Mitigating School Closures:

Parental Responses and Tutoring Vouchers *

Hyunjae Kang TaeYoung Kang Sunham Kim

May 10, 2025

Abstract

Can private tutoring vouchers mitigate the impact of school closures? This paper examines how parental responses to reduced in-person school days shape educational inequality. Leveraging quasi-experimental, school-level variations in in-person learning days in South Korea, we first estimate how school closures affect parental investment and within-school test score disparities. Using a linked administrative–survey panel dataset, we find that losing an additional 10 days of in-person instruction leads to a 4% increase in parental spending on private tutoring and a 25% rise in within-school inequality. We develop a household decision model in which both formal schooling and private tutoring jointly determine student test scores. Estimates indicate that private tutoring substitutes for school days, with low-income households disproportionately affected by school closures. Policy simulations show that a targeted voucher scheme can achieve comparable outcomes to a universal voucher policy.

JEL Classification: I24, I28, I38
Keywords: school closure, parental investment, learning loss, education voucher

* H. Kang: Kyoto University; T. Kang: KAIST; Kim: Korea Development Institute. We are grateful to Steven Stern, Juan Pantano, Miguel Sarzosa, Chong Xiang, and the seminar participants at GEPS2021, SK-CSES2021, MEA2022, AMES2022 for comments and discussions. We also thank Gyeonggi Institute of Education, Gyeonggi Province Office of Education, and Busan Metropolitan City Office of Education for sharing data. This research is supported by Krannert Doctoral Research Funds (Purdue University, 2022). Any errors are our own. Previously titled "School Closure and Educational Inequality: Parental Investment in the Pandemic."

1. Introduction

School closures increase educational inequality across socioeconomic groups, whether caused by forced displacement (Kim 2024), natural disasters (Sacerdote 2012; Andrabi et al. 2023), or a global pandemic (Agostinelli et al. 2022; Jack et al. 2023; The Economist 2021). Households compensate for lost in-person schooling with market substitutes such as private tutoring, but unequal access to these options across socioeconomic groups widens achievement gaps. While recent studies have used economic models to examine school closures' impact on educational outcomes (Grewenig et al. 2021; Agostinelli et al. 2022; Fuchs-Schündeln et al. 2022), evidence remains limited on how reduced school days lead to different parental responses across socioeconomic groups, how these differences contribute to educational inequality, and which interventions can mitigate the resulting inequality.

In this paper, we address this gap in three steps. First, we empirically identify the impact of lost in-person learning days on private tutoring expenditure and academic achievement, using a novel panel dataset from South Korea.[1] Second, we build and estimate a tournament model of parental investment in which the number of school days is augmented as an input of the education production function. We then use the estimated model to assess the learning loss of low-income households and quantify the size of the educational voucher that could fully mitigate their educational loss.

We choose South Korea and its response to the COVID-19 school closure as a laboratory to facilitate causal interpretation. The country had extensive school closures, with only 6.2% of school days being fully in-person during two years of the pandemic.[2] Schools were fully closed during 24.7% of school days, and 67.6% days were partially closed. The exogenous nature of school closure reduces concerns about reverse causality. The country has been successful in "flattening the curve" with no blanket lockdown. This success makes our estimates less contaminated by the aggregate environment, compared to other developed countries. Moreover, the fact that South Korea is known for education fervor and has had no absenteeism alleviates the concern of selection bias.

We construct a novel panel dataset by linking administrative and survey sources. Specifically, we combine two longitudinal student surveys, the Gyeonggi Educational Panel Survey (GEPS) and the Busan Educational Longitudinal Survey (BELS), with SCHOOLINFO, the admin-

[1] We use the terminology "private tutoring" to encompass various forms, including one-on-one private tutoring, group private tutoring, and hagwon (private tutoring institutions in South Korea). In the education literature, this broad category is often referred to as "shadow education" (Bray 1999).

[2] South Korea recorded the highest school closure ratio among Asia & Pacific countries that are included in World Bank High Income Groups.

istrative census of all Korean secondary schools. The resulting dataset is a panel of students from multiple adjacent cohorts and contains student-level information on private tutoring expenditure, standardized national test scores, socioeconomic background, including household income and parental education, and the number of in-person school days at the school level. This structure enables us to track within-student changes, while also exploiting quasi-random variation in school reopening decisions and comparing COVID-affected cohorts to pre-pandemic cohorts.

We employ a continuous difference-in-differences design that exploits policy-driven, quasi-random variation in in-person school days across schools. At the individual level, we find that losing 10 additional in-person days leads to a 4.1% increase in private tutoring expenditure. This response is significantly stronger for students with low past test scores and those from low-income households. Despite these compensatory investments, we find that the impact of a 10-day loss on individual standardized test scores remains small and statistically insignificant. However, the within-school standard deviation of test scores increases by 0.3 SD per 10-day loss in in-person days, while there is little impact on the within-school dispersion of private tutoring expenditure. These findings suggest that while households compensate for lost school days through private tutoring, this compensation is uneven across socioeconomic groups, potentially widening educational inequality.

Motivated by our quasi-experimental findings, we develop a structural model to address key questions: What factors drive differential impacts across these groups? What interventions might effectively mitigate educational inequality during public schooling disruptions? We adopt a tournament framework of Kang (2024), where parents choose private tutoring to maximize expected utility. This structure lets us capture the incentive to finance a child's education, the way private tutoring and in-person school days jointly determine test scores, and the extent to which the two inputs substitute for one another. A college-admission tournament sets rewards by rank, while a CES production function combines previous test score, school days, and tutoring, allowing us to identify both the marginal effect of each input and the substitution elasticity between the number of school days and private tutoring expenditure.[3]

We estimate the model by constrained maximum likelihood, which ensures simulated data reproduce our TWFE estimates for mean tutoring, within-school score dispersion, and the

[3] An alternative model assuming individuals care about absolute performance (the "Max-Score" model) substantially overpredicts parental investment levels.

counterfactual parallel-trend restriction. This approach guarantees that the structural model is estimated under the same empirical assumptions as our reduced-form analysis.

Our structural estimates imply that private tutoring and school-days are gross substitutes. Based on the CES specification, substitutability implies that the marginal effects of school days are higher for low SES households. This finding aligns with our TWFE results showing stronger tutoring responses among low-SES households despite their more binding resource constraints.

We use the estimated model to evaluate private tutoring voucher policies. First, we quantify that restoring pre-COVID school days would reduce within-school score inequality by approximately 8% while only modestly affecting mean performance. Using this full-reopening scenario as a benchmark, we evaluate voucher policies based on (i) voucher size and (ii) eligibility criteria.

Voucher size has diminishing returns on inequality reduction. Beyond a threshold, larger vouchers exacerbate inequality. Regarding eligibility, the effects of vouchers in inequality improvement become stronger as vouchers become less selective, but these effects also diminish beyond a certain threshold. Overall, consistent with the heterogeneous effects observed during school closures, our analysis demonstrates that the targeted voucher policies are more cost effective than universal approaches while achieving comparable effects in mitigating educational inequality that rose during school closures.

Specifically, our findings show that a means-tested tutoring voucher of approximately $440 for households below the 80th income percentile would fully mitigate the inequality caused by COVID-19 school closures. Moreover, this targeted approach achieves the same inequality improvements at half the cost of a universal voucher program. Our structural analysis underscores the equalizing role of public education and advocates targeted voucher policies as an effective alternative when full school reopening is not feasible.

Our paper contributes to several strands of literature. First, we provide causal estimates of the impact of reduced in-person school days on parental investments and educational inequality, extending the literature on educational disruption. Previous work has documented learning losses after shocks such as natural disasters (Sacerdote 2012; Andrabi et al. 2023), school displacement (Kim 2024), and pandemic-era school shutdowns (Bacher-Hicks et al. 2021; Maldonado and De Witte 2022; Jack et al. 2023; Hahn et al. 2023). These studies consistently find declines in average test scores and increasing test score disparities, but do not investigate parental behavioral responses. We leverage quasi-experimental variation in school closures and the longitudinal dataset with individual fixed effects to quantify how households differentially adjust educational investments when faced with school disruptions.

Second, our paper contributes to the literature on ex-ante policy evaluation aimed at mitigating educational inequalities when public education is disrupted. Our production function estimates embedded in the structural model imply that private household investments and public educational inputs are substitutes. We use this structural model to assess the effectiveness of a targeted tutoring voucher policy in offsetting learning losses. Closest to our approach, Agostinelli et al. (2022) show that low-income children lost substantially more ground during COVID-19 because affluent households could better compensate for lost schooling, and discuss potential policy interventions. In line with their approach, we construct a structural model to map households' heterogeneous ability to replace lost instructional time, investigating mechanisms through which school closures magnify educational inequality and evaluating policy tools to mitigate these effects.

Third, our estimation approach contributes to recent empirical work integrating structural models with experimental identification (Heckman 2010; Galiani and Pantano 2021; Todd and Wolpin 2023), and in particular, quasi-experimental variation (Busso et al. 2013; Voena 2015; Blundell et al. 2016; D'Haultfœuille and Février 2020; Alves et al. 2023). Building on this line of research, we formally incorporate the parallel-trends assumption from our continuous difference-in-differences design into the maximum likelihood estimation of the structural model. By doing so, our method ensures internal consistency between reduced-form evidence and model-based counterfactual analysis.

The rest of the paper is organized as follows. Section 2 outlines the background and data. Section 3 describes our empirical framework, followed by our results in Section 4. Section 5 defines the tournament model. Section 6 elucidates how we implement the structural estimation and reports our results. Section 7 conducts counterfactual policy experiments based on the estimated structural model.

2. Institutional Background and Data

2.1 Background on the Korean Education

South Korea operates a highly centralized education system with near-universal enrollment and a standardized 6-3-3 structure: six years of elementary school, three years of middle school, and three years of high school. The government requires 190 school days annually and mandates adherence to the National Curriculum for Primary and Secondary Schools.[4] Local school boards exercise only limited authority in educational policy. Parental involvement plays a minimal role in school governance compared to countries like the United States. Opting out of the formal school system is virtually impossible, with homeschooling both legally prohibited and rarely practiced.

This centralized system reflects the country's famous 'education fever' – an exceptional cultural emphasis on schooling as the primary engine of economic growth and social mobility. This national obsession translates into consistent educational participation, with South Korea maintaining the OECD's lowest absenteeism rates (OECD 2019). The government's commitment to educational continuity has been remarkable—prior to COVID-19, schools remained open through multiple national crises, including military coups (1961, 1979) and financial crises (1997, 2007). Notably, the government made no exceptions for previous epidemics that threatened public health, such as MERS (2015), H1N1 (2009), SARS (2004), and even the 1969 cholera outbreak – which posed a serious health challenge during a period when the country was still economically underdeveloped.

Parallel to the centralized schooling system exists an extensive private tutoring market. 'Hagwon' (cram schools) provide supplementary tutoring services widely utilized across all school levels.[5] These services primarily target college admission preparation, which directly impacts lifetime earnings prospects. Though accessible to all, quality varies by cost, earning private tutoring the label of "Great Unequalizer." The government has repeatedly attempted to regulate this market, most notably through business hour restrictions that survived constitutional challenges. Despite these regulatory efforts, the private tutoring market continues to thrive, offering students a variety of options to fuel their education fever beyond the formal system.

2.2 The COVID Responses in Early 2020

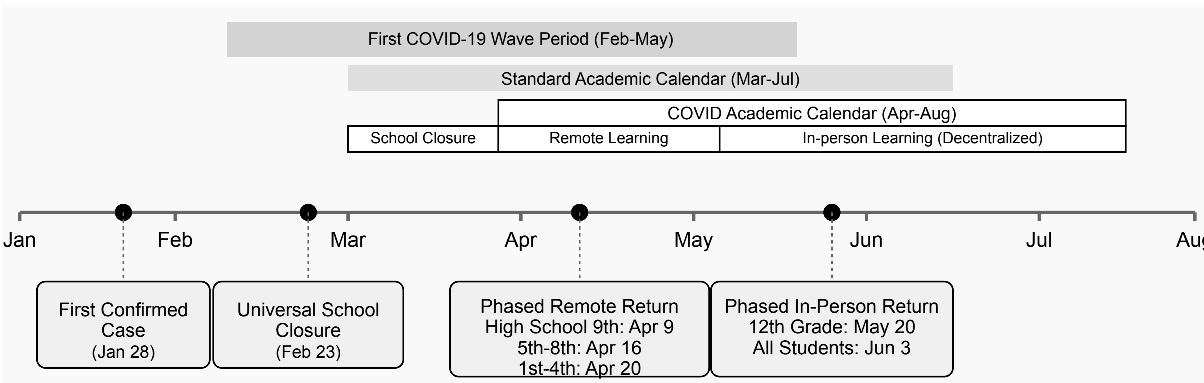
The COVID-19 pandemic brought an unprecedented challenge to South Korea's education system, triggering the first nationwide school closure in the country's history. When the first case was reported in January 2020, the Ministry of Education maintained the normal academic calendar, which starts on March 2nd. However, as the first wave arrived in mid-February, the Ministry declared the first-ever nationwide school closure on February 23rd, postponing the

[4] It applies to both public and private schools. A few selective high schools (autonomous private schools and special purpose schools) may customize the curriculum.

[5] According to Ministry of Education data (2010), 45% of elementary, 50% of middle school, and 84% of high school students participate in private tutoring services.

March 2nd school opening. This "temporary" closure extended five weeks, transitioning to a fully remote opening with a phased schedule: grades 9-12 returned April 9th, grades 5-8 on April 16th, and grades 1-4 on April 20th. Figure 1 illustrates this timeline of closures and phased reopenings.

Figure 1: South Korea's COVID-19 Education Response Timeline (Early 2020)



As the first wave subsided, the Ministry planned in-person reopening. A phased return began with 12th graders on May 20th, with all students back by June 3rd. This marked the end of nationwide closures, which never placed again. Instead, the Ministry issued guidelines but delegated open-or-remote decisions to regional education offices and school principals. This decentralized approach created the school-level variation in in-person school days that our identification strategy relies on. Figure A.2 shows the spatial variation in the regions in our dataset.

