School Closure and Educational Inequality: Parental Investment in the Pandemic*

Hyunjae Kang Tae Young Kang Sunham Kim

November 8, 2022– Version 1.0
Preliminary

Abstract

Compulsory public education is a common policy tool to increase participation in education, particularly from marginalized groups. The COVID-19 pandemic brought unprecedented school closures around the globe. Households seek their own mitigation strategy against education loss via private substitutes. We use the COVID-19 response of South Korea as a laboratory for investigating the effects of pandemic-induced school closure on parental investment and educational outcomes. Specifically, we exploit the school-level variations in in-person learning days driven by government and local authorities. Combining administrative data and longitudinal survey, our fixed effects regression identifies that losing 10 additional days implies a 9.5% increase in private education expenditure. In addition, we build and estimate a structural model tournament to quantify the impact of school closure on educational inequalities. A counterfactual experiment suggests that a $450 private tutoring voucher for low-income students may compensate for the test score loss caused by a 20-day loss of in-person schooling. Our results shed light on how households make schooling decisions and have implications on the longer-run educational inequality induced by the pandemic. Methodologically, we propose an estimation algorithm that tightly links the natural experiment estimates and the structural model.

JEL Classification: D30, I18, I21, I24, I28
Keywords: school closure, private education, learning disparity, COVID-19

*H. Kang: Stony Brook University; T. Kang: KAIST; Kim: Purdue University. 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.

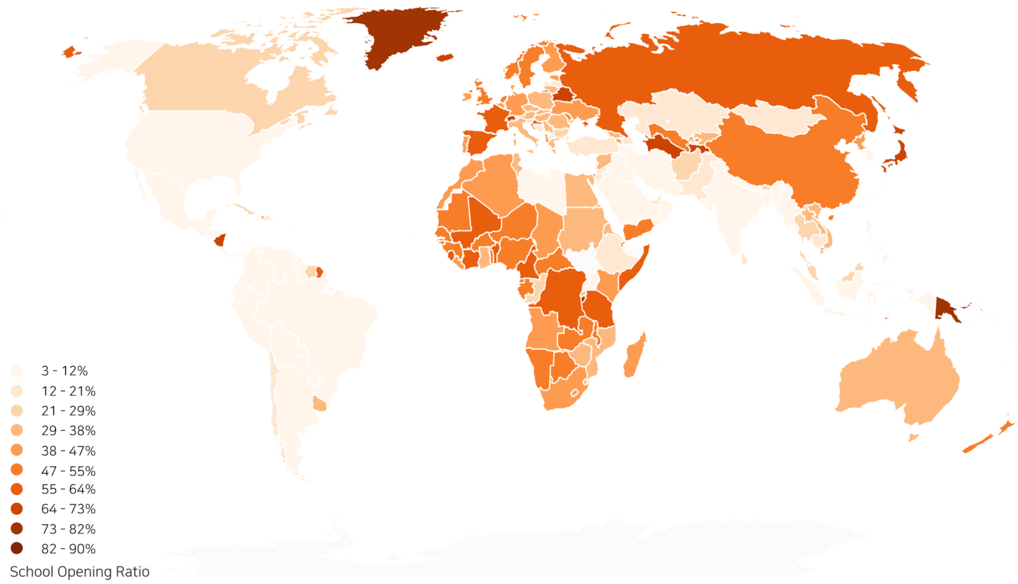
1. Introduction

Compulsory public education is a common policy tool to increase participation in education, particularly from marginalized groups. It operates through exposing students to standardized educational environment for certain amount of time. Although the state-of-the-art educational environment is provided, the education effect would not take place if students do not attend school. In general, absenteeism, either voluntary or involuntary, is reported to deteriorate the academic achievement of students. (Dobkin et al. 2010; Goodman 2014; Gottfried 2009, 2014; Aucejo and Romano 2016; Liu et al. 2021).

The COVID-19 pandemic brought unprecedented school closure around the globe. (Figure 1) "The Great Equalizer shuts down" (Agostinelli et al. 2022), households first seek their own mitigation strategy to offset education loss via private substitutes, which was remarkably heterogeneous by socioeconomic status, and likely to incur a divergence in academic performance. (The Economist 2021) While recent studies have used economic models to examine the impact of the COVID-19 school closure on educational outcomes and potential mechanisms (e.g., Agostinelli et al. 2022; Fuchs-Schündeln et al. 2020; Grewenig et al. 2021), systematic evidence from the real world remains to be provided. We first identify, empirically, that the impact of lost in-person learning days on private education expenditure and academic achievement using a novel panel dataset from South Korea. Second, we build and estimate a tournament model of parental investment in which the number of school days is augmented as an input of education production function. We then use the estimated model to quantify the impact of school closure on educational inequality and conduct counterfactual analyses.

We start by documenting the empirical findings. In order to study the link between in-person instructional days and educational inequality, we leverage

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



Data: World Bank

a novel administrative-survey linked data, the Gyeonggi Educational Panel Survey (GEPS) from South Korea. We exploit the policy-driven school-level variation of in-person days to establish causality. In the data, we identify that losing 10 additional in-person days implies a 3% increase in private education expenditure, after controlling for individual fixed effects. The estimate is higher for low-income households due to the base effect, which suggests the additional spending was more burdensome to them. Further, the 10-day lose is associated with a 0.7%p decrease in a percentile score from a national standardized test.

We choose South Korea as a laboratory to facilitate casual interpretation. The country is one of the countries with extensive school closure, allowing only 6.2% of its entire school days to be fully opened during two years of the pandemic.[1] The schools were fully closed during 24.7% of school days and

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

67.6% days were also under partial closure status. The exogenous nature of school closure dwarfs the caveat of reverse causality. The country has been successful in "flattening the curve" with no blanket lockdown. It makes our estimate less contaminated by the aggregate environment, compared to the ones from other developed countries. Moreover, the fact that the country is known for education fervor and has have no absenteeism alleviates the concern of selection bias. Armed with the causal interpretation, we develop a model that can feature the empirical facts.

Our framework takes the competition between students toward finite seats for colleges as the main driver of parental investment in our setting, which motivates the use of the tournament model of parental investment.[2] (Lazear and Rosen 1981; Grau 2018; Tincani et al. 2021; Kang 2022) To capture the substitution between public education and private parental investment, our test score function is specified as the constant elasticity of substitution (CES) production function, following recent advances in the early childhood development literature. (Cunha et al. 2010; Agostinelli and Wiswall 2016) The maximum likelihood estimation of the structural model fits the data well.

Using the estimated model, we quantify the educational inequality induced by the Covid-19 school closure. Simulating the model with the pre-Covid in-person learning days increases the mean of test score percentile by 8.9% and decreases the standard deviations of test score percentile by 23.1%, which suggests a negative impact of the Covid-19 school closure on the academic performance, and the educational inequalities. Furthermore, the estimated model suggests that the students from low-income households undergo a more drastic decline in the test score. Taking a step further, motivated by the results, we conduct a counterfactual policy of private tutoring

[2]An alternative model with an assumption of individual caring about the absolute performance severely overpredicts the choice of parental investment. [TO BE ADDED IN THE APPENDIX]

vouchers targeting low-income households. A private tutoring voucher of $2,100 may let the households fully offset the decrease in the test score on average, which suggests the substantial effects of the school closure. Our results shed light on how households make schooling decisions and has implication on the longer-run educational inequality induced by the pandemic. These results underscore the importance of public education as the balance wheel of the social machinery.

Related Literature This paper is connected to four different strands of literature. First, it is related to the literature of empirical studies of the effects of school closures on education. Most previous papers focus on the impact of school closures on educational outcomes, particularly test scores. We bring novel longitudinal data to the literature, which allows us to identify the effect of school closures on parental investment and educational outcomes controlling for unobserved individual heterogeneities. The granularity of the school closure measure varies among the empirical studies. Our attendance data is measured school level, which helps control the factors that otherwise would be unobservable. Secondly, this paper speaks to the literature of structural models of school closure and parental inputs. we estimate a structural model using the longitudinal data from the real world. Thirdly, there are many micro and macro papers that constructed structural models to tackle the question of COVID-19 and education. Due to the data limitation, they are more in the flavor of calibration instead of estimation. We complement them by combining our panel estimation and the structural model. Last but not least, this paper is linked to the papers trying to integrate the design-based and structural approaches. (Todd and Wolpin forthcoming; Galiani and Pantano 2021)

The rest of the paper is organized as follows. Section 2 outlines the background and empirical framework. Section 3 describes the data we use. Section 4 presents the reduced-form results. Section 6 defines the tournament model framework. Section 6 elucidates how we implement the structural estimation and reports the results. Section 7 conducts some counterfactual policy experiments based on the structural model. Section 8 concludes.

2. Institutions and Empirical Framework

2.1 Background on the Korean Education

South Korea has a highly centralized education system. All pre-college students follow a single track "6-3-3" system, which denotes six years of elementary school, three years of middle school, and three years of high school.[3] Most schools follow the same official curriculum, the National Curriculum for the Primary and Secondary Schools.[4] The government requires all primary and secondary schools to have 190 instructional days. Local school boards exert limited power on local education policy. Parental involvement is not a major decision-making factor in school as well, compared to the United States. Almost all youths are enrolled in such official education enrollments. Homeschooling, a form of decentralized education, is prohibited by law, so a tiny fraction of student chooses to be a homeschooler.

