DISCUSSION PAPER SERIES

IZA DP No. 15751

Do Class Closures Affect Students' Achievements? Heterogeneous Effects of Students' Socioeconomic Backgrounds

Masato Oikawa
Ryuichi Tanaka
Shun-ichiro Bessho
Akira Kawamura
Haruko Noguchi

NOVEMBER 2022

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15751

Do Class Closures Affect Students' Achievements? Heterogeneous Effects of Students' Socioeconomic Backgrounds

Masato Oikawa
Waseda University and WISH

Ryuichi Tanaka
University of Tokyo and IZA

Shun-ichiro Bessho
University of Tokyo

Akira Kawamura
Waseda University

Haruko Noguchi
Waseda University and WISH

NOVEMBER 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Do Class Closures Affect Students' Achievements? Heterogeneous Effects of Students' Socioeconomic Backgrounds

This paper examines how class closures affect the academic achievements of Japanese students in primary and middle schools, with a special focus on the heterogeneous effects of the socioeconomic backgrounds of students' households. Utilizing the administrative data of students from a city in the Tokyo Metropolitan Area, we estimated the effects of class closures due to flu epidemics, on the students' language and math test scores. We find that class closures adversely affect math test scores of economically disadvantaged students. The magnitudes of the negative effects on disadvantaged students are heterogeneous by subject, grade in school, gender, timing of class closures, and students' achievements at the beginning of the school year. Male students from economically disadvantaged households are more susceptible to class closures, and those with relatively low achievements before class closures suffer more seriously from them. The deleterious effects among economically disadvantaged male students are driven not only by reductions in class hours in school, but also by increases in time spent watching TV and playing video games. We also find that school resources can mitigate the negative impact of class closure among economically disadvantaged students. These results indicate the importance of public programs in preventing a negative temporal shock to student learning environments.

JEL Classification: I20, I24

Keywords: class closures, flu epidemic, students' achievements, students' usage of time, instruction time

Corresponding author:
Ryuichi Tanaka
Institute of Social Science
The University of Tokyo
7-3-1 Hongo
Bunkyo-ku
Tokyo 113-0033
Japan

E-mail: ryuichi.tanaka@iss.u-tokyo.ac.jp

1 Introduction

The COVID-19 pandemic has affected lives across the globe, and its adverse effects are more serious for disadvantaged households. Deaths attributed to COVID-19 are concentrated more in groups with low socioeconomic status (e.g., Mena et al., 2021). This is partly because accessibility to healthcare services during a pandemic is limited for those with low incomes (Arnault et al., 2021; Czeisler et al., 2020; Gonzalez et al., 2020; Kranz et al., 2021). Difference in work environments is another reason for this. Since not all jobs are suited for working from home (Boeri et al., 2020; Dingel and Neiman, 2020; Okubo, 2020; Rahman, 2020), and because the opportunities to work from home are limited for female, young, low-educated, and low-paid employees (Bonacini et al., 2021), the adverse effects of the pandemic are more serious for people from lower socioeconomic backgrounds. The pandemic is not a great leveler, but a great divider.

The effects of the pandemic are not limited to adults. Children are affected in various aspects as well. One important issue for children is the disruption of learning opportunities due to the pandemic. During the early wave of the pandemic, strict policies restricting interactions among individuals were implemented, and schools were closed as a result. According to an estimate by the UNESCO, as of March 30, 2020, 167 countries had closed schools nationwide, in the face of COVID-19.[1] Losses in school instruction time, associated with school closures, adversely affect students' access to education.[2] Thus, the adverse effects of reduced instruction time can be serious for children from households with low socioeconomic status. Rich households could afford to spend more on education for their children, to compensate for losses in school instruction time, while poor households could not. In fact, during a lockdown due to the pandemic in the United Kingdom, primary school students reduced their total learning time, but the reduced amount of learning time was smaller among students from a high income bracket "(Andrew, 2020). In the United States, the demand for online learning resources increased after a lockdown, compared to

[1] See https://en.unesco.org/covid19/educationresponse for more details. (accessed on June 22, 2021)

[2] Gaps in student achievement based on students' socioeconomic status become more severe during the summer vacation(Alexander et al., 2007; Downey et al., 2004). In the Japanese context, Kawaguchi (2016) found that the 2002 Japanese education reform, which decreased in-school instruction time, simultaneously increased socioeconomic gaps in test scores of the OECD's Programme for International Student Assessment (PISA). Kubota (2016) also analyzed the effects of the 2002 reform, and found that households with higher socioeconomic backgrounds increased their spending on supplementary after-school education. On the other hand, Motegi and Oikawa (2019) analyzed the 2002 reform from the perspective of school resources, and found that school instruction time becomes more effective when combined with better teachers, and that their positive effects were greater for students from lower socioeconomic backgrounds.

before it, especially in areas with residents from higher socioeconomic status (Bacher-Hicks et al., 2021).[3,4] The negative impact of class closures on students' achievements, especially among students from lower socioeconomic status, has also been observed (Chetty et al., 2020; Engzell et al., 2021; Maldonado et al., 2020). Since it is known that education in early life has long-term effects (e.g., Card and Krueger, 1992; Jaume and Willén, 2019; Kikuchi, 2014; Meghir and Palme, 2005; Meyers and Thomasson, 2021), and that socioeconomic status transmits intergenerationally (Black and Devereux, 2010), the impact of school closures on students' achievements could not only be a temporary shock, but also a long term one.[5]

This paper aims to comprehensively analyze the effects of class closures, due to epidemics, on students' achievements with a special focus on: (1) students who are susceptible to class closures; (2) the possible reasons for the negative impact of class closures, and (3) ways to mitigate this impact. For this purpose, we employ the data of class closures due to seasonal flu outbreaks from 2015 to 2017, taken from administrative data of a large city in the Tokyo metropolitan area in Japan. Although our data do not allow us to evaluate the direct impact of the ongoing COVID-19 pandemic, they enable us to analyze the effects of closed classes on various students' outcomes at least in short-run, including academic outcomes, and students' behaviors and attitudes. In addition, the data enable us to study the relationship between school resources and the adverse effects of class closures. Our study provides useful suggestions on how to allocate school resources among students from varied backgrounds, particularly those susceptible to the negative impact of class closures. An additional advantage of our data on seasonal flu in Japan is that we can estimate the effects of class closures by fixing school-year factors, using variations in class closures by school and year. This kind of analytical strategy is almost impossible in the case of the COVID-19 pandemic, because most countries had closed schools nationwide, and the construction of control groups for students who experienced school closures is extremely difficult.[6] Hence, lessons from past experiences obtained from this study will be beneficial for policy-makers in implementing measures to mitigate students'

[3] Ikeda and Yamaguchi (2021) estimated the effects of school closure on students' study time, due to the current pandemic in Japan, using an online learning platform. They found that the study time on the platform increased from the beginning of the school closure. They focused on the heterogeneity of students' responses, based on their earlier access to the platform, and the schools' quality, rather than based on students' socioeconomic status.

[4] In sociological literature, there are studies on the socioeconomic disparities in students' learning environments during school closures due to the COVID-19 pandemic (e.g., Bol, 2020; Dietrich et al., 2021; Grätz and Lipps, 2021; Reimer et al., 2021).

[5] In the case of the COVID-19 pandemic, Fuchs-Schündeln et al. (2020) quantified the long-term effect of class closures due to the pandemic on children's welfare using a structural life-cycle model. They found that the negative impact of class closures is more severe for students whose parents have lower educational attainment and lesser assets.

[6] Engzell et al. (2021) and Maldonado et al. (2020) utilized the data from before the pandemic as control groups for data gathered during the pandemic. This also makes it possible for the results to capture the effects of factors other than the nationwide school closures, related to the pandemic.

future adverse impact, in situations where the pandemic is still ongoing and data on students during the COVID-19 period is not fully accessible.

We estimated the effects of class closures on students' test scores in the year following class closures, with controls for class characteristics, school fixed effects (FEs), and year FEs. We found a negative impact on students' achievement in mathematics among the economically disadvantaged elementary school students. The negative impact of class closures on disadvantaged students is heterogeneous by subject, their grade in school, gender, the timing of class closures, and students' achievements at the beginning of the school year. The effects range from 0.06 to 0.13 of the standard deviation of test scores. Elementary school boys from economically disadvantaged households are more susceptible to the negative impact of class closures, especially those with relatively low achievements before closure. The negative impact on economically disadvantaged elementary school boys could be caused by both reductions in schools' instruction times and students' behavioral changes, such as an increase in the amount of time spent watching TV and playing video games after school, and decreased sleeping time on weekdays.[7] The estimation results are consistent with studies that found that the negative impact of school closures during the COVID-19 pandemic especially impacted students from lower socioeconomic backgrounds or low achievers (Chetty et al., 2020; Engzell et al., 2021; Grewenig et al., 2010; Maldonado et al., 2020). We also found that school resources could mitigate the negative impact of class closures among economically disadvantaged students. This indicates that public programs such as supplemental remedial education could be a way to prevent a temporary shock to students' learning environment.

This paper is related to the strand of the literature of education economics, with a special focus on the effects of instruction time. Our estimation results are consistent with previous studies, which found positive effects of longer instruction time on student achievement. Previous literatures conduct two types of identification strategies using "scheduled" versus "unscheduled" variations of instruction times. The former studies utilize remedial education programs and policy revisions of education systems (for example, Andrietti, 2015; Aucejo and Romano, 2016; Battistin and Meroni, 2016; Bellei, 2009; Bessho et al., 2019; Carlsson et al., 2015; Cattaneo et al., 2017; Cortes and Goodman, 2014; Fitzpatrick et al., 2011; Huebener et al., 2017; Jensen, 2013; Kawaguchi, 2016; Lavy, 2012, 2015; Motegi and Oikawa, 2019; Pischke, 2007; Rivkin and Schiman, 2015; Taylor, 2014; Wößmann, 2003; Zimmer et al., 2010).[8,9]

[7] The finding of Takaku and Yokoyama (2021) that a nationwide school closure in Japan increased children' weight is indicative of the impact of school closures on the change of students' behavior at home.

[8] There are studies that found insignificant effects of instruction time on student achievement (Dahmann, 2017; Meyer and Van Klaveren, 2013).

[9] Our results are also consistent with studies suggesting that mathematics scores are more sensitive to instruction time than language arts test scores(Battistin and Meroni, 2016; Jensen, 2013; Marcotte, 2007; Zimmer et al., 2010).

On the other hand, there are few studies using an "unscheduled" variation of instruction time, such as class closures due to epidemics of infectious diseases.[10,11] Chetty et al. (2020) compared educational progress among students using an online educational platform using data related to schools' share of students eligible for free and reduced price lunches during the COVID-19 pandemic. The study found that students in lower-income areas had lower numbers of completed lessons compared to students in higher-income areas during the pandemic. Engzell et al. (2021) and Maldonado et al. (2020) utilized data during the COVID-19 pandemic, and found the negative impact of school closures on students' test scores, especially among disadvantaged students. Since the school closures due to the pandemic of COVID-19 were nationwide, however, all of these studies rely on the comparison of before and after the COVID-19 pandemic as their identification strategy. Another previous studies analyzed the effects of school closures using a variation of instruction time reduced by class closures, based on historical pandemics such as the 1918-1919 influenza and the 1916 polio epidemic in the United States (Ager et al., 2020; Meyers and Thomasson, 2021). But these studies focused on factors that were not directly related to the timing of school closures, such as the years of schooling and labor market outcomes. Therefore, the mechanism behind their effects is unclear in these studies.

In contrast to these previous studies, this study has a couple of strengths. First, we estimate the effects of class closures as "unscheduled" event, utilizing both time-series and cross-sectional variations of students' outcome by controlling school and year fixed effects. Second, we consider test scores in the year following class closures as direct outcomes, enabling us to investigate the mechanism behind the causal relationship between class/school closures and students' achievements. This is extremely important in the context of unscheduled changes in instruction time, because the effect of instruction time on students' behaviors may differ substantially in cases of scheduled changes. As discussed, we find that class closures affect students' behaviors, leading to a deterioration in their achievements. Therefore, investigating direct mechanisms is important.

The remainder of this paper is organized as follows: section 2 explains the institutional background. Section 3 describes the econometric model. Section 4 discusses the data and descriptive

[10] Studies have analyzed the effects of students' voluntary absences on students' achievements, using pandemics as instrumental variables, to control for the endogeneity of absences rather than estimating the effects of school closures(Aucejo and Romano, 2016; Goulas and Megalokonomou, 2016). Aucejo and Romano (2016) found a statistically significant negative impact of students' absence, induced by flu epidemics, on their reading scores but not their mathematics scores, using the number of influenza-like illness cases by county and month as an instrument. Goulas and Megalokonomou (2016) used a one-time policy in Greece that relaxed class attendance requirements as a measure against swine flu outbreaks, and found a statistically significant positive effect of absences on academic performance among high school students in Greece.

[11] There are studies that have analyzed the effects of reduced due to snowfall instruction time on students' achievements (Goodman, 2014; Hansen, 2013; Marcotte, 2007; Marcotte and Hemelt, 2008).

statistics. Section 5 discusses the estimation results. Section 6 provides some additional remarks. Section 7 concludes this paper with suggestions for future research.

2 Institutional Background

In this section, we briefly summarize the institutional settings related to: 1) class closures in Japan, 2) a financial support program for school attendance, and 3) an annual survey conducted by the city we analyzed (hereafter referred to as City X).

2.1 Class Closures in Japan

In Japan, the School Health and Safety Act stipulates that school administrators (typically municipal education boards) have the discretion to shut down primary and middle schools, grades, and classes under their jurisdictions to prevent the spread of viral infections. The national government does not provide explicit criteria for judging the timing of a shutdown. In some cases, prefectural or municipal education boards set criteria for judgment, in their public schools. The criteria vary across prefectures and municipalities, but in many cases, education boards decide to close classes, grades, or schools when the rate of absentees reaches 20%. According to the Act, in the case of public primary and middle schools, the education board of local municipalities makes the final decision to officially consult school principals and call for a shutting down of classes. According to a news article, in some municipalities in Tokyo, school principals make the "final" decision of the closure, and in the others, they advise the necessity of class closure to the municipal education board (Nippon Hoso Kyokai, 2019).