While schools closed, private tutoring institutions remained open throughout this period. South Korea implemented its 'Trace, Test, and Treat (3T)' strategy rather than imposing blanket lockdowns.[6] This approach prevented the Ministry of Education from targeting hagwons with restrictions beyond the social distancing measures applied to other businesses. Online private tutoring services operated with even fewer limitations, as they were not constrained by physical distancing requirements. Meanwhile, with parents still required to work in-person due to the absence of workplace lockdowns, many households lacked capacity to supervise children's education at home, making private tutoring their only viable option. In effect, the pandemic shut down the 'Great Equalizer' (public schools) while unleashing the 'Great Unequalizer'.

[6]See Bicker (2020) for details.

2.3 Data

Our analyses use three cohort groups from two regionally representative longitudinal surveys of South Korean students. Each cohort is defined by their grade level during the COVID-19 pandemic in 2020: 8th graders, 11th graders, and 12th graders. The 8th and 11th grade cohorts come from Busan, Korea's second-largest city, while the 12th grade cohort comes from Gyeonggi, the country's most populous province.[7] These datasets provide rich information at student, parent, and school levels, including test scores, private education expenditure, and school budgets. We link these surveys with SCHOOLINFO, an administrative dataset containing in-person school days for all primary and secondary schools.

We combine the cohorts who experienced COVID school closures to form our treatment group (also referred to as the COVID group). Each survey includes multiple cohort waves, allowing us to construct a corresponding control group of students who attended the same grade levels before the pandemic: 8th graders from 2017, 11th graders from 2013, and 12th graders from 2014. These control cohorts, unexposed to the pandemic, provide baseline information on student behavior and learning outcomes in the absence of school closures. Our identification strategy leverages in-person school day variations within the COVID group, using the Control group for comparison.[8] Table 1 summarizes our data structure.

Table 1: Definition of COVID and Control groups

Survey Year(s)	8th graders		11th graders		12th graders	
	COVID	Control	COVID	Control	COVID	Control
2012						10th
2013						11th
2014						12th
2015						
2016		7th		10th		
2017		8th		11th		
2018	6th		9th		10th	
2019	7th		10th		11th	
2020	8th		11th		12th	

[7] These data come from the Gyeonggi Educational Panel Survey (GEPS) and the Busan Education Longitudinal Survey (BELS). Both surveys share a common design based on the Korean Education Longitudinal Study (KELS), which adapted the Education Longitudinal Study (ELS) of the United States.

[8] Due to the survey timing, the "-2" years' information of 8th and 11th graders is unavailable for the Control group.

We focus on parental investment and student academic achievement. Our main measure of parental investment is private tutoring expenditure.[9] In both surveys, parents report monthly spending (in Korean won) on private tutoring across three core subjects: Korean language, Mathematics, and English. Since the first semester of the Korean academic year runs from March to July, with surveys administered between late July and mid-August, the response effectively captures private tutoring during the spring semester. We take a sum of expenditures across all three subjects as our main measure of parental investment.

For academic achievement, we use scores from national-level standardized tests, which allow for valid comparisons of student performance across different schools and regions. Students took the exam at the end of the semester. The raw test scores range from 0 to 100 for each subject: Korean, English, and Mathematics. These scores are standardized to have a mean of 100 and standard deviation of 20 within each cohort and grade. We use the average of these standardized scores across all three subjects as our main measure of academic achievement.[10] Table A.1 presents the descriptive statistics of the working sample.

SCHOOLINFO provides data on total in-person school days for all primary and secondary schools by school and semester for 2019 and 2020. We link these figures from spring semesters to the schools in our sample. Consequently, while our parental investment and test score measures exist at the individual student level, in-person school days are measured at the school level. This variable is not included in the dataset for years prior to 2019 because there was minimal variation in school days under the centralized system, making it of little academic or policy interest. For these earlier years, we assume all students attended 95 days (half of the annual legal minimum), based on the observed 2019 pre-COVID data, which shows little variation (mean: 95.8 days, standard deviation: 2.0).

3. Empirical Strategy

We estimate the impact of reduced in-person school days on parental investment and test scores. The identifying shock is the decentralized variation at the school level in the number of days schools were closed. We estimate the following two-way fixed effect model:

$$Y^c_{ikg} = \beta_0 + \beta_1 schdays^c_{kg} + \Gamma Z^c_{ikg} + \mu_i + \eta_g + \varepsilon^c_{ikg}. \quad (1)$$

[9]We do not consider time investment of parents, as it becomes less relevant once students enter secondary school in South Korea (See Section 3 of (Kang 2024)) .

[10]For 12th graders in the COVID group, scores are provided in stanine format. We map each stanine bin to original scores using external data sources that provide the cut points for this exam.

In the above equation, Y_{ikg}^{c} is an outcome variable for student i in school k, grade g, and cohort c. The primary outcomes we examine are private tutoring expenditure and standardized test scores.[11] We denote $g = 0$ as the reference grade in which COVID cohorts experienced the pandemic-induced educational disruption. All other grades are indexed relative to this reference point across both treatment and control groups. For instance, $g = 0$ corresponds to the 2020 academic year for 8th graders in the COVID group, whereas it represents 2017 for their counterparts in the control group. The variable $schdays_{kg}^{c}$ represents the number of in-person school days measured in tens of days. The vector Z_{ikg}^{c} is a set of time-varying individual-level controls. We control for an individual fixed effect μ_i and a grade fixed effect η_g. We also conduct analyses by subgroup based on student and parent characteristics.

This is a standard continuous difference-in-differences design where we use the cumulative number of in-person school days during Spring 2020 ($g = 0$) as our treatment intensity. This treatment affects only the COVID cohort, and our data directly provides these aggregate attendance measures rather than day-by-day closure records. Since we have only this single treatment period, the post-treatment period dummy is omitted. Unlike staggered adoption designs with varying treatment timing across units, our approach examines treatment intensity within a uniform timeframe. Our identification relies on the key assumption that without COVID-induced school closures, outcome trends would have evolved parallel across the COVID and Control groups, conditional on controls and fixed effects. The validity of these assumptions is discussed in detail below.

We also examine whether school closures exacerbated educational inequality in dimensions not fully captured by changes in individual means. To address this question, we estimate a school-level model analogous to Equation (1):

$$Y_{kg}^{c} = \beta_0 + \beta_1 schdays_{kg}^{c} + \Gamma Z_{kg}^{c} + \mu_k + \eta_g + \varepsilon_{kg}^{c}. \tag{2}$$

Here, Y_{kg}^{c} represents the standard deviation of test scores within school k. This measure allows us to assess how school closures affected the dispersion of academic performance, given our premise that public education serves as a 'Great Equalizer.'

For both Equations (1) and (2), the coefficient of interest β captures the change in the outcome variable due to a ten-day increase in in-person school days. Since our focus is on school closures (i.e., decreases in school days), we interpret the coefficient with the opposite

[11] We apply the inverse hyperbolic transformation for those who spend zero.

sign for ease of interpretation. We report standard errors clustered at the city level throughout the analysis.

We discuss our four identifying assumptions and two supporting facts below. The validity of our empirical approach rests on these.

Assumption 1 *Variation in in-person school days induced by the closure is random, conditional on time-invariant individual characteristics.*

The key variation for identification is the number of in-person school days. This was determined by school principals and the Regional Offices of Education, as discussed in Section 2.2. No individual student could predict or affect the number of in-person school days. One of the major complaints from students and parents during this period was the uncertainty about whether schools would open the following week or, in extreme cases, the very next day.

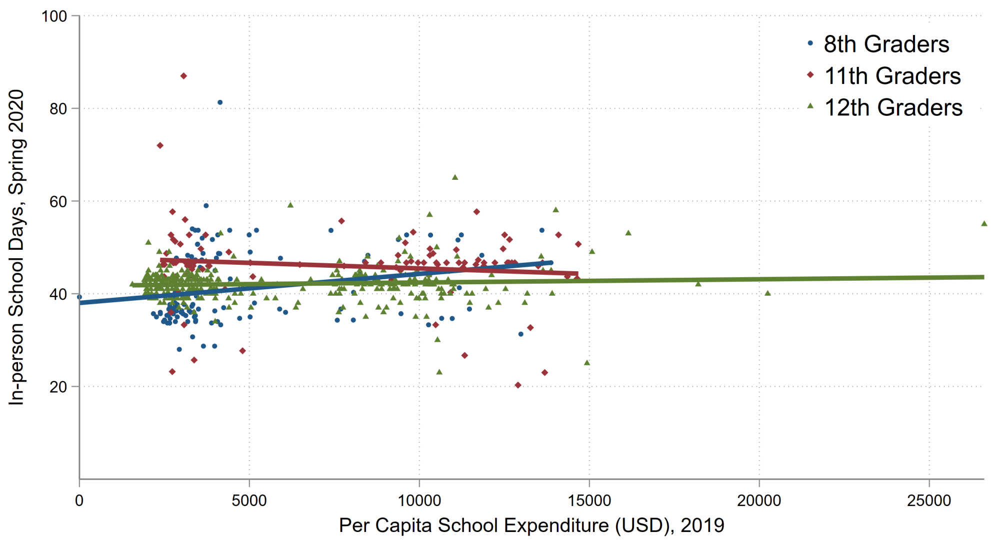
In principle, reverse causality could arise if reopening decisions were affected by households or correlated with components of the individual-level error term.[12] However, this is unlikely in our context for two reasons. First, although the opening decision was decentralized, it remained in the hands of policymakers familiar with the top-down approach, not with parents or students. Second, potential channels for reverse causality, such as selective schools having different incentives for reopening, appear absent in the data. For instance, selective schools might prefer reopening to provide intensive in-person instruction, or conversely, might remain closed if their students can effectively access private tutoring.[13] To investigate this possibility, we examined whether school quality, proxied by per capita spending, predicts the number of in-person days (Figure 2). There are no systematic correlations, and conditioning on school characteristics such as public-private status does not change the result.

There could be measurement error due to COVID infection. If a student was infected, there would have been additional losses in in-person days at the individual level, beyond the school closure policy. This would constitute nonclassical measurement error since it correlates with the idiosyncratic error term: sick students cannot participate in private tutoring. While this type of measurement error is a valid concern for COVID-era data, it is also unlikely in our context. First, the number of school-age cases was limited in our sample provinces, Gyeonggi and Busan. As of July 31st, 2020, there were 80 out of 1,537 total cases in Gyeonggi (0.07% of the

[12] Parolin and Lee (2021) reports that school closures from September to December 2020, in the United States, were more common in schools with higher shares of students from racial and ethnic minorities, who are likely to have poor socioeconomic status.

[13] For non-Korean readers, the second scenario considers the following story. Suppose that the objective function of a school is to maximize the number of students who gain admission to top colleges. Private tutoring is generally more "effective" for preparing for exams. In this case, schools have incentives to close in order to let students maximize their private tutoring hours.

Figure 2: In-person School Days and Per-capita school expenditure



Data: GEPS, BELS, and SCHOOLINFO. The per capita school expenditure is converted from Korean Won (KRW) to US Dollar (USD) assuming 1 USD = 1,200 KRW.

teenage population). For Busan, while we do not have the number of cases by age, there were only 168 cases in total by July 31st, 2020 (0.05% of the population)—this represents the upper bound of school-age cases in Busan. Furthermore, we found no association from a regression of in-person days on the number of cumulative cases in subregions of the two provinces.

Assumption 2 *Parallel trends hold for parental investment, test scores, and the school level standard deviation of test scores.*

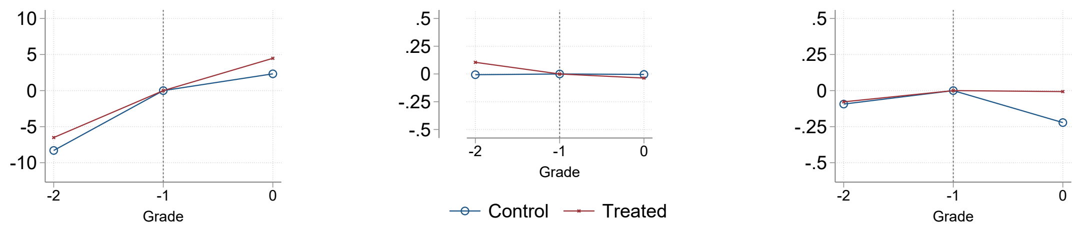
The COVID shock is an aggregate shock that absorbs into the time dimension. In order to identify the COVID effect, we assume that the annual growth of private tutoring expenditure is the same across the treated and control cohorts. That is, the growth due to the grade effect is time invariant. This parallel trends assumption enables us to use an earlier cohort in the dataset as a control group, attributing all 2020-specific outcome variations to the changes in in-person school days. Figure 3 visualizes the average tutoring expenditure growth of the treated and control cohorts between $g = -1$ and $g = 0$.

Assumption 3 *The individual fixed effects absorb most unobserved heterogeneities, and bias due to time-varying confounders is limited.*

First, we follow the standard two-way fixed effects framework. The individual fixed effects control for all time-invariant characteristics at the student level, including demographics,

Figure 3: Trends of Outcome Variables

(a) Individual Expenditure (USD) (b) Individual Std. Test Score (c) School-level Std. Test Score Standard Deviation



Note: All outcomes are normalized to have zero mean by sample group at $g = -1$. The test score plots' vertical axis is measured in standard-deviation units.

household background, and baseline academic performance. The grade fixed effects account for outcome growth over time that is common across all students.

As part of Z^c_{ikg}, we control for time-varying measures such as household income in expenditure regressions and expenditure in test score regressions. Given our short data timeframe, few other relevant individual controls vary within this period. Our analysis confirms that results remain robust when including these controls, suggesting minimal bias from time-varying confounders.[14]

Assumption 4 *Treatment effects are homogeneous across treatment intensities.*

Our continuous difference-in-differences design requires that the effect of an additional ten days of in-person schooling is similar across the distribution of treatment intensity. This allows us to interpret our estimated coefficient as the average treatment effect per ten days. While recent literature highlights concerns with heterogeneous treatment effects in difference-in-differences models (Callaway et al. 2024), our non-staggered treatment setup avoids many of these concerns. Moreover, to empirically test this assumption, we estimate a modified version of Equation (1):

$$Y^c_{ikg} = \beta_0 + \beta_1 schdays^c_{kg} + \beta_2 schdays^c_{kg} \times HIGH_i + \Gamma Z^c_{ikg} + \mu_i + \eta_g + \varepsilon^c_{ikg}. \tag{3}$$

[14] Including time-varying controls in two-way fixed effects models requires strong identification assumptions that treatment should not affect such controls (Caetano et al. 2024). As we discuss in Fact 1 below, since South Korea avoided economywide lockdowns and experienced limited labor market disruption, household income likely remained relatively stable during our period of interest. Expenditure, however, did respond, and it is one of our main findings. We nonetheless include expenditure in test score regressions because of its relevance, while acknowledging this potential endogeneity.

where $HIGH_i$ is a dummy equal to 1 if $schdays^c_{kg}$ is above-median in 2020.[15] The coefficient β_2 captures any differential effect between above- and below-median treatment intensity groups. We test the school-level treatment as well, by similarly augmenting Equation (2). The results of these tests are discussed in our robustness checks section.

