WINPEC Working Paper Series No.E2104
April 2021

Estimating Sibling Spillover Effects in Academic Performance: First Evidence from Japan

Shuhei Kaneko and Haruko Noguchi

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

Estimating Sibling Spillover Effects in Academic Performance: First Evidence from Japan

Shuhei Kaneko* Haruko Noguchi†

Abstract

Sibling interaction is one of the most important and frequent communications for children. Using administrative data collected by a municipality in the Hokkaido prefecture, Japan, we find a positive and significant spillover effect in academic performance both from older to younger siblings and from younger to older siblings. Our heterogeneity analysis shows that sibling spillovers vary by family backgrounds. The spillovers from older to younger siblings are amplified among same-gender siblings, siblings 2–3 years apart, and lower-income families. On the other hand, the spillovers from younger to older siblings are significantly observed only among brother pairs; the wider the age differences, the stronger the effects; and spillover effects among poorer families are negatively estimated. Finally, we show that younger siblings in lower-income households are more influenced by their bottom-achieving older siblings, eliciting the importance of family-based interventions aimed at better quality sibling interaction at the family level.

JEL Classification: I20; I24; J24
Keywords: Sibling Spillover; Social Interaction; Human Capital; Education; Japan

*[Corresponding] Graduate School of Economics, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku, Tokyo 169-8050 Japan; Email Address: shuhei7700@ruri.waseda.jp, Phone Number: +81 (3)5272-4782

†Faculty of Political Science and Economics, Waseda University, 1-6-1 Nishi-Waseda, Shinjuku, Tokyo 169-8050 Japan.

1 Introduction

Siblings share a common background and frequently interact with each other: they are raised by the same parents in the same home, often attend the same school, and spend time together. Therefore, the effects that siblings have on each other are transmitted through many possible pathways. Siblings might model and influence both good and bad behaviors amongst themselves (Black et al. 2017). Alternatively, within a family, time or money reserved for one child might be transferred to another child, or it might be contested. Consequently, the impact of these investments could be amplified or diluted depending on the way of sibling interaction. Examining these spillovers is important for policy evaluation: policies that target one child in a family could be amplified or offset depending on the direction of sibling spillover. Furthermore, "sibling spillovers may have implications for inequality in outcomes of children from different backgrounds- for example, if, high-income children are more likely to benefit from the transmission of good behaviors than low-income children" (Nicoletti & Rabe 2019 (NR hereinafter); p.482). This study aims to provide the first evidence of sibling spillover from Japan, where the students attain the highest grades in mathematics and science among OECD countries (according to the scores of Programme for International Student Assessment)[1]

Economic literature has investigated the spillover (or peer) effect within a classroom (see, *inter alia*, Lavy et al. 2012; Abdulkadiroğlu et al. 2014; Gibbons & Telhaj 2016; Booij et al. 2017). Although the qualitative results of these studies are mixed, many of them agree that the effects received from school peers, if any, are limited. Moreover, we should keep in mind that estimating peer effect is complicated owing to the reasons mentioned in Manski (1993): endogenous effect, contextual effects, and correlated effects.

[1] The source is `https://www.nier.go.jp/kokusai/pisa/pdf/2018/01_point-eng.pdf` (accessed April 16, 2021)

Angrist (2014) reviews studies that have investigated the peer effects. He points out some potential issues that confuse the true causal peer effects and correlation within a group.

Despite the large volume of studies on school peer effects, sibling interaction has drawn surprisingly little attention. Nonetheless, sibling *correlations* are widely found in the literature; sibling correlations are strong and much larger than neighborhood correlations (Mazumder 2008, Lindahl 2011, or Nicoletti & Rabe 2013 among others). However, these studies do not successfully quantify how much of these correlations are derived by the interactions among siblings. As mentioned, the observed sibling correlations should include family background; therefore, not all of them can be simply attributed as spillover. Only recently have economic studies begun to examine the causal impacts of sibling interaction(Joensen & Nielsen 2018; Qureshi 2018; Bingley et al. 2019; Aguirre & Matta 2021 among others). These studies have generally found the positive sibling spillover effects from older to younger siblings. While these studies have advantages in terms of clear identification strategy along with natural experiments, it is generally very difficult to determine the appropriate policy changes or randomized situation suitable for research questions.

We contribute to the literature by estimating how children's school achievement transmits to their younger and older siblings. Unlike previous studies, we estimate the spillover effects in a way that does not rely on policy reforms or natural experimental situations. To separate sibling association from the correlated observed and unobserved factors as much as possible, we utilize an identification strategy developed by NR to the administrative data collected by Date city (a municipality in the Hokkaido prefecture, Japan)[2]. The data include the test scores in multiple subjects (i.e., Japanese and mathematics)

[2] A brief overview of Date city is available at `https://www.city.date.hokkaido.jp/` (In Japanese. Accessed: April 16, 2021), and we briefly discuss the demographic characteristics in Section 3.2.

and some demographic information (i.e., gender, needs for public assistance, name of household head (identifier of sibling pairs)) of all students aged 7 to 14 attending the public school located in this city. In NR, they regressed a younger sibling's test score at 16 years old on the same test score of older siblings using within-pupil between-subject estimation. The powerful advantage of this estimation method is that we can cancel out the children's unobserved fixed effect (FE), which does not vary by subjects[3].

Following this identification strategy, we regress a younger sibling's test score at a specific year on the lagged test score of older siblings and find a positive and significant spillover from older to younger siblings. Furthermore, as a novelty over NR, we also estimate the spillover effects from younger to older siblings: spillover in the opposite direction is also positive and significantly estimated. For spillover from older to younger siblings, the effects are strongly amplified among same gender siblings, siblings 2–3 years apart, and lower-income families who need public assistance from the municipality. On the other hand, while we detect positive spillovers from younger to older siblings, the heterogeneity analyses reveal that the pattern of spillover is arguably different: a significant spillover is only observed among brother pairs; the wider the age differences, the stronger the spillover effects; and spillover effects among lower-income households is negatively estimated while the coefficient is not significant.

To the best of our knowledge, this is the first study that directly estimates the bi-directional sibling spillover and its heterogeneity in academic performance. Specifically, it is noteworthy that not only have the spillover effects among households never been

[3]To further account for subject-specific inputs by family (mostly from the parental subject-specific time and/or pecuniary investment), they employ the two-stage least squares (2SLS) approach. Owing to data limitation, we cannot apply the 2SLS to our data, as discussed in Section 2.2 in detail. However, please note that, also as in the original NR, they use Ordinary Least Squares (OLS) estimation as a preferable model because they could not reject the null hypothesis in the robust Hausman test. This would imply that the endogeneity issue caused by subject-specific *family* inputs is not serious.

investigated by using Japanese data but also no past studies have used administrative educational data where the sibships are identifiable in Japan. While the results on older to younger sibling spillovers are quite similar to those obtained by NR, we newly find that there are notably different patterns in the transmission of academic performance from younger to older siblings.

The remainder of the paper is organized as follows. Section 2 explains the identification strategy of our estimation. Section 3 introduces the institutional background and compares some demographic characteristics of Date city, Hokkaido, and overall Japan to check the external validity of our results; then, we report summary statistics of the data. We explain the empirical results and conduct robustness tests in Section 4, and discuss the implications for the policy in Section 5. Finally, Section 6 concludes the paper.

2 Identification Strategy

2.1 Model

Based on the identification strategy exploited by NR, we consider the following value-added equation for the test scores of younger siblings.

$$Y_{1,isqt} = \alpha_1 + \rho_1 Y_{1,isqt-1} + \beta_1 Y_{2,is'qt-1} + \gamma_{1,F}\mathbf{I^F_{1,it}} + \gamma_{1,S}\mathbf{I^S_{1,ist}} + \gamma_{1,X}\mathbf{X_{1,it}} + \mu_{sqt} + \mu_{1,it} + \epsilon_{1,isqt}. \quad (1)$$

where

- $Y_{1,isqt}$ is the test score of the younger sibling of sibling-pair i, in school s, subject q in year t,
- $Y_{1,isqt-1}$ is the test score of the younger sibling of sibling-pair i, in school s, subject

q in a lagged year $t-1$,

- $Y_{2,is'qt-1}$ is the test score of the older siblings, who might have belonged to a different school s' in year $t-1$[4].

- $\mathbf{I^F_{1,it}}$ is a vector of family inputs for the younger child of sibling pair i in year t.

- $\mathbf{I^S_{1,ist}}$ is a vector of school s's input at period t for younger sibling in household i, which are not subject-specific.

- $\mathbf{X_{1,it}}$ is a row vector of other child, household, and school characteristics that are not captured by $\mathbf{I^F_{1,it}}$ or $\mathbf{I^S_{1,ist}}$.

- μ_{sqt} is unobservable characteristics varying by school, subjects, and year. (e.g., School-specific curriculum for math).

- $\mu_{1,it}$ is an individual FE (not captured by the above factors) affecting the test score in year t, allowing this FE to vary across year.

- $\epsilon_{1,isqt}$ is an idiosyncratic error term with mean 0.

Similarly, we also consider the spillover from *younger* to *older* sibling:

$$Y_{2,is'qt} = \alpha_2 + \rho_2 Y_{2,is'qt-1} + \beta_2 Y_{1,isqt-1} + \mathbf{I^F_{2,it}}\gamma_{2,F} + \mathbf{I^S_{2,is't}}\gamma_{2,S} + \gamma_{2,X}\mathbf{X_{2,it}} + \mu_{s'qt} + \mu_{2,it} + \epsilon_{2,is'qt}. \tag{2}$$

In (1) and (2), our primal objective is to consistently estimate the coefficient β_1 and β_2; however, this task is challenging mainly owing to two issues: (i) correlation problem and (ii) reverse causality. In our case, the second issue is not serious because we use the

[4]Because the school they attend is assigned based on the address, the siblings attend the same school in most cases. However, if the older sibling attends junior high school and younger sibling attends elementary school in a given year, we treat them as enrolled in different schools

lagged independent variable, which implies there should not be the reverse causality from $Y_{1,isqt}$ to $Y_{2,is'qt-1}$, and from $Y_{2,is'qt}$ to $Y_{1,isqt-1}$.

