When the Household Becomes the School: Siblings, Parental Attention, and School Closures

Francisco Pardo*

University of Texas at Austin

Click here for the latest version

Abstract

This paper examines how the presence of siblings affects the production of education when instructional responsibilities move from schools to households. First, I document a new global stylized fact: children with siblings experienced substantially larger learning losses than only children during prolonged school closures. Using administrative data from Peru, I show that students with siblings lost 0.05 standard deviations more in GPA and standardized test scores, with effects increasing in the number of siblings and persisting years after schools reopened. Second, to prove these differences reflect a causal effect of siblings rather than correlated factors, I leverage a difference-in-differences strategy comparing children with and without siblings before, during, and after the COVID-19 school closures. Instrumental variable estimates that exploit a preference for the sex composition of children to isolate exogenous variation in family size corroborate these results. Third, I present evidence that parental time constraints are the main mechanism: effects are strongest in families that typically invest more time in their children. Using school starting age cutoffs and a regression discontinuity design, I show that when younger siblings start school and spend less time at home, older siblings' performance and parental support improve, but these spillovers disappeared during school closures when all children remained home. Overall, these results reveal how the dilution of parental time during school closures disproportionately disadvantaged larger families and highlight the essential role of schools in mitigating household constraints in education.

JEL Codes: I21, I24, D13

* Department of Economics, University of Texas at Austin. Email: fpardo@utexas.edu.

I am deeply grateful to Leigh Linden, Scott Carrell, Richard Murphy, and Kirabo Jackson for their invaluable mentorship and guidance. I also thank Fabiola Alba, Manuela Angelucci, Peter Bergman, Diether Beuermann, Sandy Black, Marika Cabral, Eric Chyn, Mike Geruso, Brendan Kline, Gerald Oettinger, Dean Spears, Sebastian Tello-Trillo, Steve Trejo, and Cody Tuttle. Their comments and feedback have made this paper better. Finally, I am indebted to Alexander Cubas and Luis Angel Mejía from the Ministry of Education for directing the process to match the administrative records needed for this study.

I Introduction

Education is central to human capital formation and long-run outcomes, but its production depends on multiple institutions, with households and schools playing crucial yet shifting roles. Parents provide the foundation for children's early learning, shaping cognitive and non-cognitive skills through time investments and the home environment. As children reach school age, formal instruction increasingly shifts to schools, allowing households to reallocate more of their time toward other activities such as leisure, employment, or childcare and support for other children in the household.

Extended periods of school closures disrupt this normal process by abruptly shifting instructional responsibilities from schools to parents. This sudden reorganization of the learning process provides a unique opportunity to observe how families assume roles typically played by schools. However, the burden of this substitution varies systematically across family structures. During extended school disruptions, for example, having siblings may be either helpful or detrimental to students. On the one hand, siblings may complement each other's education by serving as mentors or sources of support, especially when contact with peers is limited. On the other hand, household resources, including access to technology and parental attention, are limited and can become binding constraints as siblings compete for them. If larger families struggle to replicate the role of schools, it reveals the limits of parental substitution and the distinct comparative advantages of each institution in the production of education.

In this paper, I ask whether the presence of siblings affects learning when educational instruction shifts from schools to households. The COVID-19 pandemic created an unprecedented disruption to schooling, affecting more than 1.6 billion students worldwide. In some countries, schools remained closed for up to two years, with learning losses roughly equivalent to a year's worth of schooling (Jakubowski, Gajderowicz and Patrinos, 2023). Despite extensive evidence on the magnitude of these losses, it is still not entirely clear which aspects of this change in the education production process caused them.

I document a new global stylized fact: children with siblings experienced substantially larger learning losses than only children during school closures. I explore these in detail using administrative data covering the universe of school enrollments in Peru's education system from 2014 to 2024, which includes information on all children from pre-kindergarten through grade 11, including their progression, GPA, and demographic and school characteristics. An important feature of the data is that I am able to link siblings across the entire student population, an uncommon feature in education administrative records. I complement these data with standardized national exams and with detailed large-scale parent surveys that provide measures of household resources, parental time investments, socioeconomic status, and educational expectations.