In addition to our four identifying assumptions, we leverage the following empirical facts to argue that South Korea is one of the best "laboratories" to identify the impact of COVID-induced school closure, and our estimates are less prone to aggregate-level confounders compared to the ones from other developed economies.

Fact 1 *Without economywide lockdowns, South Korea's economy underwent minimal disruption during early COVID-19, compared with other developed economies.*

While COVID-induced school closures in South Korea offer quasi-random identifying variation, isolating their effects from broader pandemic impacts requires addressing potential confounding from aggregate COVID effects. In most developed economies, parental job loss and school closures are deeply intertwined, confounding their independent effects. South Korea provides a unique research environment not only for its quasi-random school closure variation but also due to its successful 'flattening the curve' strategy that avoided economywide lockdowns while implementing targeted social distancing, testing, and financial stimulus. Consequently, employment rates remained stable, deviating only by 1-2 percentage points from pre-pandemic trends (Figure A.3), whereas U.S. employment fell by 10 percentage points between February and April 2020.[16] That is, the universal school closure was a preventative measure rather than a response to widespread outbreak.[17] This relative economic stability allows changes in parental investment to be more confidently attributed to school closures rather than labor market shocks.

Fact 2 *School attendance is universal in South Korea, and its private tutoring market is robust.*

Our research design benefits from both universal treatment compliance and wide access to educational alternatives, which are characteristic of the country's 'education fever.' South Korea's formal education system exhibits exceptionally high attendance rates, with chronic absenteeism rarely observed. It implies that when schools closed, virtually the entire student population was affected, creating a treatment group with minimal selection concerns.

[15] Using $HIGH_c$ and $HIGH_i$ yields identical results as $schdays^c_{kg}$ varies at the school level.

[16] Using weekly administrative payroll data, Cajner et al. (2020) reported that U.S. employment fell by 21% through late-April 2020, and 30 percent of the decline was driven by business shutdowns.

[17] While the first wave of infections in February triggered massive policy responses, the labor market shock was limited to Daegu, the region hit by the wave (Aum et al. 2021). Our sample does not include Daegu. See Figure A.4 for employment rates by provinces.

Simultaneously, South Korea's private tutoring market remained operational throughout our periods of interest. It offers a comprehensive range of options from free online lectures to premium one-on-one tutoring services so that households had access to viable alternatives to public schooling during closures, making substitution behaviors both feasible and observable. With approximately 75% of students already getting some form of private tutoring service before the pandemic, households were able to adjust their educational investments in response to school closures.

In sum, these assumptions and institutional features provide a research design that isolates the effect of school closures from broader pandemic disruptions. Our approach leverages South Korea's unique combination of quasi-random school reopening decisions, stable economic conditions, and accessible educational alternatives. This setting enables us to examine not only the average impact of reduced in-person schooling days on individual outcomes but also its distributional consequences for educational inequality.

4. Empirical Results

We begin by discussing our baseline results, the effect of in-person school days on individual-level outcomes. Next, we report our heterogeneity analysis for various student subgroups. We then examine school-level outcomes to assess how school closures affected educational inequality.

4.1 Individual-Level Effects

Table 2 presents our individual-level results based on Equation (1) using inverse hyperbolic sine transformation of private tutoring expenditure and standardized test score. For each outcome, we present estimates without time-varying covariates (columns 1, 4), and with covariates (columns 2, 5). We also include tests for treatment heterogeneity using Equation (3) (columns 3, 6). We consider columns (2) and (5) our preferred specifications.

For private tutoring expenditure, we find that a ten-day decrease in in-person school days increased investment by approximately 4.1%, statistically significant at the 1% level. This estimate remains unaffected when controlling for household income (column 2), addressing concerns about time-varying confounders (Assumption 3). The interaction term with above-median in-person days $\hat{\beta}_2$ in column (3) is tiny and insignificant, indicating that heterogeneity in treatment effects across schools with different levels of closure intensity is of less concern (Assumption 4).

For test scores, we find precisely estimated null effects across all specifications. A ten-day decrease in in-person schooling had virtually no impact on standardized test score: approximately 0.096/20 = 0.0048 standard deviations (column 4), with the 95% confidence interval [-0.0068, 0.0148] standard deviations. This null effect is robust against controlling for private tutoring expenditure (column 5). The interaction term estimate is small and insignificant, which mitigates concerns about treatment heterogeneity (column 6). While our findings on test scores differ from studies reporting significant learning losses during COVID school closures (Engzell et al. 2021; Maldonado and De Witte 2022; Jack et al. 2023), they align with Hahn et al. (2023), who also found null average effects using Korean data.[18]

Table 2: Impact of In-Person School Days on Individual Outcomes

	log(Parental Expenditure)			Test Score		
	(1)	(2)	(3)	(4)	(5)	(6)
In-person days	-0.040***	-0.041***	-0.041***	0.096	0.116	0.080
	(0.008)	(0.008)	(0.008)	(0.116)	(0.119)	(0.113)
log(income)		0.215***	0.215***			
		(0.053)	(0.053)			
log(expenditure)					0.487***	0.486***
					(0.114)	(0.113)
Above Median=1 × In-person days			0.002			0.097
			(0.005)			(0.077)
N	17,164	17,164	17,164	17,164	17,164	17,164

Note: Clustered standard errors in parentheses (city level, $N = 47$). ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Our findings show increased private educational investment alongside stable average test scores, suggesting a compensatory response where households increased private tutoring to offset potential learning losses. The null average effect on test scores is surprising; combined with the expenditure responses, however, these findings suggest changes in the test score distribution beyond the first moment. In our subsequent heterogeneity and school-level analyses, we examine if these averages may mask offsetting impacts across different student subgroups.

Appendix Tables B.1 and B.2 show how our estimates change across specifications. We find estimates from the COVID-sample alone differ from our full-sample results, which shows the importance of our difference-in-differences strategy in identifying school closure effects from confounding grade-specific trends and the aggregate COVID effect.

[18] Their dataset was nationally representative. Our sample is not a subset of theirs, so this consistency provides additional validation for our findings.

4.2 Heterogeneous Effects by Student Characteristics

Tables 3 and 4 examine heterogeneity in the effects of school closures across different student subgroups using our preferred specification. We mostly focus on heterogeneity in expenditure, and discuss test score results at the end of the section, relegating the corresponding tables to appendix.

4.2.1 Household Income and Academic Performance

Table 3 focuses on heterogeneity by prior household income and academic performance. We define the subgroups using terciles based on income and test score data at $g = -1$ for all students. Both dimensions show clear gradients.

For household income (columns 2-4), low-income households increased their private tutoring expenditure by 6.2% for each ten-day reduction in in-person schooling, compared to just 2.9% and 2.8% for middle and high-income households, respectively. The difference between low-income and middle/high-income households is statistically significant (p-value = 0.026). Similarly, for prior academic performance (columns 5-7), we observe that low-performing students increased private tutoring investment by 6.7% per ten fewer in-person school days, larger than the 3.8% for middle-performing students, which is significant at the 10% level (p-value = 0.068). The effect for high-performing students, 0.8%, is small and insignificant, while its difference from mid-performers is significant (p-value = 0.007).

Table 3: Impact of In-Person School Days on Expenditure: by Income and Performance

	Baseline	Prior Household Income			Prior Performance		
	(1) All	(2) Low	(3) Mid	(4) High	(5) Low	(6) Mid	(7) High
In-person days	-0.041***	-0.062***	-0.029***	-0.028***	-0.067***	-0.038***	-0.008
	(0.008)	(0.012)	(0.009)	(0.009)	(0.014)	(0.008)	(0.007)
p-value			0.026	0.900		0.068	0.007
N	17,164	5,912	6,358	4,894	5,656	5,722	5,786

Note: Clustered standard errors in parentheses (city level, $N = 47$). P-values in columns (3)-(4) and (6)-(7) test whether effects differ from their adjacent left columns. For example, the p-value in column (3) tests the difference between columns (2) and (3), while the p-value in column (4) tests the difference between columns (3) and (4). ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Households with fewer financial resources and academically vulnerable students made larger adjustments to their educational investments in response to school closures. While the percentage measure of expenditure change may partially reflect lower baseline spending ("base effect"), the magnitude of these differences suggests meaningful variation in response intensity. That is, low-income households bear a disproportionate financial burden in covering the costs of their children's education during school closures.

4.2.2 Gender and Parental Education

In contrast to the socioeconomic and performance gradients, Table 4 shows no significant heterogeneity by gender or parental education.

Female students experienced slightly larger increases in parental expenditure (4.8%) compared to male students (3.4%), but this difference is small and insignificant (p-value = 0.254). This absence of gender bias in educational investment response is consistent with our prior, given that Korea's historical son preference has declined over the last two decades (Chung and Gupta 2007; Yoo et al. 2017; Choi and Hwang 2020). Similarly, our baseline estimate is unaffected by conditioning on parental education, whether measured by mother's or father's educational attainment.

Table 4: Impact of In-Person School Days on Expenditure: by Gender and Parental Education

	Baseline	Gender		Education (Mother)		Education (Father)	
	(1) All	(2) Boy	(3) Girl	(4) Below BA	(5) BA+	(6) Below BA	(7) BA+
In-person days	-0.041***	-0.034***	-0.048***	-0.042***	-0.037***	-0.041***	-0.038***
	(0.008)	(0.008)	(0.010)	(0.009)	(0.009)	(0.011)	(0.008)
p-value			0.254		0.714		0.795
N	17,164	8,908	8,240	10,320	6,354	7,960	8,666

Note: Clustered standard errors in parentheses (city level, $N = 47$). P-values in columns (3), (5), and (7) test whether effects differ from their adjacent left columns. For example, the p-value in column (3) tests the difference between columns (2) and (3). 'BA' stands for a Bachelor's degree granted by a four-year college. ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

The absence of significant heterogeneity in gender and parental human capital stands in marked contrast to the strong gradients in socioeconomic and performance dimensions. It suggests that a household's financial capacity to access private alternatives and their perception of their child's academic needs drove compensatory behavior.

Appendix Tables B.4 and B.5 present heterogeneity analyses for test scores. These results indicate that students from low-income households and those with less-educated parents lost more from school closure than peers with high-income and more-educated parents, though the effect sizes are modest (approximately 0.01-0.02 standard deviations per ten days).

This suggests, qualitatively, disadvantaged students rely more heavily on in-person public education, supporting the notion of "Great Equalizer."

While parental education did not significantly affect expenditure responses, it did affect test score outcomes, suggesting parental education operates through channels beyond just financial expenditure. Our data does not allow us to directly examine all potential channels, such as parental time use or specific tutoring service choices. Given South Korea's legally prohibited homeschooling and stable labor market during the pandemic (Fact 1), differences in tutoring service selection–rather than direct parental instruction – may explain this finding, though the modest effect sizes suggest limited economic significance.

4.3 School-Level Effects

In this section, we turn to distributional outcomes using Equation (2). Table 5 presents how school closures affected educational inequality by using within-school standard deviations of expenditure and test scores.[19] For time-varying controls, we use school-level averages of parental expenditure and household income after log transformation. As with our individual-level results, our preferred specifications include time-varying controls (columns 2, 5). Appendix Tables B.6 and B.7 report how these school-level estimates change with different specifications.

In-person school days had little impact on the standard deviation of parental expenditure within schools. The effect is insignificant, but remains unaffected when controlling for average income and above-median indicator. This suggests that while the average level of private tutoring expenditure increased with school closures (as shown in Table 2), the distribution of this spending within schools remained relatively stable.

In contrast, they significantly affected test score dispersion: a ten-day decrease in in-person schooling increased the within-school standard deviation of test scores by approximately 0.3 standard deviations. This effect is statistically significant at the 5% level with controls, robust against controlling for average tutoring expenditure (column 5), and the interaction term with above-median in-person days indicator is virtually zero (column 6). That is, even if average test scores remained unaffected during school closures (Section 4.1), the distribution of achievement significantly widened.

One caveat to our findings is student sorting across schools. We showed in-person school days and school quality, proxied by per capita spending, are not correlated in Figure 3. More-

[19] We add the city fixed effects on top of school and grade fixed effects.

Table 5: Impact of In-Person School Days on School-level Standard Deviation

	Expenditure			Test Score		
	(1)	(2)	(3)	(4)	(5)	(6)
In-person Days	0.251	0.251	0.271	-0.270*	-0.289**	-0.289**
	(0.245)	(0.245)	(0.257)	(0.141)	(0.139)	(0.141)
log(income)		0.683	0.599			
		(7.655)	(7.675)			
log(expenditure)					-0.023*	-0.023*
					(0.013)	(0.013)
Above Median=1 × In-person Days			-0.059			-0.000
			(0.240)			(0.070)
N	1,208	1,208	1,208	1,208	1,208	1,208

Note: Clustered standard errors in parentheses (city level, $N = 47$). The outcome variables are standard deviation at the school level. The controls are log of school-level average. ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

over, South Korea's school equalization policies mitigate the selection concern.[20] Students are assigned to middle schools (our 8th graders) and high schools (our 11th and 12th graders) via random lottery, conditional on residence.[21] While this system does not eliminate all forms of student sorting, particularly geographic sorting by residence, our results remain robust to controlling for region fixed effects. Moreover, we observe a tiny fraction of moving students during their secondary school period.

4.4 Implications

Our findings point to two key facts. First, when schools closed, households increased private tutoring expenditure, treating it as a substitute; the magnitude of this response varied with income and prior test scores. Second, despite these adaptations, achievement gaps widened. This indicates that private tutoring, though an effective individual substitute, cannot replicate the equalizing role of public schooling. It suggests the returns to public schooling are heterogeneous across student characteristics.