On the other hand, because siblings share many common characteristics (e.g., environments in households, socioeconomic status, or school characteristics), it is difficult to avoid the first correlation issue. To mitigate this, by obtaining the test scores in multiple subjects for each observation, we can cancel out the unobserved characteristics that do not vary by subjects.

We transform the dependent variable into

$$DevY_{1,isqt} \equiv Y_{1,isqt} - \sum_{k=Japanese,Math} Y_{1,iskt}/2,$$

and we apply the same transformation for Equation (2). After we also apply the analogous transformation to the right-hand side, we obtain [5]

$$\begin{aligned} DevY_{1,isqt} &= \rho_1 DevY_{1,isqt-1} + \beta_1 DevY_{2,is'qt-1} + Dev\mu_{sqt} + Dev\epsilon_{1,isqt} \\ DevY_{2,is'qt} &= \rho_2 DevY_{2,is'qt-1} + \beta_2 DevY_{1,isqt-1} + Dev\mu_{s'qt} + Dev\epsilon_{2,is'qt}. \end{aligned} \tag{3}$$

As suggested above, the advantage of this transformation is that the unobserved heterogeneity does not complicate the estimates unless it depends on subject-specific factors. Because much of the heterogeneity is not highly likely to be subject-specific (i.e., health, cognitive and non-cognitive endowments of children, or the household status), this strategy should be powerful in this context.

However, we still might face the issue of how we deal with the unobserved heterogeneity in subject-specific inputs in year t. In particular, we might be concerned about

[5] To avoid the complication, we only discuss the spillover from older to younger sibling in the remainder of this section.

subject-specific *school* inputs. Because older and younger siblings attend the same school in almost all the cases, the school's subject-specific inputs that are opaque to us should cause bias to our estimates. Regarding this issue, we partial out the shared subject-specific school characteristics by adding School-Year-Subject FE in Equation (3).

We impose an assumption that, after controlling for the School-Year-Subject FE, there is only negligible (or none at all) endogeneity between $DevY_{2,is'qt-1}$ and the residual in Equation(3): unobserved subject-specific *family* investments in year t, if any, do not influence the results of our analysis. Although this assumption seems to be restrictive, we should note that such family characteristics up to $t-1$ are largely wiped out because we control for the younger siblings' lagged test score in Equation (3). Furthermore, the main variable in our study, $DevY_{2,is'qt-1}$, is already determined at the time of parental subject-specific investment between years $t-1$ and t, which is a different point from NR. Therefore, parental subject-specific unobserved investment for *younger* siblings at t does not influence the *older* siblings' lagged test score (i.e., the variable in interests).

2.2 Difference from NR

While we basically follow the empirical strategy adopted by NR, a major difference is the time gap between the dependent variable and the lagged independent variables. Using the data for annually conducted exam scores, we take one year lag of independent variables (excluding School-Year-Subject FE) in (3). On the other hand, in NR, they estimated the sibling spillover by regressing the younger siblings' test scores at age 16 (the time of secondary school graduation (denoted as $DevY_{1,isq16}$ in NR)) on the same test score of older siblings ($DevY_{2,is'q16}$), while controlling for the younger siblings' test score at age 11 (the time of the transition from primary to secondary school) as a lagged score (they denote it as $DevY_{1,isq11}$). Therefore, they had to deal with the issue of unobserved subject-specific family inputs for *five* years (between ages 11 and 16). Moreover, in their data structure and sample restriction[6], $DevY_{2,is'q16}$ is not pre-determined with regard to the residuals ($Dev\epsilon_{1,isq16}$) because they use sibling pairs at most 3 years gap: the subject-specific family characteristics can affect the older sibling's exam score at 16 years old. Therefore, in NR, the endogeneity between the older sibling's test score at age 16 and unobserved subject-specific family inputs could be more apparent than ours. To fully tackle this issue, they adopted instrumental variable estimation where the subject-specific test scores of the older sibling at age 16 are instrumented by the predetermined average attainment of the school-by-cohort peers of the older sibling. They used the students encountered by the older sibling for the first time in a secondary school as peers to avoid the reflection issue in the first stage regression.

Ideally, we would employ the same strategy for the endogeneity issue. However, in Date city, because the reshuffling of students is not conducted every year in many schools and the number of students per grade in each school is not large, sometimes no classmates are newly encountered. Therefore, the strategy of using the newly encountered peers to construct an instrument was not an option in this study. Despite this, it should be noteworthy that the exogeneity of older siblings' test scores was not statistically rejected in NR; therefore, NR interpreted the estimation without the instrumental variable as a preferred one and they used it for their subsequent analysis. Considering that $DevY_{2,is'qt-1}$ in (3) would be less likely to be affected by the *family level unobserved subject-specific* inputs than in the case of NR because it is predetermined with regard to the residual, this result would ensure, to some extent, that our main variable in interest is not empirically endogenous after washing out the children and school-year-subject FE.

[6] See Section 3.5 of NR for the details.

3 Background and Data

3.1 Institutional Background (Standardized Examinations)

Date City conducts standardized tests (hereafter, Original Test) to measure children's academic ability every December, in addition to the national-level academic achievement test (NAAT) (*Zenkoku Gakuryoku Gakushu Jokyo Chousa* in Japanese) led by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) in every April. While NAAT is offered to the sixth and ninth graders (i.e., the highest grade in elementary and junior high schools, respectively), Original Tests are offered to all students from the 1st to the 8th grade enrolled in public schools in Date City.

Only the abilities in Japanese and mathematics are measured by NAAT. However, a science exam was newly added in 2012, and it has been conducted once every three years; English was added in 2019 and it is also going to be conducted once every three years.[7].[8] The Japanese and mathematics exams consist of two parts, A and B. The former mainly asks basic questions, while the latter mainly asks questions that require application skills to practical issues.

As we discuss in detail in the following section, all students enrolled in the public elementary and junior high schools in Date City take the Japanese and mathematics exams in the Original Test, and the science exam is added when the children reach the 4th grade. The exams are comprised of (i) basic questions (*Kiso* in Japanese) and (ii) advanced questions (*Katsuyou* in Japanese), whose problem structure is very close to that of NAAT. In Table A1 of the Appendix, we show detailed results of the Original Test by

[7] Please see `https://www.mext.go.jp/a_menu/shotou/gakuryoku-chousa/zenkoku/1344101.htm` (in Japanese) for more information on NAAT. NAAT is very similar to Key Stage 2 (taken at 11 years old) and Key Stage 4 (taken at 16 years old) analyzed by NR. Unfortunately, we are not eligible to access the individual data of NAAT

[8] The results of NAAT for students in Date City are available at `https://www.city.date.hokkaido.jp/kyoiku/detail/00000882.html` (in Japanese, accessed April 16, 2021)

year and contents.

3.2 Comparison between Date City, Hokkaido, and Overall Japan

We compare some demographic variables of Date City, Hokkaido, and overall Japan to check the external validity of our estimates. Panel A of Table 1 shows the following information: (i) number of children under 15 years old, (ii) gender ratio of children under 15 years old, (iii) total number of households, (iv) share of intact families (two parents and child(ren)), (v) share of single-parent families, which is sourced from the latest Census conducted in 2015[9]. In addition, we sourced the information on public assistance Level 1 and Level 2 (defined in Section 3.4) from the survey conducted in 2015 by MEXT[10] (Panel B of Table 1).

From Panel A, we observe that the gender ratio in Date city is slightly higher than that of Hokkaido and overall Japan, and the share of intact households (single-parent households) is slightly smaller (larger) than that of overall Japan. Panel B shows that, while the share of children with public assistance needs in Hokkaido is remarkably higher than that of Date city and overall Japan, the shares in Date city are very similar to that of overall Japan.

3.3 Data Description

The longitudinal data used in this study were constructed from administrative data created by the Education Committee of Date City. This city is a relatively small municipality with a population of approximately 35,000 and an area of approximately 450 square kilo-

[9] Available at `https://www.e-stat.go.jp/stat-search/files?page=1&toukei=00200521&tstat=000001049104` (in Japanese, accessed April 16, 2021)

[10] Available at `https://www.mext.go.jp/component/a_menu/education/detail/__icsFiles/afieldfile/2018/02/02/1632483_17_1.pdf` (in Japanese, accessed April 16, 2021)

meters in southwestern Hokkaido. The data in our study cover four years (i.e., 2012, 2014, 2015, and 2016) of all students enrolled in public primary (1st—6th grade) and junior high (7th-–9th grade) schools, which comprise approximately 7–8% of the total population of the city.[11]. The test scores for Japanese and mathematics subjects of all students are available. The test scores for science are available only of students in the fourth grade or above.

The data of test scores were merged with the students' family information (obtained from the school register), including zip-code, the name of the family head, need for public assistance, and the students' status of mental/intellectual disability. While we cannot directly identify the sibling relationship only from this information, the name of the family head can be used as a matching variable for siblings; the names of the family head are represented by Chinese characters (*Kanji*) and we did not detect any duplicates in their names. To the best of our knowledge, there have been no past studies conducted in Japan that used administrative educational data where the sibships are identifiable.