My analysis consists of three parts. First, I document that children with siblings experienced larger learning losses than only children during school closures. Second, I show that these differences reflect the causal effect of siblings rather than preexisting household differences. Finally, I examine several mechanisms underlying this effect, identifying parental time constraints as the most plausible channel.

I begin by documenting that children with siblings experienced larger declines in academic performance during the pandemic than only children. Similarly, children with more siblings performed worse than those with fewer siblings. Before the pandemic, both groups followed parallel trends in academic performance, which diverged sharply during and after school closures. Leveraging this similarity before the school closures, I use a difference-in-differences (DiD) framework that compares changes in performance between families of different sizes who, by the nature of their family composition, faced different childcare and time shocks. On average, children with siblings experienced learning losses 0.04 standard deviations (SD) greater than only children, and the effects reached up to 0.06 standard deviations among those with two or three siblings. These results hold for both GPA measures during and after closures and for standardized exams taken after schools reopened.

I then examine whether this divergence in performance is due to the number of children or to other factors correlated with family size. To verify this, I first show that the effects remain stable when controlling for observable parental and household characteristics, such as the mother's age, education, baseline socioeconomic status, and baseline student achievement, and allowing the impact of these factors to vary over time. As an additional test on the identification, I complement this analysis with a standard instrumental variables (IV) approach, using the sex composition of the first two children as an exogenous predictor of family size.[1] Within this framework, parents with two same-gender children are more likely to have an additional child, and the results confirm that an exogenous increase in family size reduces performance during school closures.

These effects are not limited to GPA, whose decline could reflects reduced homework completion or lower effort among students with siblings during school closures. They also have lasting impacts, reducing standardized test scores three years after schools reopened by 0.05 SD in both reading and math and extend beyond immediate test scores to encompass broader educational trajectories. Using large-scale survey data on educational expectations, I find that parents with multiple children became systematically more pessimistic about their children's long-term educational prospects, lowering their expectations that their children would attain a graduate degree by 6% when schools reopened. These shifts in expectations suggest that learning losses may shift families' educational investment trajectories.

Finally, I investigate why additional siblings made families less adaptable to school closures, examining whether the effects vary by birth order, technological access, sibling age gaps, grade level, and parental socioeconomic status. I find no systematic differences by birth order or technological access, suggesting that neither differences among later-born children, nor competition for technological resources, explain the effects. Instead, the results point to parental time constraints as the key mechanism: the effects are largest among elementary school children, where parental involvement is most intensive, and among families in the top three socioeconomic quartiles, where parents devote more time to children's learning. This pattern is consistent with lower socioeconomic

[1] See Black, Devereux and Salvanes (2005), Angrist, Lavy and Schlosser (2010), and Black, Devereux and Salvanes (2010).

status families investing little parental time regardless of family size. Additionally, using school starting-age cutoffs and a regression discontinuity approach that holds family characteristics constant, I show that when younger siblings start school, reducing the time they spend at home, older siblings' performance improves. I find a similar, though less robust, pattern for parental time investments. This occurs in a context where families expected to offload responsibilities to schools, an expectation that was violated during closures. In that case, older children no longer benefit when younger siblings start school. Using a difference-in-RD approach, I show that the negative spillovers of having a younger sibling emerged precisely during school closures, when all children remained at home and parental time became a binding constraint.

This paper contributes to three strands of research. First, it examines how parents substitute for schools in the education production process. Parents reduce effort when their children attend a better school (Pop-Eleches and Urquiola, 2013) or face increased school resources (Houtenville and Conway, 2008; Fredriksson, Öckert and Oosterbeek, 2016). Other work shows that parents reduce private educational spending in response to anticipated school grants but not in response to unanticipated ones (Das et al., 2013). In contrast, this paper provides the first causal evidence on how siblings mediate learning losses in a setting where schools' instructional role dramatically shifts to households. While parents can partially substitute for schools, this substitution is constrained by family structure, with each additional child reducing educational quality. This reveals the limits of household substitution as an input for the production of education, especially when facing short-run disruptions, and underscores the essential role that schools normally play in mitigating these constraints.