These facts lead to several questions about mechanisms and policy responses that call for a systematic framework to capture both substitution and heterogeneous effects: Which specific factors drive differential returns across socioeconomic groups? Which interventions can best mitigate inequality when schooling is disrupted? And, when tutoring support is offered, do

[20]This feature has been exploited by various papers in the literature. See, for example, Kang (2007), Park et al. (2013), and Park et al. (2018)

[21]Specialized high schools can select students, but they account for just 3% of total enrollment.

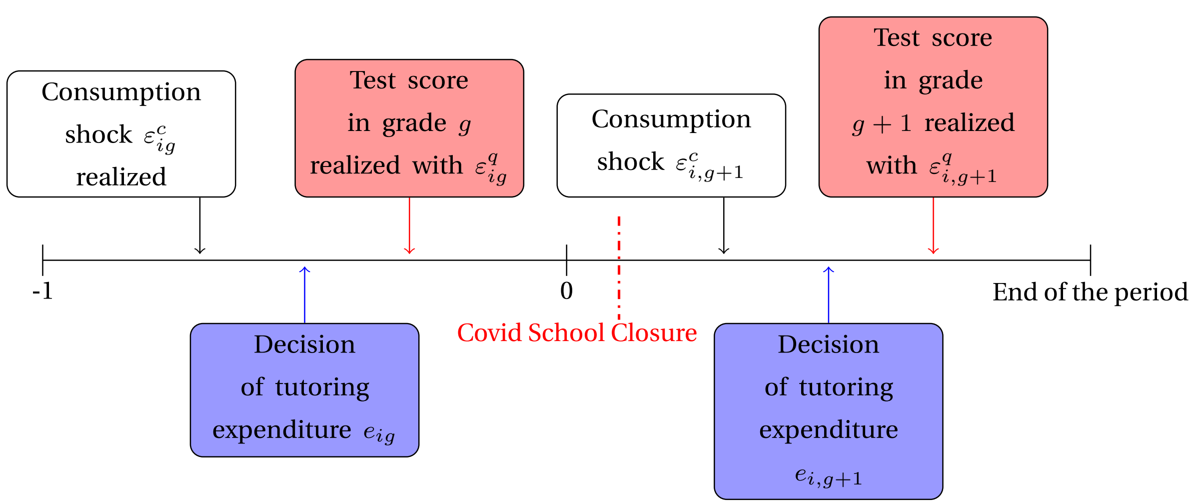
targeted vouchers outperform universal subsidies? We develop a structural model to address these questions in the following section.

5. Theoretical Framework

Motivated by the empirical results, we employ a tournament model of Kang (2024) in which each household makes investments on their child. We extend this framework to capture three key elements: (1) the underlying incentives driving demand for private tutoring, (2) the effects of private tutoring and school attendance on test scores, and (3) the substitution relationship between these educational inputs. The tournament structure represents the competitive college admissions process, where parental investments are driven by the potential returns to educational achievement through access to selective institutions. Test score production follows a Constant Elasticity of Substitution (CES) function incorporating three inputs: the child's previous academic performance, in-person school days, and private tutoring expenditures. This specification enables us to identify both the marginal effects of each educational input and the elasticity of substitution between public schooling and private educational investments when schooling availability changes.

5.1 Environment

Figure 4: Model Timeline



Consider an economy composed of N households with a unitary decision making process. Figure 4 presents our model timeline. The model begins in period -1, the year preceding the school-closure year (for treated groups only). The initial conditions are household income and the test score from the previous grade.

Denote g as the school grade of the student. Each household i derives flow utility from current consumption and the perceived returns to academic outcomes, R_{ig}, which is a function of the test score $q_{i,g+1}$ and the reward determined following tournament structure (Kang 2024). The sequence within each period is as follows: First, a consumption shock ε_{ig}^{c} is realized. Based on this shock and expectations about the test score shock ε_{ig}^{q}, each household makes a private tutoring investment decision e_{ig}. After this investment decision, the test score $q_{i,g+1}$ is produced with the realization of the time-specific test score shock ε_{ig}^{q}. This generated test score $q_{i,g+1}$ then serves as an input for producing the subsequent test score in the next period.

5.2 Household and the Tournament Structure

Each period, household i derives utility from consumption and the test score produced in grade g. The utility function is defined as:

$$\varepsilon_{ig}^{c}\ln(c_{ig}) + \ \alpha_g R(q_{i,g+1}), \tag{4}$$

where c_{ig} is household consumption, $R(q_{i,g+1})$ is the return from test score $q_{i,g+1}$, and ε_{ig}^{c} is a shock to the marginal utility of consumption.

We adopt a static tournament model to capture the underlying demand for parental investment.[22] Each household decides how much to spend on private tutoring expenditure e_{ig} to maximize utility defined in equation (4), subject to the budget constraint:

$$c_{ig} + e_{ig} \leq w_{ig} \tag{5}$$

where w_{ig} is household income. The tournament structure is realized through the return function $R(q_{i,g+1})$, which is specified as:

$$R(q_{i,g+1}) = \sum_{j=1}^{J} \ln(v_j) \times Prob(\ln \tilde{Q}_{j-1} \geq \ln q_{i,g+1} \geq \ln \tilde{Q}_j \Big| \Gamma_{ig}) \tag{6}$$

where $\tilde{Q}_j$ is the cutoff for college tier j and Γ_{ig} is the information set of household i when their child is in grade g. Competition between households occurs around the cutoff $\tilde{Q}_j$. Denoting n_j as the number of seats for college tier j, the cutoff $\tilde{Q}_j$ is the test score of the marginal student

[22]We find that an alternative model, where a household cares about the *absolute* performance of the student, fails to fit the parental investment distribution of households. The comparison is illustrated in Figure C.1.

at position M_j, where $M_j = \sum_{k=1}^{j} m_k$. That is, $\tilde{Q}_j$ is the lowest test score among students accepted into college tier j. This return function assumes that each household cares about its chance of getting into a better college tier. Households vary in their probability of securing college admission. A household with a more capable child has a higher chance of securing admission to a top-tier college, while a household with a less capable child has a higher chance of gaining admission to a lower-tier college.

5.3 Test Score Production Function

To map parental investment and school days to test scores, we specify a production function:

$$q_{i,g+1} = f(\theta_g, q_{ig}, e_{ig}, schdays_{ig}, \varepsilon^q_{ig}) \tag{7}$$

where θ_g is a set of relevant parameters that may change over school grade (age). To identify substitution between parental investment and school days, our functional form follows the constant elasticity of substitution (CES) production function (Cunha et al. 2010):

$$q_{i,g+1} = A_g \left[\delta_{q,g} q_{ig}^{\phi} + \delta_{s,g} schdays_{ig}^{\phi} + \delta_{eg}(1+e_{ig})^{\phi}\right]^{\frac{1}{\phi}} \varepsilon^q_{ig}, \tag{8}$$

where A_g is the total factor productivity, ϕ is the substitution parameter, $\delta_{q,g}$ is the marginal effect of the previous test score, $\delta_{s,g}$ is the marginal effect of school days, δ_{eg} is the marginal effect of private tutoring expenditure, and ε^q_{ig} is the time-varying shock to the test score.

6. Structural Estimation

In this section, we first explain the estimation strategy of constrained (penalized) maximum likelihood. Then, we discuss the identification of the structural parameters. Finally, we present the structural estimates and model fit.

6.1 Maximum Likelihood

Individual likelihood contribution The individual likelihood contribution is composed of the contributions regarding private tutoring expenditure e_{ig} and test scores q_{ig} for $g = -1, 0$.

Denoting θ as the set of parameters, each household i's likelihood contribution is

$$\mathcal{L}_i(\theta|q_{i,-1}\, w_{i,-1}, w_{i0}) = \prod_{g=-1}^{g=0} \mathcal{L}_{ig}(\theta|q_{ig}, w_{ig}), \tag{9}$$

and

$$\begin{aligned}\mathcal{L}_{ig}(\theta|q_{ig}, w_{ig}) = & \Big[f_{e_{ig}}(e_{ig}) \cdot f_{q_{i,g+1}}(q_{i,g+1}|e_{ig}) \Big]^{d^e_{ig}} \\ & \times \Big[\Pr(e_{ig} = 0) \cdot f_{q_{i,g+1}}(q_{i,g+1}|e_{ig}) \Big]^{(1-d^e_{ig})}\end{aligned}$$

where d^e_{ig} is an indicator function of tutoring participation decision (i.e. $e_{ig} > 0$ if $d_{ig} = 1$). The unrestricted part of the likelihood function is

$$\mathcal{L}(\theta) = \prod_{i=1}^{N} \mathcal{L}_i(\theta_g|q_{i,-1}, w_{i,-1}, w_{i0}).$$

Appendix D provides technical details on evaluating the likelihood contributions using the corresponding shocks and the Jacobian transformation.

Linking TWFE with the Structural Model We impose three constraints to ensure that the structural model aligns with our quasi-experimental evidence: (i) the TWFE estimates of how changes in the number of school days affect average private tutoring expenditure, (ii) the TWFE estimates of how these changes affect within-school test score inequality, and (iii) the parallel trends assumption underlying the TWFE estimation itself. For (i) and (ii), we run nested TWFE regressions using simulated data to fit reduced form TWFE estimates. For (iii), we simulate a counterfactual behavior of the treated group in the absence of treatment, ensuring that the treatment group behaves like control groups. We incorporate these three constraints into the structural estimation via an objective function that augments the likelihood with penalty terms. [23]

Finally, we solve

$$\widehat{\theta} = \arg\max_{\theta} \Big[\log \mathcal{L}(\theta) \;-\; \lambda\, \mathcal{C}_{\text{TWFE}}(\theta) \Big]$$

where λ is the positive penalty term. The functions $\mathcal{C}_{\text{TWFE}}(\theta)$ is a distance between the structural model's predictions and the relevant TWFE estimates and parallel constraint terms.

[23] More broadly, a handful of papers specify objective function that mimics likelihood function with aggregate moments as a penalty term (e.g., Kang (2016))

6.2 Identification

Our specified parametric form for both the production function and the flow utility function provides a framework that balances flexibility, allowing for heterogeneity across cohorts and grades, while imposing necessary restrictions on parameters to ensure model identification.

For the household preference, α_g is identified by the average level of private tutoring expenditure, which is allowed to differ by cohorts. The unobserved heterogeneity in consumption, σ_c, is identified by the residual variation of the difference between income and private tutoring expenditure that is not explained by the model.

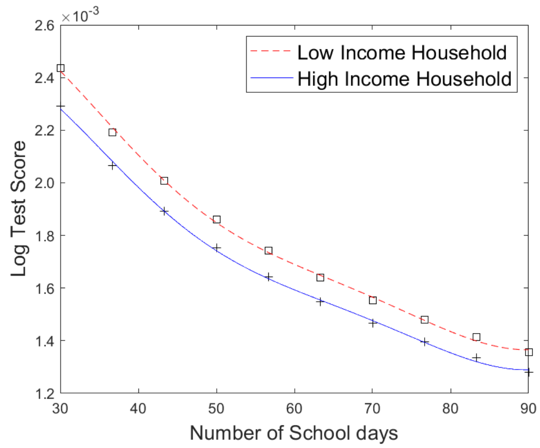
In the production function, we assume that the substitution parameter ϕ, the relative effect of school days κ, and the standard deviation of unobserved heterogeneity in test scores σ_q are constant across the population. The covariation between average tutoring expenditure and the number of school days, and the covariation of these inputs with subsequent test scores jointly identify ϕ. The relative weight of school days compared to private tutoring κ is identified by the covariation of subsequent test scores and the number of school days. The unobserved heterogeneity in test scores, σ_q, is identified by the residual variation not explained by the observable inputs in the CES production function in equation (7).

To capture time-varying heterogeneity across different grade levels and cohorts, we allow several production-function parameters to vary by grade g and cohort c. Specifically, we allow the constants (A), the CES weight of tutoring (δ_e), and the CES weight of school days (δ_s) to vary by cohorts and grades. These parameters are identified by covariation between the inputs and subsequent test scores, for which we have year-cohort variations. We however maintain that the ratio of these parameters remains constant over time, so that $\delta^c_{sg}/\delta^c_{eg} = \kappa$. This restriction stems from the limited variation available for cohorts who did not experience the Covid-19 school closure.

6.3 Structural Estimates and Model Fit

We present the structural estimates, explain what they imply for (i) substitution between school days and tutoring expenditure, and (ii) cohort-invariance in effects of tutoring expenditure on test scores. We then examine the model fit.

Figure 5: Marginal Effects of School Days



Note: Marginal effects are the first-order derivative of log test score with respect to schdays. Effects are depicted separately for high-income households (above the 80th income percentile) and low-income households (below the 20th income percentile).

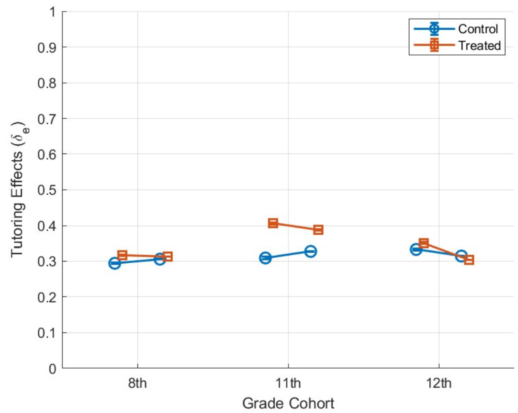
Table 6: Structural Parameters

Parameter	**Description**	**Estimate (S.E.)**
ϕ	Substitution between a school day and tutoring expenditure	0.517 (0.001)
κ	Relative weight of one school day to tutoring expenditure	0.322 (0.001)
σ_q	Test score shock standard deviation	0.262 (0.000)
σ_c	Consumption standard deviation	0.811 (0.003)

Note: Presented parameter are common across cohorts and grades. Standard errors in parentheses computed using delta method.

Table 6 reports the estimates of time-invariant structural parameters. The CES substitution elasticity between school days and tutoring expenditure ϕ is 0.517, which indicates the two inputs are gross substitutes and is consistent with our reduced-form results that the school closure raises tutoring spending. Intuitively, this implies that an additional school day is worth more to households that are constrained for monetary investment. A positive substitution parameter therefore magnifies the returns to school days for low-income households. Figure 5 illustrates this intuition: the marginal product of an additional school day, calculated from the first-order derivative of the CES production function conditional on tutoring expenditure and student-household characteristics (previous test scores and income), is consistently higher for low-income than for high-income households. This mechanism again aligns with our empirical finding that low-income households increased tutoring expenditure more than high-income households, as they suffer from larger losses for which they attempt to compensate.