Seasonal flu is a major cause for class closures in Japan. Since flu infection worsens during winters, most of the class closures are observed in this season. In the years 2018-2019 in Tokyo, the number of flu infections increased during December, peaked in January, and became closer to zero in February. Class closure is mainly observed in January and February.[12] Once students get infected by flu, they are prohibited from going to school for five days after developing a fever, and an additional two days after their temperature lowers.[13] In some cases, students infected by flu lose more instruction days than those affected by class closures.

[12] http://idsc.tokyo-eiken.go.jp/diseases/flu/flu2018/

[13] See https://www.mhlw.go.jp/bunya/kenkou/kekkaku-kansenshou01/qa_eng.html for more details.

2.2 Financial Support Program for School Attendance

In Japan, students from economically disadvantaged families receive financial support from local governments for attending school.[14] "'Public assistance beneficiaries' as prescribed in the Public Assistance Act" and "'school attendance support beneficiaries,' who are recognized by the municipal board of education to be facing a degree of hardship equivalent to public assistance beneficiaries" have the right to receive the assistance program (Noguchi et al., 2020, p.4). The municipalities provide financial support for students' expenditures, such as school supplies and lunches. Households with annual incomes below a certain level, called the "minimum cost of living" which is set by the national government and which matches the minimum cost of living, after subtracting the household income, can receive this public assistance. In the case of a household that consists of a married couple in their 40s, a junior high school student, and an elementary school student who lives in the special wards of Tokyo, the minimum cost of living is calculated to be a monthly income of 0.205 million JPY,[15] and accordingly, the household receives an assistance of up to 2.46 million JPY annually. In the Tokyo Metropolitan Area, the annual income for school attendance support differs among municipalities, ranging from 3.99 million to 4.19 million JPY.[16] According to a nationwide survey in 2016, households with an annual income that falls below 4.5 million yen accounted for the bottom 25.9% in the distribution of households with children.[17] Households receiving financial assistance programs are likely to belong to this bottom 25% of the income distribution. We utilize the financial support status as a proxy for the economic condition of the students' households.

2.3 Annual Survey Conducted by City X

City X conducted an original survey on students' academic achievements in the middle of April, in addition to surveys by national and prefectural governments. The targets of the survey by conducted City X were students from the 2nd to 9th grades, attending public schools in City X. This survey collected information on students' achievements, lifestyles, and study habits. Students are required to take examinations that cover the contents they have learned until the previous academic year. Elementary school students take mathematics and language arts tests, and middle school students take mathematics, language arts, and English tests. Information on the test scores and student

[14] Section II of Noguchi et al. (2020) provides more detailed information about the financial assistance program.

[15] Page 22 of https://www.mhlw.go.jp/content/12002000/000488808.pdf.(in Japanese)(accessed on March 1st, 2021)

[16] The types of income utilized for calculating the income criteria are different among municipalities. For example, some municipalities utilize taxable income (Noguchi et al., 2020).

[17] Table 8 of the Excel file named "Statistical Table" on the web page, https://www.mhlw.go.jp/english/database/db-hss/cslc-report2016.html, provides a percentage distribution of household income (accessed on March 1st, 2021).

behavior obtained from this survey was primarily used for the analysis. Section 4 explains the details of the data.

3 Econometric Model

We estimate the following regression model to examine the effect of class closure on students' academic outcomes:

$$ZS^s_{ijt+1} = \alpha_0 + \alpha_1 Closed_{jt} + f_1(ZS^M_{ijt}, ZS^L_{ijt}) + X'_{ijt}\delta_1 + \eta_{1t} + \phi_{1k} + \lambda_{1g} + u_{1ijt} \tag{1}$$

where i, k, g, j, t, and s are the indices for students, school, grade in school, class, year, and subject, respectively. The index of subject s is either mathematics (M) or language (L). The variable ZS^s_{ijt} is the standardized test score of subject s in year t, for student i in class j. The dependent variable was the standardized score for the school year following class closure. The dummy variable ($Closed_{jt}$) takes one if class j is shut down because of the flu epidemic in year t.

To control for the students' previous academic achievements, a cubic function of the standardized scores in the year of class closure, $f_1(\cdot, \cdot)$, is included. The vector X_{ijt} consists of a set of covariates, such as dummy variables for financial support status[18] and female students, class size of j and its squared term, size of grade g in school k, and its squared term. The parameters η_{1t}, ϕ_{1k}, and λ_{1g} are the year, school, and grade FEs, respectively. These covariates and fixed effects control for potential endogeneity bias, as discussed in the next section. u_{1ijt} is the idiosyncratic error term. Estimated coefficients are reported with standard errors that were robust against class-level clustering.

We split the sample based on whether students received financial support, as a proxy for the economic condition of the households, in order to examine the heterogeneous impacts of the class closure. We define students who receive public assistance for school attendance as those from "low-income" households and those who do not receive it as those from "middle-to-high-income" households.[19] In addition, we estimate the following equation, with an interaction term, using the entire sample:

$$\begin{aligned} zscore^s_{ijt+1} =& \beta_0 + \beta_1 Closed_{jt} + \beta_2 Low\text{-}income_{it} + \beta_3 Closed_{jt} \times Low\text{-}income_{it} \\ &+ f_2(zscore^M_{ijt}, zscore^L_{ijt}) + X'_{ijt}\delta_2 + \eta_{2t} + \phi_{2k} + \lambda_{2g} + u_{2ijt}, \end{aligned} \tag{2}$$

[18] We use the four financial support status dummy variables: "public assistance beneficiaries," "school attendance support beneficiaries," "rejected applicants," and "non-applicants".

[19] We implement a robustness check against the definition of the support condition in Appendix B.

where $Low\text{-}income_{it}$ is i's status of household income in t, taking a value of 1 if the student receives the financial support.

When we estimate the effect of class closures on students' academic achievement, there are some potential threats to identification. Figure 1 summarizes our identification strategy. Our identifying assumption is that a flu epidemic, in a given year, and in a given region where a school is located, is random by conditioning on region and year characteristics. In general, the characteristics of residents and/or healthcare resources in a certain area should be determinants of the flu epidemic in the area. However, with our data from one municipality where public schools are close to each other,[20] it is plausible that the characteristics affecting flu spread outside of school are almost the same for all schools in the municipality. Nonetheless, we add school fixed effects and year fixed effects to control for regional characteristics, and annual trends of flu outbreaks to control for time-invariant differences of the schools in the municipality.

Even though the flu epidemic in a region is assumed to be conditionally independent, there are at least two potential sources of endogeneity. The first possible source is the characteristics of classes. As will be discussed in the next section (Section 4), there is a positive relationship between the incidence of class closures and class size.[21] Since it is also possible that the class size affects the students' academic achievements, it could be the source of the bias of the estimate of the effect of class closure, if the characteristics of classes are uncontrolled. We control for class characteristics, especially class size, to obtain an estimate of the causal effects.

The second potential source of endogeneity is the characteristics of school principals. As explained in Section 2, if a class suffers from the flu, the administrator of the school, such as the municipal education board, decides whether to shut down classes or not, after consulting with the school principal. A risk-averse school principal may strongly advise the school administrator to halt classes to prevent an epidemic in the entire school. The principal may also try to pay careful attention to students' behaviors to prevent risky behaviors, and this effort may have a positive impact on students' academic achievements. Since principals transfer to other schools infrequently in our sample, school FEs capture principals' characteristics to a large extent.[22]

[20]The city has about two public schools every square kilometer. The area is about $50km^2$, and there are approximately 100 public schools, as of 2015.

[21] A causal effect of class sizes on the incidence of class closures is found by Oikawa et al. (2022).

[22] We implement a robustness check against the principals' characteristics by adding school-year FEs into the estimation model.

4 Data

We use school administrative data collected by the Education Board of City X in the Tokyo Metropolitan Area in this study. The city is a large municipality, with more than 300,000 households, and a population of more than 600 thousand in 2015. In 2015, it had approximately 70 public elementary schools with almost 1000 classes and more than 30,000 students, and around 40 public middle schools with about 400 classes and about 14,000 students.

The data used in this analysis contains information related to educational administration, such as the number of students in each classroom, scores of the test conducted by the city, and the number of instances of class shutdowns, for all public elementary and middle schools operated by the city. As explained above, the education boards of local governments make the final decision on class closure in Japanese public schools, and the Education Board of the city keeps the records on school shutdown in the schools operated by them. If all classes in a particular grade are closed, they are called grade closures, as observed in our dataset. In this paper, we use both types of closures, which for simplicity, has been referred to as class closures. While the data contain students' information for the period up to 2018, information about class closures due to flu epidemics was only available for three school years 2015-2017. Using the data, we constructed a three-year student-level panel data which includes students' characteristics, incidences of class closure, and class characteristics. We restrict the sample to the 2nd to 8th grades, at the time of class closures, because the estimation procedure requires test scores from two consecutive years the score in a school year with an incident of class closure and the score in the following school year. Additionally, we excluded classes with less than 17 students, and grades with fewer than 30 students.

4.1 Descriptive Statistics

Table 1 shows the summary statistics for students in public schools operated by City X in 2015-2017. Column (1) shows the mean and standard deviation for the elementary school students (2nd to 6th graders), and column (2) shows the same for middle school students (7th and 8th graders). We standardized test scores, with a mean of 50 and a standard deviation of 10, within subject, grade, and year (Z-scores). About 9 % of the students in elementary schools have experienced class closure in these three years, and the ratio for middle school students is about 8%. The proportion of students from low-income households is approximately 30% in elementary schools and about 38% in middle schools. More than 30% enrolled students in public schools in City X are from economically disadvantaged backgrounds. According to the Japanese Ministry of Education, Culture, Sports, Science, and Technology, this proportion was about 15.23% in 2015, for all of Japan.[23] Therefore, City X has a higher ratio of economically disadvantaged students, compared to the national ratio. This ratio is also higher for middle schools than for elementary schools. This may be because some students with economic advantages drop out and enroll in private middle schools.

Table 2 summarizes the average z-scores by subject and the students' household income, for both mathematics and language, respectively. The average z-scores for students from low-income households are approximately 7% lower than those for students from middle-to-high-income households. These numbers suggest the presence of a socioeconomic gap in students' academic achievements in public schools in City X.

Figure 2 depicts the distribution of class closures due to flu epidemics in public schools operated by City X across months. According to Figure 2, these class closures occurred mainly between December and March of the following year. Since the Japanese school year is set from April to March, the class closure was concentrated in the last four months of an academic year. In particular, there seems to be a severe time constraint that makes it difficult to organize supplementary lessons for students experiencing closure, especially in March. Therefore, the loss of instruction time due to the flu may substantially affect the test scores in April in the following year.[24] The average days of the class closures by month, due to the flu epidemic, from 2015 to 2017, ranged from two to three days in both elementary and middle schools, as shown in Figure 3.

To see whether there are systematic differences in the characteristics of classes with and without closures, Table 3 reports the mean values of students' and classes' characteristics before class closures, to compare them with characteristics of students and classes that experience closure, relative to the rest. Columns (2) and (1) show mean values for students and classes that experience class closures in a school year ("closed") and those that do not ("not closed"), respectively.[25] Column (3) shows the raw differences between Columns (1) and (2). Column (4) shows the differences between Columns (1) and (2) after controlling for school, year, and grade fixed effects.[26] Column (5) shows the percentage differences in the characteristics, in comparison to the mean for students and classes that did not experience closures. According to Table 3, most of the characteristics do not

[23] Upper part of page 2 in https://www.mext.go.jp/content/20200327-mxt_shuugaku-100001991_2.pdf (in Japanese)(accessed on March 4th, 2021)

[24] Class closures occurred in April 2016, which was the month when the test was conducted by City X. However, since the closure started on April 26, that is, after the test, there was no closure before the test in that particular academic year.

[25] We calculate the mean values by pooling the data for the three school years. The students and the classes that experience class closures during a year are categorized as "closed" in that year. We categorize by using information for a single year.

[26] As a difference in two groups, closed and not closed, we report the coefficients of the class closure dummy as results of regressing each students' and classes' characteristics on the class closure dummy, school, year, and grade fixed effects.

have significant statistical differences between the two groups, except for the class size variables, after controlling for the fixed effects. Overall, the characteristics of students and classes are balanced between those that experienced closure and those that did not. The incidence of class closures is as good as random after controlling for the fixed effects and class size variables.[27]

5 Results

5.1 Effects of Class Closures on Students' Achievements

The estimates of the effect of class closures on students' academic outcomes are reported in upper panel A of Table 4 for elementary school students, and lower panel B for middle school students. Column (1) reports the estimate of the coefficient on the class closure dummy ($Closed$) in equation (1), with the mathematics test score as a dependent variable, and using a subsample of students who do not receive school financial assistance. Column (2) reports the same, but uses a subsample of students who receive school financial assistance. The results for elementary school students show that the estimated coefficient of the class closure dummy is negative, and statistically significant at the 5% level, only among students from low-income households, indicating that class closures negatively affect math test scores of elementary school students from low-income households. The estimated coefficient for the students from low-income households is -0.535, implying that the experience of the class closure decreases the test score in the following school year by 0.0535 of the standard deviation. The magnitude of the estimate for students from low-income households is about ten times larger than for those from middle-to-high-income households, even though both of the estimates are negative.[28]

To test these differential effects of class closures on students' math test scores between low- and middle-to-high-income households, we estimate equation (2). The estimates are reported in Columns (3) for mathematics. In addition, Columns (3) reports the estimates of the sum of the coefficients of $Closed$ and those of $Closed \times Low\text{-}income$ ($\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$), which corresponds to the effect of class closures on the students from low-income households, and the p-value of a test statistically confirming whether the sum is zero ($p\text{-}value$).[29] According to Column

[27] We also implement the same analysis for the classes' characteristics after class closures in Appendix A, and all of the characteristics do not have significant statistical differences between the two groups. The result suggests no special class arrangement for students who have experienced class closures that could make interpretation of our results difficult.

[28] Section 6.1 discusses the sizes of the coefficients in comparison with previous studies.

[29] The coefficient of $Closed \times Low\text{-}income$ corresponds to the difference in the coefficients of class closure dummy between low-income and middle-to-high-income students. On the other hand, the sum of the coefficients of $Closed$ and those of $Closed \times Low\text{-}income$ corresponds to the effect of class closures on the students from low-income households.

(3) in Panel A, the difference between the two estimates is statistically significant at the 1% level, and the estimate of the sum of the coefficients is -0.581 with a p-value of 0.019.