The paper also contributes to the literature on family size and education quality, providing new evidence on when a trade-off arises. Research shows this trade-off often does not exist (Becker and Tomes, 1976; Black, Devereux and Salvanes, 2005; Angrist, Lavy and Schlosser, 2010), yet unexpected shocks can cause such trade-offs to emerge (Black, Devereux and Salvanes, 2010; Olof Åslund and Hans Grönqvist, 2010). For instance, parents can usually plan and adapt to having a new child in a way that does not affect the quality of education received by their other children, but having twins may alter that balance. This paper contributes to this literature by showing that school closures created an analogous, unanticipated increase in parental time requirements, activating otherwise latent trade-offs between family size and educational quality.

Finally, this paper contributes to the growing literature on learning losses from school closures by uncovering a key mechanism behind these effects. Prior work documents large and persistent losses, particularly among vulnerable populations (Haelermans et al., 2022; Singh, Romero and Muralidharan, 2022; Jakubowski, Gajderowicz and Patrinos, 2023;; Jack et al., 2023). While other studies find positive non-cognitive effects of siblings during school closures (Hughes et al., 2023; Lampis et al., 2023), as they provided emotional and linguistic support that buffered the loss of peer interactions, my results indicate that the effect on academic outcomes was negative.

The rest of the paper proceeds as follows. Section II describes the Peruvian education system and how school closures where implemented, and Section III describes the data used. Section IV presents a new stylized fact about divergence in performance between siblings and only children. Section V establishes the causal relationship described in the previous section by discussing the empirical strategy and main results and Section VI the mechanisms. Section VII presents evidence of a similar relationship occurring in the rest of the world, and Section VIII concludes.

II Education System in Peru and School Closures

In this section, I describe the school education system in Peru, how it is structured, some overall statistics and then provide a description of how schools operated during closures and how remote learning was implemented in them.

Peru's basic education system consists of two mandatory levels: six years of primary education (grades 1–6) and five years of secondary education (grades 7–11). Around 5.5 million students are enrolled in these grades each year. According to household surveys, 99% of children aged 7–11 were enrolled in primary education, and 85% of those aged

12–16 were enrolled in secondary education in 2019. Most students are enrolled in public school, though private education is high compared to other developing countries, with almost 15% of students both in primary and secondary education enrolled in private institutions.

When the COVID-19 pandemic forced schools to close in March 2020 at the beginning of the academic year, Peru's Ministry of Education rapidly launched "Aprendo en Casa" (Learn at Home), a national strategy to sustain educational services remotely. The strategy relied on three primary channels to reach students across diverse geographic and socioeconomic contexts: a web platform, television broadcasts, and radio programming. This multi-channel approach was designed to address Peru's significant digital divide, recognizing that while 86.6% of households had access to television and 48.8% had access to radio, only 37.6% had internet connectivity, with only 6% in rural areas. The Ministry partnered with national broadcasters and produced educational content aligned with the National Basic Education Curriculum, covering all levels of basic education.

In practice, the program emphasized flexibility, and public education during school closures was designed to be delivered asynchronously. Teachers were expected to adapt centrally produced experiences to local contexts and individual student needs, maintaining regular communication with families through available means, primarily WhatsApp, text messages, and phone calls. The strategy emphasized student autonomy in learning, family accompaniment (rather than substitution of teacher roles), and flexible differentiation based on each student's access conditions and circumstances.

Implementation revealed significant structural challenges that shaped how education was actually delivered. By July 2020, approximately 71% of students accessed content via television, with teachers predominantly using cellphones to distribute materials, provide guidance, and maintain student connection. Often, the cellphones belonged to parents, who also used them for work, further limiting children's access to learning materials and communication with teachers. The shift placed extraordinary demands on families, particularly mothers, 63.7% of whom reported accompanying their children while they watched the content. Regional governments and local municipalities supplemented national efforts by distributing printed materials, providing connectivity support, and developing complementary content. While many of these limitations also applied to those in private schools, the measures taken by them would be carried out independently, with some of them likely implementing remote learning with synchronous virtual classes. All of these environments potentially relied on higher parental involvement for their effectiveness.