We limited the analytic sample using the following procedure. First, children with disabilities were excluded because many of them are enrolled in special needs schools (i.e., the contents learned are different from those learned by other students) and their test scores are often missing. Second, children without any siblings were excluded because they are not of interest in this study. We also excluded, from the data, siblings in the same academic year (mostly twins or triplets) because we cannot generally identify the older and younger siblings from the data set. When we had multiple pairs of siblings from one family, we only employed the two oldest students to avoid any multiplier spillover effects: this strategy is similar to that in NR.

[11] Similar type of data in Japan is also used in, say, Oikawa et al. (2020), Bessho et al. (2019) among others

3.4 Descriptive Characteristics of Our Data

Table 2 shows the mean and standard deviation of key variables and demographic variables used in the heterogeneity analysis.

Panel A of Table 2 represents the raw test scores of each subject for younger and older siblings. We observe that the average score for Japanese is slightly higher than that for the other two subjects. In the following analyses, we use the standardized scores so that they have 0 mean and 1 standard deviation[12] The deviation variables, $DevY$, are constructed by taking the deviation from the averaged standardized test scores. For instance, suppose that student A's standardized test scores for each subject is $\{Japanese, Math\} = \{0.4, -0.1\}$. Then, the average of these three outcomes is 0.15; therefore, $\{DevY_{jpn}, DevY_{math}\} = \{0.25 - 0.25\}$ are obtained.

An important demographic variable is a proxy for household income level. In Japan, there are two levels of assistance for poor families to alleviate the burden of educational expenses[13]. First, the most urgent group is the families that are eligible to receive public assistance from the government; hereinafter, we call this group "Level 1." The second classification is the families that are poor, but whose economic situation is not too severe to receive governmental public assistance. These families obtain a certain level of support for children's school activities (e.g., lunch, commuting, extracurricular activity, medical expenses, and expenses for any items necessary for school attendance) from each municipality's support plan. Hereinafter, this group is called "Level 2." We provide the detailed contents of supports provided by Date City in Appendix A.2. As shown in Table

[12]The test scores are separately standardized by year, grade, and subjects.

[13]You can find a detailed explanation from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan at `https://www.mext.go.jp/a_menu/shotou/career/05010502/017.htm` (In Japanese. Accessed: April 16, 2021) The Level 1 group is called "Yo-Hogo" (which means assistance ("Hogo") is necessary (Yo)) and the Level 2 group is called "Jun Yo-Hogo" ("Jun" in Japanese corresponds to the prefix "semi-" in English)

2, the families classified in Level 1 are few (0.7%) in this sample, whereas the Level 2 families account for 13.5% of all families in these data. We implemented the subsample analysis by combining these categories.

As for the pattern of the siblings' gender combination, we found that the pair of older sister and younger brother is slightly more common than other groups, while the four subgroups generally occur equally. The average age difference between the older and younger siblings is approximately 2.7 years and ranges from 1 to 7 years.

Table 3 shows the descriptive statistics of test scores divided by year for the older and younger siblings. We see that the average and standard deviation of the test scores taken by younger and older siblings are quite similar in all years, which implies that the annual exams children in Date city take are highly comparable in terms of difficulty or quality.

4 Results

4.1 Baseline Result

As a first analysis, we examined the role of controlling for the children's FE by showing the scatter plot of raw correlation of $Y_{1,isqt}$ and $Y_{2,is'qt-1}$, and the scatter of $DevY_{1,isqt}$ and $DevY_{2,is'qt-1}$.

In Panel A of Figure 1, while we find the positive and sufficiently significant correlation[14] between sibling test scores in both cases, the difference in their magnitudes (0.369 versus 0.124) indicates the naive OLS without considering that $\mu_{1,i}$ leads to an excessive overestimation of the spillover effect. As the siblings have similar or same (family level) characteristics in common, we can easily assume that unobserved heterogeneity of a younger sibling $\mu_{1,i}$ positively correlates with the older sibling's performance. In

[14] Pairwise correlation for the analytic sample of the main analysis was computed.

Panel B, we also examine the sibling spillover from younger to older siblings. While the magnitudes (0.324 versus 0.063) are smaller than those observed in Panel A, we find the tendency is highly close. The regression results of Equation (3) are reported in Table 4. The result in column (1) implies that the increase in an older sibling's lagged test score by 1 standard deviation (SD) leads to approximately 11.3% of an SD increase in the younger sibling's test score when controlling for the younger sibling's lagged test score. After adding School-Year-Subject FE, the estimated coefficient slightly decreases to 10.1 % of an SD. In columns (3) and (4) of Table 4, we conduct the same estimation for the spillover from younger to older siblings. After controlling for the older sibling's lagged test score and the set of FE, while the magnitude is approximately 0.6 times of impacts from older to younger siblings, we obtain the significant spillover effects: 1 SD increase in younger sibling's test score is linked to 5.8 % of an SD increase in older sibling's test score in the next year.

In all columns, we find that the coefficient ρ has a highly similar magnitude regardless of the age of the children: younger siblings and older siblings have similar persistence in cognitive abilities.

4.2 Robustness

4.2.1 Reverse Regression

We begin our robustness tests by swapping the role of younger and older siblings' test scores in Equation (3), that is, we consider estimating the following equations.

$$DevY_{2,is'qt-1} = \tilde{\rho}_1 DevY_{2,is'qt-2} + \tilde{\beta}_1 DevY_{1,isqt} + Dev\mu_{sqt} + Dev\epsilon_{2,is'qt} \tag{4}$$

$$DevY_{1,isqt-1} = \tilde{\rho}_2 DevY_{1,isqt-2} + \tilde{\beta}_2 DevY_{2,is'qt} + Dev\mu_{s'qt} + Dev\epsilon_{1,isqt}, \tag{5}$$

where we estimate the significance of coefficients $\tilde{\beta_1}, \tilde{\beta_2}$. If our baseline results indicate a correlation originated from the unobserved subject-specific family inputs, not a causal effect, we should obtain a similar result in these regressions. Moreover, because the lagged variables are regressed on the variables at t, we should not obtain the significant effect. In columns (2) and (4) of Table 5, we find that the positive and significant spillover effect disappears in this specification conditional on controlling for the subject-specific school inputs by year. In column (1), we obtain the marginally insignificant coefficient (p-value = 0.104) with a magnitude of 4.7%; this result would reflect the possibility that both siblings share many unobserved school inputs and leaving them uncontrolled would cause bias to our estimates.

4.2.2 Permutation Test

To further check the robustness of our estimates, we perform permutation tests with a random sampling of the observations under the following setting. To alleviate the concern that the sibling spillover effects are confounded with the family characteristics, we assign an unrelated child who shares the same characteristics (gender, enrolled school, age, and need for public assistance) to be the older sibling; then, we estimate the placebo spillover effects. If the residual endogeneity caused by unobserved subject-specific family inputs is largely entangled with specific family characteristics, our placebo estimates would have a magnitude close to our baseline results. We run this procedure 1,000 times and summarize the results in panel A of Figure 2, where the histogram of 1000 point estimates is reported. A dashed black lines represent the 95 percentile of the distribution, and the red line shows the point estimate we obtained from our baseline analysis (in Table 4). The figure suggests that the true sibling spillover is almost unlikely to be derived by the similarity in the family backgrounds: the effect is larger than the 95 percentile of the placebo estimates. This result corroborates that uncontrolled subject-specific family inputs do not seem to seriously cause bias to our baseline estimates.

Also, we conducted the same analysis for spillovers from younger to older siblings (see panel (B) of Figure 2). The point estimate of our baseline analysis is larger than 95 percentile of the distribution, implying again that spillover effects would not confound with the family backgrounds. In Appendix B, we report the results of an additional permutation test where we randomly sample a classmate and then assign him/her to be the sibling. The results suggest that the true sibling spillover is not likely to be derived from the classmate (i.e., the remaining school level endogeneity).

4.2.3 Exclusion of Data in 2012

As described in Section 3.3, our data are composed of students' exam scores in 2012, 2014, 2015, and 2016. Since our specification assumes that the lags between t and $t-1$ are constant over time, dropping data collected in 2012 should not largely change our results. To confirm this, we conducted the same analysis excluding the data in 2012. Table 6 shows the results: while the point estimates vary to some extent, all the newly estimated sibling spillovers are within 1 standard error from the baseline results and the significance of spillover is maintained. Whether we exclude the partial data does not change the qualitative results of our analysis.

We also conducted the reverse regression for the 2014–2016 restricted data: the outcome variables in 2015 are regressed on siblings' test score in 2016, children's own test score in 2014, and the set of School-Year-Subject FE. The results shown in Table 7 imply that the siblings' future test scores do not significantly affect the outcome variables, which corroborates the validity of our main specification.

4.2.4 Further Threats

Further, we conducted two checks to verify the reliability of our estimation.

First, our identification strategy relies on the assumption that the coefficient β in Equation (3) is identical across all the subjects: Japanese and mathematics. To test this, we run the following regression and conduct F-tests with the null hypothesis: $\delta_1 = \delta_2$.

$$Y_{1,isqt} = \alpha + \delta_1 Y_{2,isJt-1} + \delta_2 Y_{2,isMt-1} + \omega Y_{1,isqt-1} + \mu_{1,it} + \mu_{sqt} + e_{1,isqt}, \tag{6}$$

where the definition of each variable is identical to that in Equation (3) and the subscripts J, M represent Japanese and Math, respectively. We also estimate the same model for spillover in the opposite direction.[15] We also test the equivalence of the persistence parameter (ρ) across subjects by allowing for different ω by subjects. We test $\omega_1 = \omega_2$ in the following equation:

$$Y_{1,isqt} = \alpha + \delta_1 Y_{2,isJt-1} + \delta_2 Y_{2,isMt-1} + \omega_1 Y_{1,isJt-1} + \omega_2 Y_{1,isMt-1} + \mu_{1,it} + \mu_{sqt} + e_{1,isqt}, \tag{7}$$

In Table 8, we report the estimation results where we allow for the varying spillover effects across subjects. In columns (1) and (2), coefficients corresponding to each subject (δ_1 and δ_2) are very similar and we do not reject the equality of the two coefficients along with the F-test whose p-value is approximately 0.88. After allowing for the different persistent parameters, we find that the spillover and persistence parameters are very close in all subjects, which strongly supports our assumption. We also find that the spillover effects

[15] In estimating (7), we also run the regression where we do not control for the children's own lagged test score to avoid the possible complication of "Nickell Bias" (Nickell 1981). The results' magnitudes of coefficients are similar and we do not reject all the F-tests under this specification.

are almost the same across subjects for younger to older siblings in columns (4) to (6) of Table 8. Overall, these results support the representation in Equation (1) or (2), where the parameters β and ρ are assumed to be identical across different subjects to obtain Equation (3) after taking the deviation.