In contrast to the case of mathematics, the estimated effect on language scores for elementary school students is statistically insignificant. However, the coefficient is negative only for students from low-income households (Columns (4) and (5) in Panel (A)). Indeed, the difference between the two estimates is statistically significant at the 10% level (Column (6) in Panel A). Although neither of the effects of class closures is statistically significant for those from low- and middle-to-high-income households, this differential result may be a reflection of the fact that the students from relatively rich households have some advantages to compensate for the loss caused by the closures when compared to the economically disadvantaged students.

Finally, we discuss the results for the middle school students reported in Panel B. Among the middle school students, no significant effects were observed in either group or subjects. The following subsection discusses some heterogeneity of the effects of class closures and a possible reason for the insignificant effects among middle school students.

5.2 Heterogeneity of Effects of Class Closures

In the previous subsection, we confirmed that the effects of class closures are different between those from low- and middle-to-high-income households. In this subsection, we explore heterogeneous effects of class closures across other dimensions.

5.2.1 Gender

First, we discuss the heterogeneity of the effects based on the gender of students.

Table 5 reports the estimated effects of class closure on students test score by grade in school, gender, and subject, with equation (2). According to Table 5, the significant negative impact on mathematics test scores, observed among the economically disadvantaged elementary students, is driven by the negative impact among economically disadvantaged boys. Among disadvantaged elementary school boys, the coefficient of the interaction term between the class closure dummy and the low-income dummy is negative and statistically significant, and the sum of these two coefficients is also statistically significant (Column (1)). For languages, the sum is negative, but its p-value is over 10%, while the estimate of the interaction term is negative and statistically significant at the 10% level (Column (2)).

Among girls, there were no statistically significant effects for either the class closure dummy or the interaction term (Columns (3) and (4)). Additionally, the magnitudes of the coefficients for boys from disadvantaged households are about three times larger than those for girls from disadvantaged households. These results suggest that the damage to class closures is serious only among economically disadvantaged elementary school boys.

5.2.2 Timing of Class Closure

Next, we discuss the heterogeneous effects by the timing of the class closures. It may be more difficult for students experiencing class closures at the end of the school year to catch up with the students not experiencing it, because they do not have enough time to prepare for the test in April in the subsequent year. We estimate the different effects of class closures by the timing of the closures using two dummy variables: the dummy takes the value one if students experience the class closures between February and March, that is, the last two months of school year ($Closed(Feb. - Mar.)$), and the dummy takes the value of one if students experience the class closures, but the timing is not the last two months of the school year ($Closed(Others)$).

Table 6 shows the results of the heterogeneous effects by the timing of the class closures for the elementary school students. Estimates of the coefficient of the interaction term of $Closed(Feb. - Mar.)$, and the low-income dummy are negative and statistically significant for boys for both mathematics and language, while the sum is statistically significant only for mathematics. In contrast, the interaction terms of $Closed(Others)$ and the low-income dummy are negative but statistically insignificant for both subjects, and the sums of these coefficients are statistically insignificant. For mathematics test scores, the magnitude of the sum of $Closed(Feb. - Mar.)$ is approximately 40% larger than that for $Closed(Others)$. These results suggest that class closures at the end of the school year have stronger adverse effects on boys.

Note that a positive impact of class closures on or before January is estimated only among girls from middle-to-high-income households. The coefficient of $Closed(Others)$ for languages is estimated positively and statistically significant (Column (6)). This may be because the students suffering from the class closures at earlier timings have more time until the test in the following school year than those experiencing closures at the end of the school year to study more to recover from the negative effects of the closures. In addition, the girls who suffered from class closures may spend more time studying the language arts than studying mathematics.

The heterogeneous effect of the timings of class closures is one possible explanation for the insignificant impact among middle school students in our sample. Among the middle school students, there were no class closures in March, which was the last month of the school year. The peak of the closures is observed in January of 2017 and 2018, while the elementary school students have faced some closures in February and March (Figure 2). Therefore, middle school students usually have sufficient time to recover from the damage of class closures, and test scores in the following school year may not be affected by class closures.

5.2.3 Achievement at the Beginning of the School Year

Finally, we estimated the differential effects of class closures by students' achievements at the beginning of the school year. Since, as discussed in Section 4, the class closure was concentrated in the last four months of the school year, students' achievements at the beginning of the school year can be interpreted as the achievements before the flu epidemic seasons. Table 7 shows the estimation results with the subsample related to gender and Z-scores at the beginning of school year, t. We divided the sample into two groups using the mean value of z-scores, 50: students whose scores for both mathematics and language arts were less than 50 ($ZS < 50$), and those whose scores were more than or equal to 50 for at least one subject ($ZS \geq 50$).

The negative impact of class closures was more severe among economically disadvantaged boys with relatively low achievements at the beginning of the school year. Among elementary school boys whose test scores at the beginning of the school year are less than 50, the estimate of the interaction term of the closed and the low-income dummies on mathematics z-scores is negative and statistically significant. Their sum is about -1.302 with a p-value of 0.02. For boys with relatively high test scores at the beginning of the school year, the sum is -0.557 with a p-value of 0.088 (Columns (1) and (2)). The magnitude of the sum for boys with relatively low achievements is about 2.3 times more than that for boys with relatively high achievements. For language, the same tendency was observed, while the sums were statistically insignificant. These results suggest that boys who are not good at studying are susceptible to a temporal shock on schools' instruction time, such as class closures.

To sum up, among the economically disadvantaged students, the negative impact of class closures on students' achievement is statistically significant for mathematics test scores. This is consistent with previous studies suggesting that mathematics scores are more affected by instruction time than language arts test scores(Battistin and Meroni, 2016; Jensen, 2013; Marcotte, 2007; Zimmer et al., 2010). The adverse impact on economically disadvantaged students is heterogeneous by subject, grade, gender, timing of class closures, and students' achievements before the flu epidemic seasons. The estimation results also imply that boys with relatively low achievements before the flu epidemic seasons from disadvantaged households are the most susceptible to class closures.

6 Discussion

This section discusses three issues related to the adverse impact of class closures on economically disadvantaged students: 1) sizes of impact of class closures and possible mechanisms behind the negative impact; 2) ways to mitigate the impact; and 3) the persistence of the impact.

6.1 Interpretation of the Effects and their Mechanisms

To interpret the magnitude of the effects found in our analysis, we compare our results with those of previous studies that analyzed the effects of instruction time on students' achievements. We convert the effect of class closures into effects per hour for comparability.

Table 8 summarizes the standards for the numbers of classes of mathematics and language for elementary school students, and the number of days in a school year for an elementary school operated by City X. The average number of classes per school day ranges from 0.7 to 0.9 for mathematics, and from 0.8 to 1.5 for languages, where the instruction time for each class in elementary schools is set at 45 minutes. Suppose that a one-day class closure reduces both mathematics and language arts classes by one. Then, students experiencing class closures lose a total of 45 minutes of instruction time, for each subject. As discussed in Section 4, the average number of days for class closures is two or three days. Therefore, if students face a three-day class closure, they lose 2.25 hours of the schools' instruction time for mathematics and language arts. Additionally, if students get infected by the flu, they are prohibited from going to school for five days after developing a fever, and for two more days after their body temperature lowers. If a student's temperature is lesser on the fourth day after developing a fever, they can go to school six days after the onset of infection. This means that the student loses a maximum of five days of classes. In this case, the student loses 3.75 hours of the schools' instruction time for each subject.

We calculate the effects of class closures per hour using the estimates from these two scenarios: the 2.25 hours reduction; and the 3.75 hours reduction. According to the estimates discussed in Section 5, the impact of class closures on mathematics test scores ranges from 0.0557 to 0.13 of a standard deviation for economically disadvantaged students; thus, the negative impact per hour can be calculated to be between 0.0152 and 0.0578 of the standard deviation.

The magnitude of the effects per hour estimated in this study (are larger than those suggested by some previous studies. For example, the effect of instruction time per hour can be calculated to be between 0.0014 and 0.0094, using Battistin and Meroni (2016), Cattaneo et al. (2017), Jensen (2013), Lavy (2015), and Motegi and Oikawa (2019).[30] Aucejo and Romano (2016) found

[30] Battistin and Meroni (2016) analyzed the effect of a remedial education program in Italy, which is on average about

that a one-day reduction in absences increases mathematics test scores by 0.0055 of a standard deviation. Andrietti (2015) and Huebener et al. (2017) analyzed education reforms in Germany, which expanded weekly instruction time by two hours for 5th to 9th graders using PISA, and found that the reform increases math test scores by 0.095 of the standard deviation, and by 0.053 of the international standard deviation, respectively. On the other hand, the magnitude of our estimate is in line with Goodman (2014), who found that an additional day of absence induced by moderate snowfall decreases mathematics test scores by 0.05 of the standard deviation, among 3rd to 8th graders and 10th graders.

To investigate the reason why the magnitude of the effects per hour estimated in this paper is larger than those of the previous studies, we examine how class closures affect the students' activities after school and on holidays, such as studying, watching TV, playing video games, and sleeping. We estimate the effects of class closures on students' usage of time, in the same manner as estimations for students' test scores: estimating equation (2) using students' usage of time in the following school year as dependent variables. In this analysis, we also utilize students' usage of time for the year they experience class closures as control variables.[31] Our analysis also helps to explain the heterogeneous impact by gender and students' achievements at the beginning of the school year. Appendix C explains the variables for students' usage of time.

50 hours of afternoon activities for mathematics and language arts. They found the effect on mathematics test scores to be 0.296 of a standard deviation among lower secondary school students. We can therefore calculate the effect per hour as 0.0059 (= 0.296/50). Cattaneo et al. (2017) analyzed the effect of instruction time on test scores in Switzerland using PISA, and found that adding one hour of instruction time per week increases PISA scores from 0.05 to 0.06 of the standard deviation. By dividing the effect by 37, which is the OECD average of the number of instruction weeks in 2006(OECD, 2008), the effect per hour can be calculated to be between 0.0014 and 0.0016 of the standard deviation. Here, we utilize the OECD average, instead of the data for Switzerland, because the data for Switzerland is not available. Jensen (2013) found that each one-hour of the instruction time can increase the mathematics test score by 0.0021 of the standard deviation. According to Lavy (2015), who analyzed the effect of instruction time on test scores using a sample from OECD countries of PISA, each one-hour weekly instruction time increases the PISA test score by 0.058 of the standard deviation. According to OECD (2008), the OECD average of the number of weeks of instruction for upper secondary education was 37 weeks in 2006, and the effect of instruction time per hour can be calculated as 0.0016 of the standard deviation. In the case of Japan, Motegi and Oikawa (2019) found that adding one hour is effective by 0.33 standard deviations of test scores for 8th graders taught by teachers with a postgraduate degree, as opposed to those without the degree. As the number of weeks of instruction is 35, according to OECD (2008), the effect of instruction time per hour can be calculated to be 0.0094 of the standard deviation.

[31] The additional control variables are as follows: Dummy variables for hours of time spent studying at home on a weekday (none, more than zero but less than a half hour, more than or equal to a half hour but less than one hour, more than or equal to one hour); dummy variables for hours of time spent studying at home during a holiday (same categories as dummies for studying on a weekday); interaction terms of the dummies for study times on a weekday and during a holiday; dummy variables for hours of time spent watching TV on a weekday (never or hardly, about 30 minutes, about one hour, one and a half hours, about two hours, about two and a half hours, about three hours, more than or equal to three hours); dummy variables for hours spent playing video games on a weekday (same categories as dummies for watching TV on a weekday); interaction terms of the dummies for watching TV and for playing video games; and dummy variables for the amount of sleep time (330 min., 390 min., 420 min., 450 min., 480 min., 510 min., 540 min., 570 min., 600 min., 630 min., 660 min., 690 min., 720 min.).

Tables 9, 10, and 11 summarize the estimation results of the amount of time for studying, watching TV or playing video games, and sleeping, respectively. We construct dummy variables for usage of time variables, and use them as dependent variables. We generate dummy variables indicating whether students study on a weekday, spend time for TV or video games on a weekday for three and a half hours or more, and whether students sleep for eight and a half hours or more on a weekday.

According to Table 9, we could not observe a systematic change in the effect of class closures on study time among economically disadvantaged boys whose mathematics scores are negatively affected by class closures. However, among other groups, some positive impacts on the amount of time spent studying at home are observed. Among boys from middle-to-high-income households, the experience of class closures slightly increases their proportion of studying at home during a holiday (*Closed* in Column (1) of Panel B), while there are no changes among boys from low-income households ($\hat{\beta}_{Closed}+\hat{\beta}_{Closed\times Low\text{-}income}$ in Column (1) of Panel B). The same tendency is observed among students with low achievement at the beginning of the school year (Columns (1) and (2) of Panel B).

For girls from low-income households and low achievement at the beginning of the school year, study time tends to increase. There is a slight increase in the proportion of studying on a weekday with a p-value of 0.085 (Column (5) of Panel A), and an increase in the proportion of girls studying at home for one hour or more on either a weekday or a holiday by 6 percentage points. This corresponds to an increase of approximately 13% compared to the overall mean (with p-value of 0.063 in column (5) of Panel C). Moreover, a proportion of girls who were not good at studying and from middle-to-high-income households increased their studying at home for one hour or more on both a weekday and a holiday by 5 percentage points, which corresponds to an increase of approximately 27% compared to the overall mean (Column (5) of Panel D). These tendencies of increasing study time for girls may help to mitigate the adverse effect of class closures resulting in insignificant effect for girls.

Next, Table 10 shows the effect of class closure on leisure activities at home. There is an increase in the proportion of boys who spend more time pursuing leisure activities in the following school year, such as watching TV and/or playing games, and especially among those from economically disadvantaged households. Among boys with low achievement at the beginning of the school year from low-income households, there is an increase in the proportion who spend time on TV and video games, for more than or equal to seven hours, by 4.7 percentage points, which is substantial and corresponds to the overall mean of 11.4%. This contrasts to the case for boys from middle-to-high-income households who decreased this proportion by 2.5 percentage points (Column (2) of Panel D). In an elementary school operated by City X, all classes ended at 15:05, and students usually arrived at their home by 16:00. If students spend seven hours on leisure activities after arriving home, their time to sleep would be 23:00 or later, and thus this reduces the amount of sleeping time. Among girls, there are some decreases in the proportion of long periods of time spent on leisure. These results suggest that the increase in the amount of study time may be associated with the decrease in the amount of time spent on TV and video games, and vice versa among girls.