Moreover, during school closures, the Ministry of Education promoted every student regardless of performance. A different grading scale was also adopted. In elementary schools, where students are graded A through D, students were only given grades A through C during closures. In secondary school, they were graded from 11 to 20 instead of 0 to 20. To make grades comparable, I assign a C or 11 to those who received a D or a grade between 0 and 10 during school normal operation.[2]

III Data

I estimate the effects of family structure on educational outcomes before, during, and after school closures using the national population of enrolled students tracked through administrative records, combined with standardized exams and parental survey data covering 2014–2024. This comprehensive dataset allows me to examine effects even after schools reopened. Siblings are identified across sources using the mother's national identification number.

A. Administrative data on school progression and GPA (SIAGIE)

The SIAGIE (Sistema de Información de Apoyo a la Gestión Educativa) is a comprehensive administrative database maintained by Peru's Ministry of Education that tracks enrollment and academic records for all students in the country's education system. From 2014 to 2024, the system captured data across pre-K through grade 11 in both

[2] This change in grading policy is less relevant for elementary school given that only about 1% of students get a D. In secondary schools about 10% of students get a grade of 10 or lower, so the adjustment helps make grades comparable. However, unadjusted grades lead to similar results.

public and private institutions. The database contains detailed information on each student's school enrollment, grade level, grades by subject, passing status,[3] and sex, as well as parents' education levels and date of birth. Unique student identifiers allow individual students to be tracked across years, and I have access to the birth dates of all students enrolled in 2024.

B. *Sibling Identification*

To identify students with siblings, I use parental identification numbers to link students who share the same mother, which serves as a proxy for living in the same household. This information is available for 98% of students enrolled in pre-K through grade 11.

Table 1 presents descriptive statistics by sibling composition and key demographic and academic characteristics. I show that 38% of students are only children, 32% have one sibling, 20% have two siblings, and 10% have three siblings. Because sibling identification relies on enrollment data, the number of siblings may be understated, potentially attenuating the results.[4] Panel A shows similar characteristics across the first two columns, with 79% of students in urban areas, two-thirds enrolled in public schools, an average class size of 24 students, and 39% of mothers with complete secondary education. Panel B shows academic characteristics based on the SIAGIE data, such as grade promotion and standardized GPA at the class-year-school level.

C. *Standardized National Exams (ECE)*

To measure academic performance beyond GPA, I use national standardized exams conducted by the Peruvian Ministry of Education, also known as ECE (Evaluacion Censal de Estudiantes), which evaluate students in math and reading skills. These evaluations were implemented across different academic years and grade levels. Grade

[3] I do not explore school progression as an outcome because the Ministry of Education adopted universal progression during the COVID-19 pandemic.

[4] Chetty and Hendren (2018) similarly limit sibling identification to children whose parents filed a tax return and claimed them as dependents. Still, to have a better sense of the degree to which I might be missing siblings in 2020-2021 when using data up to three years later in 2024, I identify siblings considering only data up to 2021 and compare both cases three years earlier, in 2018. With the full data up to 2024, 60.7% of students have siblings, while that is the case for 56.4% when using data up to 2021. Results are consistent when using the more restrictive measure of sibling identification.

2 students were tested from 2007 to 2016 nationally and then in smaller representative samples in 2019 and 2022; grade 4 students were tested nationally in 2016, 2018, and 2024 and in smaller representative samples in 2019, 2022, and 2023.[5] This is a low stakes test for the students and measures their basic competencies in math and language at the end of the school year.

I standardize these tests to have a mean 0 and SD 1 in the base year of 2007 to ensure comparability across time. Table 1 shows that in grade 2 math, only children score 0.1 SD lower than children with one sibling and 0.02 SD higher than children with two siblings. In grade 2 reading, only children score 0.04 SD lower than children with one sibling and 0.12 SD higher than those with two siblings.

D. Surveys

In addition to standardized exams, starting in 2015, the Ministry of Education surveyed parents. These surveys include information about socioeconomic status, parents' mother tongue, expectations for educational attainment, parental investment in education, and access to internet and a computer. Socioeconomic status is reported in a standardized index with mean 0 and SD 1. This is based on materials in the household, access to services, assets owned, and parents' education level.