Second, our concern is that the unit of a cluster, sibling-pair level, might be insufficient in that it might overlook the possible correlation of the error term within a specific group. Our baseline estimates assume that observable characteristics (included as $Dev\epsilon_{.,isqt}$) of children are independent of each other after controlling for all the FEs we discussed. However, for instance, if a test for a specific grade in a specific year is anomalously difficult, this might provide a biased estimate of standard errors because the students taking the same test score would be correlated in some way. As a further test, we compute the standard errors clustered at the school level and school-year level. Since the resulting number of clusters is few, the unadjusted results could underestimate the standard errors; therefore, we apply the wild bootstrapping method with 10000 times replications (Angrist & Pischke 2008 Cameron et al. 2008) and Webb (2013) type weight. The results are shown in Table 9, where we see that the qualitative results of spillover effects (both directions) are robust to the changes of clustering units.

4.3 Heterogeneity

We conduct heterogeneity analyses by interacting some characteristics with $DevY_{1,isqt-1}$ or $DevY_{2,is'qt-1}$ in Equation (3):

$$\begin{aligned} DevY_{1,isqt} &= \rho_1 DevY_{1,isqt-1} + \beta_{1,Z} * DevY_{2,is'qt-1} * Z_{2,iqt} + Dev\mu_{sqt} + Dev\epsilon_{1,isqt} \\ DevY_{2,is'qt} &= \rho_2 DevY_{2,is'qt-1} + \beta_{2,Z} * DevY_{1,isqt-1} * Z_{1,iqt} + Dev\mu_{s'qt} + Dev\epsilon_{2,is'qt}. \end{aligned} \tag{8}$$

where $(Z_{1,iqt}, Z_{2,iqt})$ is a characteristic that possibly varies by the household, subjects, and year. We report the estimation results of the parameter $\beta_{1,Z}$ and $\beta_{2,Z}$ in the following discussion.

4.3.1 Gender Combination

Intuitively, we may expect that the effects are stronger among same-gender siblings. Dunn & Kendrick (1979) found that same-gender sibling pairs show a higher percentage of positive interaction (e.g., laughing, smiling, joint physical play)[16]. However, at the same time, Abramovitch et al. (1980) suggested that the level of interaction is large even among mixed-gender siblings. Hence, the literature on development psychology acknowledges that the gender composition itself is not an important factor while significant levels of sibling interaction are observed.

As for spillover from older to younger siblings, the effects are stronger among same-gender siblings: The effects are 17.7 % of an SD for brother pairs and 18.2 % of an SD for sister pairs, and these effects are much larger than those of mixed-gender siblings (See the top panel of Table 10). While we find that the spillover from younger to older siblings among brother pairs is also significant, the spillovers among sister pairs are arguably smaller (5.0 % of an SD) and insignificant.

4.3.2 Age Spacing

The second panel of Table 10 suggests that the spillover from older to younger siblings is nonlinear in the degree of age spacing: the magnitude of effects is increasing until 3 years of difference and then the effects become insignificant if the age difference is 4 years or

[16]We should note that their study subjects are very young siblings. "The ages of the first child when the second child was born ranged from 18 to 43 months," (p.144, Dunn and Kendrick 1979) and the follow-up survey was conducted at the time when the younger sibling became 14 months old.

larger. Borrowing the discussion in the past literature in psychology or sociology (Minnett et al. 1983 among others), we could interpret this result as a trade-off between the close relationship and conflict among siblings. While the sibling relationship of siblings with 1 year age gap is intensive, this causes the conflict at the same time; therefore, these two factors might offset each other.

On the other hand, the patterns of spillover in opposite direction (i.e., younger to older siblings) are largely different: we found that the wider the age difference, the greater the effects. The effect among siblings with 4 years or larger age differences is 9.7 % of an SD, which is approximately 8 times bigger than that among siblings with 1 year age gap.

4.3.3 Needs for Public Assistance (Income Level)

In panel (C) of Table 10, we investigate the heterogeneous impacts depending on the needs for public assistance. While the point estimate is marginally insignificant, the spillover effect from older to younger siblings is slightly larger among the low-income households: 1 SD increase in older sibling's test score leads to 12.7% of SD increase in younger sibling's performance in low-income households, while this effect is 9.5% among other households. This result would indicate that sibling interaction plays a more important role if the household is comparatively poor. Because poor households should be less likely to have educational resources for children (e.g., expenditures for cram schools and extra-curricular activities), the effects from older siblings could be more important inputs in these households.

The magnitude of sibling spillover from younger to older siblings is very similar but these have different signs: insignificant -12.1% of an SD for lower-income households, and significant 9.0% of an SD for other households. Since the point estimate is not statistically significant, we should be careful to interpret the results; however, the negative and larger coefficients for lower-income households would imply that the good performance of the younger sibling might stress out the older siblings through, say, sense of inferiority.

4.3.4 Sibling's Lagged Test Score

Finally, we consider the heterogeneity based on the performance of lagged siblings' test score: $Y_{1,isqt-1}$ or $Y_{2,is'qt-1}$. We simply split the sample based on the percentile scores. In our setting, the students are classified into lower half if their score is less than 50 percentile and upper if their score is beyond 50 percentile. For instance, assume that the older sibling's (standardized) test scores at $t-1$ are 0.5 for Japanese and -0.3 for mathematics, and the medians of both tests are zero. Then, in Equation (8), $Z_{2,iJt}$ is classified into "Upper" half, and $Z_{2,iMt}$ is classified into "Lower" half. Hereinafter, we interchangeably use the words "Upper half" and "Top-achieving", and "Lower half" and "Bottom-achieving".

Panel D of Table 10 shows that the spillover from top-achieving older sibling to younger is about half as high as that from a bottom-achieving older sibling (the estimated magnitude is 6.3% versus 12.1%); spillover from a top-achieving younger sibling is nearly twice as high as that from a bottom-achieving younger sibling (the estimated magnitude is 8.5 % versus 4.4%). Our results suggest that the older siblings can play a more encouraging role if their performance is not great; however, to put it another way, it can be interpreted that younger siblings are more susceptible to the deterioration of the non-top achieving older siblings' grades. On the other hand, older siblings are more likely to be inspired by the younger sibling's achievement if the younger sibling is a good-performing student.

5 Discussion

One of the major concerns of policymakers in the educational sector is how to reduce the educational gap between affluent and disadvantaged households. Affluent households should be able to invest in their children's education by pecuniary and non-pecuniary aspects, while the disadvantaged households may be constrained in providing education for their children. Many policies intending to provide educational opportunities to children from disadvantaged households have been formulated (e.g., the Head Start policy). The literature has evaluated the effectiveness of these policies (see, *inter alia*, Currie & Thomas 1995; Currie & Thomas 1999). Our above analysis supplements the studies on these aforementioned educational policies by examining the (positive) externality through sibling spillover effects.

In the bottom panel of Table 10, we estimated the spillover effects across older (and younger) sibling's attainment (upper half and lower half). We additionally split the sample based on the household income level, and then, estimated spillover effects for each subgroup in Table 11.

The results imply that younger siblings in higher-income households are likely to benefit from their top-achieving older siblings than the younger siblings in low-income households (where the effects are insignificant), and less likely to be affected by bottom-achieving older siblings. Based on these estimates, we discuss the effectiveness of school-level educational intervention focusing on less achieving students from low-income households.

What if the intervention could increase the test score of bottom-achieving students in lower-income families by 10 % of an SD? This virtual policy is similar in spirit to the Head Start policy in the United States or the Pupil Premium policy in the United

Kingdom[17] among other similar policies targeting low-income households, and it can be interpreted as remedial education (analyzed in Bessho et al. 2019 for instance). The results suggest that this policy can efficiently reduce the attainment gap through large spillover from older to younger siblings: this intervention would increase the younger siblings' performance by 1.60% (0.160*10) of an SD.

However, it is noteworthy that this intervention would not have positive spillover to older siblings if it focuses on younger siblings: panel B of Table 11 suggests that the increase in the performance of bottom-achieving younger sibling in poor families by 1 SD would lead to 15.1 % (marginally insignificant) decrease of older sibling's performance.

We should also consider the possibility of family-based interventions, especially for the spillover from older to younger siblings. In our study, the average performance of children in low-income families is arguably lower than that of children in higher-income families as many studies have found. Also, as we documented above, the younger siblings in low-income families are affected by spillover from bottom-achieving students rather than the top-achieving ones. Taken together, the attainment gap between these two groups could be widened by the quality of communication that siblings have in their families.