The centralized system stems from a long-standing fervor for education of Koreans. Education is regarded as the main driver of economic growth and the key determinant of individual career and social mobility. The lowest absence rate and highest literacy in OECD are the byproducts of the fervor. In order to meet the demanding public's expectations, especially to the fair and dependable education service, the government keeps control over education

[3] Elementary school is often referred to as primary school. Secondary school consists of middle and high school.

[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 to any meaningful extent.

and puts top priority on it. Schools have never been closed before the COVID pandemic under the various national crises after the Korean War, including military coups (1961, 1979) and financial crises (1997, 2007). No exception was taken to many epidemics before the COVID-19 such as MERS (2015), H1N1 (2009), SARS (2004), and even to cholera (1969), which was a serious threat to public health of the country before its economic takeoff.

While the official schooling system is centralized, education institutions are not limited to schools. There are numerous private academic institutes, called cram schools or "hagwon", that offer a variety of tutoring services as a complement of school education. The private education market is accessible to anyone while the service quality may vary by cost. According to the Ministry of Education (2010), 45% of elementary, 50% of middle school, and 84% of high school. Private education is often blamed being the "Great Unequalizer," as the main incentive for hagwon enrollment is to "win" the competition in the early stages of the game. The government has tried to shackle the market, based on concerns about its toll on households' finances and students' well-being. A law enacted in the late 2000s, which known as "curfew', even set a restriction on the business hours of hagwons. In sum, education is tauted as the "Great Equalizer" in South Korea, as in the United States. The government plays a major role as a service provider and wanted to control the private education market, the "Great Unequalizer."

2.2 The COVID Responses in Early 2020

The pandemic shock challenged the uninterrupted centralized system. When the first confirmed case was reported in January 2020, the Ministry of Education stuck to the normal academic calendar. As the situation escalates rapidly after the notorious super-spreader "Patient 31" and the first wave arrived, however, the MoE declared the unprecedented universal school closure on February 23rd and postponed the school opening from March 2nd to 9th. It took five more weeks to lift the closure, and the opening was fully remote. All high school students and 9th graders went back to school on April 9th. Younger groups, 5-8th and 1-4th graders did on April 16th and 20th, respectively. The 2020 national academic calendar had to be overhauled.

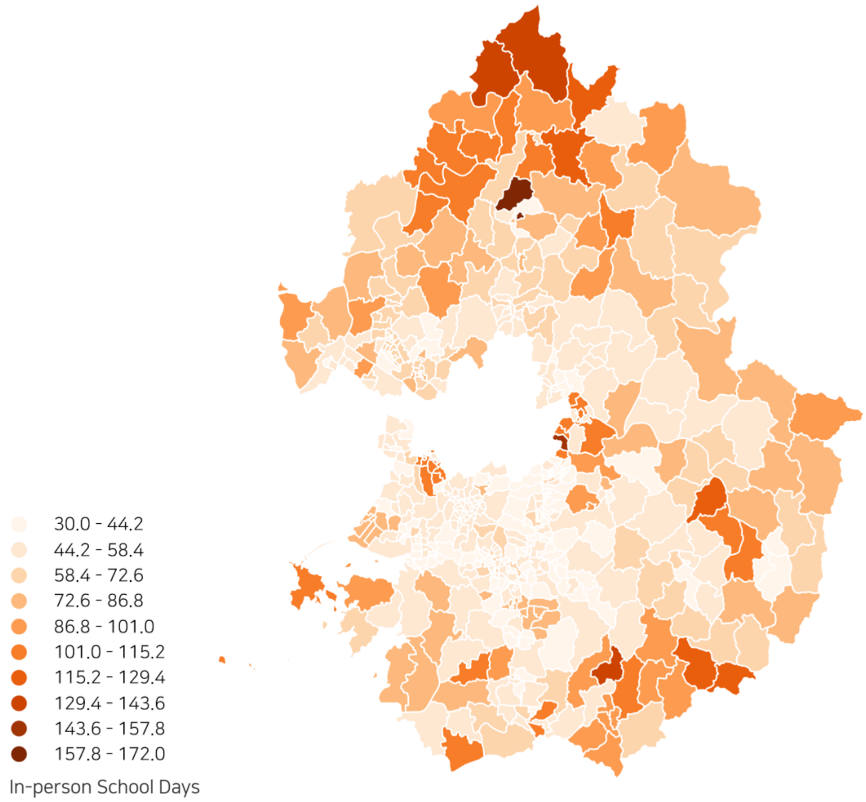
As the country passes the peak of the first wave, the government tried to open schools in person on May 13th. The schedule was postponed by a week as the second wave arrived in early May. Eventually, the 12th graders went back to school and all the other graders did in order by June 3rd. No nationwide school closure is placed ever since this full reopening. Providing some guidelines, the MoE left the open-or-remote decision to regional offices and school principals. This decentralized school-level variation is central to our identification strategy.

Importantly, private education institutions were open while schools were closed. The Korean government never placed a blanket lockdown, boasting their effective 'Trace, Test, and Treat (3T)' strategy. All private businesses were open with certain degrees of social distancing, so were hagwons. The MoE was not able to close hagwons as long as all the other private sectors were in business. Limited by social distancing, hagwons could leverage the existing online lecture programs. The online programs were out of control. Moreover, the no lockdown policy kept parents in workplaces so prevented them from monitoring and educating their children. They had no other option but private education. In other words, the virus shut down the "Great Equalizer" and unleashed the "Great Unequalizer."

2.3 Measurement Framework and Identification

We identify the effect of reduced in-person school days on parental investment. The treatment is the decentralized variations at the school level. In

Figure 2: In-Person School Days in Gyeonggi Province



Data: SCHOOLINFO and administrative records.

order to evaluate the impact of the short in-person school days, we estimate the following linear panel model:

$$y^c_{ikg} = \delta_0 + \delta_1 schdays^c_{kg} + \Gamma Z^c_{ikg} + \mu_i + \alpha D_{g=12th} + \varepsilon^c_{ikg}, \tag{1}$$

In the above equation, y^c_{ikg} is an outcome variable for individual i in school k, grade g, and cohort c. We focus on 11th and 12th graders in 2013 (control) and 2020 (treated) cohorts. y can be private education participation dummy, log expenditure, and academic performance.[5] $schdays^c_{kg}$ is the in-person

[5]We address the log of zero problem for the expenditure variable by applying the inverse hyperbolic transformation.

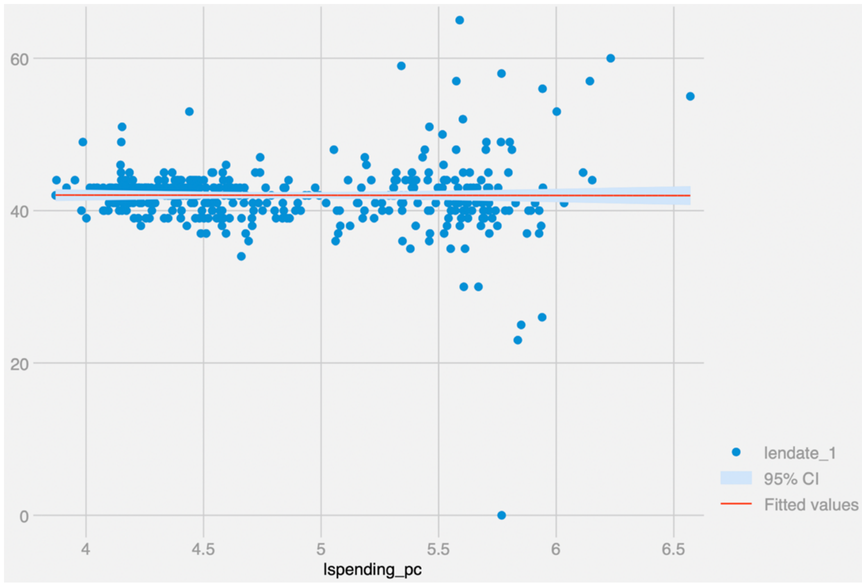
school days in Spring 2020. The variable comes from the administrative records and the unit is ten days.

We include the individual fixed effects μ_i to account for unobserved individual heterogeneity. All standard errors are clustered at the city level. We estimate all models with the cohort-invariant grade dummy $D_{g=12th}$ which equals one if student i is a 12th grader. It captures an average trend in y by grade. The grade dummy materializes the parallel trend assumption, which will be discussed below. The vector Z^c_{ikg} is a set of time-varying individual-level controls.

We state our three key identifying assumptions below.

1. In-person Days as a Random Treatment The key variation for identification is the number of in-person days. It is determined by school principals and the Gyeonggi Office of Education. No individual student could alter the decision by the nature. Still, there could be a reverse causality if the reopening is affected by households or correlated with any component of the (individual level) error term. It is unlikely to be true. First, although the decision is delegated to local authorities, it was still in the hands of policymakers who are familiar with the top-down approach, instead of parents and students. Second, if the reverse causality existed, what would be the main channel, and which schools might be affected by such bottom-up requests? The usual suspect is the student and school quality. Along the same lines, selective schools might or might not want to reopen. They might do because of the need for intensive schooling. They might not do as the students can benefit from private education. In sum, we can check if a school's per capita spending, a proxy of school quality, predicts the length of in-school days. The data exhibits no correlation. (Figure 3) Conditioning on a school characteristic (public v. private) does not change the result.