Table 1 shows that families with only one child have higher socioeconomic status on average—0.03 SD more than those with two children. Access to internet and a computer is also similar between these two groups, with around one-third of households having them. Most parents have high expectations for the maximum level of education that their children will achieve: 80% of parents of only children expect that to be college education or higher, similar to 81% of parents of children with one sibling.

Only children and children with one sibling are similar in most observable characteristics, and those with more siblings tend to have lower performance and socioeconomic status. However, these are average differences across schools. Ultimately, I compare students within schools, where characteristics are much more similar across families

[5] The national exams were administered to schools with at least five students in the respective grade.

with different numbers of children.

IV Stylized Fact: Larger Learning Losses for Children with Siblings

A. Empirical Strategy

In this section, I present a new fact about family structure and student performance during school closures: children with siblings experienced larger learning losses than those without. I begin by examining how performance trends evolved during and after an extended period of school closures, using both GPA and standardized exams. I then analyze these patterns within a DiD framework, treating children without siblings as the comparison group, to estimate the effect of having a sibling on learning losses during school closures. The identification strategy is developed further in Section V.

To make GPA measures more comparable across schools and years, I make two adjustments. First, I impose the grading scale used during school closures on all other years, censoring lower grades accordingly (as discussed in the previous section). Second, I standardize GPAs at the school-grade-year level to account for variation in teachers' grading standards and leniency.[6] Standardized exam scores are already comparable across schools and over time, so changes across years directly reflect gains or losses in performance.

To estimate the magnitude of these differential learning losses, I implement a DiD specification comparing children with and without siblings during school closures. This approach identifies the effect of the period of school closures on relative performance of children with and without siblings. Even though these families are different, this approach addresses that in two ways. First, by adding school fixed effects, I focus on variations within schools, where family characteristics are more homogeneous while variation in fertility remains significant. Second, it accounts for differences that are constant over time. The DiD especification is

[6]These adjustments aid interpretation of the results, though the effects are consistent with using unadjusted measures.

$$Y_{isgt} = \alpha + \delta Sib_i + \boldsymbol{\beta} Post_{it} Sib_i + \lambda_s + \mu_g + \tau_t + \varepsilon_{isgt} \tag{1}$$

Similarly, the corresponding estimate for the event study is

$$\begin{aligned} Y_{isgt} = \alpha + \delta_1 + \delta_2 Sib_i + \sum_{k=-5}^{-2} \delta_k (\mathbb{I}[t = 2020 + k] Sib_i) \\ + \sum_{k=0}^{4} \beta_k (\mathbb{I}[t = 2020 + k] Sib_i) + \lambda_s + \mu_g + \tau_t + \varepsilon_{isgt} \end{aligned} \tag{2}$$

Y denotes the standardized GPA, Sib is an indicator variable equal to one if the individual has siblings (or if the individual has one, two, or three siblings), and $Post$ is an indicator variable equal to one if the year is 2020 and over, marking the beginning of school closures. I also include school (λ), grade (μ), and time (τ) fixed effects. The coefficient of interest, β, represents the difference in achievement gaps between children with and without siblings. Standard errors are clustered at the school level.

B. *Performance before, during, and after school closures*

Figure 1 presents trends in academic performance for children with and without siblings. Panel (a) shows trends in grade 4 standardized math exams, which provide an absolute measure of learning that can be compared consistently across years. These exams were administered each November from 2016 to 2024 (except in 2017 and the years of school closures). Before closures, both groups performed similarly and experienced learning gains at comparable rates. After schools reopened, students were tested again in 2022 and 2023 in a representative sample, and all grade 4 students were assessed in 2024. The emergence of a gap between children with siblings and only children is evident. This is particularly striking because not all of these students were enrolled in school when closures began: the 2024 cohort started grade 1 in 2021, yet even they show larger losses among those with siblings.

While standardized scores capture overall learning losses, they are measured as early as a year after schools reopened, when some recovery may have already occurred. To examine changes in real time, I next turn to GPA measures, which are available for every school year, including those affected by closures, and standardize GPAs within each school-grade-year.