Figure 6: Tutoring Effects by Grade Cohort and Treatment Group



Note: This figure illustrates the estimated tutoring effects (δ_e) across three grade cohorts (8th, 11th, and 12th grades). Estimates for control and treated groups at two periods (g and $g+1$) are presented side by side, with distinct marker shapes indicating grades (circles for 8th, squares for 11th, triangles for 12th). Error bars represent standard errors of the estimates.

The estimate of the relative CES weight of school days to tutoring expenditure, $\hat{\kappa}$, is 0.322 (Table 6). Figure 6 shows that the marginal productivity of tutoring expenditure, $\hat{\delta}_e$, lies between 0.30 and 0.40 in every cohort–grade cell. Multiplying these $\hat{\delta}_e$ values by $\hat{\kappa}$ implies an associated marginal effect of a school day, $\hat{\delta}_s$, of about 0.10–0.13. In words, we do not observe any higher marginal effect of tutoring for the cohorts that faced COVID-19 school closures ("treated") compared to others ("control"). It suggests that the tutoring response to school closures is driven by input substitution under binding budgets rather than cohort heterogeneity in tutoring effect. The full set of structural estimates is reported in Table D.1.

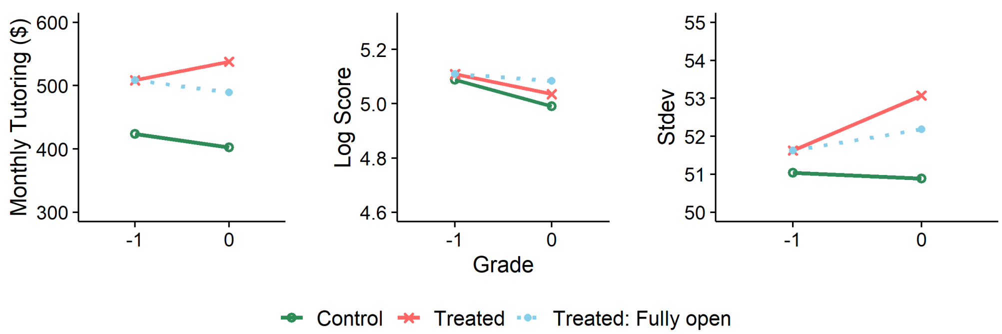
Table 7: TWFE Moments Fit

	Data	Model
Coefficient on schdays in **log tutoring**	-0.041	-0.045
Coefficient on schdays in **log test score**	0.002	0.002
Coefficient on schdays in **stdev**	-0.289	-0.253

Note: This table presents the fit of the TWFE regression coefficients for data and simulation. The reduced form estimates were presented earlier part of the paper in Tables 2, 5, and B.3.

Finally, we examine the model fit and the implementation of the parallel trends assumption. First, the model fits the TWFE estimates. Table 7 presents the fit of the TWFE estimates of the previous section and the simulated TWFE within the model, which shows an excellent fit. The model also captures the key data pattern for different cohort and treated control groups, which is reported in Table D.2. Secondly, imposing parallel trend clearly demonstrates the difference between the treated vs control groups. Figure 7 presents the simulated tutoring expenditure with the parallel trends constraint. For the cohort who experience the COVID-19 school closure, tutoring expenditure shows a steeper increase, and the test score inequality worsens. The dotted lines represent the counterfactual scenario if the schools had remained fully open. Tutoring expenditure would have decreased significantly, following the trends of the control group, and school-level inequality would not have risen as substantially in the absence of the school closures.[24]

Figure 7: Trends of Simulated Outcome Variables



Note: This figure presents the implementation of a parallel trend in the structural estimation results using simulated monthly tutoring expenditure. Specifically, it shows the trends of average tutoring expenditure for three different cohorts as students progress from g to $g+1$.

7. Policy Experiments

After-school tutoring voucher is increasingly used as a tool to help disadvantaged students recover from learning disruptions. The UK's National Tutoring Programme and South Korea's 2020 Future Education Center budget exemplify real-world initiatives expanding access to tutoring. Using the estimated model, we first simulate a counterfactual full-open scenario in order to quantify the impact of school closures on students' educational outcome. We then conduct ex-ante evaluations of voucher program, and assess the current and alternative tutoring voucher policies. In evaluating the baseline and alternative policies, we use test score standard deviation under the full-open scenario as a benchmark.

[24]We discuss this as a counterfactual scenario in Section 7

7.1 School Reopening and Education Inequalities

Table 8: Baseline vs. Full Open (All Individuals)

	Baseline	Full Open
Tutoring Expenditure (Average)	63.14	57.41
Log Test Score (Mean)	5.0360	5.0853
Log Test Score (Stdev)	0.3961	0.3649

Note: This table presents a comparison between the baseline simulation (Baseline) and the counterfactual simulation where in-person school days total 95 days (Full Open).

We first simulate a counterfactual case with no school closures by increasing the number of school days to pre-pandemic level of 95 days. Table 8 presents the results of this full-open counterfactual. Compared to the baseline model with actual school closures, the Full Open simulation decreases test score standard deviation by 7.8% and increases the average log score by 0.9%. These findings suggest that school closures widened education inequality with a modest impact on the mean of the test scores, which is consistent with our TWFE estimates. The relatively modest impact on mean scores compared to the larger effect on dispersion aligns with our earlier findings on differential marginal effects. As demonstrated in Figure 5, low-income households experience steeper marginal effects from school days, making them more vulnerable to closures. While the mean test score change is modest because both household types experience negative effects from school closures and attempt to compensate through tutoring, the inequality widens due to fundamental differences in their capacity to compensate. Low-income households face income constraints that limit their ability to fully offset learning losses despite their efforts, explaining why closures significantly increased test score variance without dramatically reducing mean scores.

7.2 Targeted Private Tutoring Voucher

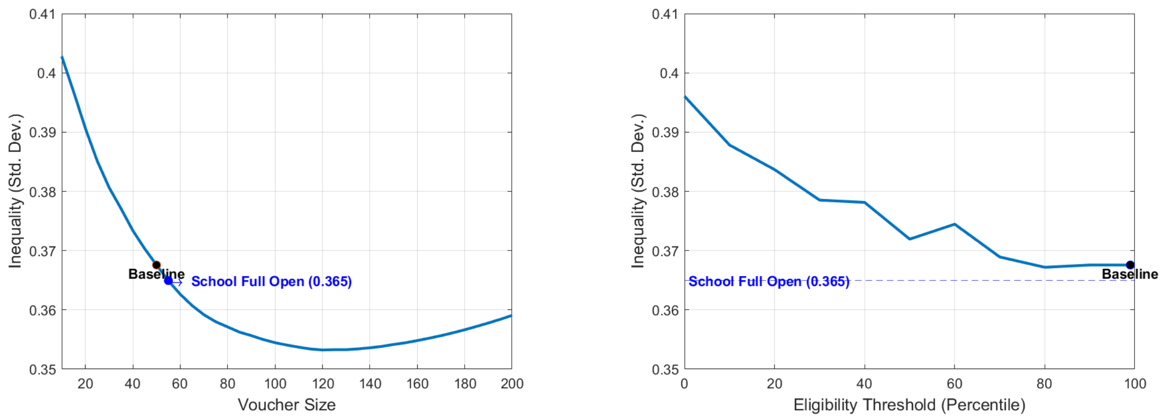
Baseline Voucher Policy: Our initial policy evaluation assesses the Future Education Center tutoring voucher program implemented by the Korean Ministry of Education to help students recover from pandemic-related disruptions. This universal voucher lacked explicit means-testing criteria. The government funded a group tutoring program between students and tutors, subsidizing 500K KRW (around $400 USD) per student.[25]

Our simulation exercise demonstrates that universal voucher policy is inefficient. As shown in Figure 8b, which illustrates the effect of varying eligibility thresholds at a fixed voucher

[25] Tutors were primarily college students. Most groups were matched four students to each tutor.

Figure 8: Inequality Reduction under Baseline Voucher Policy

(a) Impact of Voucher Size (b) Impact of Eligibility Threshold



Note: The horizontal dashed line represents inequality under full school reopening (standard deviation of 0.365). "Baseline" indicates current policy scenario.

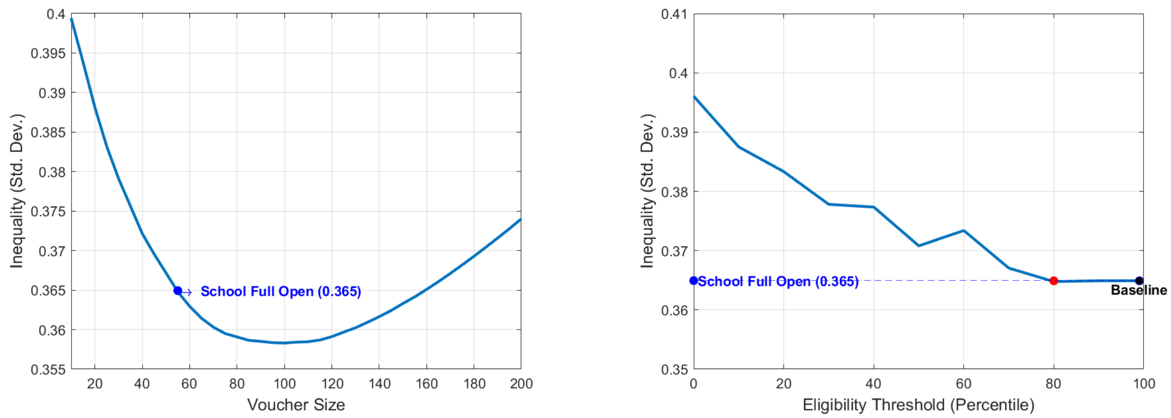
size of 500,000 KRW, inequality never reaches the benchmark level achieved with full school reopening (0.365). This indicates the baseline voucher size is insufficient regardless of targeting distribution. Eligibility thresholds are based on household income percentiles, where a threshold of 90% means households up to the 90th percentile are eligible. We observe minimal improvement when eligibility extends beyond the 80th percentile, revealing diminishing returns from broader coverage and substantial inefficiency in an untargeted system.

The voucher policy's effect on reducing inequality is concave, indicating diminishing marginal returns. Figure 8a demonstrates how educational inequality responds non-linearly to increasing voucher size under the universal scheme. Voucher amounts initially increase, inequality decreases sharply, due to higher marginal returns in voucher size less than 1M KRW. It leads to substantial benefits for disadvantaged students. However, we observe diminishing returns beyond a certain point, with inequality actually rising when voucher sizes exceed 1.2M KRW (approximately 1,000 USD), suggesting potential inefficiencies at very high subsidy levels.

Alternative Voucher Policy: We propose a targeted voucher program with a size of 550K KRW and eligibility threshold at the 80th income percentile as a cost-improving alternative policy. Figure 9 presents the results of this alternative policy simulations. Figure 9a shows how inequality responds to variations in voucher size when eligibility is fixed at the 80th percentile. Figure 9b illustrates the impact of varying the eligibility threshold while holding the voucher amount fixed at 550K KRW ($\approx$440 USD).

Figure 9: Inequality Reduction under Alternative Voucher Policies

(a) Impact of Voucher Size (b) Impact of Eligibility Threshold



Note: The horizontal dashed line represents inequality under full school reopening (standard deviation of 0.365). "Baseline" indicates the scenario without voucher intervention.

The simulation results demonstrate that a 550K KRW voucher eligible for households below the 80th percentile is sufficient to fully restore educational inequality to pre-pandemic levels. Beyond this size of voucher, additional funding yields diminishing and eventually negative equity returns, suggesting potential distortions from excessively large subsidies. Figure 9b reveals minimal inequality reduction when eligibility extends beyond the 80th percentile, confirming significant efficiency gains from targeted rather than universal distribution.

Discussion: Our analysis demonstrates that a targeted voucher program below the 80th income percentile effectively restores educational inequality to pre-pandemic levels while optimizing cost efficiency. This finding aligns with our production function estimates, which reveal higher marginal effects of school attendance for lower-income students (Figure 5), making them more responsive to educational interventions. This heterogeneity renders universal voucher schemes inefficient. As Figure 9a demonstrates, voucher effectiveness displays a non-monotonic relationship with inequality reduction, with excessive subsidies actually exacerbating inequalities. Targeting vouchers to households below the 80th income percentile captures nearly all potential equity gains at substantially lower costs, offering a viable alternative when full school reopening is infeasible.

8. Conclusion

This paper examines the role of the tutoring voucher as a potential mitigator of educational inequality caused by COVID-induced school closures. We identify causal effects that demonstrate significant impacts on parental investment and student outcomes. Building on these empirical findings, we develop and estimate a tournament model to conduct policy counterfactual experiments. Our analysis demonstrates that targeted voucher policies are more cost effective than universal approaches while achieving comparable effects in mitigating educational inequality that rose during school closures.

The cost of school closure is substantial and disproportionately burdens vulnerable households. Our analysis underscores the equalizing role of public education and advocates targeted voucher policies as an effective alternative when full school reopening is not feasible. These findings underscore the need to account for heterogeneity when designing compensatory education policies, and offer evidence to inform the broader debate on crisis-time schooling interventions.

References

Agostinelli, Francesco, Matthias Doepke, Giuseppe Sorrenti, and Fabrizio Zilibotti, "When the Great Equalizer Shuts Down: Schools, Peers, and Parents in Pandemic Times," *Journal of Public Economics*, February 2022, *206*, 104574.

Alves, Guillermo, William H Burton, and Sebastián Fleitas, "Difference-in-differences in equilibrium: Evidence from placed-based policies," 2023.

Andrabi, Tahir, Benjamin Daniels, and Jishnu Das, "Human Capital Accumulation and Disasters: Evidence from the Pakistan Earthquake of 2005," *Journal of Human Resources*, 2023, *58* (4), 1057–1096.

Aum, Sangmin, Sang Yoon (Tim) Lee, and Yongseok Shin, "COVID-19 Doesn't Need Lockdowns to Destroy Jobs: The Effect of Local Outbreaks in Korea," *Labour Economics*, 2021, *70*, 101993.