Finally, Table 11 reports the effects on the amount of sleep time on a weekday. Among boys with low achievement at the beginning of the school year from economically disadvantaged households, the proportion of boys sleeping for eight and a half hours or more on a weekday decreases by approximately 4.7 percentage points, which corresponds to a decrease of about 6% compared to the overall mean (Column (2) of Panel A). Since about 80% of students sleep for longer than eight and a half hours, this effect could be substantial. There were also statistically significant decreases in the proportions of boys who sleep for nine hours or more and for nine and a half hours or more, but only among those from middle-to-high-income households. Meanwhile, there were no statistically significant changes in the amount of sleeping time among girls.

To sum up, it is possible that economically disadvantaged boys, namely the only group whose mathematics scores are negatively affected by class closures, increase the time they spend on TV and video games, reduce their duration of sleeping hours, and do not change the amount of study time. These changes may be a reason for the negative impacts on test scores, and could be one mechanism behind the larger magnitudes of class closure effects in this study than in previous studies. The results could also explain the reason of the heterogeneous effects by gender. Girls increase study times at home, perhaps trying to compensate for the decrease in school instruction time. These results are consistent with Grewenig et al. (2010), who analyzed students' usage of time before and during school closures due to COVID-19 in Germany, and found that lower achievers replaced their learning times with leisure activities, including TV and computer games, rather than with activities beneficial for child development, and that boys were more susceptible to this behavior than girls.

6.2 Can We Mitigate The Impact?

We discuss how to mitigate the impact of class closures by focusing on school resources. As shown in Motegi and Oikawa (2019), it is possible that high-quality teachers efficiently reallocate instruction time after experiencing class closures, and that the negative impact due to the class closures is mitigated by such a reallocation. To examine the heterogeneous effects by school resources, we added the interaction terms of the closed dummy, the low-income dummy, and schools' resources into equation (2). We used class sizes, homeroom teachers' age, and homeroom teachers' teaching experience (in years) as measures of school resources. Unfortunately, data on homeroom teachers were only available up to 2016, reducing the samples only for 2015 and 2016. Appendix D explains the details of the data on homeroom teachers, and the estimation procedures used in this analysis.

Table 12 summarizes estimation results reporting only estimates of the interaction terms. Columns (1) to (5) are the results for mathematics scores, and Columns (6) to (10) are those for language arts scores. For mathematics scores, Column (1) reports the result reported in the Column (3) of Table 4, and Column (3) reports the result of the estimation with the specification of Column (1) using students in 2015 and 2016 only. Columns (2), (4), and (5) show the results of estimating equation (2) with various interaction terms with school resources: only the small class size dummy for Column (2), three homeroom teachers' characteristics for Column (4), and the small class size dummy and the homeroom teachers' characteristics for Column (5). The estimation results for language arts are reported in the same manner as Columns (1) to (5) (Columns (6) to (10)).

Table 12 reports that, for mathematics test scores, smaller class sizes and longer teaching experience (in years)could mitigate the negative impacts of class closures among economically-disadvantaged students, while an increase in teachers' age boost the negative impacts (Columns (4) and (5)). When we add only a dummy indicating that the size of class is 30 or less, the estimate of class closure and low-income dummies are negative and statistically significant, while the interaction term of the class closure, low-income, and class size dummies are positively and statistically significantly estimated. The magnitude of the estimate of the triple interaction term is almost the same as that of the double interaction term (Column (2)). Estimates of the interaction terms of the class closure dummy, the low-income dummy, and teaching experience (in years)are positive, while the estimate for only the teaching experience (in years)in current schools is statistically significant (Column (4)). An estimate of the interaction terms of the class closure dummy, the low-income dummy, and the adjusted teachers' age is negative and statistically significant (Column (4)). These results are robust after adding the class size dummy to Column (4) (Column (5)). Table D.3 in Appendix D details the effects of class closures on students from low-income households using the estimates in Column (5) for some scenarios by changing values of the school resource variables. The results of Table D.3 suggest that an increase in teachers' age boosts the negative impacts among economically disadvantaged students, and increases with teaching experience (in years), in both the prefecture where City X is located and the current schools, mitigating the negative impacts of class closures.

6.3 Persistency of The Negative Impact

Finally, we discuss the persistence of the negative impacts of class closures on test scores. We use data of students whose test scores were available for two years after class closures. Since data on test scores were available up to 2018, the analyzed sample consisted of students who experienced class closures in 2015 and 2016.

Tables 13 and 14 show estimation results for elementary school boys and girls, respectively. Columns (1) and (3) are the estimation results of the effects of class closures on z-scores, one year after the closures, for mathematics and language arts, respectively. Columns (2) and (4) show z-scores two years later.

Among elementary school boys from economically disadvantaged households, the estimated effect of class closures on mathematics test scores one year later is negative, and statistically significant, with a p-value of 0.011 in column(1) of Table 13). The scores for two years are negative but statistically insignificant, with a p-value of 0.148 ($\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$ in columns and (2) of Table 13. Moreover, the magnitude of the estimate for the scores one year after the closures is about 1.8 times larger than for the scores two years after the closures. These results suggest that the negative impacts of class closures on students' academic achievements could be mitigated over time, at least in schools operating in City X. In fact, remedial education programs were conducted by City X during the sample period. City X provides 3rd and 4th graders, with relatively low academic achievements, with remedial education programs for both mathematics and language arts, for at least 45 hours in a school year. Students impacted by class closures may attend remedial programs in the next school year and catch up with other students to a certain extent.[32] These public programs could be one way to save students from a temporary shock to their learning environment.

7 Conclusion

This paper utilizes the administrative data of students from City X in the Tokyo Metropolitan area, and analyzes the effects of class closures on students' test scores in the following years, focusing on the heterogeneous effects of students' economic backgrounds. According to our results, among the economically disadvantaged students, the negative impact of class closures on mathematics achievement is statistically significant. The negative impact on disadvantaged students is heterogeneous by subject, grade of school, gender, timing of class closures, and students'

[32] The positive effect of remedial education programs is observed among elementary students in Japan. Bessho et al. (2019) analyzed an effect of remedial education programs on mathematics and language arts test scores for elementary students in a city in Tokyo, and found a positive impact on the test scores of language arts.

achievements before class closure. Boys from economically disadvantaged households are more susceptible to class closures, especially those with relatively low achievements before closures. The negative impact on economically disadvantaged boys could be caused by reductions in school instruction times, and increases in amounts of time spent watching TV and playing video games and a decrease in the amount of sleeping time. These results suggest that students from economically disadvantaged households are more vulnerable to temporal shocks to their learning environment. We also found that school resources could mitigate damages among economically disadvantaged students. These results indicate that public programs, such as the remedial education programs conducted by City X, might play a role in mitigating the damage from class closures and saving students from a temporary shock to their learning environment.

Appendix

A Additional Descriptive Statistics

Table A.1 reports the mean values of classes' characteristics after class closures, to compare them with post-closure characteristics of classes that experience closure, relative to the rest. As post-closure characteristics of classes, we use the girl ratio within classes, the ratio of students from low-income households within classes, class size variables, grade size, and homeroom teachers' characteristics at period $t + 1$, the following school year of class closures. Columns (2) and (1) show mean values for students and classes that experience class closures in a school year ("closed") and those that do not ("not closed"), respectively. Column (3) shows the raw differences between Columns (1) and (2). Column (4) shows the differences between Columns (1) and (2) after controlling for school, year, and grade fixed effects and class characteristics at period t. Column (5) shows the percentage differences in the characteristics, in comparison to the mean for students and classes that did not experience closures. According to Table 3, all of the characteristics do not have significant statistical differences between the two groups. The result suggests no special class arrangement for students who have experienced class closures that could make interpretation of our results difficult.

B Robustness Checks

Table B.1 reports two robustness checks: 1) adding school-year fixed effects (FEs) in the estimation model; and 2) changing the definition of the students' status of the household income. Panel B of Table B.1 summarizes the estimation results that replicate Table 5 with school-year FEs into the model to control for the unobserved heterogeneity of a school in a year, such as school principals' characteristics. Panel C of Table B.1 shows the estimation results that replicate Table 5 by changing the definition of the low-income dummy variable from the dummy taking value one if students receive the support in the survey year to taking value one if students receive the support in all three years. Panel A reports Table 5 again.

In the estimation, we control for the school FEs to capture the school principals' characteristics, because principals do not transfer to other schools frequently, and our sample period is just three years. However, one could argue that they could transfer to other schools in the three years, and that the estimation models could not control for the principals' characteristics. To address this concern, we re-estimate Table 5 with school-year FEs to control for the unobserved heterogeneity of a school in a year, such as school principals' characteristics. According to Panel B of Table B.1, the estimation results are robust. The results have the same tendency as those in Panel A, and the magnitudes of the estimates are comparable to those in Panel A.

Since the household income status for a student could change from year to year, it is possible that the student is handled as both "low-income" and "middle-to-high-income" in an estimation. We use the dummy taking a value of one if students' household income is certified as low in the entire sample period instead of taking a value of one if students' household income is certified as low in the survey year to examine whether the change in the household income status affects the estimation results. According to Panel C of Table B.1, the estimation results are robust against the definition of the low-income dummy variable, the results have the same tendency as those in Panel A, and the magnitudes of the estimates are comparable to those in Panel A.

C Data on Students' Activities

Information on students' activities after a school day and on a holiday is collected by a survey conducted in City X. We focus on two measures of usage of time among students: amount of time spent watching TV and playing video games on a weekday, the amount of time spent studying at home on both a weekday and a holiday, and amount of sleeping time on a weekday. This is because the contents of the survey differ according to grade, and the information on those is relatively compatible across grades.

A variable of the amount of time spent watching TV and playing video games on a weekday is constructed with two variables: amount of time spent watching TV and amount of time spent playing video games (including handheld game consoles). Students were asked two questions: how many hours do they usually watch TV on a weekday, and how many hours do they usually play video games on a weekday? For both questions, students had eight options: 1) never or hardly, 2) about 30 minutes, 3) about one hour, 4) one and a half hours, 5) about two hours, 6) about two and a half hours, 7) about three hours, and 8) more than or equal to three hours. We let the first and the eighth options be 0 hours and three and a half hours, and converted the categorical variables to continuous variables.[33]

[33] Note that these variables were asked through the entire sample period and to all grades, while sentences and options of questions are slightly different according to school grade. For 3rd graders or higher, for questions on amounts of time spent watching TV and playing video games, students had eight options: 1) never or hardly; 2) about 30 minutes; 3) about one hour; 4) one and a half hours; 5) about two hours; 6) about two and a half hours; 7) about three hours; and 8) more than or equal to three hours. On the other hand, for only 2nd graders, the eighth option was not the same. Instead it was "more than three hours." Moreover, 2nd and 3rd graders were asked about time spent watching TV, while 4th graders or higher were asked about time for watching TV, videos, and DVDs. The question for 2nd and 3rd graders was "how many hours a weekday (Monday through Friday) do you usually watch TV?" For 4th graders or higher the

The survey also asked two closed-ended questions to students about the amount of time spent studying at home: amount of time on a weekday, and those on a holiday. Since there are some slight differences in the content of the questions, to maintain comparability, we generate three types of dummy variables for both studying time on a weekday and on a holiday: a dummy indicating whether students study at home; indicating whether they study at home for half an hour or more; and indicating whether they study at home for an hour or more.[34] We generate a variable of amounts of sleeping time in different ways between less than 4th graders and 4th graders or higher, because the survey does not ask 2nd and 3rd graders about the amount of sleeping time but asks them about wake-up times and bedtimes. The survey asked 4th graders or higher close-ended questions about the amount of sleeping time with the following options: "10 hours or more," "9 hours or more and less than 10 hours," "8 hours or more and less than 9 hours," "7 hours or more and less than 8 hours," "6 hours or more and less than 7 hours," "less than 6 hours." For 2nd and 3rd graders, there were survey questions about wake-up times and bedtimes, and we calculate the amount of sleeping time for them using these variables. Options for the questions about wake-up times are "before 6:00," "about 6:00," "about 6:30," "about 7:00," "about 7:30," and "after 7:30," and those about bedtimes were "about 20:00," "about 20:30," "about 21:00," "about 21:30," "about 22:00," "after 22:00." We let the options "before 6:00," "after 7:30" and "after 22:00" be 5:30, 8:00, and 22:30, and calculated the amount of sleeping time.

Table C.1 shows the summary statistics for the usage of time variables among elementary school students. For the amounts of time for TVs and Video games and those of sleeping times, we report some percentiles in addition to mean and standard deviation. Panels A and B shows the statistics

same question was phrased as "how many hours a weekday (Monday through Friday) do you usually watch TV, videos, and DVDs?" For simplicity, we treat the eighth option asked to 2nd graders the same as that for 3rd graders or higher, and treat the question for 2nd and 3rd graders the same as that for 4th graders or higher.

[34] For these questions, there are notes on comparability over grade. First, the survey asks 6th graders or lower how many hours students study on a holiday at home, while that asks 7th graders or higher about total time for studying on both Saturday and Sunday. For 7th graders or higher, we calculate average amounts of time for studying on holidays, and treat the average amounts the same as the amount of time answered by 6th graders or lower.

Second, options for the questions are different by grade. The options of the questions for both a weekday and a holiday for 2nd and 3rd graders were as follows: "none," "about an hour," and "more than an hour." For 4th graders, the options for both a weekday and a holiday are "none," "a half hour or less," "a half hour or more and less than an hour," "an hours or more and less than two hours," and "two hours or more." Fifth and sixth graders have the same options as 4th graders for questions about studying time on a weekday, while the options for a holiday are different from those for 4th graders and were as follows: "none," " half an hour or less," " half an hour or more and less than an hour," "an hour or more and less than two hours," "two hours or more and less than three hours," and "three hours or more." Among 7th graders or higher, the options for studying time on a weekday and those for total time for studying on Saturday and Sunday were different. The options for the former are the same as the options for the question about studying time on a holiday for 5th and 6th graders, and those for the latter are as follows: "none," "an hour or less," "an hour or more and less than two hours," "two hours or more and less than four hours," "four hours or more and less than six hours," and "six hours or more."

for boys and girls, respectively. According to Table C.1, elementary school boys tend to spend more time watching TV and/or playing video games than elementary school girls. The average amount of time for TV and Video games among boys was about 36% more than that among girls. There are few differences between boys and girls in terms of time spent studying and sleeping.