These suggest that the policymakers should also be worried that the poor quality of sibling interaction among low-income families could lead to a larger attainment gap. Since the school-based interaction cannot fully fix this issue, it would be worthwhile to consider interventions directed to individual families. A promising fact is that the past literature mainly in psychology implies the validity of family-based interventions. For instance, Brotman, Dawson-McClure, et al. (2005) evaluated a family-based intervention (conducted in New York area)[18] aimed at preventing conduct problems of pre-school

[17] See `https://www.gov.uk/government/publications/pupil-premium/pupil-premium` (Accessed April 16, 2021) for the general information of the policy

[18] The detailed description of the prevention program is documented by Brotman, Gouley, et al. (2005).

siblings of delinquent youths. This study finds that the antisocial behavior (e.g., parent-rated defiant behavior, physical aggression) of untargeted older siblings was decreased 8 months after the intervention[19]. The authors inferred that improved parenting skills for a specific (targeted) child were likely to overflow to other children. The result of this study implies the possibility that sibling interaction could be ameliorated through better parenting. Curiously, the policy implications of our study are arguably different from NR: they find that the magnitude of the spillover does not vary by older sibling's attainment between poor and affluent background families, letting them conclude that the policymakers do not need to be anxious about the quality of sibling interaction too much.

6 Conclusion

In this paper, we apply the identification strategy proposed by NR to the unique data collected by a municipality in Japan, and provide supportive evidence on the sibling spillover. Furthermore, not only the effect from the older siblings, which have been examined by some past studies, we also investigate the sibling spillover from younger to older siblings. An increase in the test score of an older sibling leads to 10.1% of an SD in a younger sibling's test score, and the increase in younger sibling's test score by 1 SD induces 5.8 % of an SD increase in older sibling's test score in the next year. The heterogeneity analysis implies that the spillover effect varies depending on the family backgrounds. For spillovers from older to younger siblings, the impact is larger among same-gender siblings, siblings with 2–3 years gap, and low-income families. Conversely,

[19]Journal of Family Psychology, Vol 19 (4) is a special issue focusing on sibling interaction and its effects on family welfare. The title is "Special Issue: Sibling Relationship Contributions to Individual and Family Well-Being."

we found that the spillovers from younger to older siblings are significant only among brother pairs, and these spillover effects become stronger as the age difference becomes wider. Surprisingly, spillover effects among lower-income families are negatively estimated while the coefficient is not significant.

After interacting the sibling's lagged test scores (i.e., above or below 50 percentile) and household income (i.e., with or without need for public assistance), we found that the younger siblings in lower-income households are more likely to be affected by the bottom-achieving older siblings, suggesting that quality of sibling interaction might be different between wealthy and poor households. This would indicate that the attainment gap between low and high-income families can be partly attributed to the difference in the quality of sibling interaction.

Combining the above results, the policy implication from our study is twofold. First, the school-based intervention targeting a certain set of children would have non-trivial positive externalities through the sibling spillover. Besides, more importantly, the family-based intervention for poor households would also be effective in improving the attainment gap, especially through the spillover from older to younger siblings.

Despite the novel findings, we should note that there are some limitations to our study. First, the sample size of our study is not large because we rely on data collected from one municipality (e.g., the National Pupil Database used by NR contains over 230,000 sibling pairs). Second, while we check that the demographic information in Date city is not very different from that in overall Japan: especially, the share of poor families in need of public assistance is highly comparable with that of the total population of Japan (see Panel B of Table 1), it is not completely clear if our results are universal in other countries. In fact, our results have some qualitative difference from those of NR despite that we generally follow their empirical strategy. This clearly reflects that the sibling spillover and its heterogeneity can vary depending on the cultural, economic, or religious backgrounds.

While it is beyond the scope of this study to answer the question "Why does the policy implication vary depending on the context?", this study would provide important implications on educational policies or interventions. Especially, our findings based on Japanese students, who are comparatively successful in terms of academic skills among developed countries, should be universally valuable. Furthermore, the methodology in this study can be applied to a very wide range of children's outcomes. For example, we would also be able to estimate the effects on non-cognitive skills, which recently have gathered close attention by researchers not limited to economics. Estimating the precise spillover effects in both these skills by a sophisticated econometric strategy will help researchers to synthesize the findings in economics and other related fields of social science.

Declarations of Interest

None

Acknowledgement

This work was supported by a Grant-in-Aid for JSPS Research Fellow [Shuhei Kaneko: grant number 20J20017]; the MEXT KAKENHI Grant-in-Aid for Scientific Research (A) (General) project "Creation and implementation of a scientific basis for children's human capital: Development of policy evaluation process by government-academic collaboration" [PI: Haruko Noguchi: grant number 19H00602]. This study benefits from the comments and suggestions by the seminar participants of the Tokyo Labor Economics Conference in 2020. The authors have declared that no conflicting interests exist. All remaining errors are ours.

References

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at boston and new york exam schools. *Econometrica*, *82*(1), 137–196.

Abramovitch, R., Corter, C., & Pepler, D. J. (1980). Observations of mixed-sex sibling dyads. *Child Development*, 1268–1271.

Aguirre, J., & Matta, J. (2021). Walking in your footsteps: Sibling spillovers in higher education choices. *Economics of Education Review*, *80*, 102062. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0272775720305483` doi: https://doi.org/10.1016/j.econedurev.2020.102062

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion.* Princeton University Press.

Bessho, S., Noguchi, H., Kawamura, A., Tanaka, R., & Ushijima, K. (2019). Evaluating remedial education in elementary schools: Administrative data from a municipality in japan. *Japan and the World Economy*, *50*, 36–46.

Bingley, P., Lundborg, P., & Lyk-Jensen, S. V. (2019). Brothers in arms: spillovers from a draft lottery. *Journal of Human Resources*, 0317–8646R3.

Black, S. E., Breining, S., Figlio, D. N., Guryan, J., Karbownik, K., Nielsen, H. S., … Simonsen, M. (2017). *Sibling spillovers* (National Bureau of Economic Research (NBER) Working Paper 23062.) National Bureau of Economic Research.

Booij, A. S., Leuven, E., & Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies*, *84*(2), 547–578.

Brotman, L. M., Dawson-McClure, S., Gouley, K. K., McGuire, K., Burraston, B., & Bank, L. (2005). Older siblings benefit from a family-based preventive intervention for preschoolers at risk for conduct problems. *Journal of Family Psychology*, *19*(4), 581.

Brotman, L. M., Gouley, K. K., Chesir-Teran, D., Dennis, T., Klein, R. G., & Shrout, P. (2005). Prevention for preschoolers at high risk for conduct problems: Immediate outcomes on parenting practices and child social competence. *Journal of Clinical Child and Adolescent Psychology*, *34*(4), 724–734.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, *90*(3), 414–427.

Currie, J., & Thomas, D. (1995). Does head start make a difference? *American Economic Review*, *85*(3), 341–364.

Currie, J., & Thomas, D. (1999). Does head start help hispanic children? *Journal of public Economics*, *74*(2), 235–262.

Dunn, J., & Kendrick, C. (1979). Interaction between young siblings in the context of family relationships. In *The child and its family* (pp. 143–168). Springer.

Gibbons, S., & Telhaj, S. (2016). Peer effects: Evidence from secondary school transition in england. *Oxford Bulletin of Economics and Statistics*, *78*(4), 548–575.

Joensen, J. S., & Nielsen, H. S. (2018). Spillovers in education choice. *Journal of Public Economics*, *157*, 158–183.

Lavy, V., Silva, O., & Weinhardt, F. (2012). The good, the bad, and the average: Evidence on ability peer effects in schools. *Journal of Labor Economics*, *30*(2), 367–414.

Lindahl, L. (2011). A comparison of family and neighborhood effects on grades, test scores, educational attainment and income—evidence from sweden. *The Journal of Economic Inequality*, *9*(2), 207–226.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, *60*(3), 531–542.

Mazumder, B. (2008). Sibling similarities and economic inequality in the us. *Journal of Population Economics*, *21*(3), 685–701.

Minnett, A. M., Vandell, D. L., & Santrock, J. W. (1983). The effects of sibling status on sibling interaction: Influence of birth order, age spacing, sex of child, and sex of sibling. *Child Development*, 1064–1072.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society*, 1417–1426.

Nicoletti, C., & Rabe, B. (2013). Inequality in pupils' test scores: How much do family, sibling type and neighbourhood matter? *Economica*, *80*(318), 197–218.

Nicoletti, C., & Rabe, B. (2019). Sibling spillover effects in school achievement. *Journal of Applied Econometrics*, *34*(4), 482–501.

Oikawa, M., Tanaka, R., Bessho, S., & Noguchi, H. (2020). *Do class size reductions protect students from infectious disease? lessons for covid-19 policy from flu epidemic in tokyo metropolitan area* (IZA Discussion Paper No. 13432).

Qureshi, J. A. (2018). Additional returns to investing in girls' education: Impact on younger sibling human capital. *The Economic Journal*, *128*(616), 3285–3319.

Webb, M. D. (2013). *Reworking wild bootstrap based inference for clustered errors* (Queen's Economics Department Working Paper No. 1315). Kingston (Ontario). Retrieved from `http://hdl.handle.net/10419/97480`

Tables and Figures

Table 1: Comparison of Characteristics

	Date City	Hokkaido	Overall Japan
Panel A			
Number of Children under 15	3,966	608,296	15,886,810
Number of Children under 15 (Male)	2,056	310,387	8,133,536
Number of Children under 15 (Female)	1,910	297,909	7,753,274
Gender Ratio of Children under 15 (Male/Female)	1.076	1.042	1.049
Total number of households	14,953	2,438,206	53,331,797
Share of Intact households	0.222	0.227	0.268
Share of Single-parent households	0.095	0.093	0.089
Panel B			
Number (Public Assistance Level 1)	22	13,358	136,798
Share (Public Assistance Level 1)	0.009	0.035	0.014
Number (Public Assistance Level 2)	351	69,230	1,329,336
Share (Public Assistance Level 2)	0.141	0.181	0.138
Share (Level 1 + Level 2)	0.150	0.216	0.152

Panel (A) Notes: We source the data from Census in 2015 for the variables in panel (A).
Panel (B) Notes: We refer to the survey conducted in 2015 by MEXT for the information on Hokkaido and overall Japan. Since this survey does not contain the municipality-level information, we summarize the same information from the school register of Date City in 2015.