Figure 3: In-person School Days (vertical) and log spending per capita (horizontal) by school

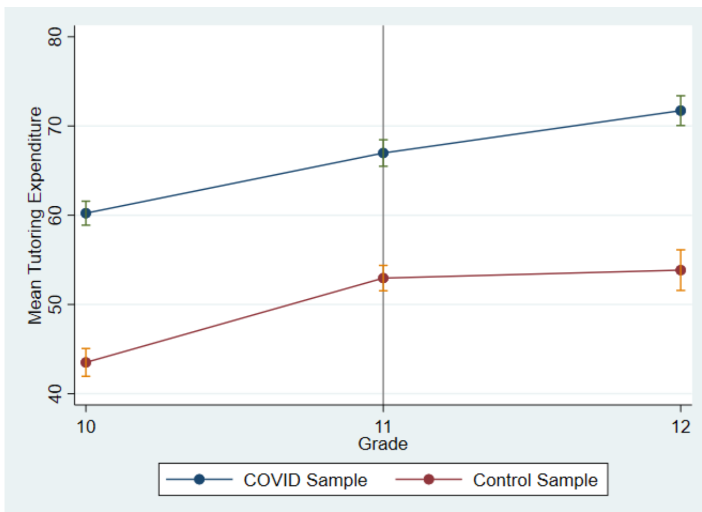


Data: GEPS and administrative records.

There could be a measurement error due to COVID infection. Say, if a student and/or family members were infected, there would have been losses in in-person days at the individual level. The error is nonclassical as it is correlated with the idiosyncratic error term: sick students cannot participate in private education. It is also unlikely. The number of school-age cases was limited: 80 out of 1,537 total cases in Gyeonggi as of July 30th 2020 (0.07% of teenager population). We regressed the in-person days on the number of cumulative cases of each city and found no association.

2. Parallel Trends in Expenditure Growth Fundamentally, 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 expenditure is the same across cohorts. That is, all seemingly time effects are time-invariant grade effects. This parallel trend assumption enables us to use an earlier cohort in the dataset as a control group and attributes all 2020-specific outcome variations to the changes in school days. Figure 4 visualizes the average tutoring expenditure growth of the COVID and control cohorts between 10^{th} and 12^{th} grades.[6]

Figure 4: The Tutoring Expenditure Growth (10^{th} -12^{th} Graders) by Cohort



Data: GEPS and SCHOOLINFO.

3. Individual Fixed Effects We study changes from 2019 to 2020. Given the short timeframe, there are not many relevant individual controls which vary over time. We assume that the individual fixed effects absorb all unobserved heterogeneities, occasionally controlling for household incomes. We present a set of subgroup analyses.

On top of the three identifying assumptions, we leverage the following empirical facts to argue that South Korea is one of the best "laboratories" to

[6]It is unclear that the canonical pre-trend plot can support our version of parallel trend assumption. It could have been better support if there were more data points. Unfortunately, the control cohort data is available from the 10th grade.

tease out the school closure effect from the COVID impacts, and our estimates are less prone to the aggregate confounder.

4. Massive Policy Responses and Limited Economic Damage All who travel in the rain get wet. It is hard to disentangle the school closure effect from the aggregate COVID effect. Parental job loss and school closure are intertwined in the other developed economies, for example. Thanks to the successful "flattening the curve" without economywide lockdowns, the country's macroeconomic response to the COVID shock was moderate and the labor market disruption was minimal. While avoiding lockdowns, the policymakers combined social distancing, massive testing, financial stimulus, and other policies to keep the economy on track. The first universal school closure in history was part of the overall policy package aimed at fighting against COVID-19.

As a result, the monthly trend of the employment rate was not different from usual, though the level was slightly lower, whereas the U.S labor market response was wild. (Figure 5) This empirical fact suggests that the changes in parental investment are likely to be driven by school closure, being less contaminated by indirect effects fromz'z the aggregate shock.

5. "Education Fever" as a Full Compliance Condition Consider our environment as a language of the local average treatment effect framework. The treatment is the school closure. The COVID sample is the treatment group and the earlier sample is the control group. Alway-takers are the students who are chronically absent. Never-takers are the ones who come to school regardless of its operating status. Defiers are the ones who come to closed school but never attend to opened school.

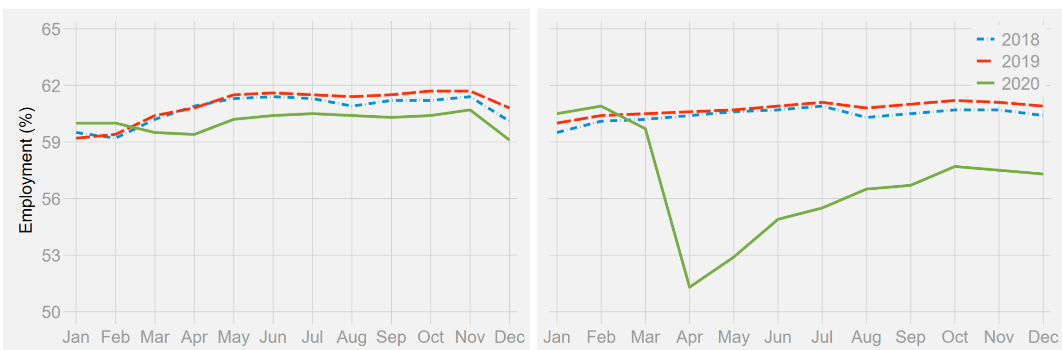
The nature of (centralized) school education eliminates the never-taker and defier groups. The low historical absence rate, which applies to our control group as well, suggests that the fraction of potential always-takers is likely to be infinitesimal. As a result, we have nearly the whole student population as a complier group.

Meanwhile, the private education market was open and thick. The market offers a wide range of services from free online lectures to expensive one-on-one tutoring. Students (and parents) could find many viable options. In other words, they had access to market alternatives. If there were no such market, they would not be able to spend more money, at least in the short run.

Figure 5: Monthly Employments by Country, 2020

(a) South Korea

(b) United States



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

Under the three identifying assumptions and two supporting facts, $\hat{\delta}_1$ captures the causal effect of in-person school days. As the regressor is the total length of in-person school days, $\hat{\delta}_1$ is the change in y due to one unit *increase* in $schdays$.

3. Data

3.1 The Gyeonggi Educational Panel Survey (GEPS)

We use the Gyeonggi Educational Panel Survey (GEPS).[7] The GEPS is a Korean longitudinal survey collected from Gyeonggi Province, the most populated province in the country.[8] The regionally representative survey started in 2012, sampling three cohorts of Gyeonggi students: 4^{th} year elementary-school students, 1^{st} year middle-school students, 1^{st} year high-school students. (4^{th} , 7^{th} , and 10^{th} graders, respectively) They have been followed up through 9 years of primary and secondary education, being interviewed once a year. As this is a critical period regarding adolescent development, the GEPS offers decent opportunity to understand decisions and behavior of the surveyed population.

Each year, information was collected in four separate questionnaires for: teenager, parents, teacher, and school principal. The survey is merged with the administrative school information (the number of teachers, students, classrooms, campus size, etc.) This design covers household information (household composition, parental education, occupation, income, etc.), covering information related to household composition and their parents education, occupation and income.

The GEPS pays special interest in the context of such decisions and collects data on learning behaviors. In particular, it asks detailed information about enrollment in private academies, their type (e.g., group meetings, one-on-one, online), their topic (e.g., Korean, English, math), the hours spent on them and their cost. Importantly, the GEPS offers information about academic achievement of adolescents, measured by a dedicated cognitive tests

[7]The survey is based on Korean Education Longitudinal Study (KELS), which benchmarked the Education Longitudinal Study (ELS) of the United States.

[8]A quarter of the Korean population lives in the province as of 2020. The country has 17 provinces.

(elementary and middle school) and national-level achievement test (high school). It enables us to translate the student decisions to the changes in academic achievement.

Our COVID sample is the elementary student cohort, who became 12th graders in 2020, the first year of the pandemic. Students were initially sampled according to the proportion of the fourth year elementary-school students present in each city. Two supplemental samples were added in the 4^{th} and 7^{th} waves, respectively. Subjects were consistently interviewed for nine waves (six or three for each supplemental sample). The Control sample is the high school student cohort, who was 12th graders in 2014. In sum, the working sample consists of 4,500 youths and their parents or guardians (see descriptive statistics in Table 1).

Private education measures. As with all other personal characteristics collected in the GEPS, private education experience is self-reported by the students and parents. Parents, the payers, report participation and monthly expenditure (Korean won). Given that the GEPS collects its data during late July and the first semester of Korean academic year runs from March to June, one can interpret the information as asking for private education experience during the spring semester.