Bacher-Hicks, Andrew, Joshua Goodman, and Christine Mulhern, "Inequality in Household Adaptation to Schooling Shocks: Covid-Induced Online Learning Engagement in Real Time," *Journal of Public Economics*, January 2021, *193*, 104345.

Bicker, Laura, "Coronavirus in South Korea: How 'Trace, Test and Treat' May Be Saving Lives," *BBC*, March 2020. https://www.bbc.com/news/world-asia-51836898.

Blundell, Richard, Monica Costa Dias, Costas Meghir, and Jonathan Shaw, "Female Labor Supply, Human Capital, and Welfare Reform," *Econometrica*, 2016, *84* (5), 1705–1753.

Bray, TM, *The shadow education system: Private tutoring and its implications for planners*, UNESCO International Institute for Educational Planning., 1999.

Busso, Matias, Jesse Gregory, and Patrick Kline, "Assessing the incidence and efficiency of a prominent place based policy," *American Economic Review*, 2013, *103* (2), 897–947.

Caetano, Carolina, Brantly Callaway, Stroud Payne, and Hugo Sant'Anna Rodrigues, "Difference in Differences with Time-Varying Covariates," 2024.

Cajner, Tomaz, Leland D Crane, Ryan A Decker, John Grigsby, Adrian Hamins-Puertolas, Erik Hurst, Christopher Kurz, and Ahu Yildirmaz, "The U.S. Labor Market during the Beginning of the Pandemic Recession," Working Paper 27159, National Bureau of Economic Research May 2020.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C Sant'Anna, "Difference-in-differences with a Continuous Treatment," Working Paper 32117, National Bureau of Economic Research February 2024.

Choi, Eleanor Jawon and Jisoo Hwang, "Transition of Son Preference: Evidence From South Korea," *Demography*, April 2020, *57* (2), 627–652.

Chung, Woojin and Monica Das Gupta, "The Decline of Son Preference in South Korea: The Roles of Development and Public Policy," *Population and Development Review*, 2007, *33* (4), 757–783.

Cunha, Flavio, James J Heckman, and Susanne M Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 2010, *78* (3), 883–931.

D'Haultfœuille, Xavier and Philippe Février, "The provision of wage incentives: A structural estimation using contracts variation," *Quantitative Economics*, 2020, *11* (1), 349–397.

Engzell, Per, Arun Frey, and Mark D. Verhagen, "Learning Loss Due to School Closures during the COVID-19 Pandemic," *Proceedings of the National Academy of Sciences*, 2021, *118* (17), e2022376118.

Fuchs-Schündeln, Nicola, Dirk Krueger, Alexander Ludwig, and Irina Popova, "The Long-Term Distributional and Welfare Effects of Covid-19 School Closures," *The Economic Journal*, 04 2022, *132* (645), 1647–1683.

Galiani, Sebastian and Juan Pantano, "Structural Models: Inception and Frontier," Working Paper 28698, National Bureau of Economic Research April 2021.

Grewenig, Elisabeth, Philipp Lergetporer, Katharina Werner, Ludger Woessmann, and Larissa Zierow, "COVID-19 and Educational Inequality: How School Closures Affect Low- and High-Achieving Students," *European Economic Review*, November 2021, *140*, 103920.

Hahn, Youjin, Hyuncheol Bryant Kim, and Hee-Seung Yang, "Impacts of In-Person School Days on Student Outcomes and Inequality: Evidence from Korean High Schools During the Pandemic," *Unpublished Manuscript*, 2023.

Heckman, James J, "Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic literature*, 2010, *48* (2), 356–398.

Jack, Rebecca, Clare Halloran, James Okun, and Emily Oster, "Pandemic Schooling Mode and Student Test Scores: Evidence from US School Districts," *The American Economic Review: Insights*, June 2023, *5* (2), 173–190.

Kang, Changhui, "Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea," *Journal of Urban Economics*, 2007, *61* (3), 458–495.

Kang, Hyunjae, "Dynamic Competition in Parental Investment and Child's Efforts," *Unpublished Manuscript*, 2024.

Kang, Karam, "Policy influence and private returns from lobbying in the energy sector," *The Review of Economic Studies*, 2016, *83* (1), 269–305.

Kim, Jeonghyeok, "The Long Shadow of School Closures: Impacts on Students' Educational and Labor Market Outcomes," *Unpublished Manuscript*, 2024.

Maldonado, Joana Elisa and Kristof De Witte, “The Effect of School Closures on Standardised Student Test Outcomes,” *British Educational Research Journal*, 2022, *48* (1), 49–94.

OECD, *PISA 2018 Results (volume III): What school life means for students’ lives*, PISA, OECD Publishing Paris, 2019.

Park, Hyunjoon, Jere R Behrman, and Jaesung Choi, “Causal effects of single-sex schools on college entrance exams and college attendance: Random assignment in Seoul high schools,” *Demography*, 2013, *50*, 447–469.

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

Parolin, Zachary and Emma K Lee, “Large Socio-Economic, Geographic and Demographic Disparities Exist in Exposure to School Closures,” *Nature Human Behaviour*, 2021, *5* (4), 522–528.

Sacerdote, Bruce, “When the Saints Go Marching Out: Long-Term Outcomes for Student Evacuees from Hurricanes Katrina and Rita,” *American Economic Journal: Applied Economics*, 2012, *4* (1), 109–135.

The Economist, “The pandemic will spur the worldwide growth of private tutoring,” *The Economist*, October 2021.

Todd, Petra E and Kenneth I Wolpin, “The Best of Both Worlds: Combining Randomized Controlled Trials with Structural Modeling,” *Journal of Economic Literature*, March 2023, *61* (1), 41–85.

Voena, Alessandra, “Yours, mine, and ours: Do divorce laws affect the intertemporal behavior of married couples?,” *American Economic Review*, 2015, *105* (8), 2295–2332.

Yoo, Sam Hyun, Sarah R Hayford, and Victor Agadjanian, “Old Habits Die Hard? Lingering Son Preference in an Era of Normalizing Sex Ratios at Birth in South Korea,” *Population Research and Policy Review*, 2017, *36* (1), 25–54.

A. Data Appendix

A.1 Descriptive Statistics

Table A.1: Descriptive Statistics

	8th graders		11th graders		12th graders	
	Treated	Control	Treated	Control	Treated	Control
N (Students)	3,572	3,045	2,063	1,353	2,740	1,760
N (Schools)	121	56	102	41	379	63
Female (%)	46.9	45.9	47.5	53.5	50.5	51.5
Mother's Education (%)						
2-year College or Less	61.51	68.03	65.30	55.70	61.51	68.03
4-year College or More	38.49	31.97	34.70	44.30	38.49	31.97
Average HH Income	4,657	4,516	5,141	4,880	5,109	4,351
Private Tutoring Expenditure						
Korean	42.9	45.9	97.8	112.2	133.0	106.6
Mathematics	175.6	138.9	207.4	205.8	216.3	193.8
English	172.4	140.7	250.2	241.0	264.4	232.4

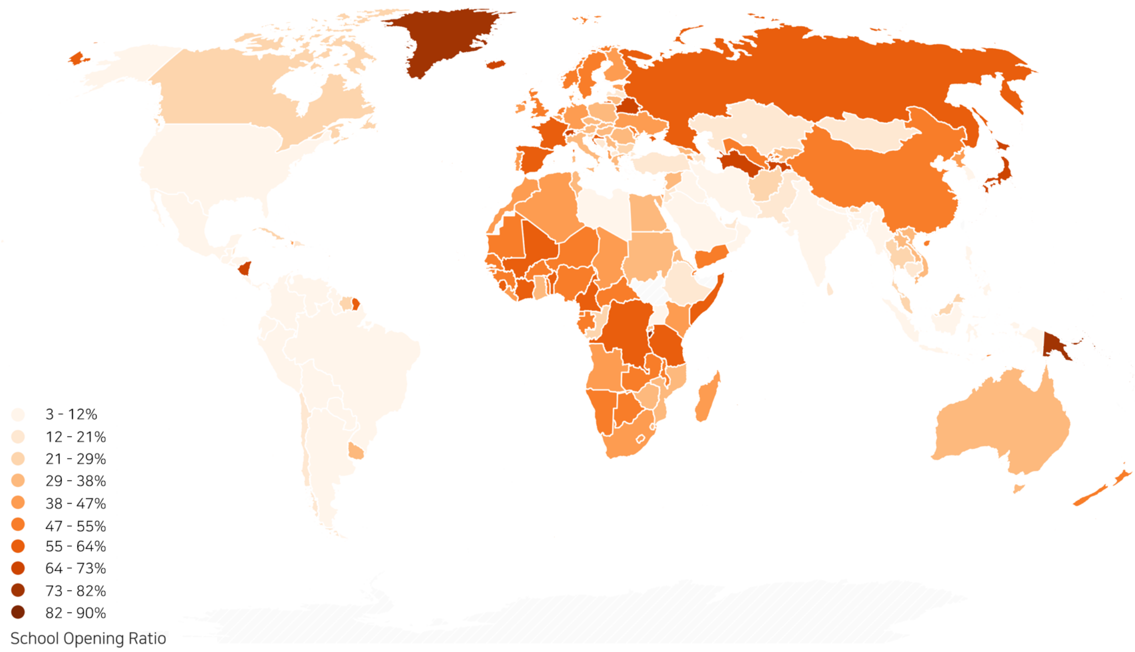
Note: Standard deviations in parentheses. All USD values are converted from KRW. (1 USD = 1,200 KRW) HH income and private tutoring expenditure are monthly values. "Selective school" is a sum of "autonomous" and "special purpose" schools. All numbers are calculated from the final year of each cohort.

A.2 Additional Data Description

Each year, for both surveys, information was collected through four separate questionnaires: one for teenagers, one for parents, one for teachers, and one for school principals. The survey data is then merged with administrative school information, including the number of teachers, students, classrooms, and campus size, among other details. This comprehensive design covers household information such as parental education, occupation, and income, providing valuable insights into the relationships between household composition and the educational factors of parents, including their education, occupation, and income.

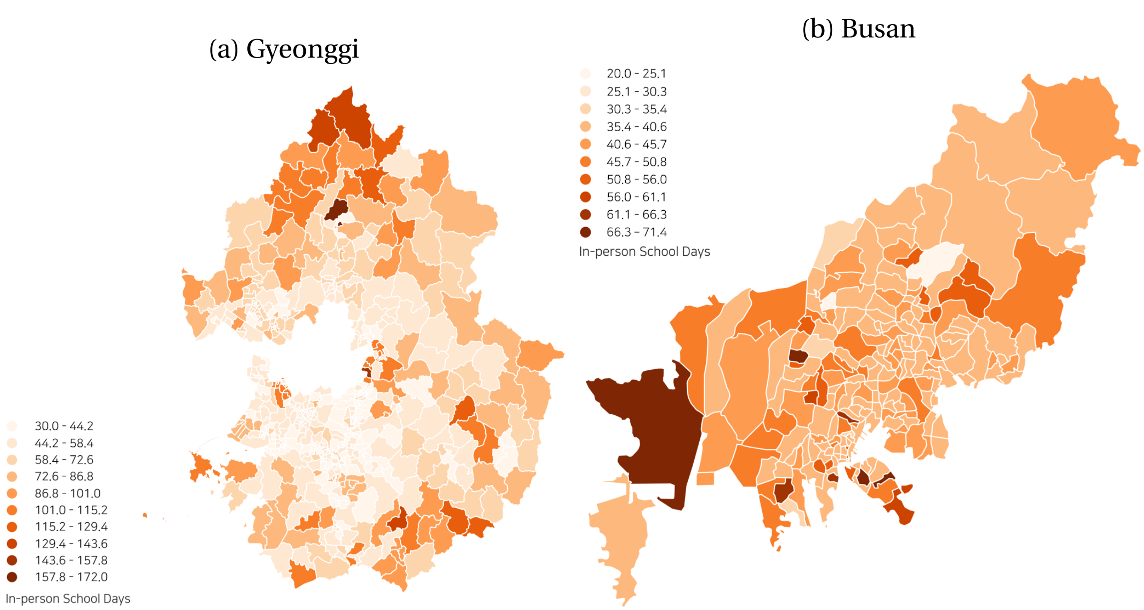
The datasets contains detailed information on students' learning behaviors. In particular, they ask detailed information about participation in private tutoring, their types (e.g., group meetings, one-on-one, online), their topics (e.g., Korean, English, math), the hours spent on them and their cost. Importantly, the datasets offer information about academic achievement of adolescents, measured by a dedicated cognitive tests (Middle School Sample) and national-level achievement test (High School Sample). It enables us to map the student decisions to the changes in academic achievement.

Figure A.1: Global School Opening Ratio During the Pandemic (2020-2021)



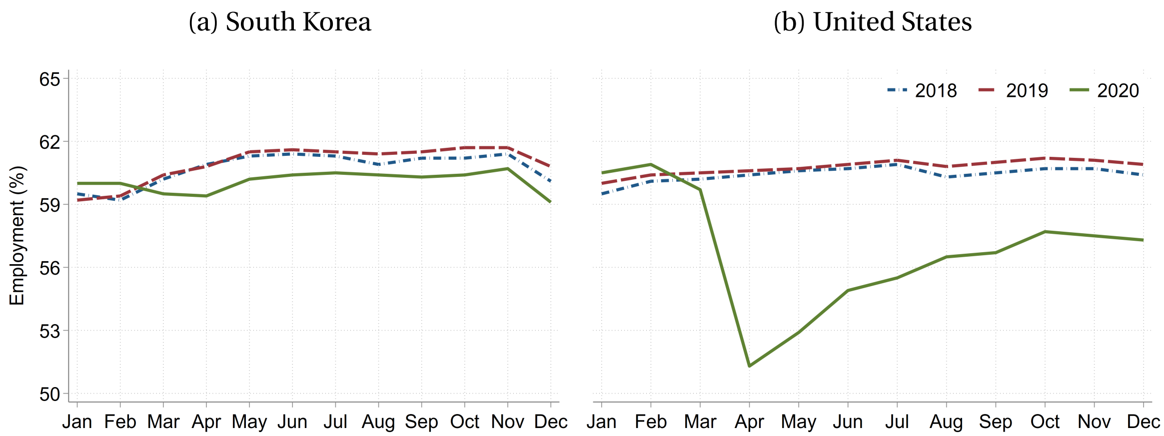
Data: World Bank

Figure A.2: Spatial variations in In-Person School Days



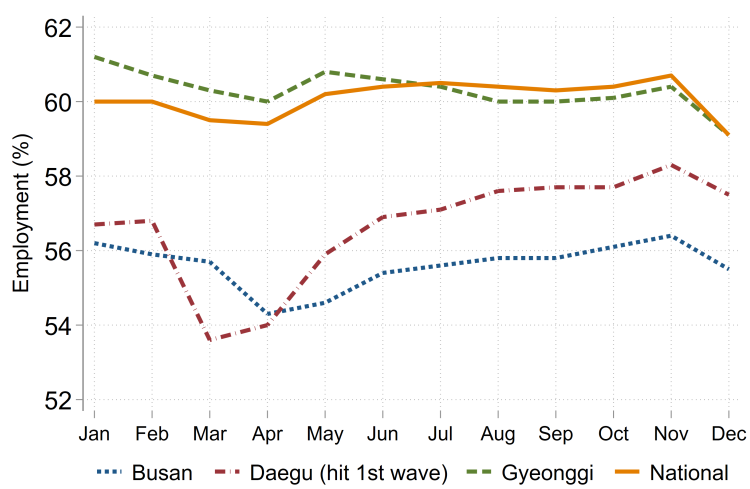
Data: SCHOOLINFO and administrative records.