Table 2: Summary Statistics

	Mean	Std. Dev.	Min	Max	Obs.
A: Test Score (Unadjusted)					
Test Score (Japanese, Youger Sib.)	73.134	19.253	0	100	1928
Test Score (Math, Youger Sib.)	68.853	20.748	0	100	1928
Test Score (Japanese, Older Sib.)	72.249	17.713	0	100	1935
Test Score (Math, Older Sib.)	67.390	20.869	0	100	1936
B: Demographic Variables					
Older Bro, Younger Bro	0.260	0.439	0	1	1931
Older Bro, Younger Sis	0.239	0.426	0	1	1931
Older Sis, Younger Bro	0.283	0.450	0	1	1931
Older Sis, Younger Sis	0.218	0.413	0	1	1931
Age Difference	2.712	1.239	1	7	1931
Public Assistance (Level 1)	0.007	0.083	0	1	1436
Public Assistance (Level 2)	0.135	0.342	0	1	1436

Notes:We used longitudinal data constructed from the administrative records of academic test scores and the students registry created by Date City(in 2012, 2014, 2015, 2016). The descriptive table is over the analytic sample for the estimation of Equation (3). The number of observation of Public Assistance is smaller because this information is only available since 2014.

Table 3: Exam Score by Years

	Younger Sibling					Older Sibling				
	Mean	Std. Dev.	Min	Max	Observation	Mean	Std. Dev.	Min	Max	Observation
Japanese										
In 2012	70.292	19.836	0.000	100.000	494	69.641	18.400	3.571	100.000	496
In 2014	74.053	19.723	10.000	100.000	475	72.167	17.942	12.903	100.000	475
In 2015	73.799	19.186	6.897	100.000	480	73.532	16.925	6.452	100.000	481
In 2016	74.485	17.955	9.259	100.000	479	73.731	17.279	0.000	100.000	483
Mathematics										
In 2012	66.518	21.868	0.000	100.000	494	65.998	21.586	0.000	100.000	494
In 2014	72.403	19.831	5.263	100.000	475	69.288	20.529	5.263	100.000	476
In 2015	68.885	19.853	0.000	100.000	480	67.396	20.480	8.824	100.000	482
In 2016	67.707	20.923	0.000	100.000	479	66.936	20.766	0.000	100.000	484

Notes: We used longitudinal data constructed from the administrative records of academic test scores and the students registry created by Date City(in 2012, 2014, 2015, 2016). The descriptive table is over the analytic sample for the estimation of Equation (3).

Table 4: Baseline Results

	Older to Younger		Younger to Older	
	(1)	(2)	(3)	(4)
β_1, β_2 (Spillover)	0.113**	0.101**	0.042	0.058^{+}
	(0.033)	(0.033)	(0.030)	(0.031)
ρ_1, ρ_2 (Lagged Score)	0.371**	0.362**	0.376**	0.385**
	(0.033)	(0.035)	(0.034)	(0.034)
School-Year-Subject FE	No	Yes	No	Yes
Adjusted R^2	0.151	0.155	0.131	0.154
Number of Clusters (Sibling Pairs)	600	600	600	600
Total Observations	1948	1948	1950	1950

Notes: Robust standard errors clustered at sibling level are reported in parentheses. We used longitudinal data (2012, 2014, 2015 and 2016) constructed from the administrative records of academic test scores and the students registry created by Date City. In columns(1) and (2), the dependent variable is younger siblings' attainment at year t. In column (3) and (4), spillover in inverse direction (from lagged younger siblings' test score to older siblings' attainment) is estimated: the dependent variable is older siblings' attainment at year t. $^{+}: p < 0.1$, *: $p < 0.05$, **: $p < 0.01$.

Table 5: Reverse Regression

Dependent Variables	(1) $DevY_{2,isq,t-1}$	(2) $DevY_{2,isq,t-1}$	(3) $DevY_{1,isq,t-1}$	(4) $DevY_{1,isq,t-1}$
$DevY_{1,isqt}$	0.047	0.034		
	(0.029)	(0.029)		
$DevY_{2,isqt-2}$	0.270**	0.264**		
	(0.037)	(0.038)		
$DevY_{2,isqt}$			0.032	0.045
			(0.047)	(0.048)
$DevY_{1,isqt-2}$			0.381**	0.375**
			(0.042)	(0.045)
School-Year-Subject FE	No	Yes	No	Yes
Adjusted R^2	0.085	0.094	0.151	0.187
Number of Clusters (Sibling Pairs)	530	530	351	351
Total Observations	1612	1612	868	868

Notes: Robust standard errors clustered at sibling level are reported in parentheses.We used longitudinal data (2012, 2014, 2015 and 2016) constructed from the administrative records of academic test scores and the students registry created by Date City. In columns(1) and (2), the dependent variable is lagged test score of older siblings; In columns (3) and (4), the dependent variable is lagged test score of younger siblings. $^{+}: p < 0.1$, *: $p < 0.05$, **: $p < 0.01$.

Table 6: Results Excluding 2012 Data

	Older to Younger		Younger to Older	
	(1)	(2)	(3)	(4)
β_1, β_2 (Spillover)	0.093*	0.079$^+$	0.058$^+$	0.080*
	(0.040)	(0.041)	(0.035)	(0.035)
ρ_1, ρ_2 (Lagged Score)	0.375**	0.359**	0.409**	0.416**
	(0.038)	(0.040)	(0.040)	(0.041)
School-Year-Subject FE	No	Yes	No	Yes
Adjusted R^2	0.151	0.151	0.146	0.160
Number of Clusters (Sibling Pairs)	479	479	478	478
Total Observations	1434	1434	1436	1436

Notes: Robust standard errors clustered at sibling level are reported in parentheses.We used longitudinal data (2014, 2015 and 2016) constructed from the administrative records of academic test scores and the students registry created by Date City. In columns(1) and (2), the dependent variable is younger siblings' attainment at year t. In column (3) and (4), spillover in inverse direction (from lagged younger siblings' test score to older siblings' attainment) is estimated: the dependent variable is older siblings' attainment at year t. $^+$: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$.

Table 7: Reverse Regression Excluding 2012 Data

	(1)	(2)	(3)	(4)
Dependent Variables	$DevY_{2,isq,2015}$	$DevY_{2,isq,2015}$	$DevY_{1,isq,2015}$	$DevY_{1,isq,2015}$
$DevY_{1,isq2016}$	0.057	0.037		
	(0.042)	(0.041)		
$DevY_{2,isq2014}$	0.361**	0.354**		
	(0.050)	(0.050)		
$DevY_{2,isq2016}$			0.045	0.052
			(0.058)	(0.061)
$DevY_{1,isq2014}$			0.359**	0.342**
			(0.053)	(0.058)
School-Year-Subject FE	No	Yes	No	Yes
Adjusted R^2	0.135	0.155	0.140	0.122
Number of Clusters (Sibling Pairs)	420	420	276	276
Total Observations	840	840	552	552

Notes: Robust standard errors clustered at sibling level are reported in parentheses. We used longitudinal data (2014, 2015, and 2016) constructed from the administrative records of academic test scores and the students registry created by Date City. In columns(1) and (2), the dependent variable is lagged test score of older siblings; In columns (3) and (4), the dependent variable is lagged test score of younger siblings. $^+$: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$.

Table 8: Model Allowing for the Different Spillover Effects across Subjects

	Older to Younger			Younger to Older		
	(1)	(2)	(3)	(4)	(5)	(6)
δ_1	0.111**	0.103**	0.102**	0.042	0.059$^+$	0.058$^+$
	(0.037)	(0.037)	(0.038)	(0.032)	(0.032)	(0.033)
δ_2	0.115**	0.099**	0.100**	0.043	0.058$^+$	0.061$^+$
	(0.034)	(0.034)	(0.035)	(0.034)	(0.034)	(0.035)
ω	0.371**	0.363**		0.376**	0.385**	
	(0.033)	(0.034)		(0.034)	(0.034)	
ω_1			0.365**			0.391**
			(0.036)			(0.037)
ω_2			0.360**			0.380**
			(0.039)			(0.036)
$\mathbf{1}(Test = Math)$		-0.087	-0.086		0.191$^+$	0.194$^+$
		(0.067)	(0.068)		(0.115)	(0.116)
School-Year-Subject FE	No	Yes	Yes	No	Yes	Yes
F test ($\delta_1 = \delta_2$)	0.024	0.022	0.004	0.005	0.002	0.010
p-value ($\delta_1 = \delta_2$)	0.878	0.882	0.953	0.944	0.966	0.920
F test ($\omega_1 = \omega_2$)			0.035			0.138
p-value ($\omega_1 = \omega_2$)			0.851			0.710
Number of Clusters (Sibling Pairs)	600	600	600	600	600	600
Adjusted R^2	0.754	0.755	0.755	0.723	0.730	0.730
Within R^2	0.152	0.185	0.185	0.131	0.189	0.189
N	1948	1948	1948	1950	1950	1950

Notes: Robust standard errors clustered at sibling level are reported in parentheses. We used longitudinal data (2012, 2014, 2015 and 2016) constructed from the administrative records of academic test scores and the students registry created by Date City. In columns(1), (2), and (3), the dependent variable is younger siblings' attainment at year t. In column (4), (5) and (6), spillover in inverse direction (from lagged younger siblings' test score to older siblings' attainment) is estimated: the dependent variable is older siblings' attainment at year t. We test the equality of δ_1 and δ_2 in Equation (7). $^+$: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$.