Test scores. Dedicated achievement tests for earlier periods. Stannine structure for the national tests in the later periods. [to be added]

3.2 SCHOOLINFO: The In-person School Days Data

The SCHOOLINFO is a rich administrative dataset about all primary and secondary educational institutions in South Korea, which is managed by the Ministry of Education. It has the number of instructional days by school

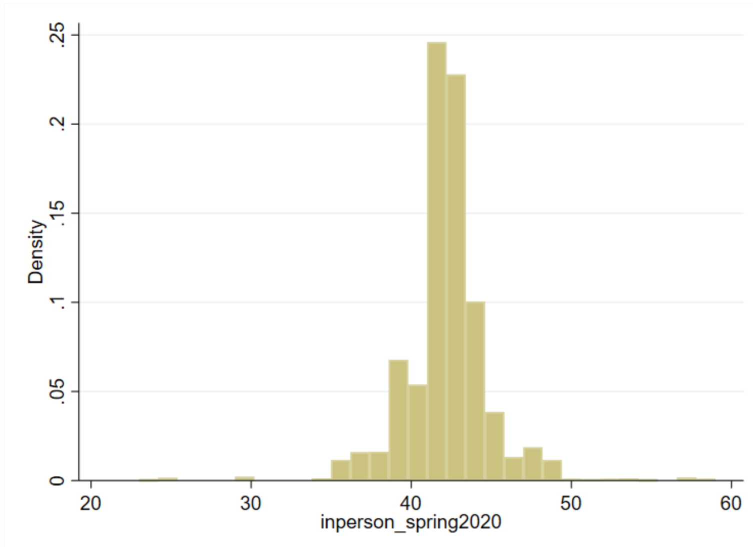
Table 1: GEPS Descriptive Statistics

	COVID cohort (2019-20)	Control cohort (2014-15)
N at 11th grade	2,738	1,763
N at 12th grade	2,740	1,760
% Female	50.51	51.5
% Ordinary school	89.1	89.2
% Selective school	6.9	4.9
Mother's Education		
2-yr college or less	60.5	66.6
4-yr college or more	39.5	33.4
Mean HH income. ($)	5,062.5	4,610.3
	(2,471.3)	(4,948.4)
% Particpation		
Korean	54.7	42.7
Math	81.7	78.0
English	85.8	83.8
$ spent on private educ.		
Korean	156.7	107.9
Math	224.9	192.8
English	268.5	224.0

Notes: Standard deviations in parentheses. All USD values are converted from KRW. (1 USD = 1,200 KRW) "Selective school" is a sum of "autonomous" and "special purpose" schools.

and semester. The schooling mode, the number of in-person instructional days, is included as part of the data after the coronavirus outbreak. The augmented information has not been disclosed to the public. We received the confidential data for Gyeonggi schools and merged it with the GEPS. (Figure 6)

Figure 6: In-person School Days in Spring 2020 (GEPS Schools)



Data: GEPS and SCHOOLINFO.

4. Empirical Results

We estimate Equation (1). Household income is the only time-varying control included. "TWFE" is our preferred specification. "OLS" is a pooled cross-section implementation. "OWFE" exploits the panel structure but does not have the grade dummy.

As a recap, Our $\hat{\delta}_1$ is the marginal effect of one unit *increase* in school days. As our focus is school closure, i.e. decrease in school days, we flip the sign for interpretation.

4.1 Take-up and Expenditure

The private tutoring participation decreased by 1%p. A ten-day decrease of in-person school days increased private education take-up by 2.6%p and expenditure by 9.5%. (Table 2)

Table 2: Impact of In-Person School Days on Tutoring Behaviors

	Take-up			log Expenditure		
	(1) OLS	(2) OWFE	(3) TWFE	(4) OLS	(5) OWFE	(6) TWFE
schdays	-0.012***	-0.007***	-0.026***	-0.109***	-0.033***	-0.095***
	(0.002)	(0.001)	(0.002)	(0.011)	(0.004)	(0.010)
log Income	0.061***	0.016	0.007	0.642***	0.157***	0.128**
	(0.011)	(0.013)	(0.014)	(0.050)	(0.054)	(0.056)
N	4295	8972	8972	4296	8990	8990
R-squared	0.083	0.620	0.643	0.193	0.760	0.766

Notes: OLS estimated using 2020 data only. Clustered Std Errors in parentheses (city level, N=30)

4.2 Subgroup Analyses

We explore socioeconomic variations in the impact of in-person learning days. In these regressions, we estimate Equation (1) by gender, parental education, lagged household income, and lagged academic performance. This allows us to examine the potential heterogeneity in parental responses to the in-person learning days shock. All reported results are TWFE.

4.2.1 Gender

The effect is similar across gender while female students are more responsive by 0.7%p and 1.1%p for take-up and expenditure, respectively. (Table 3) It is consistent with our prior as "son preference" has declined in South Korea during last two decades. (Chung and Gupta 2007, Yoo et al. 2017, Choi and Hwang 2020) We believe this pattern suggests that the estimates are more likely to capture signals rather than noises.

Table 3: Impact of in-person school days on Tutoring Behaviors, by Gender

	Take-up			log Expenditure		
	(1) All	(2) Female	(3) Male	(4) All	(5) Female	(6) Male
schdays	-0.026***	-0.029***	-0.022***	-0.095***	-0.100***	-0.089***
	(0.002)	(0.003)	(0.002)	(0.010)	(0.014)	(0.011)
log Income	0.007	-0.005	0.017	0.128**	0.071	0.175***
	(0.014)	(0.018)	(0.015)	(0.056)	(0.081)	(0.062)
N	8972	4564	4408	8990	4574	4416
R-squared	0.643	0.655	0.629	0.766	0.786	0.742

Notes: Clustered Std Errors in parentheses (city level, N=30)

4.2.2 Household Income and Parents' Education

In Table 4, we estimate Equation (1) by one-year lagged income terciles. The take-up and expenditure growth are higher for students from lower-income families. We interpret it as a result of the base effect. Before the pandemic, low-income students were less likely to participate in private education or to spend much money on it. Therefore they have a higher margin of adjustment when the pandemic hit. Given the positive correlation between income and education, the same logic applies to the estimates by parental education. (Tables 5-6) All results align with our prior.

4.2.3 Academic Performance

Lastly, we estimate the model by one-year lagged academic performance (test score terciles) and find that higher-performance groups better took advantage of the school closure. (Table 7) The take-up and expenditure growth are ordered by the lagged academic performance.

Table 4: Impact of in-person school days on expenditure, by previous year Income

	Take-up				log Expenditure			
	(1) All	(2) Low	(3) Mid	(4) High	(5) All	(6) Low	(7) Mid	(8) High
schdays	-0.026***	-0.037***	-0.021***	-0.015***	-0.095***	-0.119***	-0.088***	-0.073***
	(0.002)	(0.003)	(0.003)	(0.003)	(0.010)	(0.017)	(0.011)	(0.015)
log Income	0.008				0.132**			
	(0.013)				(0.054)			
N	8994	3242	2990	2762	9012	3256	2990	2766
R-squared	0.643	0.661	0.633	0.592	0.767	0.763	0.736	0.734

Notes: Clustered Std Errors in parentheses (city level, N=30)

Table 5: Impact of in-person school days on expenditure, by Mother's Educ

	Take-up			log Expenditure		
	(1) All	(2) 2yrCollorLess	(3) 4yrCollorMore	(4) All	(5) 2yrCollorLess	(6) 4yrCollorMore
schdays	-0.026***	-0.027***	-0.021***	-0.095***	-0.099***	-0.076***
	(0.002)	(0.002)	(0.003)	(0.010)	(0.011)	(0.014)
log Income	0.007	0.003	0.025	0.128**	0.124*	0.210*
	(0.014)	(0.020)	(0.024)	(0.056)	(0.072)	(0.108)
N	8972	5338	3124	8990	5350	3126
R-squared	0.643	0.666	0.588	0.766	0.765	0.764

Notes: Clustered Std Errors in parentheses (city level, N=30)

5. Theoretical Framework

5.1 From Quasi-Experimental Evidence to a Structural Model

We utilize the structural model allowing for both parental investment and the number of school days affecting the academic achievement. The first purpose of employing a structural model is to evaluate the counterfactual policies that

Table 6: Impact of in-person school days on expenditure, by Fathers' Educ

	Take-up			log Expenditure		
	(1) All	(2) 2yrCollorLess	(3) 4yrCollorMore	(4) All	(5) 2yrCollorLess	(6) 4yrCollorMore
schdays	-0.026***	-0.030***	-0.021***	-0.095***	-0.112***	-0.074***
	(0.002)	(0.003)	(0.002)	(0.010)	(0.015)	(0.011)
log Income	0.007	0.001	0.018	0.128**	0.079	0.208***
	(0.014)	(0.020)	(0.015)	(0.056)	(0.064)	(0.065)
N	8972	3950	4370	8990	3964	4374
R-squared	0.643	0.681	0.570	0.766	0.771	0.752

Notes: Clustered Std Errors in parentheses (city level, N=30)