Figure A.3: Monthly Employment by Country, 2020



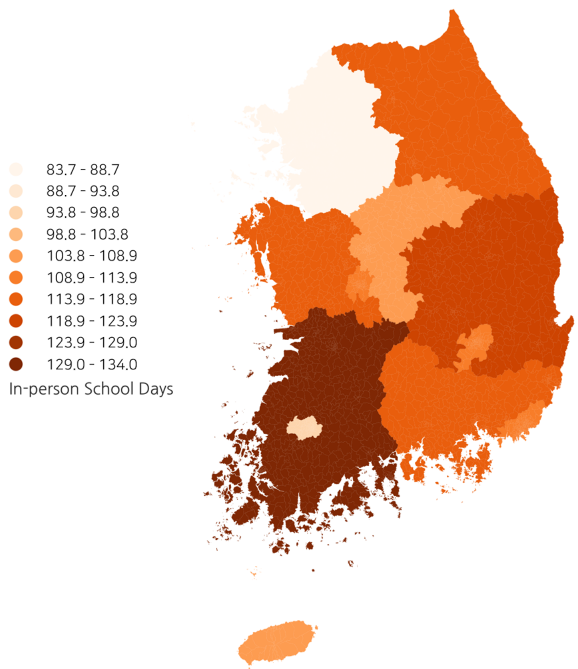
Data: Economically Active Population Survey (KR), CPS (US).

Figure A.4: 2020 Monthly Employment of South Korea, by Province



Data: EAPS.

Figure A.5: In-person School Days in South Korea (Province Level)



Data: SCHOOLINFO and administrative records.

B. Additional Empirical Results

Table B.1: Impact of In-Person School Days on Parental Expenditure

	COVID Sample			All Sample					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
In-person days	-0.017***	0.033	-0.017	-0.068***	-0.109***	-0.089***	-0.040***	-0.041***	-0.041***
	(0.003)	(0.031)	(0.030)	(0.005)	(0.010)	(0.007)	(0.008)	(0.008)	(0.008)
log(income)						0.889***		0.215***	0.215***
						(0.053)		(0.053)	(0.053)
Above Median=1 × In-person days									0.002
									(0.005)
N	9,700	9,700	9,700	17,164	17,164	17,164	17,164	17,164	17,164
Grade FE		✓	✓		✓	✓	✓	✓	✓
Student FE			✓				✓	✓	✓

Note: Clustered standard errors in parentheses (city level, $N = 47$). ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Table B.2: Impact of In-Person School Days on Test Scores

	COVID Sample			All Sample					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
In-person days	0.153**	0.147	-0.084	-0.668***	-1.220***	-0.674***	0.096	0.116	0.080
	(0.074)	(0.569)	(0.202)	(0.146)	(0.242)	(0.153)	(0.116)	(0.119)	(0.113)
log(expenditure)						1.107***		0.487***	0.486***
						(0.235)		(0.114)	(0.113)
Average Std. Score, 1-year lagged						0.341***			
						(0.030)			
Above Median=1 × In-person days									0.097
									(0.077)
N	9,700	9,700	11,443	17,164	17,164	17,164	17,164	17,164	17,164
Grade FE		✓	✓		✓	✓	✓	✓	✓
Student FE			✓				✓	✓	✓

Note: Clustered standard errors in parentheses (city level, $N = 47$). ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Table B.3: Impact of In-Person School Days on Log Test Scores

	COVID Sample			All Sample					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
In-person days	0.001*	0.001	-0.001	-0.007***	-0.012***	-0.007***	0.001	0.002	0.001
	(0.001)	(0.006)	(0.002)	(0.002)	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)
log(expenditure)						0.011***		0.006***	0.006***
						(0.002)		(0.001)	(0.001)
Average Std. Score, 1-year lagged						0.003***			
						(0.000)			
Above Median=1 × In-person days									0.001
									(0.001)
N	9,700	9,700	11,443	17,164	17,164	17,164	17,164	17,164	17,164
Grade FE		✓	✓		✓	✓	✓	✓	✓
Student FE			✓				✓	✓	✓

Note: Clustered standard errors in parentheses (city level, $N = 47$). ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Table B.4: Impact of In-Person School Days on Test Score: by Income and Performance

	Baseline	Prior Household Income			Prior Performance		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Low	Mid	High	Low	Mid	High
In-person days	0.116	0.325***	-0.103	-0.006	0.036	0.038	-0.622***
	(0.119)	(0.102)	(0.143)	(0.151)	(0.199)	(0.115)	(0.148)
p-value			0.015	0.644		0.994	0.000
N	17,164	5,912	6,358	4,894	5,656	5,722	5,786

Note: Clustered standard errors in parentheses (city level, $N = 47$). P-values in columns (3)-(4) and (6)-(7) test whether effects differ from their adjacent left columns. For example, the p-value in column (3) tests the difference between columns (2) and (3), while the p-value in column (4) tests the difference between columns (3) and (4). ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Table B.5: Impact of In-Person School Days on Test Score: by Gender and Parental Education

	Baseline	Gender		Education (Mother)		Education (Father)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	All	Boy	Girl	Below BA	BA+	Below BA	BA+
In-person days	0.116	0.169	0.054	0.225*	-0.111	0.268**	-0.075
	(0.119)	(0.130)	(0.172)	(0.123)	(0.136)	(0.113)	(0.139)
p-value			0.594		0.067		0.056
N	17,164	8,908	8,240	10,320	6,354	7,960	8,666

Note: Clustered standard errors in parentheses (city level, $N = 47$). 'BA' stands for a Bachelor's degree granted by a four-year college. P-values in columns (3), (5), and (7) test whether effects differ from their adjacent left columns. For example, the p-value in column (3) tests the difference between columns (2) and (3). ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Table B.6: Impact of In-Person School Days on Expenditure Stdev at School Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
In-person Days	-0.849***	-0.680***	0.386	0.271	0.251	0.251	0.251
	(0.145)	(0.131)	(0.249)	(0.257)	(0.245)	(0.245)	(0.245)
log(income)		23.078***	25.095***	0.599			0.683
		(3.535)	(3.570)	(7.675)			(7.655)
Above Median=1 × In-person Days				-0.059			
				(0.240)			
N	1,208	1,208	1,208	1,208	1,208	1,208	1,208
Grade FE			✓	✓	✓	✓	✓
School FE				✓	✓	✓	✓
City FE				✓		✓	✓

Note: Clustered standard errors in parentheses (city level, $N = 47$). ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

Table B.7: Impact of In-Person School Days on Test Score Std Dev at School Level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
In-person Days	0.040	-0.044	-0.291**	-0.270*	-0.270*	-0.289**	-0.289**
	(0.039)	(0.041)	(0.134)	(0.141)	(0.141)	(0.139)	(0.141)
log(expenditure)		-0.043***	-0.046***			-0.023*	-0.023*
		(0.008)	(0.008)			(0.013)	(0.013)
Above Median=1 × In-person Days							-0.000
							(0.070)
N	1,208	1,208	1,208	1,208	1,208	1,208	1,208
Grade FE			✓	✓	✓	✓	✓
School FE				✓	✓	✓	✓
City FE					✓	✓	✓

Note: Clustered standard errors in parentheses (city level, $N = 47$). ***/**/* indicate estimate is significantly different from 0 at the 1/5/10 percent level.

C. Structural Model Appendix

C.1 Construction of the prize term (v_j)

In this section I explain how the prize term v_j is constructed. The only difference with the Section 3.1 of Kang (2022) is that the discount factor is changed to 0.98, which reflects the change of interest rate. The prize term v_j differs by the college tiers. Conditional on the college-tier, the member colleges are assumed to share the same v_j. The tier-specific lifetime income v_j is defined as a discounted sum of the predicted income of the graduates, which is specified as

$$v_j = \sum_{t=T+1}^{T^*} \beta^{t-T} \hat{y}_{jt},$$

where $\hat{y}_{jt}$ is the income of the alumni of college tier j, and T^* is the retirement age. $\hat{y}_{jt}$ is the tier-specific annual income at time t. The predicted lifetime income $\hat{y}_{jt}$ is predicted using the regression equation,

$$\ln y_{it} = \sum_{j=1}^{J} (\beta_j + \delta_j \cdot age_{it}) D_{i,j}^{Tier} + Z_{\mathrm{i}}\gamma + \varepsilon_{it}, \tag{10}$$

where $D_{i,j}^{Tier}$ is the dummy variable indicating that person i graduated from a tier j college, and Z is the set of explanatory variables including squared age, birth year, and gender of person i. See Kang (2024) for the estimates of the Pooled-OLS estimates, which is used for the prediction. The resulted vector $\underset{5\times 1}{\mathcal{V}}$ is

$$\underset{5\times 1}{\mathcal{V}} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix} = \begin{pmatrix} 357554.68 \\ 295126.48 \\ 203498.74 \\ 172310.24 \\ 122007.46 \end{pmatrix}$$

C.2 Comparison: the “Max Log-score” model and the tournament model

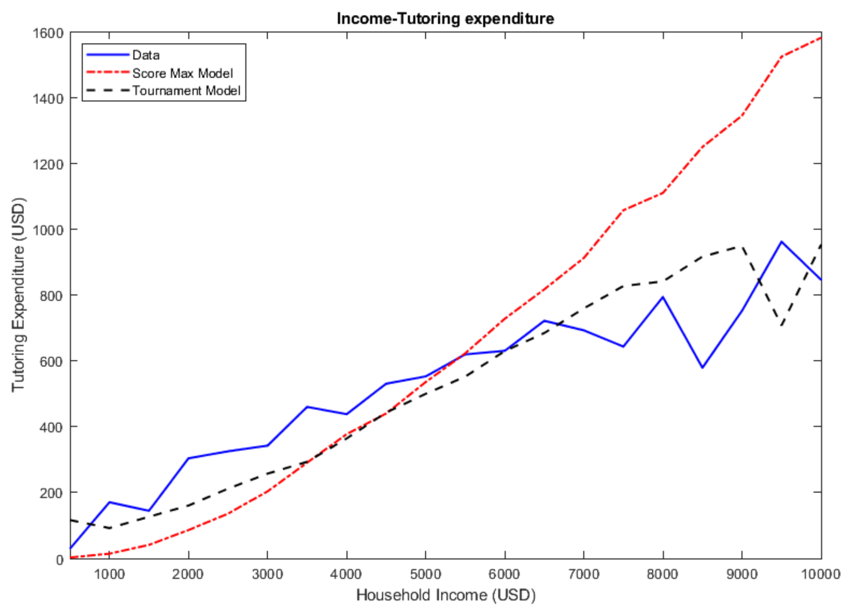


Figure C.1: Comparison with Max-Score Model

Figure C.1 presents the comparison between the ”Max-score” model and the tournament model against the empirical data. The Max-score model defines the household’s terminal value as the logarithm of the final test score. While this model differs from the tournament model only in its terminal value function (R_g), it significantly overestimates household expenditure on tutoring.