Panel (b) of Figure 1 shows how this measure evolved for children with and without siblings. The gap widens markedly during school closures and remains partially persistent after schools reopened. Because the variable is standardized within each year, the overall mean is always zero; therefore, the apparent increase in scores among only children should be interpreted as a relative improvement, not an absolute gain in performance. In addition to larger losses during school closures, the gap remains wider than it was before schools reopened.

Figure 2 presents results from the event-study specification. Panel (a) shows that during school closures, children with siblings experienced larger learning losses of about 0.04 SD and then remained 0.01 SD lower once schools reopened. Panel (b) shows that the effects are larger for those who had more siblings, though the main change occurs when going from zero to one sibling or from one to two; having more than two is associated with a similar learning loss.[7]

V The Causal Effect of Siblings

Having established in Section IV that children with siblings experienced larger learning losses than only children, this section examines whether those differences reflect a causal effect of family size. I first describe the empirical strategy, combining DiD models and an IV approach, and then present the main results and their persistence beyond GPA using standardized exams and parental expectations.

[7] To have a sense of the size of this effect at the school level, 0.04 SD in GPA is associated with 1 SD increase in socioeconomic status or half the gap between the bottom and top quartile of the socioeconomic status.

A. Empirical Strategy

Families with different numbers of children may also differ in socioeconomic status, preferences for education quality, mothers' age, and other characteristics. To address this, I first use a DiD strategy that controls for potential heterogeneity from observed characteristics. I then use an IV approach that exploits the same-sex composition of the first two children to generate exogenous variation in family size.

A.1. DiD with Observed Heterogeneity

One potential concern is that differences that remain within schools could interact with school closures and generate heterogeneous effects, for example, if wealthier parents, or those more concerned about education quality, have fewer children and cope better with remote learning. To address this, I include an extensive set of controls and their interactions with *Post* to capture potential observed heterogeneities. Assuming these characteristics are correlated with unobserved ones (e.g., socioeconomic status correlated with preferences for educational quality), I can assess the degree of potential endogeneity by unobserved characteristics based on how the estimated coefficient changes, as proposed by Oster (2019). Data on mothers' education and date of birth are available for the full sample, and to include a more comprehensive set of controls, I also incorporate information from baseline surveys on socioeconomic status and standardized exam achievement.

I augment equations (1) and (2) by including the term X_{ist} and its interaction with $Post_{it}$ to account for heterogeneous effects on observable characteristics. The extended DiD specification is

$$Y_{isgt} = \alpha + \delta Sib_i + \boldsymbol{\beta} Post_{it} Sib_i + \gamma_1 X_{ist} + \gamma_2 Post_{it} X_{ist} + \lambda_s + \mu_g + \tau_t + \varepsilon_{isgt} \tag{3}$$

Similarly, the corresponding event-study specification is

$$Y_{isgt} = \alpha + \delta_1 + \delta_2 Sib_i + \sum_{k=-5}^{-2} \delta_k(\mathbb{I}[t = 2020 + k]Sib_i)$$

$$+ \sum_{k=0}^{4} \beta_k(\mathbb{I}[t = 2020 + k]Sib_i) + \gamma_1 X_{ist} + \gamma_2 Post_{it} X_{ist} + \lambda_s + \mu_g + \tau_t + \varepsilon_{isgt} \tag{4}$$

If the endogeneity that jointly affects fertility decisions and performance during school closures is partly correlated with observable characteristics, controlling for these variables should mitigate some of the resulting bias.

A.2. Sex Composition as an Instrument for Family Size

In addition to controlling for heterogeneity in family characteristics, I also address the endogeneity of fertility decisions directly. Previous research shows that parents have preferences for variety and are more likely to have a third child if their first two are of the same sex (Angrist and Evans, 1998). If sex composition is random, families with two same-sex children are therefore more likely to have a third child, making sex composition a potential instrument that exogenously induces variation in family size. The two-stage least squares (2SLS) equations using the indicator for the first two children being of the same sex as an instrument for family size are

$$FamSize_{it} = \delta_0 + \delta_1 SameSex_i + \delta_2 X_{it} + \nu_{it} \tag{5}$$

$$Y_{it} = \beta_0 + \beta_1 \widehat{FamSize}_i + \beta_2 X_{it} + \epsilon_{it} \tag{6}$$

where Y_{it} is an educational outcome, which in this case is standardized GPA; $FamSize$ is the number of children in the family; $SameSex$ is a dummy variable that equals one if the first two children are of the same sex; and X_{it} is a set of demographic characteristics such as mother's education level, age, and age at first birth, as well as a birth order dummy for the second-born child.

