**	**District 6**	**District 7**	**District 8**	**District 9**	**All**
A. Boys							
Single-sex school	0.286	0.353	0.357	0.529	0.474	0.417	0.409
	[0.469]	[0.493]	[0.497]	[0.514]	[0.513]	[0.515]	[0.494]
Private school	0.214	0.412	0.571	0.706	0.526	0.500	0.495
	[0.426]	[0.507]	[0.514]	[0.470]	[0.513]	[0.522]	[0.503]
School establishment year	1982.9	1988.9	1965.9	1980.1	1968.0	1975.6	1977.0
	[14.5]	[19.0]	[35.5]	[23.6]	[28.4]	[15.8]	[25.0]
Class size	35.9	35.0	36.3	35.1	34.3	34.7	35.2
	[2.4]	[1.9]	[2.7]	[2.6]	[2.5]	[2.2]	[2.5]
Total enrollment	1,356.8	1,381.1	1,711.4	1,404.3	1,429.1	1,238.1	1,422.8
	[214.8]	[321.8]	[233.4]	[261.9]	[213.0]	[208.0]	[279.2]
Number of teachers	83.5	80.8	96.0	78.8	86.1	76.3	83.6
	[9.2]	[15.2]	[12.2]	[15.1]	[9.1]	[10.0]	[13.5]
Fraction of female teachers	0.497	0.446	0.375	0.312	0.391	0.388	0.400
	[0.180]	[0.133]	[0.170]	[0.194]	[0.194]	[0.230]	[0.190]
Fraction of regular teachers	0.924	0.899	0.897	0.884	0.868	0.913	0.895
	[0.046]	[0.055]	[0.070]	[0.072]	[0.082]	[0.052]	[0.067]
Annual school spending	1,597,245	2,245,422	2,429,417	2,512,820	2,150,629	1,794,521	2,146,877
(in 1000's of 2009 KRW)	[535,718]	[1,317,903]	[985,202]	[1,238,362]	[912,145]	[769,896]	[1,052,292]
Number of schools	14	17	14	17	19	12	93
B. Girls							
Single-sex school	0.286	0.353	0.400	0.467	0.412	0.364	0.382
	[0.469]	[0.493]	[0.507]	[0.516]	[0.507]	[0.505]	[0.489]
Private school	0.143	0.471	0.533	0.600	0.529	0.455	0.461
	[0.363]	[0.514]	[0.516]	[0.507]	[0.514]	[0.522]	[0.501]
School establishment year	1984.7	1991.6	1972.5	1983.9	1963.9	1972.3	1978.3
	[16.0]	[17.2]	[30.1]	[23.6]	[30.4]	[27.0]	[25.9]
Class size	36.1	35.1	36.5	35.6	34.8	34.8	35.5
	[2.3]	[1.8]	[2.9]	[2.8]	[2.9]	[2.6]	[2.6]
Total enrollment	1,311.1	1,329.9	1,700.5	1,429.3	1,411.3	1,144.0	1,398.7
	[217.8]	[235.6]	[217.9]	[351.3]	[198.0]	[197.0]	[288.6]
Number of teachers	80.6	79.9	96.2	80.8	84.3	72.5	82.8
	[12.6]	[13.5]	[12.4]	[20.0]	[8.3]	[11.9]	[15.0]
Fraction of female teachers	0.606	0.533	0.470	0.530	0.534	0.527	0.533
	[0.077]	[0.056]	[0.157]	[0.088]	[0.122]	[0.165]	[0.120]
Fraction of regular teachers	0.924	0.871	0.913	0.892	0.869	0.892	0.892
	[0.042]	[0.071]	[0.046]	[0.064]	[0.079]	[0.061]	[0.065]
Annual school spending	1,493,008	2,169,527	2,225,870	2,177,840	2,182,203	1,666,303	2,014,230
(in 1000's of 2009 KRW)	[625,483]	[1,117,155]	[649,094]	[1,044,735]	[811,303]	[681,503]	[893,447]
Number of schools	14	17	15	15	17	11	89

Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The unit of observation is school for time-invariant school-level variables and school-year for time-varying school-level variables. 1,000 KRW is worth approximately 1 USD. Standard deviations in brackets.