D. Likelihood Function Details

D.1 Tournament model

First Order Conditions The first order condition with respect to e_{ig} is

$$-\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}} + \alpha_g \frac{\partial}{\partial e_{ig}} \sum_{j=1}^{J} \ln(v_j) \times Prob(\ln \widetilde{Q}_{j-1} \geq \ln q_{i,g+1} \geq \ln \widetilde{Q}_j \Big| \Omega_{iT}, e_{ig}) = 0. \tag{11}$$

Equation 11 cannot be analytically solved in terms of e_{ig}, but the likelihood contribution is transformed from the analytical form of ε^c_{ig}. Denoting the argument of the production function as B for simplicity, the first order condition g is specified as

$$g = -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}} + \alpha_g\Bigg(\frac{\partial}{\partial e_{ig}}\sum_{j=1}^J \ln(v_j)\times Prob(\ln\widetilde{Q}_{j-1}\geq \ln q_{i,g+1}\geq \ln\widetilde{Q}_j)\Bigg)$$

$$= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}}$$

$$+ \alpha_g\frac{\partial}{\partial e_{ig}}\Bigg\{\sum_{j=1}^J \ln(v_j)\times\Bigg(\Phi(\frac{\ln\widetilde{Q}_{j-1}-\widehat{\ln q_{i,g+1}}}{\sigma_q}) - \Phi(\frac{\ln\widetilde{Q}_{j-1}-\widehat{\ln q_{i,g+1}}}{\sigma_q})\Bigg)\Bigg\}$$

$$= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}}$$

$$+ \alpha_g\Bigg[\sum_{j=1}^J \ln(v_j)\times\Bigg(\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_{j-1}-\widehat{\ln q_{i,g+1}}}{\sigma_q}) - \frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_{j}-\widehat{\ln q_{i,g+1}}}{\sigma_q})\Bigg)(-\frac{\nu_g\delta_{eg}(1+e_{ig})^{\phi_g-1}}{B})\Bigg]$$

$$= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}}$$

$$+ \alpha_g\Bigg[\ln(v_1)\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_0-\widehat{\ln q_{ig}}}{\sigma_q}) - \frac{1}{\sigma_q}\sum_{j=1}^{J-1}\Bigg(\ln(v_j)-\ln(v_{j+1})\Bigg)\phi(\frac{\ln\widetilde{Q}_j-\widehat{\ln q_{ig+1}}}{\sigma_q}) - \frac{1}{\sigma_q}\ln(v_J)$$

$$\phi(\frac{\ln\widetilde{Q}_J-\widehat{\ln q_{ig+1}}}{\sigma_q})\Bigg]\times(-\frac{\delta_3(1+e_{ig})^{\phi_g-1}}{B}), \text{ where } B = \delta_q q^\phi_{ig} + \delta_s Schdays^\phi_{ig} + \delta_e(1+e_{ig})^\phi$$

$$g = -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}} + \alpha_g\Bigg(\frac{\partial}{\partial e_{ig}}\sum_{j=1}^J \ln(v_j)\times Prob(\ln\widetilde{Q}_{j-1}\geq \ln q_{i,g+1}\geq \ln\widetilde{Q}_j)\Bigg)$$

$$= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}} + \alpha_g\frac{\partial}{\partial e_{ig}}\Bigg[\sum_{j=1}^J \ln(v_j)\times\Bigg\{\Phi\Bigg(\sigma_q^{-1}\ln\frac{\widetilde{Q}_{j-1}}{\widehat{q_{i,g+1}}}\Bigg) - \Phi\Bigg(\sigma_q^{-1}\ln\frac{\widetilde{q}_{i,g+1}}{\widetilde{Q}_j}\Bigg)\Bigg\}\Bigg]$$

$$= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}}$$

$$+ \alpha_g\Bigg[\sum_{j=1}^J \ln(v_j)\times\sigma_q^{-1}\Bigg\{\phi\Bigg(\sigma_q^{-1}\ln\frac{\widetilde{Q}_{j-1}}{\widehat{q_{i,g+1}}}\Bigg) - \phi\Bigg(\sigma_q^{-1}\ln\frac{\widetilde{q}_{i,g+1}}{\widetilde{Q}_j}\Bigg)\Bigg\}\Bigg(-\frac{\nu_g\delta_{eg}(1+e_{ig})^{\phi_g-1}}{B}\Bigg)\Bigg]$$

$$= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}} + \alpha_g\Bigg[\ln(v_1)\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_0-\widehat{\ln q_{ig}}}{\sigma_q}) - \frac{1}{\sigma_q}\sum_{j=1}^{J-1}\Bigg(\ln(v_j)-\ln(v_{j+1})\Bigg)\phi(\frac{\ln\widetilde{Q}_j-\widehat{\ln q_{ig+1}}}{\sigma_q}) - \frac{1}{\sigma_q}\ln(v_J)$$

$$\phi(\frac{\ln\widetilde{Q}_J-\widehat{\ln q_{ig+1}}}{\sigma_q})\Bigg]\times(-\frac{\delta_3(1+e_{ig})^{\phi_g-1}}{B}),$$

Since $\ln\widetilde{Q}_0 = \infty$ and $\ln(\widetilde{Q}_j) = -\infty$, the first order condition H is specified as follows.

$$H = -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}} + \alpha_g\Bigg[-\sum_{j=1}^{J-1}\Bigg(\ln(v_j)-\ln(v_{j+1})\Bigg)\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_j-\widehat{\ln q_{i,g+1}}}{\sigma_q})\Bigg](-\frac{\nu_g\delta_g(1+e_{ig})^{\phi_g-1}}{B})$$

$$= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}} + \alpha_g\underbrace{\Bigg[\sum_{j=1}^{J-1}\Bigg(\ln(v_j)-\ln(v_{j+1})\Bigg)\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_j-\widehat{\ln q_{i,g+1}}}{\sigma_q})\Bigg]}_{\varphi}(\frac{\delta_g(1+e_{ig})^{\phi_g-1}}{B}) = 0$$

The likelihood contribution of e_{ig} is transformed from the specified shock $\ln\tilde{\varepsilon}^c_{ig}$ that makes $H = 0$. In particular,

$$\ln\tilde{\varepsilon}^c_{ig} = \ln\alpha_g + \ln\delta_{eg} + \log\varphi + (\phi_g - 1)\ln(1+e_{ig}) + \ln(w_{ig}-e_{ig}) - \ln B.$$

On the other hand,

$$\ln \tilde{\varepsilon}^q_{ig} = \ln q_{i,g+1} - \left(\delta_{0g} + \frac{1}{\phi_g}\ln(\delta_{qg} q_{ig}^{\phi_g} + \delta_{sg} schdays_{ig}^{\phi} + \delta_{eg}(1+e_{ig})^{\phi})\right).$$

D.2 Likelihood Evaluation details

Individual likelihood contribution The individual likelihood contribution is composed of the contributions regarding private tutoring expenditure e_{ig} and test score $q_{i,g+1}$.

For the test scores, the likelihood contribution is a probability density function (PDF) for all individuals. For private tutoring expenditure, the likelihood contribution differs depending upon whether the household participates in private tutoring activities. For the participants, likelihood contribution is a PDF since tutoring expenditure is a positive continuous variable. On the other hand, the non-participants' likelihood contribution is a cumulative distribution function, as it needs to integrate all possible shocks that make individual not participate. Denoting Θ^c_g as the set of parameters, each household i's likelihood contribution is

$$\mathcal{L}_i(\Theta^c_g | \{q_{ig}\}_g^{g+1}, \{w_{ig}\}_g^{g+1}) = \prod_g^{g+1} \int_{\varepsilon^q_{ig}} \mathcal{L}_{ig}(\theta | q_{ig}, w_{ig}) f(\varepsilon^q_{ig}) d\varepsilon^q_{ig}, \tag{12}$$

and

$$\begin{aligned}\mathcal{L}_{ig}(\Theta^c_g | q_{ig}, w_{ig}) = & \Big[f_{e_{ig}}(e_{ig}) \cdot f_{q_{i,g+1}}(q_{i,g+1}|e_{ig}) \Big]^{d^e_{ig}} \\ & \times \Big[\Pr(e_{ig}=0) \cdot f_{q_{i,g+1}}(q_{i,g+1}|e_{ig}) \Big]^{(1-d^e_{ig})};\end{aligned}$$

where d^e_{ig} is an indicator function of tutoring participation decision (i.e. $e_{ig} > 0$ if $d_{ig} = 1$). The unrestricted part of the likelihood function is

$$\mathcal{L}(\Theta^c_g) = \prod_{i=1}^{N} \mathcal{L}_i(\theta^c_g | \{q_{ig}\}_g^{g+1}, \{w_{ig}\}_g^{g+1}).$$

Jacobian Transformation I denote $\tilde{\eta}^c_{it} = \ln \tilde{\varepsilon}^q_{ig}$. For the participant of the tutoring expenditure, the Jacobian transformation is

$$
\begin{aligned}
|J_1| &= \det|\frac{\partial(\tilde{\eta}^c_{it}, \tilde{\eta}^q_{it})}{\partial(e_{ig}, q_{i,g+1})}| \\
&= \frac{\partial\tilde{\eta}^c_{it}}{\partial e_{ig}}\frac{\partial\tilde{\eta}^q_{it}}{\partial q_{i,g+1}} - \frac{\partial\tilde{\eta}^q_{it}}{\partial e_{ig}}\frac{\partial\tilde{\eta}^c_{it}}{\partial q_{i,g+1}} \\
&= \frac{\partial\tilde{\eta}^c_{it}}{\partial e_{ig}}\frac{\partial\tilde{\eta}^q_{it}}{\partial q_{i,g+1}}
\end{aligned}
$$

From $\frac{\partial}{\partial x}\frac{1}{\sigma}\phi(\frac{x}{\sigma}) = -\frac{x}{\sigma^2}(\frac{1}{\sigma}\phi(\frac{x}{\sigma}))$, we have $\frac{\partial\eta^c_{ig}}{\partial e_{ig}} = \frac{1}{\varphi}\frac{\partial\varphi}{\partial e_{ig}} + \frac{\phi-1}{(1+e_{ig})} - \frac{1}{w_{ig}-e_{ig}} - \frac{1}{\delta_q q^{\phi}_{ig}+\delta_s Schdays^{\phi}_{ig}+\delta_e(1+e_{ig})^{\phi}}\delta_e\phi(1+e_{ig})^{(\phi-1)}$, where $\varphi = \sum_{j=1}^{J-1}\Bigg(\ln(v_j)-\ln(v_{j+1})\Bigg)\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_j-\ln\widehat{q_{i,g+1}(e_{ig})}}{\sigma_q})$ and $\frac{\partial\varphi}{\partial e_{ig}} = \sum_{j=1}^{J-1}\Bigg(\ln(v_j)-\ln(v_{j+1})\Bigg)\{-\frac{\ln\widetilde{Q}_j-\ln\widehat{q_{i,g+1}(e_{ig})}}{\sigma^2_q}\}\frac{1}{\sigma_q}\phi(\ln\widetilde{Q}$ $= \sum_{j=1}^{J-1}\Bigg(\ln(v_j)-\ln(v_{j+1})\Bigg)\{\frac{\ln\widetilde{Q}_j-\ln\widehat{q_{i,g+1}(e_{ig})}}{\sigma^2_q}\}\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_j-\ln\widehat{q_{i,g+1}(e_{ig})}}{\sigma_q})(\frac{\partial\ln\widehat{q_{i,g+1}(e_{ig})}}{\partial e_{ig}})$ $= \sum_{j=1}^{J-1}\Bigg\{\Bigg(\ln(v_j)-\ln(v_{j+1})\Bigg)\{\frac{\ln\widetilde{Q}_j-\ln\widehat{q_{i,g+1}(e_{ig})}}{\sigma^2_q}\}\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_j-\ln\widehat{q_{i,g+1}(e_{it})}}{\sigma_q})\Bigg\}(\frac{\delta_e\phi(1+e_{ig})^{(\phi-1)}}{B})$. For the non-paritcipants, the Jacobian transformation term is

$$
\begin{aligned}
|J_2| &= \det|\frac{\partial\tilde{\eta}^q_{it}}{\partial q_{i,g+1}}| \\
&= \frac{\partial\tilde{\eta}^q_{it}}{\partial q_{i,g+1}}
\end{aligned}
$$

The likelihood contribution uses the equilibrium conditions of the theoretical framework. I denote $\tilde{\varepsilon}^c_{it}$ and $\tilde{\varepsilon}^q_{it}$ as the particular points of the shocks where the utility of the household is maximized. They are assumed to be jointly normal, and the likelihood function is based on $\eta^z_{it} = \ln\varepsilon^z_{it}$ for $z = c, q$. In particular,

$$
\begin{aligned}
\mathcal{L}_{ig}(\Theta^c_g|q_{ig}, w_{ig}) = &\Bigg[f_{\eta^c_{ig}}(\tilde{\eta}^c_{ig})\cdot f_{\eta^q_{ig}}(\tilde{\eta}^q_{ig}|\eta^c_{ig})|J^1_i|\Bigg]^{d_{ig}} \\
&\times\Bigg[\int_{\tilde{\eta}^c_{ig}}^{\infty} f_{\eta^c_{ig}}(\eta^c_{ig})\cdot f_{\eta^q_{ig}}(\tilde{\eta}^q_{ig}|\eta^c_{ig})d\eta^c_{ig}|J^2_i|\Bigg]^{(1-d^e_{ig})}
\end{aligned}
$$

where $|J^j_i|_{j=1,2}$ is the corresponding Jacobian transformation term. For the non-participants of the tutoring activities, they do not spend on tutoring expenditure if they are above the threshold of the consumption shock $\tilde{\eta}^c_{ig}$, which is the minimum amount of the shock that makes household stop spending on the tutoring expenditure. Note that θ^c_g depends on the student's grade g and the cohort the student belongs to.

D.3 Structural Estimates

Table D.1: Grade-Specific Structural Parameters

	Grade	8th		11^{th}		12^{th}	
		Control	Treated	Control	Treated	Control	Treated
Tutoring Effects	$g-1$	0.294	0.316	0.309	0.406	0.333	0.351
		(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
	g	0.306	0.313	0.328	0.388	0.315	0.304
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
School Days Effects*	$g-1$	0.095	0.102	0.100	0.131	0.107	0.113
		(-)	(-)	(-)	(-)	(-)	(-)
	g	0.099	0.101	0.105	0.125	0.101	0.098
		(-)	(-)	(-)	(-)	(-)	(-)
Altruism Parameters	$g-1$	0.342	0.369	0.694	0.955	0.953	1.568
		(0.005)	(0.010)	(0.014)	(0.025)	(0.021)	(0.026)
	$g+1$	0.433	0.522	0.738	1.349	0.765	1.004
		(0.005)	(0.013)	(0.004)	(0.023)	(0.016)	(0.018)
Previous Test Scores	$g-1$	1.120	0.538	1.077	0.761	0.768	0.750
		(0.002)	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
	g	0.330	0.414	0.316	0.502	0.250	0.172
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Note: Standard errors in parentheses computed using delta method. The effects of school days are implied estimates using estimated κ, thus indicated by "-".

Table D.2: Year-Specific Fit: Expenditure and Log Test Scores (Data vs. Model)

	Data		Model	
	$g=-1$	$g=0$	$g=-1$	$g=0$
Panel A: Tutoring Expenditure				
Treated 12th:	70.37	75.63	70.40	78.65
Control 12th:	53.63	55.10	57.28	56.34
Treated 11th:	62.41	64.31	61.89	60.75
Control 11th:	55.52	58.55	63.06	54.65
Treated 8th:	43.16	49.41	48.93	49.86
Control 8th:	35.46	37.94	37.94	37.63
Panel B: Log Test Scores				
Treated 12th:	4.68	4.66	5.27	5.07
Control 12th:	4.51	4.51	5.16	5.03
Treated 11th:	4.63	4.63	5.17	5.15
Control 11th:	4.60	4.61	5.13	5.05
Treated 8th:	4.63	4.63	4.92	4.92
Control 8th:	4.62	4.62	5.01	4.93

Note: Unit of tutoring expenditure is 10,000KRW $\approx$ 8 USD. Each cohort is named after their grade in $g=0$. For example, 12th grade cohort refers to students who were in 12th grade at $g=0$)