D Heterogeneous Effects By School Resources

The characteristics of homeroom teachers are available from a list of teachers in schools operated by City X. Unfortunately, since the list of teachers was only had data up to 2016, the merged data covers students' data in only two years: 2015 and 2016. The list includes teachers' characteristics such as age, years of teaching experience, and IDs of classes they were in charge of. We merged the teachers' data with the students' data using the class-id. Using the merged data, we analyzed the heterogeneities of the effects of class closures by school resources variables.

To examine the possibility of heterogeneous effects caused by school resources that students face, we estimate the following equation:

$$\begin{aligned} zscore^{s}_{ijt+1} &= \gamma_0 + \gamma_1 Closed_{jt} + \gamma_2 Low\text{-}income_{it} + \gamma_3 Closed_{jt} \times Low\text{-}income_{it} \\ &\quad + \sum_{k=1}^{K} \big[\gamma_{4k} SR_{kjt} + \gamma_{5k} Closed_{jt} \times SR_{kjt} + \gamma_{6k} Low\text{-}income_{it} \times SR_{kjt} \\ &\qquad + \gamma_{7k} Closed_{jt} \times Low\text{-}income_{it} \times SR_{kjt}\big] \\ &\quad + X'_{ijt}\delta_3 + f_3(zscore^{M}_{ijt}, zscore^{L}_{ijt}) + \eta_{3t} + \phi_{3k} + \lambda_{3g} + u_{3ijt}, \end{aligned} \tag{C.1}$$

where SR_{kjt} is the k-th index of school resources for students belonging to class j in the school year of t. In this analysis, a dummy indicating that the size of class is less than or equal to 30, homeroom teachers' age, homeroom teachers' years of experience in the prefecture where City X is located, and in schools where they currently belong, are utilized as measures of school resources. We adjusted the homeroom teachers' age by subtracting 22, which is the age just after graduating from college to interpret intercepts more easily.

Table D.1 summarizes the estimation results for mathematics test scores for elementary school students, and reports not only estimates of the interaction terms of class closure and low-income dummies and those of the triple interaction terms of the two dummies and school resources, which are reported in Table 12, but also those of class closure dummy itself and those of interaction terms of class closure dummy and school resources. Column (1) reports the result reported in Column (3) of Table 4, and Column (3) reports the result with the specification of Column (1) using students in

2015 and 2016 only. Columns (2), (4), and (5) show the results of estimating equation (C.1) with various interaction terms with school resources: only the small class size dummy for Column (2), three homeroom teachers' characteristics for Column (4), and the small class size dummy and the homeroom teachers' characteristics for Column (5). The estimation results for language arts are reported in Table D.2 in the same manner as in Table D.1.

To interpret the results in Tables D.1 and D.2 as effects for students from low-income households, we let SR_1, SR_2, and SR_3 be the teachers' age, years of experience in the prefecture where City X is located, and in current schools, respectively, and calculate the effects using some scenarios for specific values of school resources.

We substitute the 25th, 50th, and 75th percentiles of an interested variable for the variable and compare the calculated effects among those. Here, the other variables are fixed as median values. The following is an effect calculated by substituting SR_1 for a:

$$\begin{aligned} \widehat{Effect(SR_1} = a) = (\hat{\beta}_{Closed} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income}}) \\ + (\hat{\beta}_{\substack{Closed \\ \times SR_1}} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income \times SR_1}}) \times a \\ + (\hat{\beta}_{\substack{Closed \\ \times SR_2}} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income \times SR_2}}) \times median(SR_2) \\ + (\hat{\beta}_{\substack{Closed \\ \times SR_3}} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income \times SR_3}}) \times median(SR_3), . \end{aligned} \tag{C.2}$$

where $\hat{\beta}_x$ is the estimated coefficient of a variable x in columns (5) of Tables D.1 and D.2.

Table D.3 summarizes the calculated effects for students from low-income households in the same manner as equation C.2. Columns (1) and (2) show the calculated effects for the mathematics and language arts, respectively. The value of the first cell in Column (1), -0.231, is the calculated effect on mathematics scores for students from low-income households by substituting Age-22 for 25 percentile, and the other variables for median values.

Acknowledgments

This study is conducted as a part of the Project Microeconometric Analysis of Education Policy with Large Administrative Data undertaken at the Research Institute of Economy, Trade and Industry (RIETI). This study utilizes a school administrative data collected by the Education Board of a city. This study was funded by the Japan Society for the Promotion of Science (JSPS) KAKENHI through a Grant-in-Aid for Scientific Research (A) "Scientific Evidence on Children's Human Capital Creation and Its Implementation: Development of Policy Assessment Process by Government-Academia Collaboration" (Project Number: 19H00602) and its former fund (Project Number: 16H03636). In addition, Tanaka received a fund from the JSPS (Project Number:20H05629). We would like to express our deepest gratitude to the City Office of City X and the Board of Education for their assistance in conducting this study. This study was approved by the Ethical Committee of City X on November, 2016. We thank Koichi Ushijima for his effort to construct the basis of the dataset. We also thank Yoko Yamamoto for their steady administrative support. We would also like to thank Chishio Furukawa, Daiji Kawaguchi, Masayuki Morikawa, Masanori Tsuruoka, and the seminar participants of the research meeting for the project and Discussion Paper seminar participants at RIETI and Yokohama National University Data Science Mini Conference 2022 for their helpful comments. The contents and opinions of this article are solely the personal views of the authors. We affirm that all remaining errors are our own.

References

Ager, Philipp, Katherine Eriksson, Ezra Karger, Peter Nencka, and Melissa Thomasson, "School Closures During the 1918 Flu Pandemic," dec 2020.

Alexander, Karl L., Doris R. Entwisle, and Linda Steffel Olson, "Lasting consequences of the summer learning gap," *American Sociological Review*, 2007, *72* (2), 167–180.

Andrew, Alison, "Inequalities in children's experiences of home learning during the COVID-19 lockdown in England," 2020, p. 35.

Andrietti, Vincenzo, "The Causal Effects of Increased Learning Intensity on Student Achievement: Evidence from a Natural Experiment," *SSRN Electronic Journal*, 2015, (June).

Arnault, Louis, Florence Jusot, and Thomas Renaud, "Social inequalities in access to healthcare among the population aged 50+ years during the COVID-19 pandemic in Europe," 2021.

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

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*, 2021, *193*, 104345.

Battistin, Erich and Elena Claudia Meroni, "Should we increase instruction time in low achieving schools? Evidence from Southern Italy," *Economics of Education Review*, 2016, *55*, 39–56.

Bellei, Cristián, "Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile," *Economics of Education Review*, 2009, *28* (5), 629–640.

Bessho, Shun-ichiro, Haruko Noguchi, Akira Kawamura, Ryuichi Tanaka, and Koichi Ushijima, "Evaluating remedial education in elementary schools: Administrative data from a municipality in Japan," *Japan and the World Economy*, 2019, *50* (April), 36–46.

Black, Sandra and Paul Devereux, "Recent Developments in Intergenerational Mobility," 2010.

Boeri, Tito, Alessandro Caiumi, and Marco Paccagnella, "Mitigating the work-safety trade-off," *Covid Economics*, 2020, (2), 60–66.

Bol, Thijs, "Inequality in homeschooling during the Corona crisis in the Netherlands. First results from the LISS Panel.," 2020.

Bonacini, Luca, Giovanni Gallo, and Sergio Scicchitano, "Working from home and income inequality: risks of a new normal' with COVID-19," *Journal of Population Economics*, 2021, *34* (1), 303–360.

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

Carlsson, Magnus, Gordon B. Dahl, Björn Öckert, and Dan-Olof Rooth, "The Effect of Schooling on Cognitive Skills," *Review of Economics and Statistics*, jul 2015, *97* (3), 533–547.

Cattaneo, Maria A., Chantal Oggenfuss, and Stefan C. Wolter, "The more, the better? The impact of instructional time on student performance," *Education Economics*, 2017, *25* (5), 433–445.

Chetty, Raj, John Friedman, Nathaniel Hendren, Michael Stepner, and The Opportunity Insights Team, "The Economic Impacts of COVID-19: Evidence from a New Public Database Built Using Private Sector Data," jun 2020.

Cortes, Kalena E. and Joshua S Goodman, "Ability-Tracking, Instructional Time, and Better Pedagogy: The Effect of Double-Dose Algebra on Student Achievement," *American Economic Review*, may 2014, *104* (5), 400–405.

Czeisler, Mark É., Kristy Marynak, Kristie E.N. Clarke, Zainab Salah, Iju Shakya, JoAnn M. Thierry, Nida Ali, Hannah McMillan, Joshua F. Wiley, Matthew D. Weaver, Charles A. Czeisler, Shantha M.W. Rajaratnam, and Mark E. Howard, "Delay or Avoidance of Medical Care Because of COVID-19Related Concerns United States, June 2020," *MMWR. Morbidity and Mortality Weekly Report*, 2020, *69* (36), 1250–1257.

Dahmann, Sarah C., "How does education improve cognitive skills? Instructional time versus timing of instruction," *Labour Economics*, 2017, *47* (April), 35–47.

Dietrich, Hans, Alexander Patzina, and Adrian Lerche, "Social inequality in the homeschooling efforts of German high school students during a school closing period," *European Societies*, 2021, *23* (S1), S348–S369.

Dingel, Jonathan and Brent Neiman, “How many jobs can be done at home?,” *Covid Economics*, 2020, (1), 16–24.

Downey, Douglas B., Paul T. von Hippel, and Beckett A. Broh, “Are schools the great equalizer? Cognitive inequality during the summer months and the school year,” *American Sociological Review*, 2004, *69* (5), 613–635.

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 of the United States of America*, 2021, *118* (17).

Fitzpatrick, Maria D., David Grissmer, and Sarah Hastedt, “What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment,” *Economics of Education Review*, 2011, *30* (2), 269–279.

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

Gonzalez, Dulce, Stephen Zuckerman, Genevieve M Kenney, and Michael Karpman, “Almost Half of Adults in Families Losing Work during the Pandemic Avoided Health Care Because of Costs or COVID-19 Concerns,” *Urban.Org*, 2020.

Goodman, Joshua S, “Flaking out: Student absences and snow days as disruptions of instruction time,” 2014.

Goulas, Sofoklis and Rigissa Megalokonomou, “Swine Flu and The Effect of Compulsory Class Attendance on Academic Performance Swine Flu and The Effect of Compulsory Class Attendance on Academic Performance 1,” *MPRA Paper No. 75395*, 2016.

Grätz, Michael and Oliver Lipps, “Large loss in studying time during the closure of schools in Switzerland in 2020,” *Research in Social Stratification and Mobility*, 2021, *71* (September 2020), 0–4.

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* number 8648 2010.

Hansen, Benjamin, “School Year Length and Student Performance: Quasi-Experimental Evidence,” *SSRN Electronic Journal*, 2013.

Huebener, Mathias, Susanne Kuger, and Jan Marcus, "Increased instruction hours and the widening gap in student performance," *Labour Economics*, 2017, *47* (April), 15–34.

Ikeda, Masato and Shintaro Yamaguchi, "Online learning during school closure due to COVID-19," *The Japanese Economic Review*, 2021, (0123456789).

Jaume, David and Alexander Willén, "The long-run effects of teacher strikes: Evidence from Argentina," *Journal of Labor Economics*, 2019, *37* (4), 1097–1139.

Jensen, Vibeke Myrup, "Working longer makes students stronger? The effects of ninth grade classroom hours on ninth grade student performance," *Educational Research*, 2013, *55* (2), 180–194.

Kawaguchi, Daiji, "Fewer school days, more inequality," *Journal of the Japanese and International Economies*, 2016, *39*, 35–52.

Kikuchi, Nobuyoshi, "The effect of instructional time reduction on educational attainment: Evidence from the Japanese curriculum standards revision," *Journal of the Japanese and International Economies*, 2014, *32*, 17–41.

Kranz, A. M., G. Gahlon, A. W. Dick, and B. D. Stein, "Characteristics of US Adults Delaying Dental Care Due to the COVID-19 Pandemic," *JDR Clinical and Translational Research*, 2021, *6* (1), 8–14.

Kubota, Kohei, "Effects of Japanese compulsory educational reforms on household educational expenditure," *Journal of the Japanese and International Economies*, 2016, *42*, 47–60.

Lavy, Victor, "Expanding School Resources and Increasing Time on Task: Effects of a Policy Experiment in Israel on Student Academic Achievement and Behavior," *NBER Working Paper No.18369*, sep 2012, *Volume 201* (Number 95).

_ , "Do Differences in Schools' Instruction Time Explain International Achievement Gaps? Evidence from Developed and Developing Countries," *Economic Journal*, 2015, *125* (588), F397–F424.

Maldonado, Joana Elisa, Kristof De Witte, and Joana Maldonado, "The effect of school closures on standardised student test outcomes," *FEB Research Report Department of Economics*, 2020, (September).

Marcotte, Dave E., "Schooling and test scores: A mother-natural experiment," *Economics of Education Review*, 2007, *26* (5), 629–640.

_ **and Steven W. Hemelt**, "Unscheduled School Closings and Student Performance," *Education Finance and Policy*, 2008, *3* (3), 316–338.

Meghir, Costas and Marten Palme, "Educational reform, ability, and family background," *American Economic Review*, 2005, *95* (1), 414–424.

Mena, Gonzalo E., Pamela P. Martinez, Ayesha S. Mahmud, Pablo A. Marquet, Caroline O. Buckee, and Mauricio Santillana, "Socioeconomic status determines COVID-19 incidence and related mortality in Santiago, Chile," *Science*, 2021, *372* (6545), eabg5298.

Meyer, Erik and Chris Van Klaveren, "The effectiveness of extended day programs: Evidence from a randomized field experiment in the Netherlands," *Economics of Education Review*, 2013, *36*, 1–11.

Meyers, Keith and Melissa A. Thomasson, *Can pandemics affect educational attainment? Evidence from the polio epidemic of 1916*, Vol. 15, Springer Berlin Heidelberg, 2021.

Motegi, Hiroyuki and Masato Oikawa, "The effect of instructional quality on student achievement: Evidence from Japan," *Japan and the World Economy*, 2019, *52* (January).

Nippon Hoso Kyokai, "NHK News Up: Igai To Abauto!? Gakkyu Heisa," 2019.

Noguchi, Haruko, Shun-ichiro Bessho, Koichi Ushijima, Akira Kawamura, and Ryuichi Tanaka, "Relationship between School Attendance Support and Academic Performance: An Analysis of Panel Data from Adachi Ward," *Public Policy Review*, 2020, *16* (6), 1–22.