Table 7: Impact of in-person school days on expenditure, by lagged Performance

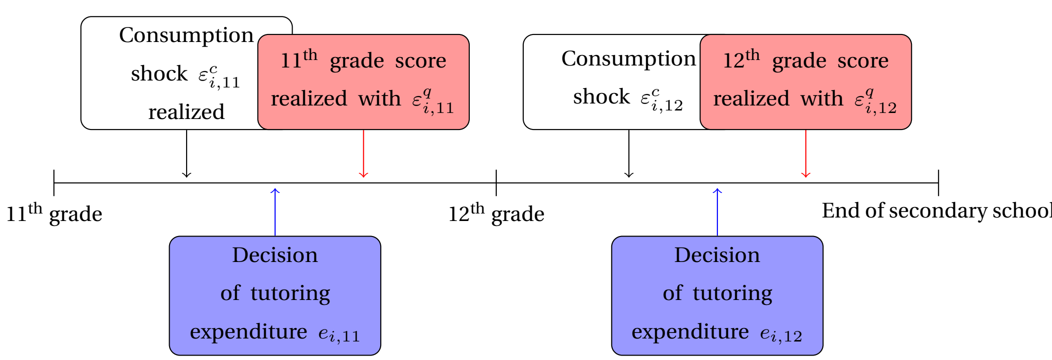
	Take-up				log Expenditure			
	(1) All	(2) Low	(3) Mid	(4) High	(5) All	(6) Low	(7) Mid	(8) High
schdays	-0.026***	-0.014***	-0.027***	-0.033**	-0.095***	-0.065***	-0.120***	-0.194***
	(0.002)	(0.003)	(0.002)	(0.004)	(0.010)	(0.012)	(0.015)	(0.042)
log Income	0.008	0.153	0.034	0.015	0.132**	0.139	0.049	0.409***
	(0.013)	(0.092)	(0.099)	(0.083)	(0.054)	(0.087)	(0.093)	(0.106)
N	8994	5542	2276	694	9012	5542	2276	694
R-squared	0.643	0.822	0.788	0.818	0.766	0.737	0.741	0.755

Notes: Clustered Std Errors in parentheses (city level, N=30)

can help compensate the negative impact caused by the pandemic-induced school closure. Having a model enables to conduct ex-ante policy evaluation which accounts for the behavior of the targeted households. Second, bringing an structural model makes it explicit what underlying assumptions are needed in identifying the relevant parameters. Third, the estimated structural model is used to quantify the impact of school closures on the educational loss and the inequalities. We employ the tournament model of parental investment (Kang 2022), and adjust the model to incorporate the notion of school closure. The choice of model is motivated by the features of secondary school education in South Korea where students do their best to get into the best college tier they can.[9] As seats for prestigious colleges are finite, the motivation of relative performance governs the decision of the household.[10] An alternative model would assume that household care about the absolute performance such as the level of human capital, rather than their relative peroramnce. We find that such a model overpredicts the parental investment of top-performing students and underpredicts the parental investment of low-performing students compared to the tournament model.

5.2 Environment

Figure 7: Model Timeline



Consider an economy composed of N households. We assume household makes a unitary decision. Figure 7 presents the timeline of the model. The model begins when the student becomes 11^{th} grade. The initial conditions are the complete income stream and the test score in 10^{th} grade. We assume

[9] See Section 2 of Kang (2022) for the relevant institutional features of secondary school education in South Korea.

[10] See Section 3 of Kang (2022) for the evidence of household using parental investment as a means to college admission competition.

households are myopic: they receive utility from the current return to academic outcome R_g, which is a function of the test score $q_{i,g+1}$. The first period starts with the realization of the consumption shock ε_{ig}^c. Each household makes a decision of private tutoring investment based on the realized consumption shock and the expectation over the test score shock ε_{it}^q. After the decision is made, test score $q_{i,g+1}$ is produced with the realization of the time-specific test score shock ε_{ig}^q. The generated test score $q_{i,12}$ is used as the input of producing the subsequent tests score in the 12$^{\text{th}}$ grade. At the end of 12$^{\text{th}}$ grade period, students compete against each other to get into better college tier based on the final test score. The return function in the final period is formulated based on the tournament structure, which is explained in Section 5.3. As a result of the college admission competition, students are assigned to one of the college tiers based on their ranking of the final test score.

5.3 Household

In each period the household gets utility from consumption and the test score, which is produced in period g. In particular, denoting g as school grade of the student of household i, the utility function of household i is specified as

$$\varepsilon_{ig}^c \ln(c_{ig}) + \ R_g(q_{i,g+1}), \tag{2}$$

where c_{ig} is household consumption, R_g is a return from test score $q_{i,g+1}$, and ε_{ig}^c is a shock to the marginal utility of consumption. Each household decides how much it spends on private tutoring expenditure e_{ig} to maximize (2) subject to the budget constraint

$$c_{ig} + e_{ig} \leq w_{ig}, \tag{3}$$

where w_{ig} is household income. For the final period when $^{\text{th}}$ grade, the test score tournament is realized through the return function $R_g(q_{i,g+1})$, which is specified as

$$R(q_{i,g+1}) = \alpha_g \sum_{j=1}^{J} \ln(v_j) \times Prob(\ln \tilde{Q}_{j-1} \geq \ln q_{i,g+1} \geq \ln \tilde{Q}_j \Big| \Gamma_{ig}) \Big\}, \tag{4}$$

where $\tilde{Q}_j$ is the cutoff for college tier j. The cutoff $\tilde{Q}_j$ is where the competition between household occurs. Denoting n_j as the seats for college tier j, the cutoff $\tilde{Q}_j$ is the test score of N_j^{th} student, where $N_j = \sum_1^j n_j$. That is, $\tilde{Q}_j$ is the lowest test score of students who made it for college tier j. Note that we assume the return function for both 11$^{\text{th}}$ and 12$^{\text{th}}$ graders. The differential of college structure is important in capturing the investment behavior of the households. We find that an alternative model, where a household cares about the *absolute* performance of the student, fails to capture the parental investment of the household.

5.4 Test Score Production Function

The channel in which compensation for school closure using parental investment occurs is the test score production function. We assume test score $q_{i,g+1}$ is a function of previous test score q_{ig}, the number of school days $schdays_{kg}$ for school k, private tutoring expenditure e_{ig}. The test score production function is specified as

$$q_{i,g+1} = f(\theta_g, q_{ig}, e_{ig}, schdays_{kg}, \varepsilon_{ig}^q),$$

where θ_g is set of the relavant parameters that may change over school grade. To identify the subsitution between parental investment and the number of school days, we adopt a constant elasticity of subsitution (CES) production function. (Cunha et al. 2010) In particular,

$$q_{i,g+1} = A_g(\delta_{qg} q_{ig}^{\phi} + \delta_{sg} schdays_{ig}^{\phi} + \delta_{eg}(1+e_{ig})^{\phi})^{\frac{1}{\phi}} \varepsilon_{ig}^{q},$$

where A_g is the total factor productivity, ϕ_g is the substitution parameter, δ_{qg} is the marignal effects of the previous test score, δ_{sg} is the marginal effects of the number of school days, δ_{eg} is the marginal effects of private tutoring expenditure, and ε_{ig}^{q} is the time-varying shock of test score.

6. Structural Estimation

6.1 Likelihood Function

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 form of probability density function (PDF) for all individuals, as it is a continuous variable. For private tutoring expenditure, the likelihood contribution differs depending upon the whether the household participates in private tutoring activities or not. For the participants, likelihood contribution is form of PDF as tutoring expenditure is a continuous variable. On the other hand, the non-participants' likelihood contribution is a cumulative distribution function. Denoting Θ_g^c as the set of parameters, each household i's likelihood contribution is

$$\mathcal{L}_i(\Theta_g^c | \{q_{ig}\}_{g=11}^{12}, \{w_{ig}\}_{g=11}^{12}) = \Pi_{g=11}^{12} \int_{\varepsilon_{ig}^{q}} \mathcal{L}_{ig}(\theta | q_{ig}, w_{ig}) f(\varepsilon_{ig}^{q}) d\varepsilon_{ig}^{q}, \tag{5}$$

and

$$\mathcal{L}_{ig}(\Theta_g^c|q_{ig},w_{ig}) = \left[f_{e_{ig}}(e_{ig})\cdot f_{q_{i,g+1}}(q_{i,g+1}|e_{ig})\right]^{d_{ig}^e} \times \left[\Pr(e_{ig}=0)\cdot f_{q_{i,g+1}}(q_{i,g+1}|e_{ig})\right]^{(1-d_{ig}^e)};$$

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

$$\mathcal{L}(\Theta_g^c) = \Pi_{i=1}^N \mathcal{L}_i(\theta_g^c|\{q_{ig}\}_{g=11}^{12},\{w_{ig}\}_{g=11}^{12}).$$

The likelihood contribution uses the equilibrium conditions of the theoretical framework. I denote $\tilde{\varepsilon}_{it}^c$ and $\tilde{\varepsilon}_{it}^q$ 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 $\eta_{it}^z = \ln\varepsilon_{it}^z$ for $z = c, q$. In particular,

$$\mathcal{L}_{ig}(\Theta_g^c|q_{ig},w_{ig}) = \left[f_{\eta_{ig}^c}(\tilde{\eta}_{ig}^c)\cdot f_{\eta_{ig}^q}(\tilde{\eta}_{ig}^q|\eta_{ig}^c)|J_i^1|\right]^{d_{ig}} \times \left[\int_{\tilde{\eta}_{ig}^c}^{\infty} f_{\eta_{ig}^c}(\eta_{ig}^c)\cdot f_{\eta_{ig}^q}(\tilde{\eta}_{ig}^q|\eta_{ig}^c)d\eta_{ig}^c|J_i^2|\right]^{(1-d_{ig}^e)}$$

where $|J_i^j|_{j=1,2}$ 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 $\tilde{\eta}_{ig}^c$, which is the minimum amount of the shock that makes household stop spending on the tutoring expenditure. Note that θ_g^c depends on the student's grade g and the cohort the student belongs to. In the subsequent section, I explain the implications of the differentiation of the parameters.