Table A2: Random Assignment Check

	Boys		Girls	
Dependent variable:	**Single-sex**	**Class size**	**Single-sex**	**Class size**
A. Random assignment sample				
9th grade test score	0.035**	0.176	0.013	0.081
	(0.015)	(0.109)	(0.017)	(0.094)
Monthly expenditures on private out-of-school	-0.054*	0.028	-0.052	0.188
education (in 2007 1 million KRW)	(0.029)	(0.143)	(0.038)	(0.219)
Monthly household income	0.001	0.008	0.000	0.003
(in 2007 1 million KRW)	(0.002)	(0.006)	(0.002)	(0.007)
Father high school graduate	-0.009	-1.109**	0.050	-0.238
	(0.049)	(0.468)	(0.071)	(0.513)
Father college graduate	-0.031	-0.855*	0.042	0.131
	(0.055)	(0.489)	(0.075)	(0.521)
Mother high school graduate	0.074	0.724	0.016	0.039
	(0.054)	(0.455)	(0.069)	(0.453)
Mother college graduate	0.111*	0.610	0.002	-0.193
	(0.061)	(0.491)	(0.076)	(0.489)
F-statistic testing all coefficients = 0	1.917	1.683	0.529	0.924
	[0.064]	[0.109]	[0.813]	[0.487]
Middle school fixed effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.407	0.402	0.405	0.410
Observations	1,144	1,144	1,110	1,110
B. No random assignment sample				
9th grade test score	0.057***	0.755***	0.051***	0.524***
	(0.014)	(0.172)	(0.016)	(0.146)
Monthly expenditures on private out-of-school	0.054	0.453	-0.031	0.371
education (in 2007 1 million KRW)	(0.033)	(0.376)	(0.032)	(0.274)
Monthly household income	-0.001	-0.017	0.001	-0.026
(in 2007 1 million KRW)	(0.001)	(0.018)	(0.002)	(0.022)
Father high school graduate	0.021	-0.149	0.077*	0.475
	(0.040)	(0.446)	(0.040)	(0.386)
Father college graduate	0.101**	0.509	0.111**	0.569
	(0.046)	(0.479)	(0.048)	(0.442)
Mother high school graduate	0.004	0.253	-0.100**	0.322
	(0.036)	(0.438)	(0.041)	(0.369)
Mother college graduate	-0.061	0.110	-0.171***	0.368
	(0.047)	(0.543)	(0.054)	(0.518)
F-statistic testing all coefficients = 0	5.770	5.507	2.881	3.237
	[0.000]	[0.000]	[0.006]	[0.002]
Middle school fixed effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.424	0.632	0.452	0.655
Observations	1,184	1,184	1,132	1,132

Notes. Data are from KELS 2005, 2007, 2008. The random assignment sample includes students attending general academic high schools subject to the random assignment lotteries. The no random assignment sample includes students attending high schools not subject to the random assignment lotteries—either schools in the non-HSEP regions or selective/vocational high schools in the HSEP regions. The 9th grade test score is the combined score on Korean, English and Math that is normalized to have zero mean and unit variance. 1 million KRW is worth approximately 1000 USD. Robust standard errors in parentheses. p-values in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: The Effect of School Inputs on the CSAT Score Using Controls in Choi et al. (2014)

	Boys				Girls			
	Single-sex		Class size		Single-sex		Class size	
A. Our approach								
All districts								
APE	0.231	(0.053)***	-0.011	(0.002)***	0.021	(0.055)	0.004	(0.002)**
By district								
District 3	0.557	(0.146)***	-0.019	(0.010)*	0.234	(0.203)	-0.012	(0.008)
District 4	0.216	(0.172)	-0.015	(0.003)***	0.197	(0.130)	0.004	(0.003)
District 6	0.044	(0.039)	-0.016	(0.003)***	-0.034	(0.046)	0.009	(0.004)**
District 7	0.311	(0.173)*	0.006	(0.006)	-0.256	(0.135)*	0.006	(0.003)*
District 8	0.214	(0.083)***	-0.012	(0.004)***	0.022	(0.152)	0.006	(0.005)
District 9	0.068	(0.083)	-0.015	(0.008)*	0.057	(0.088)	0.009	(0.005)*
Testing school input effects identical across districts								
χ^2-statistic	15.72	[0.008]	11.81	[0.037]	8.23	[0.144]	6.37	[0.272]
B. Other empirical strategies								
DFE	0.045	(0.032)	0.000	(0.007)	0.073	(0.037)*	0.004	(0.006)
OLS	0.001	(0.036)	-0.001	(0.006)	0.040	(0.049)	0.012	(0.005)**

Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The CSAT score refers to the total of Korean and English CSAT scores and is standardized to have zero mean and unit variance. For CSAT applicants who were absent for the exam, missing raw scores on Korean or English are imputed with zeros. Time-varying control variables include the fraction of students receiving free or reduced price lunch, annual development fund spending per student (in thousands of 2009 KRW), 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. Robust standard errors clustered at the school-year level in parentheses. Section 5.2 provides more details on standard error computation. p-values associated with χ^2-statistics in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A4: The Effect of School Inputs on the Korean CSAT Score

	Boys				**Girls**			
	Single-sex		**Class size**		**Single-sex**		**Class size**	
A. With controls								
All districts								
APE	-0.030	(0.081)	-0.010	(0.003)***	0.008	(0.063)	-0.009	(0.003)***
By district								
District 3	0.331	(0.156)**	0.002	(0.007)	-0.006	(0.108)	-0.002	(0.006)
District 4	-0.056	(0.269)	-0.011	(0.006)*	0.238	(0.105)**	-0.019	(0.004)***
District 6	0.002	(0.044)	-0.029	(0.007)***	-0.029	(0.083)	-0.001	(0.005)
District 7	-0.187	(0.256)	-0.005	(0.006)	-0.358	(0.265)	-0.005	(0.008)
District 8	0.100	(0.101)	-0.006	(0.006)	0.181	(0.068)***	-0.015	(0.006)***
District 9	-0.453	(0.275)*	-0.007	(0.009)	0.023	(0.193)	-0.010	(0.008)
Testing school input effects are identical across districts								
χ^2-statistic	8.44	[0.133]	10.39	[0.065]	9.61	[0.087]	11.82	[0.037]
B. No controls								
All districts								
APE	0.127	(0.022)***	-0.006	(0.002)***	0.106	(0.024)***	-0.003	(0.001)**
By district								
District 3	0.038	(0.066)	-0.004	(0.006)	0.077	(0.057)	0.001	(0.005)
District 4	0.159	(0.054)***	-0.008	(0.004)*	0.156	(0.041)***	0.002	(0.003)
District 6	0.022	(0.036)	-0.017	(0.003)***	0.021	(0.046)	-0.003	(0.002)
District 7	0.184	(0.073)**	0.007	(0.004)*	0.034	(0.088)	0.002	(0.003)
District 8	0.168	(0.039)***	-0.005	(0.005)	0.189	(0.037)***	-0.005	(0.003)*
District 9	0.177	(0.038)***	-0.013	(0.006)**	0.183	(0.058)***	-0.022	(0.005)***
Testing school input effects are identical across districts								
χ^2-statistic	14.12	[0.015]	28.08	[0.000]	11.48	[0.043]	18.64	[0.002]

Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The Korean CSAT score is standardized to have zero mean and unit variance. For CSAT applicants who were absent for the exam, missing raw scores are imputed with zeros. Time-varying control variables include total enrollment, the number of teachers, the fraction of regular teachers, the fraction of female teachers, and log annual school spending in thousands of 2009 KRW. Time-invariant control variables include a private indicator, school establishment year, and the interaction between the two. Robust standard errors clustered at the school-year level in parentheses. Section 5.2 provides more details on standard error computation. p-values associated with χ^2-statistics in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: The Effect of School Inputs on the English CSAT Score

	Boys				Girls			
	Single-sex		Class size		Single-sex		Class size	
A. With controls								
All districts								
APE	0.022	(0.083)	-0.016	(0.003)***	-0.080	(0.089)	-0.002	(0.003)
By district								
District 3	0.540	(0.172)***	-0.001	(0.010)	0.077	(0.127)	0.005	(0.006)
District 4	0.169	(0.291)	-0.018	(0.010)*	0.309	(0.156)**	0.001	(0.004)
District 6	-0.009	(0.076)	-0.014	(0.005)***	-0.110	(0.106)	0.001	(0.005)
District 7	-0.037	(0.177)	-0.031	(0.006)***	-0.865	(0.393)**	-0.024	(0.006)***
District 8	0.119	(0.110)	-0.007	(0.005)	0.126	(0.096)	0.006	(0.008)
District 9	-0.879	(0.402)**	-0.023	(0.009)***	0.130	(0.232)	0.002	(0.008)
Testing school input effects identical across districts								
χ^2-statistic	14.77	[0.011]	13.11	[0.022]	11.27	[0.046]	17.14	[0.004]
B. No controls								
All districts								
APE	0.149	(0.028)***	-0.012	(0.002)***	0.093	(0.031)***	0.012	(0.002)***
By district								
District 3	0.087	(0.082)	-0.004	(0.007)	0.081	(0.088)	0.006	(0.005)
District 4	0.236	(0.060)***	-0.017	(0.005)***	0.214	(0.057)***	0.016	(0.004)***
District 6	0.040	(0.049)	-0.012	(0.002)***	0.014	(0.054)	0.010	(0.003)***
District 7	0.171	(0.096)*	-0.005	(0.004)	0.020	(0.116)	0.012	(0.003)***
District 8	0.179	(0.049)***	-0.010	(0.004)**	0.092	(0.048)*	0.023	(0.004)***
District 9	0.176	(0.048)***	-0.032	(0.006)***	0.176	(0.052)***	-0.003	(0.004)
Testing school input effects identical across districts								
χ^2-statistic	8.47	[0.132]	14.95	[0.011]	8.89	[0.114]	25.36	[0.000]

Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The English CSAT score is standardized to have zero mean and unit variance. For CSAT applicants who were absent for the exam, missing raw scores are imputed with zeros. Time-varying control variables include total enrollment, the number of teachers, the fraction of regular teachers, the fraction of female teachers, and log annual school spending in thousands of 2009 KRW. Time-invariant control variables include a private indicator, school establishment year, and the interaction between the two. Robust standard errors clustered at the school-year level in parentheses. Section 5.2 provides more details on standard error computation. p-values associated with χ^2-statistics in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A6: The Effect of School Inputs on the Math CSAT Score

	Boys				**Girls**			
	Single-sex		**Class size**		**Single-sex**		**Class size**	
A. With controls								
All districts								
APE	-0.022	(0.088)	-0.009	(0.004)**	0.057	(0.094)	-0.003	(0.004)
By district								
District 3	0.363	(0.162)**	0.006	(0.012)	0.244	(0.110)**	0.013	(0.008)*
District 4	-0.206	(0.152)	-0.020	(0.008)***	0.410	(0.134)***	0.002	(0.011)
District 6	0.028	(0.051)	-0.011	(0.006)*	0.059	(0.126)	-0.023	(0.008)***
District 7	-0.513	(0.356)	-0.008	(0.007)	-0.648	(0.414)	-0.019	(0.007)***
District 8	0.343	(0.116)***	-0.006	(0.007)	0.162	(0.145)	0.013	(0.015)
District 9	-0.102	(0.326)	-0.008	(0.016)	0.282	(0.194)	0.005	(0.005)
Testing school input effects identical across districts								
χ^2-statistic	15.71	[0.008]	3.97	[0.554]	8.28	[0.141]	19.94	[0.001]
B. No controls								
All districts								
APE	0.122	(0.019)***	-0.007	(0.002)***	0.026	(0.025)	0.012	(0.003)***
By district								
District 3	0.072	(0.063)	0.006	(0.010)	0.088	(0.045)**	0.005	(0.009)
District 4	0.136	(0.040)***	-0.014	(0.005)***	0.131	(0.048)***	0.020	(0.006)***
District 6	0.030	(0.032)	-0.008	(0.003)***	-0.010	(0.045)	0.005	(0.005)
District 7	0.104	(0.057)*	0.001	(0.004)	-0.114	(0.078)	0.006	(0.003)**
District 8	0.214	(0.035)***	-0.013	(0.006)**	0.011	(0.068)	0.028	(0.008)***
District 9	0.174	(0.047)***	-0.014	(0.010)	0.114	(0.065)*	-0.006	(0.004)
Testing school input effects identical across districts								
χ^2-statistic	17.23	[0.004]	9.64	[0.086]	11.09	[0.050]	21.82	[0.001]

Notes. Data are from the Korean Ministry of Education. The analysis sample includes 57,443 male students (26,669 in 2008 and 30,774 in 2009) at 55 coed and 38 all-boys high schools, and 52,271 female students (24,140 in 2008 and 28,131 in 2009) at 55 coed and 34 all-girls high schools. The Math CSAT score is standardized to have zero mean and unit variance. For CSAT applicants who were absent for the exam, missing raw scores are imputed with zeros. Time-varying control variables include total enrollment, the number of teachers, the fraction of regular teachers, the fraction of female teachers, and log annual school spending in thousands of 2009 KRW. Time-invariant control variables include a private indicator, school establishment year, and the interaction between the two. Robust standard errors clustered at the school-year level in parentheses. Section 5.2 provides more details on standard error computation. p-values associated with χ^2-statistics in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$