OECD, *Education at a Glance 2008* Education at a Glance, OECD, sep 2008.

Oikawa, Masato, Ryuichi Tanaka, Shun ichiro Bessho, and Haruko Noguchi, "Do class size reductions protect students from infectious disease? Lessons for Covid-19 policy from flu epidemic in Tokyo Metropolitan Area," *American Journal of Health Economics*, feb 2022.

Okubo, Toshihiro, "Spread of COVID-19 and telework: Evidence from Japan," *Covid Economics*, 2020, (32), 1–25.

Pischke, JörnSteffen, "The Impact of Length of the School Year on Student Performance and Earnings: Evidence from the German Short School Years," *The Economic Journal*, oct 2007, *117* (523), 1216–1242.

Rahman, Ahmed S., "Why can't everybody work remotely? Blame the robots," *Covid Economics*, 2020, (36), 105–128.

Reimer, David, Emil Smith, Ida Gran Andersen, and Bent Sortkær, "What happens when schools shut down? Investigating inequality in students' reading behavior during Covid-19 in Denmark," *Research in Social Stratification and Mobility*, 2021, *71* (September 2020), 100568.

Rivkin, Steven G. and Jeffrey C. Schiman, "Instruction Time, Classroom Quality, and Academic Achievement," *Economic Journal*, 2015, *125* (588), F425–F448.

Takaku, Reo and Izumi Yokoyama, "What the COVID-19 school closure left in its wake: Evidence from a regression discontinuity analysis in Japan," *Journal of Public Economics*, 2021, *195*, 104364.

Taylor, Eric, "Spending more of the school day in math class: Evidence from a regression discontinuity in middle school," *Journal of Public Economics*, 2014, *117*, 162–181.

Wößmann, Ludger, "Schooling resources, educational institutions and student performance: The international evidence," *Oxford Bulletin of Economics and Statistics*, 2003, *65* (2), 117–170.

Zimmer, Ron, Laura Hamilton, and Rachel Christina, "After-school tutoring in the context of no Child Left Behind: Effectiveness of two programs in the Pittsburgh Public Schools," *Economics of Education Review*, feb 2010, *29* (1), 18–28.

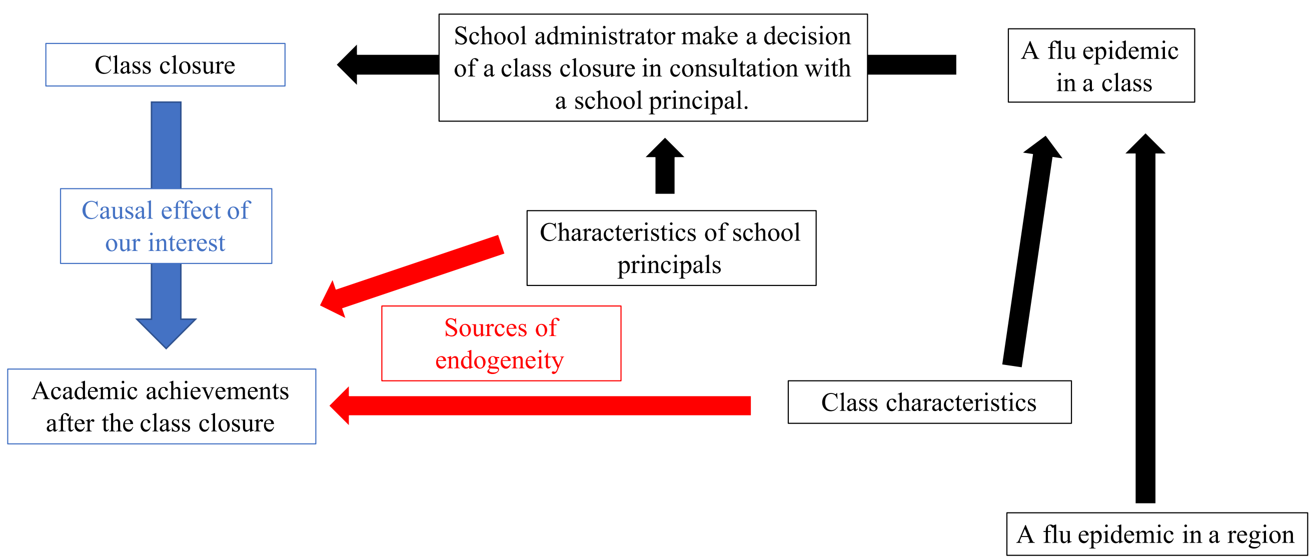
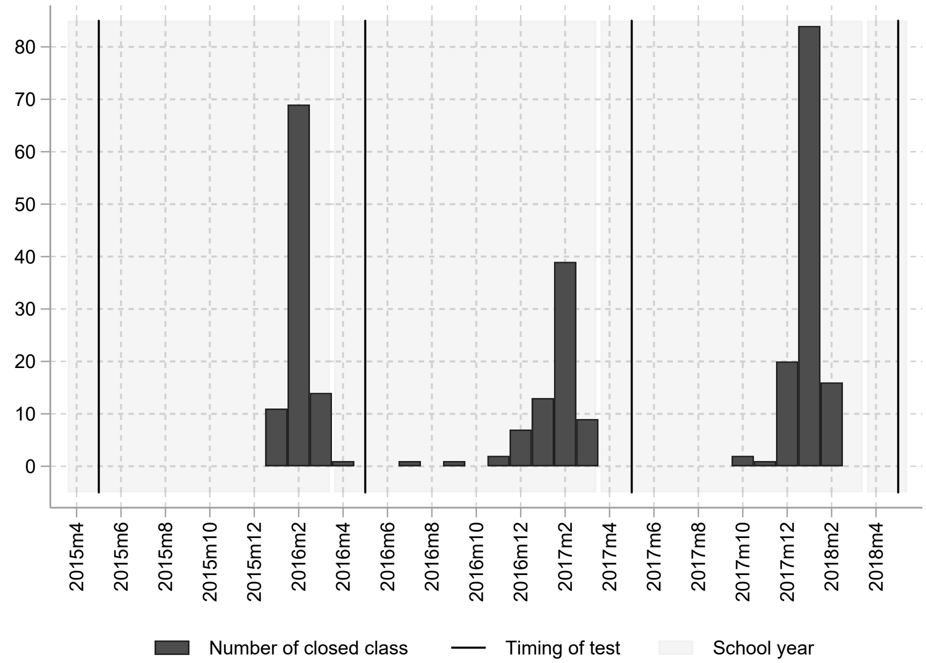
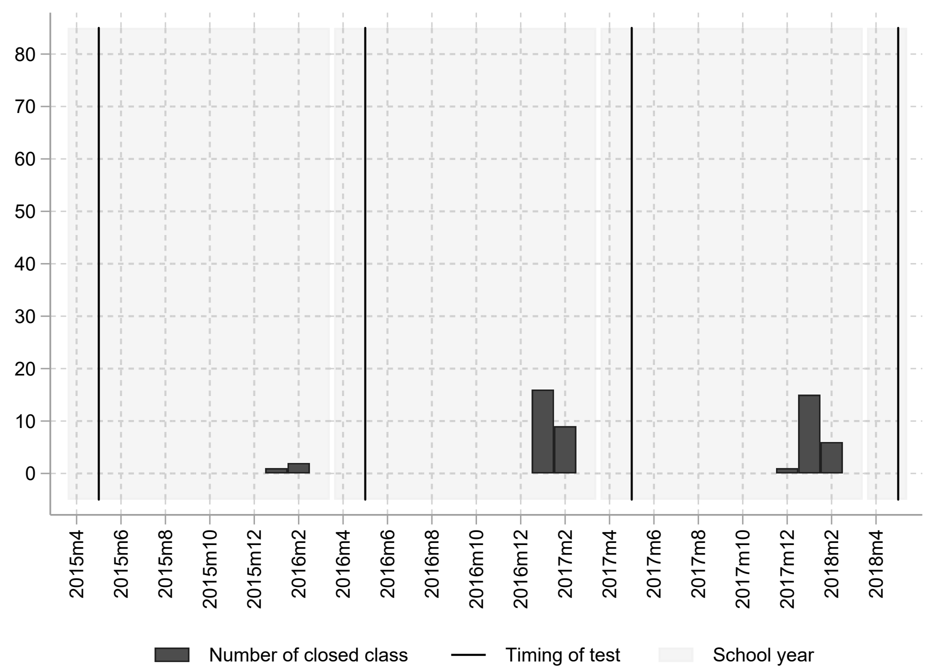


Figure 1: Image of Identification Strategy

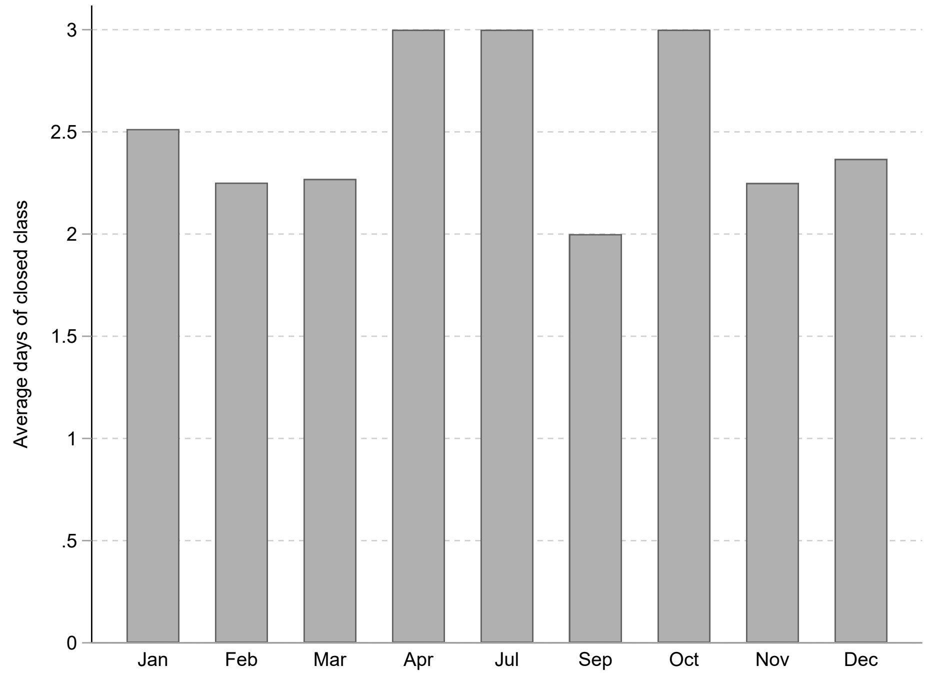


(a) Elementary Schools

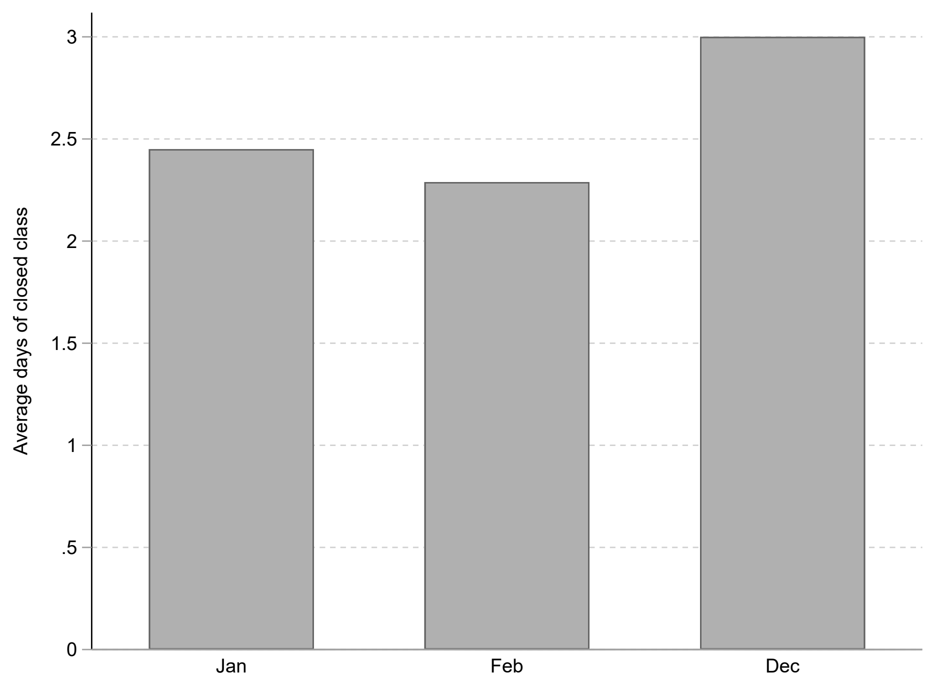


(b) Middle Schools

Figure 2: Number of Class Closures Due to Flu Epidemics, by Month



(a) Elementary Schools



(b) Middle Schools

Figure 3: Average Days of Class Closure Due to Flu Epidemic, by Month

Table 1: Summary Statistics

	Elementary Schools (1)		Middle Schools (2)	
	mean	sd	mean	sd
Z-scores				
Language arts	50.00	10.00	50.01	10.00
Mathematics	50.00	10.00	50.01	9.99
Class Closure	0.09	0.28	0.08	0.27
Girl	0.49	0.50	0.49	0.50
Low-income	0.30	0.46	0.38	0.49
Class size	31.57	4.19	33.67	3.69
School size	86.57	34.85	160.70	58.25

Table 2: Difference in Mean Z-scores Based on Household Income

	Middle-to-high -income	Low-income	Difference	%Δ from middle-to-high -income
Mathematics	51.20	47.49	-3.72	-7.26***
				(0.12)
Language arts	51.12	47.67	-3.46	-6.76***
				(0.12)

Standard errors are in parenthesis. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Students' and Classes' Characteristics Before Class Closures