6.2 Parameterization for Identification

We set the substitution parameter ϕ and the standard deviation of the unobserved heterogeneity parameters, σ_c and σ_q to be common across cohorts and grades. To identify the effects of school days, we allow the parameter governing the effects of number of schoool days δ_{sg} change by grade g, but assume that the relative size of the effects of school days to the effects of parental investment does not change over time. (i.e. $\delta_{sg}/\delta_{eg} = \kappa$, where κ is a constant.) As the only source of significant variation in school days come from 2020, it is impossible to identify δ_{eg} for different gs. Alternatively, we identify the relative effects of school days to parental investment, namely κ. The differentiation of a subset of parameters are necessary to implement the parallel trend assumption. Note that our parallel trend assumption is the growth of parental investment in 12$^{\text{th}}$ grade compared to 11$^{\text{th}}$ grade. Thus, the average cohort difference between the treatment group and control group should be captured in the structural estimation. To capture the mean difference, we differentiate the altruism parameter α_g by cohorts. The parameter governing the effects of tutoring expenditure, δ_{eg} is allowed to change for in 2020, the treatedment period. In addition, as the tests are different by cohort and grade, we differentiate the constant A_g by cohorts.

6.3 Implementation of the Parallel Trend Assumption

We propose an estimation algorithm that tightly links the natural experiment estimates with the structural model. The key assumption underlying the TWFE estimator is the parallel trend assumption. Note that the timing of the treatment (Covid-19 school closure) happens when the students are in 12th grade. The trearment is realized as the increasing amount of tutoring expenditure compared to 11th grade. Therefore, the parallel trend assumption implies the treated and the control groups would have had the same change of tutoring expenditure, if there was no school closure. denoting $\tilde{e}^{A}_{i,g}$ as the counterfactual private tutoring expenditure of household i who belongs to group A if schools were fully open and $\widehat{e}^{A}_{i,group}$ the predicted private tutoring expenditure with the actual number of school days of household i who belongs to group A the parallel trend assumption implies

$$||\frac{1}{N_{control}}\sum_{i=1}^{N^{control}}(\widehat{e}^{control}_{i,12}-\widehat{e}^{control}_{i11})-\frac{1}{N_{Treated}}\sum_{i=1}^{N^{control}}(\tilde{e}^{treated}_{i,12}-\tilde{e}^{treated}_{i,11})||=0,$$

which plays as a constraint of the estimation routine.

Finally, the objective function is a likelihood function with the counterfactual constraint. The set of structural parameters $\widehat{\theta}$ is estimated by

$$\widehat{\theta}=\arg\max\left\{\log\mathcal{L}+\lambda||\frac{1}{N_{control}}\sum_{i=1}^{N^{control}}(\widehat{e}^{control}_{i,12}-\widehat{e}^{control}_{i11})-\frac{1}{N_{Treated}}\sum_{i=1}^{N^{control}}(\tilde{e}^{treated}_{i,12}-\tilde{e}^{treated}_{i,11})||\right\},$$

where $\lambda<0$ and

$$\mathcal{L}=\prod_{i=1}^{N}\mathcal{L}_{ig}(\theta|q_{ig},w_{ig}).$$

6.4 Estimation Results

Table 8 presents the structural estimates. The estimates of the altruism parameter are similar across the cohorts and the grades. The overall difference is within the range of 0.1. The relative size of the effects of the number of school days to the effects of tutoring expenditure is 1.722. The estimated substitution parameter ϕ suggests that tutoring expenditure and school days are gross complements. In Table 9, the effects of tutoring expenditure slightly decreases in the 12$^{\text{th}}$ grade. This pattern is consistent with the early studies in the literature of parental investment. (Del Boca et al. 2019)

Table 8: Structural Estimation: Shock and Preference Parameters

A. Shock parameters			
σ_q	Test score shock std	0.463***	
		(0.008)	
σ_c	Consumpt shock std	0.725***	
		(0.006)	
κ	relative share of schdays	1.722***	
		(0.013)	
B. Preference Parameters		11^{th}	12^{th}
$\alpha^{treated}$	altruism	0.596***	0.713***
		(0.014)	(0.019)
$\alpha^{control}$	altruism	0.609***	0.657***
		(0.015)	(0.012)

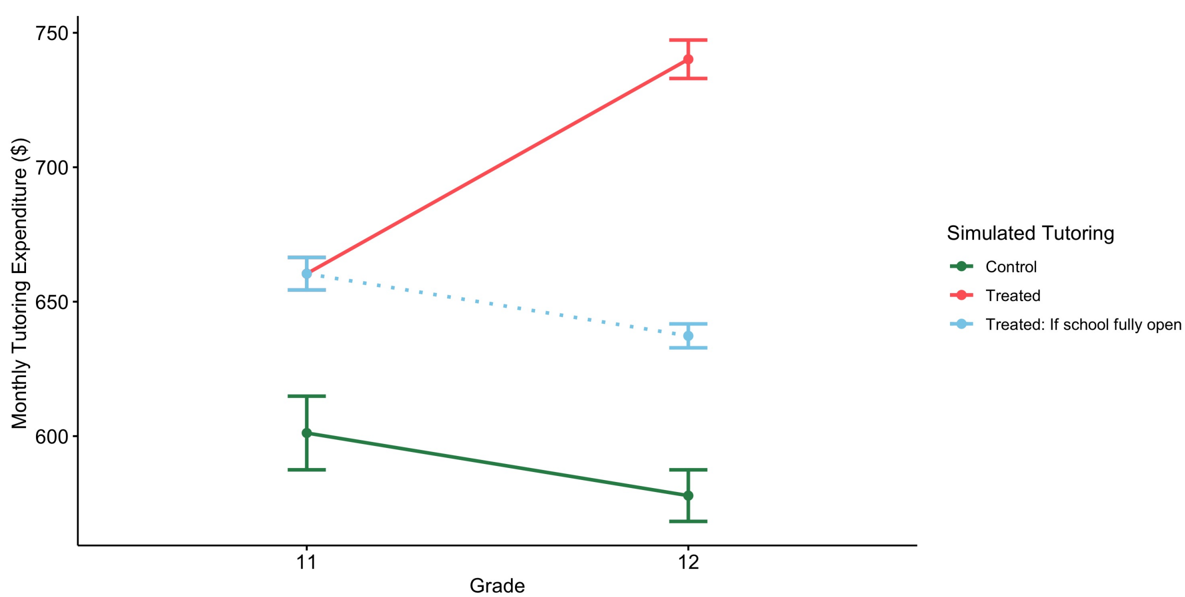
Based on the estimates, we can compute the price of 10 days school closure in terms of private tutoring expenditure. In terms of marginal contribution to the test score percentiles, 10 days of school closure is equivalent to the $172.2 in private tutoring expenditure on average. However, this is mechanical estimates using the CES production function independent of the household behavior. In the subsequent section, we introduce the simulation using structural model, which takes into account for the household behaviors.

The simulated results suggest that the parallel trend assumption is well implemented. Figure 8 presents the simulated tutoring expenditure with the parallel trend assumption. The vertical difference between solid line and dotted line for treated group is the effects of school closure on parental investment.

Table 9: Structural Estimation: Production Function

		Grades	
		11^{th}	12^{th}
Production Function			
ϕ	CES Substitution	-0.238***	
		(0.02)	
δ_e	(ptexp)	0.287***	0.272***
		(0.01)	(0.01)
δ_s	Schdays*	0.494***	0.468***
		-	-
δ_e^{covid}		-	0.301***
		-	(0.01)
δ_s^{covid}		-	0.519***
		-	-

Figure 8: Parallel Trend Assumption in the Structural Estimation



7. Counterfactual Experiment

7.1 School Open Counterfactual

To quantify the consequences of the Covid-19 pandemic educational outcomes, we simulate the structural model by imposing number of school days of pre-pandemic. Specifically, we investigate how increasing the school days to 60 and 95 (the pre-pandemic legal school days for one semester) changes the educational outcomes. Table 10 presents the result of school open counterfactual. Compared to the estimated model, Full Open simulation increases the test score by 0.61 SD, and the standard deviation decreases by 23.1%.