Table 9: Different Level of Clustering

	Older to Younger		Younger to Older	
Level of Clusters	(1): School	(2): School-Year	(3): School	(4): School-Year
β_1, β_2 (Spillover)	0.101		0.058	
Standard Error	(0.017)	(0.035)	(0.035)	(0.028)
Unadjusted p-value	[0.000]**	[0.007]**	[0.124]	[0.047]*
Wild Bootstrap p-value	[0.027]*	[0.017]*	[0.109]	[0.051]$^{+}$
School-Year-Subject FE	Yes	Yes	Yes	Yes
Number of Clusters	13	34	13	39
Total Observations	1948		1950	

Notes: Robust standard errors clustered at school level (school-year level) are reported in parentheses of columns (1) and (3) (columns (2) and (4)). We report unadjusted and wild-bootstrap p-value in brackets of each column. $^{+}: p < 0.1$, *: $p < 0.05$, **: $p < 0.01$.

Table 10: Heterogeneity Analysis

	A: Gender Combination			
	BB^a	BS^a	SB^a	SS^a
Older to Younger	0.177**	0.051	0.017	0.182**
	(0.062)	(0.071)	(0.064)	(0.063)
Younger to Older	0.135*	-0.024	0.013	0.050
	(0.056)	(0.073)	(0.055)	(0.066)
	B:Age Difference			
	1 year	2 years	3 years	4+ years
Older to Younger	0.051	0.133**	0.170*	-0.009
	(0.093)	(0.047)	(0.070)	(0.076)
Younger to Older	0.012	0.040	0.082	0.097^+
	(0.080)	(0.054)	(0.062)	(0.056)
	C: Public Assistance			
	No	Yes		
Older to Younger	0.095**	0.127		
	(0.036)	(0.078)		
Younger to Older	0.090**	-0.121		
	(0.031)	(0.096)		
	D: Sibling's Test Score at t-1			
	Upper Half	Lower Half		
Older to Younger	0.063	0.121**		
	(0.048)	(0.039)		
Younger to Older	0.085^+	0.044		
	(0.044)	(0.036)		

Notes: Robust standard errors clustered at sibling-pair level are reported in parentheses. We used longitudinal data (2012, 2014, 2015 and 2016) constructed from the administrative records of academic test scores and the students registry created by Date City. In each panel, the main explanatory variable is interacted with household characteristics. $^+$: $p < 0.1$, *: $p < 0.05$, **: $p < 0.01$.

a: BB = Older Brother and Younger Brother, BS = Older Brother and Younger Sister, SB = Older Sister and Younger Brother, SS = Older Sister and Younger Sister

Table 11: Spillover: Income Level and Sibling's Achievement Interacted

	Public Assistance	
	No	Yes
A: Older to Younger		
Older Sibling (Upper Half)	0.069	0.030
	(0.052)	(0.124)
Older Sibling (Lower Half)	0.109*	0.160^{+}
	(0.043)	(0.083)
B: Younger to Older		
Younger Sibling (Upper Half)	0.099*	-0.034
	(0.044)	(0.159)
Younger Sibling (Lower Half)	0.085*	-0.151
	(0.037)	(0.098)

Notes: Robust standard errors clustered at sibling level are reported in parentheses. We used longitudinal data constructed from the administrative records of academic test scores and the students registry created by Date City. Each panel reports the regression results with household income level × sibling attainment interaction terms. $^{+}p < 0.1$, **: $p < 0.05$, ***: $p < 0.01$

Figure 1: Simple Scatter Plot of Sibling Spillovers

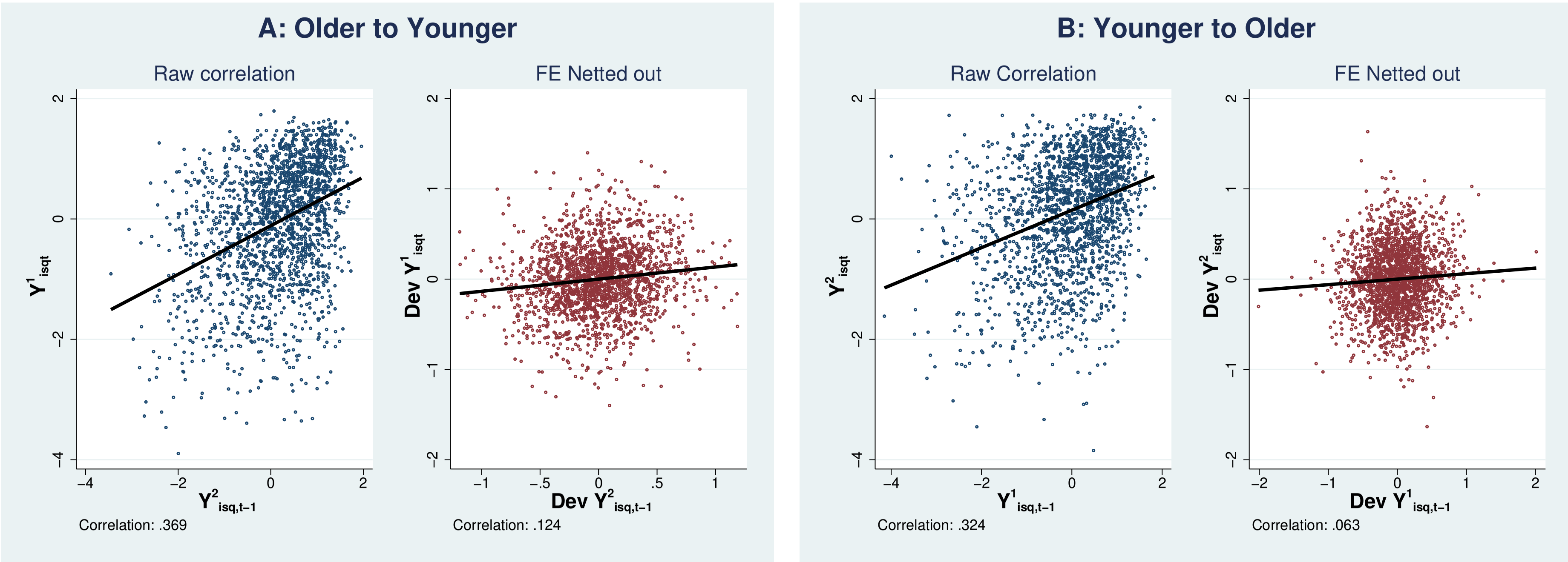
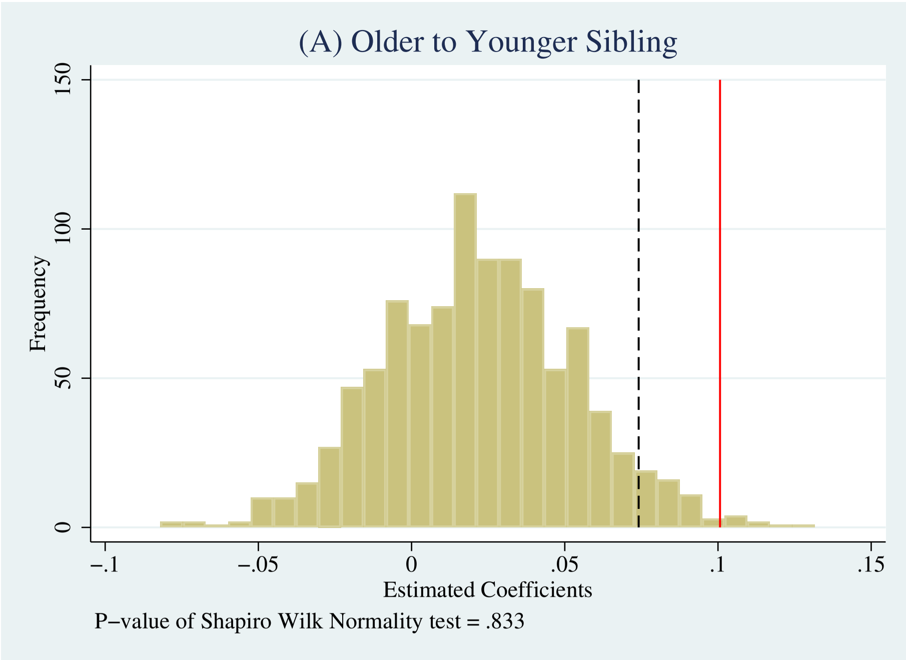
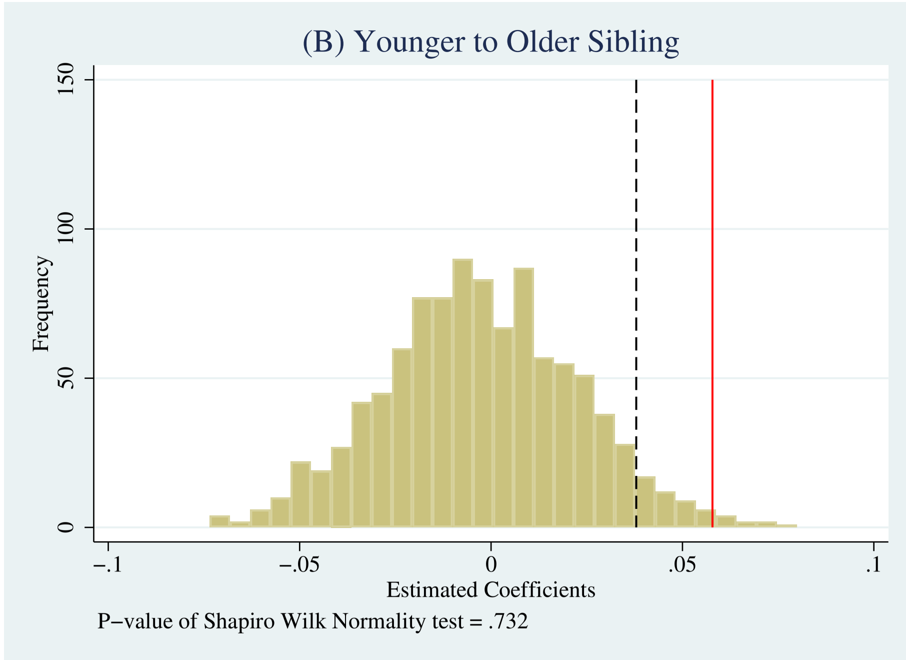


Figure 2: Permutation Test





Permutation Test Note: In panel A, we assign an unrelated child who shares the same characteristics (gender, enrolled school, age, and needs for public assistance) to be the older sibling then estimate the placebo spillover effects for 1000 times; in panel B, we conducted the same assignment for younger sibling to estimate the placebo spillover effect. The dashed black lines represent the 95 and 99 percentiles of the distribution, and the red line shows the point estimate we obtained from our baseline analysis (in Table 4).