	Mean		Difference		
	Not Closed (1)	Closed (2)	Raw (3)	Adjusted by FEs (4)	(4)/(1)×100 (5)
A.Students' characteristics					
Zscores					
Mathematics	49.99	50.19	0.20*	0.16	0.32
			(0.11)	(0.11)	(0.23)
Language	50.00	50.09	0.09	0.02	0.04
			(0.11)	(0.11)	(0.23)
Low-income	0.32	0.30	-0.02***	-0.00	-0.40
			(0.00)	(0.00)	(1.53)
Girl	0.49	0.49	-0.00	0.00	0.17
			(0.00)	(0.01)	(1.03)
Studying					
in a weekday	0.93	0.93	0.01*	0.00	0.26
			(0.00)	(0.00)	(0.33)
> 30 min on a weekday	0.73	0.70	-0.02***	0.01	0.94
			(0.01)	(0.00)	(0.68)
> 60 60 min on a weekday	0.46	0.43	-0.04***	0.00	0.11
			(0.01)	(0.01)	(1.21)
in a holiday	0.81	0.81	-0.00	0.00	0.50
			(0.00)	(0.00)	(0.57)
> 30 min on a holiday	0.54	0.52	-0.02***	0.00	0.84
			(0.01)	(0.01)	(1.06)
> 60 min on a holiday	0.30	0.27	-0.03***	-0.01	-1.81
			(0.01)	(0.01)	(1.80)
Amounts of time (min)					
for TV or video games	188.45	181.15	-7.30***	-1.24	-0.66
			(1.30)	(1.30)	(0.69)
for sleeping	513.06	525.04	11.98***	0.34	0.07
			(0.90)	(0.79)	(0.15)
B.Classes' characteristics					
Class sizes	31.54	31.56	0.02	0.44**	1.39**
			(0.24)	(0.18)	(0.57)
≤ 30	0.39	0.39	-0.00	-0.05**	-12.40**
			(0.03)	(0.02)	(5.59)
Grade sizes	105.31	101.38	-3.94	0.15	0.14
			(2.91)	(0.67)	(0.63)
Homeroom teachers' characteristics					
age - 22	15.45	15.83	0.38	0.73	4.72
			(0.78)	(0.80)	(5.17)
Years of experience					
in the prefecture where City X	11.54	11.86	0.32	0.74	6.39
			(0.78)	(0.79)	(6.89)
in the current schools	3.50	3.51	0.01	-0.00	-0.01
			(0.17)	(0.17)	(4.98)

The unit of observation is students for Panel A and classes for Panel B. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4: Effects of Class Closures on Students' Test Scores

	Mathematics			Language		
	(1)	(2)	(3)	(4)	(5)	(6)
	Middle -to-high income	Low income		Middle -to-high income	Low income	
A. Elementary school students						
Closed	-0.054	-0.535**	-0.035	0.105	-0.146	0.139
	(0.154)	(0.248)	(0.153)	(0.147)	(0.240)	(0.147)
Closed × Low-income			-0.546***			-0.371*
			(0.209)			(0.211)
Observations	49061	20777	69838	49038	20752	69790
$\hat{\beta}_{Closed} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income}}$			-0.581			-0.232
p-value			0.019			0.326
B. Middle school students						
Closed	-0.085	-0.152	-0.116	-0.107	-0.233	-0.087
	(0.271)	(0.345)	(0.270)	(0.235)	(0.316)	(0.230)
Closed × Low-income			-0.001			-0.177
			(0.269)			(0.293)
Observations	15850	9345	25195	15808	9312	25120
$\hat{\beta}_{Closed} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income}}$			-0.117			-0.263
p-value			0.712			0.383

Dependent variables are standardized test scores in the school year following class closures, with a mean of 50 and a standard deviation of 10. Standard errors that were robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 5: Heterogeneous Effects By Students' Gender

	Elementary School				Middle School			
	Boy		Girl		Boy		Girl	
	(1) Math	(2) Lang.	(3) Math	(4) Lang.	(5) Math	(6) Lang.	(7) Math	(8) Lang.
Closed	-0.027	0.164	-0.039	0.135	-0.181	-0.233	-0.080	0.050
	(0.185)	(0.187)	(0.174)	(0.157)	(0.352)	(0.336)	(0.303)	(0.236)
Closed × Low-income	-0.841***	-0.576*	-0.252	-0.193	0.321	-0.172	-0.285	-0.152
	(0.299)	(0.295)	(0.287)	(0.278)	(0.394)	(0.463)	(0.367)	(0.325)
Observations	35490	35464	34348	34326	12656	12610	12539	12510
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	-0.868	-0.412	-0.290	-0.059	0.140	-0.405	-0.365	-0.102
p-value	0.008	0.190	0.305	0.831	0.712	0.365	0.359	0.748

Dependent variables are standardized test scores in the following school year of class closures, with a mean of 50 and a standard deviation of 10. Standard errors that are robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Heterogeneous Effects By Timing of Closures Among Elementary School Students

	Mathematics			Language Arts		
	(1) Whole	(2) Boy	(3) Girl	(4) Whole	(5) Boy	(6) Girl
Closed (Feb.-Mar.)	-0.093	0.111	-0.307	0.014	0.225	-0.189
	(0.228)	(0.269)	(0.246)	(0.213)	(0.267)	(0.227)
Closed (Others)	0.021	-0.152	0.218	0.259	0.110	0.442^{**}
	(0.201)	(0.248)	(0.239)	(0.195)	(0.252)	(0.210)
Closed (Feb.-Mar.) $\times$ Low-income	-0.574^{**}	-1.117^{***}	-0.042	-0.324	-0.813^{**}	0.151
	(0.271)	(0.398)	(0.381)	(0.287)	(0.411)	(0.382)
Closed (Others) $\times$ Low-income	-0.505	-0.562	-0.435	-0.400	-0.327	-0.515
	(0.308)	(0.436)	(0.408)	(0.294)	(0.400)	(0.384)
Observations	69838	35490	34348	69790	35464	34326
Feb.-Mar.						
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	-0.667	-1.007	-0.348	-0.311	-0.588	-0.038
$p\text{-}value$	0.049	0.019	0.389	0.345	0.178	0.921
Others						
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	-0.484	-0.714	-0.217	-0.141	-0.218	-0.073
$p\text{-}value$	0.176	0.154	0.564	0.661	0.609	0.846

Dependent variables are standardized test scores in the following school year of class closures, with a mean of 50 and a standard deviation of 10. Standard errors robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 7: Heterogeneous Effects by Students' Test Scores at the Beginning of the School Yea Among Elementary School Students

	Boy				Girl			
	Mathematics		Language arts		Mathematics		Language arts	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$ZS < 50$	$ZS \geq 50$	$ZS < 50$	$ZS \geq 50$	$ZS < 50$	$ZS \geq 50$	$ZS < 50$	$ZS \geq 50$
Closed	0.208	-0.133	0.559	-0.004	0.208	-0.112	0.094	0.159
	(0.384)	(0.175)	(0.370)	(0.187)	(0.470)	(0.162)	(0.445)	(0.148)
Closed × Low-income	-1.510***	-0.424	-1.311**	-0.178	-0.272	-0.287	-0.204	-0.218
	(0.581)	(0.326)	(0.548)	(0.342)	(0.707)	(0.296)	(0.631)	(0.298)
Observations	11441	24049	11427	24037	7724	26624	7720	26606
$\hat{\beta}_{Closed} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income}}$	-1.302	-0.557	-0.753	-0.182	-0.065	-0.399	-0.109	-0.059
p-value	0.020	0.088	0.137	0.604	0.912	0.175	0.841	0.831

The dependent variables are the standardized test scores in the school year that follows the class closures, with a mean of 50 and a standard deviation of 10. Standard errors that are robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 8: Standard for The Numbers of Classes of Mathematics and Language Arts For Elementary School Students

	1st grader	2nd grader	3rd grader	4th grader	5th grader	6th grader
A. Standard number of classes						
Mathematics	136	175	175	175	175	175
Language arts	306	315	245	245	175	175
B. The number of days in a school year of an elementary school in City X						
	205	205	205	205	206	206
C. The average number of classes per day (A/B)						
Mathematics	0.7	0.9	0.9	0.9	0.8	0.8
Language arts	1.5	1.5	1.2	1.2	0.8	0.8

Source: https://www.mext.go.jp/a_menu/shotou/new-cs/youryou/syo/ (Panel A)(Japanese)(accessed on March 16, 2021).

Table 9: Effects on the Amount of Time Spent Studying At Home

	Boy			Girl		
	(1)	(2) $ZS < 50$	(3) $ZS > 50$	(4)	(5) $ZS < 50$	(6) $ZS > 50$
A. Studying on a weekday						
Closed	-0.006	-0.003	-0.007	-0.004	0.004	-0.007
	(0.005)	(0.013)	(0.006)	(0.004)	(0.012)	(0.005)
Closed × Low-income	0.019	0.030	0.010	0.006	0.023	-0.000
	(0.012)	(0.023)	(0.012)	(0.009)	(0.019)	(0.011)
Over mean	0.936	0.901	0.952	0.959	0.939	0.964
$\hat{\beta}_{Closed}+\hat{\beta}_{Closed \times Low\text{-}income}$	0.013	0.027	0.002	0.002	0.026	-0.007
p-$value$	0.231	0.140	0.833	0.817	0.085	0.494
B. Studying on a holiday						
Closed	0.015*	0.033*	0.008	0.005	0.021	0.002
	(0.009)	(0.017)	(0.010)	(0.007)	(0.020)	(0.007)
Closed × Low-income	-0.013	-0.032	-0.005	0.008	0.011	0.005
	(0.018)	(0.028)	(0.021)	(0.015)	(0.031)	(0.017)
Overall mean	0.804	0.761	0.825	0.845	0.805	0.856
$\hat{\beta}_{Closed}+\hat{\beta}_{Closed \times Low\text{-}income}$	0.002	0.001	0.003	0.013	0.032	0.007
p-$value$	0.912	0.968	0.875	0.358	0.221	0.659
C. Study for one hour or more either on a weekday or on a holiday.						
Closed	0.014	0.026	0.008	-0.005	0.036	-0.013
	(0.011)	(0.022)	(0.013)	(0.011)	(0.027)	(0.012)
Closed × Low-income	-0.007	-0.035	0.008	0.021	0.022	0.010
	(0.020)	(0.035)	(0.024)	(0.019)	(0.038)	(0.023)
Overall mean	0.460	0.408	0.483	0.490	0.448	0.501
$\hat{\beta}_{Closed}+\hat{\beta}_{Closed \times Low\text{-}income}$	0.007	-0.009	0.016	0.017	0.059	-0.003
p-$value$	0.685	0.760	0.458	0.327	0.063	0.865
D. Studying for one hour or more on both a weekday and a holiday						
Closed	0.002	0.013	-0.004	0.016	0.049**	0.010
	(0.010)	(0.017)	(0.011)	(0.011)	(0.023)	(0.012)
Closed × Low-income	-0.011	0.013	-0.032*	-0.020	-0.042	-0.018
	(0.016)	(0.027)	(0.018)	(0.016)	(0.033)	(0.020)
Overall mean	0.217	0.159	0.243	0.237	0.181	0.254
$\hat{\beta}_{Closed}+\hat{\beta}_{Closed \times Low\text{-}income}$	-0.009	0.026	-0.036	-0.004	0.007	-0.008
p-$value$	0.484	0.226	0.019	0.769	0.771	0.649

The number of observations was 32,093, 10,143, 21,950, 31,462, 6,974, and 24,488, respectively. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 10: Effects on the Amount of Time Spent watching TV and Playing Video Games on a Weekday

	Boy			Girl		
	(1)	(2) $ZS < 50$	(3) $ZS > 50$	(4)	(5) $ZS < 50$	(6) $ZS > 50$
A. Three and a half hours or more						
Closed	0.020*	0.010	0.024*	0.014	0.015	0.013
	(0.011)	(0.020)	(0.013)	(0.010)	(0.025)	(0.010)
Closed × Low-income	-0.016	0.013	-0.031	-0.050***	-0.020	-0.064***
	(0.020)	(0.034)	(0.025)	(0.018)	(0.039)	(0.021)
Overall mean	0.474	0.537	0.445	0.285	0.357	0.264
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	0.004	0.023	-0.007	-0.036	-0.005	-0.051
p-value	0.790	0.392	0.746	0.020	0.871	0.005
B. Four and a half hours or more						
Closed	0.017*	-0.007	0.027**	0.002	0.007	0.001
	(0.010)	(0.021)	(0.011)	(0.008)	(0.022)	(0.008)
Closed × Low-income	0.027	0.057	0.012	-0.024	-0.044	-0.018
	(0.019)	(0.036)	(0.023)	(0.015)	(0.034)	(0.016)
Overall mean	0.291	0.343	0.267	0.130	0.182	0.115
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	0.044	0.050	0.039	-0.023	-0.037	-0.017
p-value	0.011	0.085	0.054	0.120	0.158	0.252
C. Five and a half hours or more						
Closed	-0.001	-0.021	0.008	0.003	0.016	0.000
	(0.008)	(0.017)	(0.009)	(0.006)	(0.017)	(0.006)
Closed × Low-income	0.024	0.050	0.012	-0.024**	-0.093***	0.001
	(0.018)	(0.033)	(0.021)	(0.012)	(0.025)	(0.013)
Overall mean	0.174	0.225	0.151	0.063	0.102	0.053
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	0.023	0.029	0.020	-0.021	-0.077	0.001
p-value	0.147	0.294	0.259	0.043	0.000	0.932
D. Seven hours or more						
Closed	-0.009	-0.025**	-0.002	-0.002	-0.008	-0.000
	(0.006)	(0.012)	(0.007)	(0.003)	(0.011)	(0.003)
Closed × Low-income	0.024*	0.072***	-0.004	0.004	-0.018	0.014
	(0.013)	(0.026)	(0.014)	(0.009)	(0.016)	(0.010)
Overall mean	0.080	0.114	0.064	0.024	0.042	0.019
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	0.014	0.047	-0.006	0.002	-0.026	0.014
p-value	0.216	0.043	0.600	0.827	0.036	0.156

The number of observations was 32,073, 10,132, 21,941, 31,479, 6,980, and 24,499, respectively. Standard errors, robust against class-level clustering, are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Effects on the Amount of Sleep-Time on a Weekday