Figure 9: Experiment: School Open Simulation

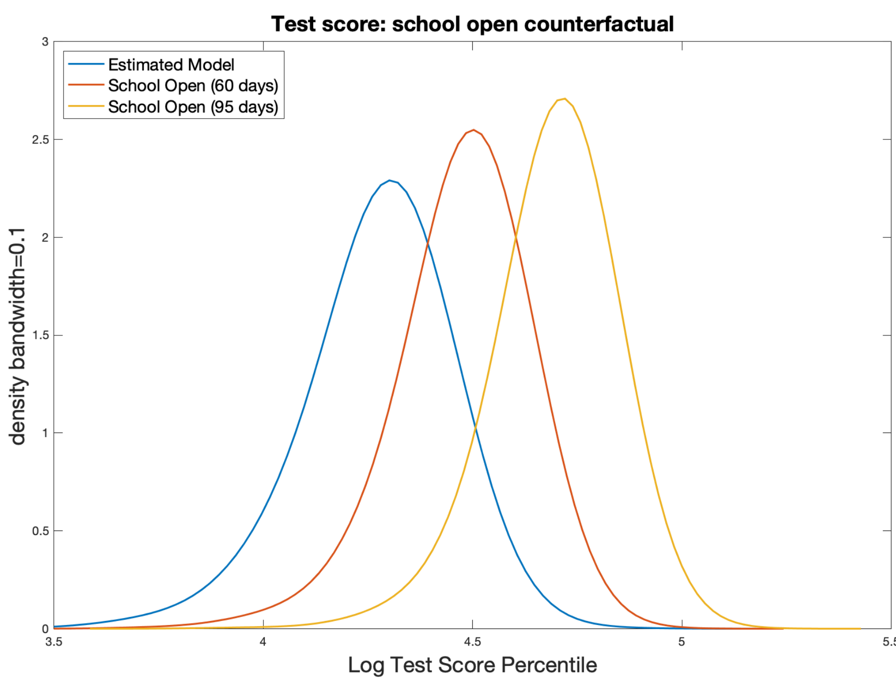
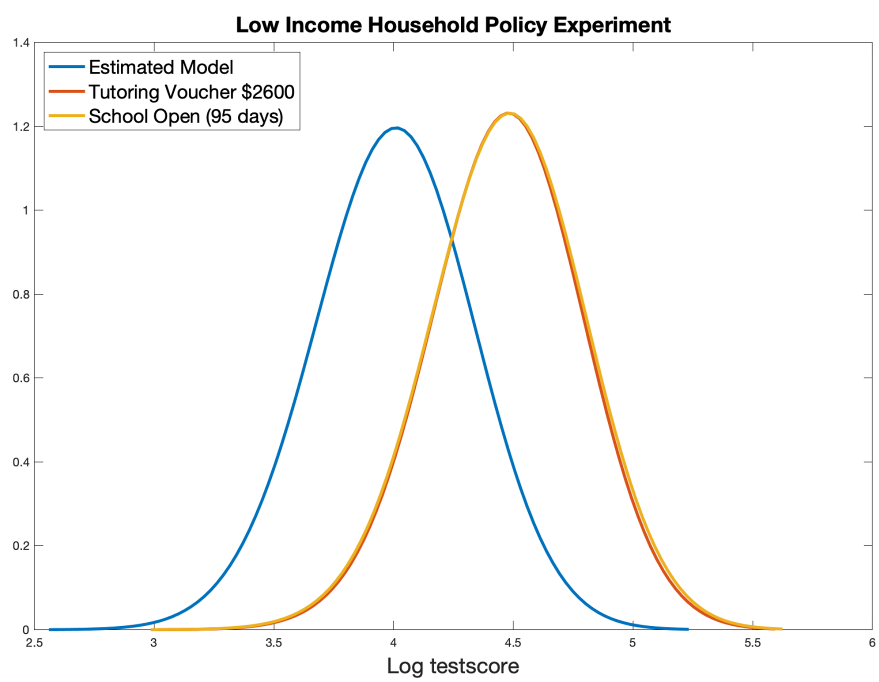


Table 10: Experiment: School Open Counterfactual

	Mean	Std. Deviation
Estimated Model	4.2737	0.1625
Partial Open (60 days)	4.4758	0.1398
Full Open (95 days)	4.6946	0.1249

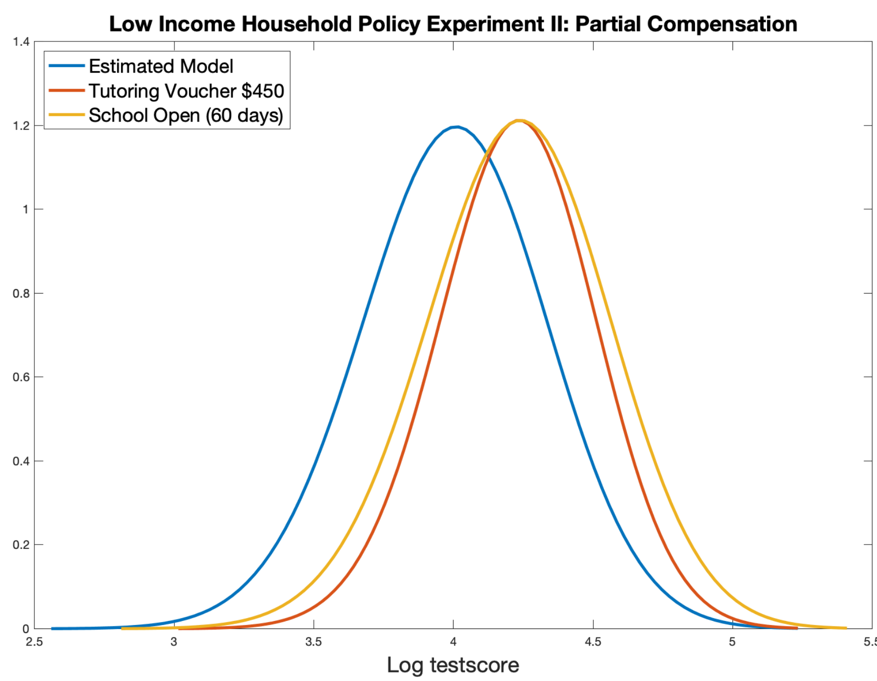
7.2 Private Tutoring Voucher Simulation

Figure 10: Compensation of schooling loss for low income household



The result of voucher simulation suggests that tutoring voucher of about $2600 may fully compensate the average loss of the test scores for low income households. As private tutoring expenditure is measured monthly, the amount of expenditure needed for compensating the school closure is substantial. Although it is fairly expensive to fully compensate the average test score loss from the reduced number of school days, tutoring expenditure could be an effective way to partially mitigate the test score loss. Since there exists diminishing marginal effects from private tutoring expenditure, the marginal returns from tutoring is greater in the low amount of tutoring expenditure. We find that the tutoring voucher of $450 can compensate the test score loss about 15 days. Figure 11 shows the comparison of the 60-days school open simulation and the $450 tutoring voucher simulation. The average test score of the two simulation is almost identical, which suggests that tutoring voucher of $450 has comparable effects of increasing the number of school days to 60 days.

Figure 11: Compensation of schooling loss for low income household



8. Conclusion

This paper examines the effect of the COVID-induced school closure on parental investment and the educational outcomes of students. We identified the causal estimates using panel regression with the novel dataset. Motivated by the empirical finding, we build and estimate a simple tournament model. The model lets us conduct consistent policy experiments.

Our findings show that the short in-person instructional days increased private education participation and expenditure. The responses were heterogeneous by income, parental education, and academic performance. It implies that the school closure would have widened educational inequality.

All in all, our results elucidate the importance of public education and compulsory schooling. The cost of school closure is huge and distributed unevenly. We hope it sheds light on the policy debate of opening schools. It is worth mentioning that our results do not deliver a clear-cut prescription alone as our framework abstracted away from the tradeoff between in-classroom transmission and learning loss. Still, it could be an informative building block for a full cost-benefit analysis of school closure.

References

Agostinelli, Francesco and Matthew Wiswall, "Estimating the Technology of Children's Skill Formation," July 2016.

_ , 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.

Aucejo, Esteban M and Teresa Foy Romano, "Assessing the effect of school days and absences on test score performance," *Economics of education review*, December 2016, *55*, 70–87.

Boca, Daniela Del, Christopher J Flinn, Ewout Verriest, and Matthew J Wiswall, "Actors in the Child Development Process," February 2019.

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 Heckman, and Susanne Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica: journal of the Econometric Society*, May 2010, *78* (3), 883–931.

Dobkin, Carlos, Ricard Gil, and Justin Marion, "Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment," *Economics of education review*, August 2010, *29* (4), 566–575.

Fuchs-Schündeln, Nicola, Dirk Krueger, Alexander Ludwig, and Irina Popova, "The Long-Term Distributional and Welfare Effects of Covid-19 School Closures," September 2020.

Galiani, Sebastian and Juan Pantano, "Structural Models: Inception and Frontier," April 2021.

Goodman, Joshua, "Flaking Out: Student Absences and Snow Days as Disruptions of Instructional Time," June 2014.

Gottfried, Michael A, "Excused Versus Unexcused: How Student Absences in Elementary School Affect Academic Achievement," *Educational evaluation and policy analysis*, December 2009, *31* (4), 392–415.

__ , "Chronic absenteeism and its effects on students' academic and socioemotional outcomes," *Journal of education for students placed at risk*, April 2014, *19* (2), 53–75.

Grau, Nicolás, "The impact of college admissions policies on the academic effort of high school students," *Economics of education review*, August 2018, *65*, 58–92.