Appendix (Not for publication)

A Data Appendix

A.1 Detailed Descriptive Statistics of Test Scores

Table A1: Detailed Descriptive Statistics of (Unadjusted) Test Scores

	2012			2014			2015			2016		
	Mean	SD	Obs.	Mean	SD	Obs.	Mean	SD	Obs.	Mean	SD	Obs.
Younger sibling												
Japanese total	70.29	19.84	494	74.05	19.72	475	73.80	19.19	480	74.49	17.96	479
Japanese basic	74.30	19.64	494	78.04	19.67	475	77.74	18.97	480	78.37	17.97	479
Japanese advanced	51.90	30.02	494	55.71	30.28	475	55.85	29.14	480	56.72	28.16	479
Math total	66.52	21.87	494	72.40	19.83	475	68.88	19.85	480	67.71	20.92	479
Math basic	69.94	21.46	494	76.33	19.29	475	73.10	19.97	480	71.72	20.88	479
Math advanced	48.33	31.42	494	52.59	31.11	475	49.32	29.13	480	49.21	29.32	479
Older sibling												
Japanese total	69.64	18.40	496	72.17	17.94	475	73.53	16.92	481	73.73	17.28	483
Japanese basic	73.01	18.33	496	76.09	17.78	475	77.52	16.83	481	76.72	17.70	483
Japanese advanced	54.70	27.23	496	55.02	27.29	475	56.18	26.10	481	59.99	25.59	483
Math total	66.00	21.59	494	69.29	20.15	476	67.40	20.48	482	66.94	20.77	484
Math basic	68.35	21.45	494	71.58	20.15	476	69.93	20.17	482	69.41	20.62	484
Math advanced	53.46	30.40	494	57.05	30.21	476	54.97	28.97	482	55.33	28.97	484

Notes: The longitudinal data constructed by the administrative record of academic test scores and the students registry (in 2012, 2014, 2015, 2016). Please note that these descriptive statistics are over all children (including single child without any siblings.)

A.2 Public Assistance Provided by Date City

In addition to the public assistance provided by the Japanese government, Date City (as many other municipalities do) originally provides financial assistance to help families with financial difficulties pay for part of their elementary and junior high school expenses. The following Table A2 summarize the contents of financial supports, and Table A3 shows the eligibility criteria with some model cases. The income criteria vary depending on the age structure of the households. Financial supports are discontinued when (i) the household moves out from Date city, (ii) the household is excluded from the eligibility due to the changes in household structure, and (iii) a parent or guardian declines to apply for schooling assistance. All the information is available at the website of Date City Board of Education (`https://www.city.date.hokkaido.jp/kyoiku/detail/00000492.html`) in Japanese.

Table A2: Financial Assistance in Date City

Items	Description
I: For Level 1 & 2 students	
School trip	Actual expense; including accomodation, transportation, travel insurance, commemoration photo etc.
Medical Expense	Copayment for diseases specified in Article 8 of School Health and Safety Act is waived: • Trachoma and conjunctivitis • Tinea, scabies and impetigo • Otitis media • Chronic sinusitis and adenoids • Tooth decay • Parasitic diseases (including ovine carriage)
II: For Level 2 students[a]	
School supplies	11,420 JPY for students in elementary school; 22,320 JPY for students in junior high school
School supplies (for newly enrolled students)	40,600 JPY for 1st grade; 47,400 JPY for 7th grade
Sport equipment	Payment in kind (e.g., Skis)
Transportation	The cost of public transportation from home to school.
Extracurricular activities	Actual expenses; educational field trip, club activities etc.
PTA membership	Actual expenses with cap (Elementary school; 3,380 JPY, Junior high school: 4,190 JPY)
Student council	Actual expenses with cap (Elementary school; 3,380 JPY, Junior high school: 5,450 JPY)
School lunch	Actual expenses

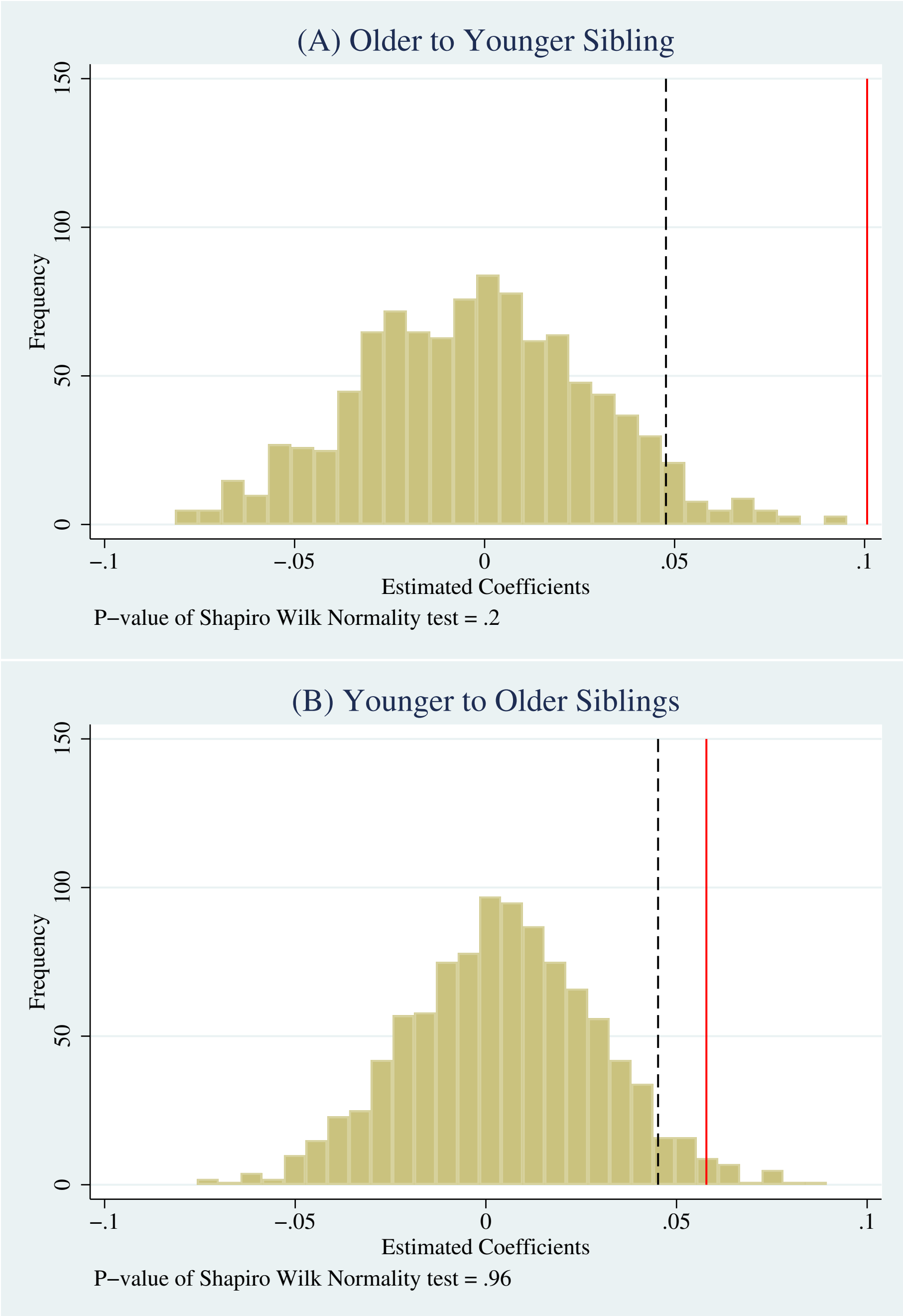
Notes: The source is `https://www.city.date.hokkaido.jp/hotnews/files/00000400/00000492/20181227115014.pdf` (in Japanese, accessed April 16, 2021)
a: For students classified into Level 1, the items listed in panel II are supported by the governmental supports

Table A3: Eligibility Criteria with Model Case

Household size	Amount of annual income criteria (JPY)	Model case of household composition (Age in parentheses)
2	1,867,819	Father or Mother (35), Child (10)
3	2,377,018	Father (35), Mother (35), and Child (10)
4	2,948,212	Father(35), Mother (30), Child (13), and Child (10)
5	3,323,501	Father(35), Mother (30), Child (13), Child (10), and Child (7)
6	3,705,140	Father(35), Mother (30), Child (13), Child (10), Child (7), and Child (6)

B Additional Permutation Test

Figure B1: Additional Permutation Test: A classmate is assigned to be a sibling



Notes: In panel (A), we randomly assign a classmate of an older sibling child to be the older sibling then estimate the placebo spillover effects for 1000 times; in panel (B), we conducted the same assignment for a younger sibling to estimate the placebo spillover effect. The dashed black line represents the 95 percentile of the distribution, and the red line shows the point estimate we obtained from our baseline analysis (in Table 4).