	Boy			Girl		
	(1)	(2) $ZS < 50$	(3) $ZS > 50$	(4)	(5) $ZS < 50$	(6) $ZS > 50$
A. Sleeping for eight and a half hours or more on a weekday						
Closed	-0.014	-0.031*	-0.006	0.011	0.012	0.012
	(0.010)	(0.019)	(0.011)	(0.008)	(0.022)	(0.009)
Closed × Low-income	-0.017	-0.016	-0.014	-0.018	-0.048	-0.010
	(0.018)	(0.033)	(0.021)	(0.017)	(0.038)	(0.020)
Overall mean	0.812	0.764	0.835	0.862	0.820	0.874
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	-0.032	-0.047	-0.019	-0.007	-0.036	0.002
p-value	0.042	0.077	0.313	0.639	0.202	0.924
B. Sleeping for nine hours or more on a weekday						
Closed	-0.015	-0.036*	-0.006	-0.000	-0.005	0.003
	(0.011)	(0.020)	(0.013)	(0.012)	(0.029)	(0.013)
Closed × Low-income	-0.009	0.000	-0.009	0.011	0.003	0.016
	(0.020)	(0.032)	(0.025)	(0.020)	(0.041)	(0.024)
Overall mean	0.570	0.542	0.583	0.638	0.617	0.644
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	-0.024	-0.035	-0.015	0.011	-0.002	0.018
p-value	0.189	0.181	0.521	0.519	0.955	0.377
C. Sleeping for nine and half hours or more on a weekday						
Closed	-0.013	-0.042**	-0.000	0.010	-0.001	0.015
	(0.012)	(0.020)	(0.013)	(0.012)	(0.029)	(0.013)
Closed × Low-income	-0.001	0.004	0.002	0.014	0.016	0.015
	(0.021)	(0.032)	(0.026)	(0.021)	(0.041)	(0.026)
Overall mean	0.493	0.475	0.500	0.557	0.556	0.557
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	-0.014	-0.038	0.002	0.024	0.015	0.030
p-value	0.462	0.150	0.932	0.160	0.641	0.172

The number of observations was 32,089, 10,153, 21,936, 31,479, 6,976, and 24,484, respectively. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 12: Heterogeneous Effects of School Resources Among Elementary School Students

	Mathematics					Lanuage Arts				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Closed × Low-income	-0.546***	-0.833***	-0.645**	-0.945*	-1.193**	-0.371*	-0.474*	-0.184	0.479	0.498
	(0.209)	(0.265)	(0.269)	(0.547)	(0.584)	(0.211)	(0.275)	(0.299)	(0.662)	(0.670)
× Class size ≤ 30		0.775*			0.781		0.255			-0.372
		(0.426)			(0.551)		(0.423)			(0.646)
× Teacher' age - 22				-0.072*	-0.090**				-0.066	-0.050
				(0.037)	(0.038)				(0.047)	(0.052)
× experience in the prefecture where City X locates				0.052	0.069				0.022	0.010
				(0.045)	(0.049)				(0.046)	(0.050)
× experience in current school				0.230**	0.235**				0.004	0.005
				(0.111)	(0.111)				(0.127)	(0.128)
Observations	69838	69838	40752	40752	40752	69790	69790	40717	40717	40717

Dependent variables are standardized test scores in the school year following class closures, with a mean of 50 and a standard deviation of 10. This table only reports the coefficients of the cross term for class closure and low-income dummies, and those of the triple cross term of the two dummies and school resources. Tables D.1 and D.2 report the coefficients of the class closure dummy itself, and those of the interaction terms of class closure dummy and school resources, in addition to the coefficients reported in this table. Since data on teachers' characteristics is available only up to 2016, the sample size for the last three columns is smaller than that for the first two columns, in which data are available from 2015 to 2017. Standard errors robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 13: Effects on Test Scores of One Year and Two Years after Class Closures among Boys in Elementary Schools

	Mathematics		Language	
	(1) One year later	(2) Two years later	(3) One year later	(4) Two year later
Closed	-0.006	-0.160	0.124	-0.237
	(0.280)	(0.238)	(0.268)	(0.279)
Closed $\times$ Low-income	-1.068***	-0.439	-0.522	-0.147
	(0.395)	(0.411)	(0.388)	(0.441)
Observations	22098	22098	22057	22057
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	-1.074	-0.599	-0.398	-0.383
p-value	0.011	0.148	0.315	0.359

Dependent variables are standardized test scores in the school year following class closures, with a mean of 50 and a standard deviation of 10. Standard errors robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 14: Effects on Test Scores One Year and Two Years after Class Closures among Girls in Elementary Schools

	Mathematics		Language	
	(1) One year later	(2) Two years later	(3) One year later	(4) Two year later
Closed	0.008	0.001	0.130	-0.220
	(0.257)	(0.229)	(0.223)	(0.223)
Closed $\times$ Low-income	-0.378	0.046	0.028	-0.003
	(0.382)	(0.379)	(0.402)	(0.353)
Observations	21533	21533	21502	21502
$\hat{\beta}_{Closed} + \hat{\beta}_{Closed \times Low\text{-}income}$	-0.370	0.047	0.158	-0.223
p-value	0.317	0.900	0.678	0.532

Dependent variables are standardized test scores in the school year following class closures, with a mean of 50 and a standard deviation of 10. Standard errors robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.1: Classes' Characteristics After Class Closures

	Mean		Difference		
	Not Closed	Closed	Raw	Adjusted by FEs and variables at t	(4)/(1)×100
	(1)	(2)	(3)	(4)	(5)
Girl ratio (t + 1)	0.49	0.49	-0.00	-0.00	-0.08
			(0.00)	(0.00)	(0.44)
Ratio of students from low-income households (t + 1)	0.30	0.29	-0.02***	0.00	0.51
			(0.01)	(0.00)	(1.47)
Class size (t + 1)	32.00	32.20	0.19	0.10	0.31
			(0.24)	(0.30)	(0.92)
≤ 30 (t + 1)	0.35	0.33	-0.02	-0.02	-5.12
			(0.03)	(0.03)	(7.53)
Grade sizes (t + 1)	109.68	105.09	-5.46*	0.23	0.21
			(2.80)	(1.19)	(1.08)
Homeroom teachers' characteristics					
age - 22 (t + 1)	15.17	15.09	-0.10	0.74	4.88
			(0.88)	(0.97)	(6.37)
Years of experience					
in the prefecture where City X (t + 1)	10.93	11.15	0.20	0.78	7.17
			(0.78)	(0.82)	(7.51)
in the current schools (t + 1)	3.47	3.79	0.31	0.40	11.52
			(0.21)	(0.25)	(7.16)

The unit of observation is classes. Column (4) shows the differences between Columns (1) and (2) after controlling for school, year and grade fixed effects and class characteristics at period t. We use the girl ratio, the ratio of students with financial supports, class size variables, grade size, and homeroom teachers' characteristics at period t for the adjustment. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.1: Robustness Checks

	Elementary School				Middle School			
	Boy		Girl		Boy		Girl	
	(1) Math	(2) Lang.	(3) Math	(4) Lang.	(5) Math	(6) Lang.	(7) Math	(8) Lang.
A. Table 5								
Closed	-0.027	0.164	-0.039	0.135	-0.181	-0.233	-0.080	0.050
	(0.185)	(0.187)	(0.174)	(0.157)	(0.352)	(0.336)	(0.303)	(0.236)
Closed × Low-income	-0.841***	-0.576*	-0.252	-0.193	0.321	-0.172	-0.285	-0.152
	(0.299)	(0.295)	(0.287)	(0.278)	(0.394)	(0.463)	(0.367)	(0.325)
$\hat{\beta}_{Closed} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income}}$	-0.868	-0.412	-0.290	-0.059	0.140	-0.405	-0.365	-0.102
p-value	0.008	0.190	0.305	0.831	0.712	0.365	0.359	0.748
B. Robustness check: with school-year FEs								
Closed	0.010	0.099	-0.086	0.125	0.226	-0.114	0.175	0.055
	(0.176)	(0.181)	(0.165)	(0.150)	(0.384)	(0.357)	(0.334)	(0.268)
Closed × Low-income	-0.704**	-0.566*	-0.208	-0.175	0.406	-0.130	-0.226	-0.149
	(0.295)	(0.296)	(0.283)	(0.276)	(0.387)	(0.457)	(0.346)	(0.331)
$\hat{\beta}_{Closed} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income}}$	-0.694	-0.467	-0.295	-0.049	0.633	-0.244	-0.051	-0.094
p-value	0.027	0.123	0.272	0.852	0.134	0.606	0.907	0.785
C. Robustness check: household income status								
Closed	-0.102	0.149	-0.038	0.093	-0.019	-0.175	-0.106	0.029
	(0.185)	(0.188)	(0.171)	(0.154)	(0.351)	(0.342)	(0.298)	(0.244)
Closed × Low-income in all the three years	-0.679**	-0.618*	-0.298	-0.068	-0.182	-0.391	-0.241	-0.106
	(0.318)	(0.323)	(0.314)	(0.282)	(0.420)	(0.527)	(0.360)	(0.392)
$\hat{\beta}_{Closed} + \hat{\beta}_{\substack{Closed \\ \times Low\text{-}income}}$	-0.781	-0.469	-0.336	0.026	-0.201	-0.566	-0.347	-0.076
p-value	0.027	0.166	0.283	0.929	0.606	0.244	0.392	0.830

The dependent variables are the standardized test scores in the school year following the class closures, with a mean of 50 and a standard deviation of 10. Standard errors robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.1: Summary Statistics For Usage of time Variables Among Elementary School Students

	(1) mean	sd	p5	p25	p50	p75	p95
A. Elementary School Boys							
Studying on a weekday	0.94	0.24	0	1	1	1	1
Study more than 30 minutes on a weekday	0.68	0.47	0	0	1	1	1
Study more than 60 minutes on a weekday	0.39	0.49	0	0	0	1	1
Studying on a holiday	0.80	0.40	0	1	1	1	1
Study more than 30 minutes on a holiday	0.52	0.50	0	0	1	1	1
Study more than 60 minutes on a holiday	0.29	0.45	0	0	0	1	1
Time for TV and video games	196.89	115.04	30	120	180	270	420
Time for sleeping	536.77	69.75	390	510	540	570	630
A. Elementary School Girls							
Studying on a weekday	0.96	0.20	1	1	1	1	1
Study more than 30 minutes on a weekday	0.72	0.45	0	0	1	1	1
Study more than 60 minutes on a weekday	0.43	0.50	0	0	0	1	1
Studying on a holiday	0.84	0.36	0	1	1	1	1
Study more than 30 minutes on a holiday	0.55	0.50	0	0	1	1	1
Study more than 60 minutes on a holiday	0.30	0.46	0	0	0	1	1
Time for TV and video games	144.63	98.59	0	60	120	210	330
Time for sleeping	546.18	62.72	450	510	570	570	630

Table D.1: Heterogeneous Effects on Mathematics Scores By School Resources Among Elementary School Students

	(1)	(2)	(3)	(4)	(5)
Closed	-0.035	-0.009	-0.058	0.556	0.508
	(0.153)	(0.197)	(0.239)	(0.462)	(0.487)
$\times$ Class size ≤ 30		-0.083			0.196
		(0.301)			(0.478)
$\times$ Age -22				-0.078	-0.081
				(0.051)	(0.051)
$\times$ Years of experience in the prefecture where City X locates				0.069	0.071
				(0.057)	(0.058)
$\times$ Years of experience in current school				-0.058	-0.059
				(0.097)	(0.097)
Closed $\times$ Low-income	-0.546***	-0.833***	-0.645**	-0.945*	-1.193**
	(0.209)	(0.265)	(0.269)	(0.547)	(0.584)
$\times$ Class size ≤ 30		0.775*			0.781
		(0.426)			(0.551)
$\times$ Age				-0.072*	-0.090**
				(0.037)	(0.038)
$\times$ Years of experience in the prefecture where City X locates				0.052	0.069
				(0.045)	(0.049)
$\times$ Years of experience in current school				0.230**	0.235**
				(0.111)	(0.111)
N	69838	69838	40752	40752	40752

Dependent variables are standardized test scores in the following school year of class closures, with a mean of 50 and a standard deviation of 10. Since data on teachers' characteristics are available only up to 2016, the sample size for the last three columns is smaller than that for the first two columns, in which data are available from 2015 to 2017. Standard errors that are robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.2: Heterogeneous Effects on Language Arts Scores By School Resources Among Elementary School Students

	(1)	(2)	(3)	(4)	(5)
Closed	0.139	0.149	-0.087	0.331	0.289
	(0.147)	(0.194)	(0.219)	(0.443)	(0.448)
× Class size ≤ 30		-0.024			0.299
		(0.284)			(0.429)
× Age - 22				-0.062	-0.070
				(0.047)	(0.048)
× Years of experience in the prefecture where City X locates				0.058	0.064
				(0.052)	(0.053)
× Years of experience in current school				-0.036	-0.038
				(0.101)	(0.101)
Closed × Low-income	-0.371*	-0.474*	-0.184	0.479	0.498
	(0.211)	(0.275)	(0.299)	(0.662)	(0.670)
× Class size ≤ 30		0.255			-0.372
		(0.423)			(0.646)
× Age				-0.066	-0.050
				(0.047)	(0.052)
× Years of experience in the prefecture where City X locates				0.022	0.010
				(0.046)	(0.050)
× Years of experience in current school				0.004	0.005
				(0.127)	(0.128)
N	69790	69790	40717	40717	40717

Dependent variables are standardized test scores in the following school year of class closures, with a mean of 50 and a standard deviation of 10. Since data on teachers' characteristics are available only up to 2016, the sample size for the last three columns is smaller than that for the first two columns, in which data are available from 2015 to 2017. Standard errors that are robust against class-level clustering are in parenthesis. All models control the cubic function of the standardized scores in the year of the class closure, financial support status dummies, female dummy, class size and its squared term, size of grade in school and its squared term, school FEs, grade in school FEs, and year FEs. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3: Effects Calculated By Substituting Percentile values for School Resources Variables Among Economically-disadvantaged Students

	(1) Math	(2) Language arts
Age - 22		
25%tile (=7)	-0.231	0.437
	(0.477)	(0.489)
50%tile (=12)	-1.083**	-0.163
	(0.451)	(0.448)
75%tile (=22)	-2.786***	-1.363*
	(0.743)	(0.803)
Years of experience in the prefecture where City X locates		
25%tile (=4)	-1.641***	-0.457
	(0.573)	(0.558)
50%tile (=8)	-1.083**	-0.163
	(0.451)	(0.448)
75%tile (=17)	0.174	0.499
	(0.598)	(0.654)
Years of experience in current school		
25%tile (=2)	-1.259**	-0.130
	(0.504)	(0.476)
50%tile (=3)	-1.083**	-0.163
	(0.451)	(0.448)
75%tile (=5)	-0.730	-0.229
	(0.446)	(0.552)

Standard errors robust against class-level clustering are in parenthesis. Inference: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.