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.

Kang, Hyunjae, "Parental investment, child's efforts, and intergenerational mobility," 2022.

Lazear, Edward P and Sherwin Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," *The journal of political economy*, 1981, *89* (5), 841–864.

Liu, Jing, Monica Lee, and Seth Gershenson, "The short- and long-run impacts of secondary school absences," *Journal of public economics*, July 2021, *199*, 104441.

Tincani, Michela M, Fabian Kosse, and Enrico Miglino, "Subjective Beliefs and Inclusion Policies: Evidence," 2021.

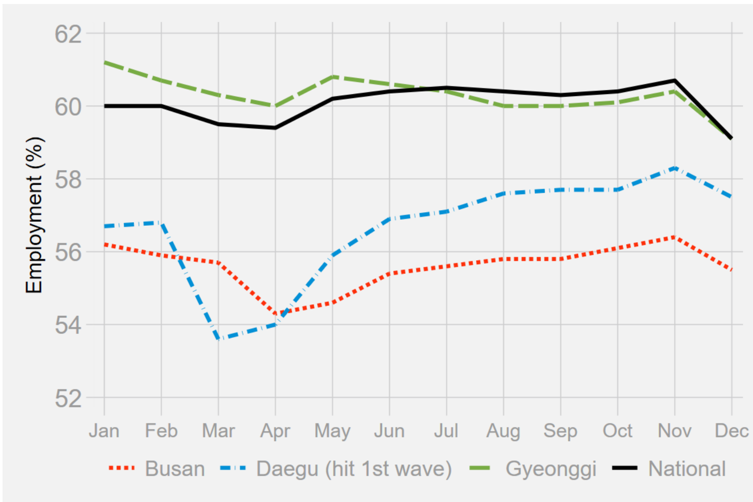
Todd, Petra E and Kenneth I Wolpin, "The Best of Both Worlds: Combining RCTs with Structural Modeling," *Journal of economic literature*, forthcoming.

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. More Discussion on Background

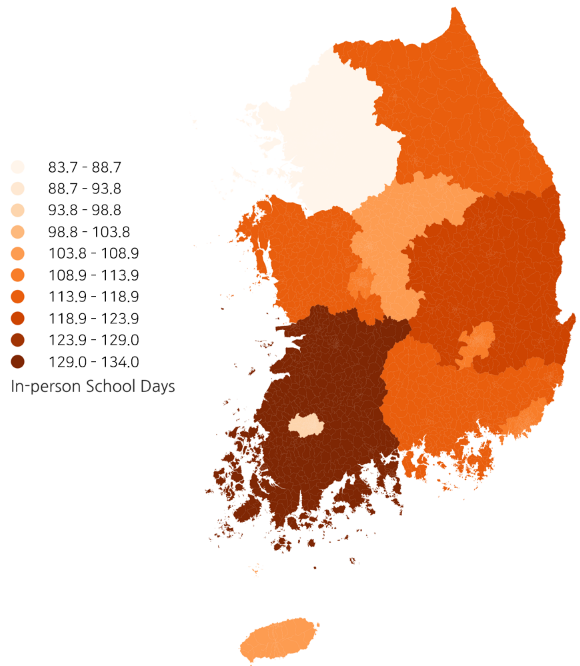
TO BE ADDED

Figure A.1: Monthly Employment by Province, 2020



Data: EAPS.

Figure A.2: In-person School Days in South Korea (Province Level)



Data: SCHOOLINFO and administrative records.

B. Comprehensive Description on All Data Sources

TO BE ADDED

C. 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,

$$\ln y_{it} = \sum_{j=1}^{J} (\beta_j + \delta_j \cdot age_{it}) D_{i,j}^{Tier} + Z_{\mathrm{i}}\gamma + \varepsilon_{it}, \tag{6}$$

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 (2022) for the estimates of the Pooled-OLS estimates, which is used for the prediction. The resulted vector $\underset{5\times 1}{\mathcal{V}}$ is

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

D. Likelihood Function Details

D.1 Tournament model

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

$$-\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}}+\alpha_g\frac{\partial}{\partial e_{ig}}\sum_{j=1}^{J}\ln(v_j)\times Prob(\ln\widetilde{Q}_{j-1}\geq\ln q_{i,g+1}\geq\ln\widetilde{Q}_j\Big|\Omega_{iT},e_{ig})=0. \tag{7}$$

Equation 7 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

$$\begin{aligned}
g &= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}}+\alpha_g\left(\frac{\partial}{\partial e_{ig}}\sum_{j=1}^{J}\ln(v_j)\times Prob(\ln\widetilde{Q}_{j-1}\geq\ln q_{i,g+1}\geq\ln\widetilde{Q}_j)\right)\\
&= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}}+\alpha_g\frac{\partial}{\partial e_{ig}}\left\{\sum_{j=1}^{J}\ln(v_j)\times\left(\Phi(\frac{\ln\widetilde{Q}_{j-1}-\widehat{\ln q_{i,g+1}}}{\sigma_q})-\Phi(\frac{\ln\widetilde{Q}_{j-1}-\widehat{\ln q_{i,g+1}}}{\sigma_q})\right)\right\}\\
&= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}}+\alpha_g\left[\sum_{j=1}^{J}\ln(v_j)\times\left(\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_{j-1}-\widehat{\ln q_{i,g+1}}}{\sigma_q})-\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_{j}-\widehat{\ln q_{i,g+1}}}{\sigma_q})\right)(-\frac{\nu_g\delta_{eg}(1+e}{B}\right.\\
&= -\frac{\varepsilon^c_{ig}}{w_{ig}-e_{ig}}+\alpha_g\left[\ln(v_1)\frac{1}{\sigma_q}\phi(\frac{\ln\widetilde{Q}_0-\widehat{\ln q_{ig}}}{\sigma_q})-\frac{1}{\sigma_q}\sum_{j=1}^{J-1}\left(\ln(v_j)-\ln(v_{j+1})\right)\phi(\frac{\ln\widetilde{Q}_j-\widehat{\ln q_{ig+1}}}{\sigma_q})-\frac{1}{\sigma}\right.\\
&\quad\left.\phi(\frac{\ln\widetilde{Q}_J-\widehat{\ln q_{ig+1}}}{\sigma_q})\right]\times(-\frac{\delta_3(1+e_{ig})^{\phi_g-1}}{B}),
\end{aligned}$$

where $B=\delta_q q^{\phi}_{ig}+\delta_s Schdays^{\phi}_{ig}+\delta_e(1+e_{ig})^{\phi}$

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

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

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,

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

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

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

Using $\frac{\partial}{\partial x} \frac{1}{\sigma} \phi(\frac{x}{\sigma}) = -\frac{x}{\sigma^2} (\frac{1}{\sigma} \phi(\frac{x}{\sigma}))$,

$$\frac{\partial \eta^c_{ig}}{\partial e_{ig}} = \frac{1}{\varphi}\frac{\partial \varphi}{\partial e_{ig}} + \frac{\phi - 1}{(1+e_{ig})} - \frac{1}{w_{ig} - e_{ig}} - \frac{1}{\delta_q q^{\phi}_{ig} + \delta_s Schdays^{\phi}_{ig} + \delta_e (1+e_{ig})^{\phi}} \delta_e \phi (1+e_{ig})^{(\phi-1)},$$

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

$$\begin{aligned}
\frac{\partial \varphi}{\partial e_{ig}} =& \sum_{j=1}^{J-1} \Bigg(\ln(v_j) - \ln(v_{j+1}) \Bigg) \{ - \frac{\ln \widetilde{Q}_j - \ln \widehat{q_{i,g+1}(e_{ig})}}{\sigma^2_q} \} \frac{1}{\sigma_q} \phi(\frac{\ln \widetilde{Q}_j - \ln \widehat{q_{i,g+1}(e_{ig})}}{\sigma_q})(\frac{-\partial \ln \widehat{q_{i,g+1}(e_{ig})}}{\partial e_{ig}}) \\
=& \sum_{j=1}^{J-1} \Bigg(\ln(v_j) - \ln(v_{j+1}) \Bigg) \{ \frac{\ln \widetilde{Q}_j - \ln \widehat{q_{i,g+1}(e_{ig})}}{\sigma^2_q} \} \frac{1}{\sigma_q} \phi(\frac{\ln \widetilde{Q}_j - \ln \widehat{q_{i,g+1}(e_{ig})}}{\sigma_q})(\frac{\partial \ln \widehat{q_{i,g+1}(e_{ig})}}{\partial e_{ig}}) \\
=& \sum_{j=1}^{J-1} \Bigg\{ \Bigg(\ln(v_j) - \ln(v_{j+1}) \Bigg) \{ \frac{\ln \widetilde{Q}_j - \ln \widehat{q_{i,g+1}(e_{ig})}}{\sigma^2_q} \} \frac{1}{\sigma_q} \phi(\frac{\ln \widetilde{Q}_j - \ln \widehat{q_{i,g+1}(e_{it})}}{\sigma_q}) \Bigg\} (\frac{\delta_e \phi (1+e_{ig})^{(\phi-1)}}{B}
\end{aligned}$$

For the non-paritcipants, the Jacobian transformation term is

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