B. Results

In this section, I present the causal results in two steps: first using DiD with observed heterogeneity and then using exogenous variation in family size.

B.1. DiD with Observed Heterogeneity

I first examine whether the estimated sibling effect remains stable after controlling for observable household characteristics. Table 2 reports estimates for grade 6 students between 2019 and 2020 using their grade 2 baseline characteristics, as students in that grade were tested and their parents surveyed in 2015 and 2016.[8] Panel A shows the effects on standardized GPA for math and Panel B for reading. The first column shows the estimate controlling for fixed effects but no other control for heterogeneous characteristics, as was done in equation (1). The size of the estimate for this sample is –0.051 SD for math and –0.034 SD for reading.

The rest of the columns progressively add other controls: a dummy for the mother having higher education, a dummy for both mother's age and age at first birth being above 30, and dummies for socioeconomic status and performance in standardized test scores being above the mean. The main coefficient is highly stable across all these estimates, with a reduction of less than 10% of its initial value. Given the consistency of the estimates when including observed heterogeneity, it is unlikely that all of the effect is driven by heterogeneous effects from unobserved characteristics.[9]

[8]I also identify two additional groups for which similar baseline information is available. I estimate a DiD model using 2018 and 2020 grade 6 students with grade 2 and grade 4 baseline characteristics, and 2019 and 2020 grade 5 students using grade 4 baseline characteristics. The results are presented in Table A.1.

[9]In Table A.1 I show this for two other samples. The second sample, which uses grade 6 students from 2018 and 2020, shows a slightly larger reduction in the coefficient, up to 20%. The third sample, which uses grade 5 students from 2019 and 2020, goes in the opposite direction, with slight increases in the coefficient.

B.2. Exogenous Change in Family Size

I then examine whether exogenous variation in family size addresses remaining concerns about unobserved heterogeneity. Table 3 presents the 2SLS estimates, along with two sets of OLS coefficients estimated on the same samples for comparison. The first OLS specification includes controls only for birth order, while the second includes the full set of control variables. I estimate this model for the years 2018–2019, when schools were operating normally and for the years 2020–2021, when schools were closed. The first-stage estimate on *SameSex* is 0.05. Similarly, Black and Devereux (2010) find a coefficient of 0.082 using Norwegian data, Angrist, Lavy and Schlosser (2010) find 0.073 using Israeli data, and Conley and Glauber (2006) find that having two same-sex children increases the probability of having a third child by about 0.07 in the United States. This shows that as in other settings, Peru also exhibits a preference for variety in the sex composition of children.

The fourth column presents the 2SLS estimate of an exogenous increase in the number of children in the family on the GPA performance of the first two children. This estimate is 0.063 SD during 2018–2019 but decreases to –0.031 during school closures. The difference between these two periods (–0.094) is significant at the 1% level and captures the increased penalty of larger families during closures—the key effect highlighted in this paper.

It is worth noting that the point estimates from the DiD and IV strategies need not be identical. The DiD captures an average treatment effect, while the IV identifies a local average treatment effect for families whose fertility decisions are influenced by having same-sex children, typically families of higher socioeconomic status, for whom the cost of an additional child is lower. Moreover, the IV estimate is based on families with at least two children and measures the effect of an additional child, whereas the DiD compares families with one versus multiple children. Despite these conceptual differences, both results point in the same direction: having siblings negatively affected learning outcomes during school closures.

C. Persistence of Effects and Impacts Beyond GPA

The previous results focused on GPA outcomes during school closures. Here, I examine the persistence of these effects for younger children after schools reopened and whether they extend to standardized exams and parents' educational expectations for their children.

C.1. GPA

I have already shown in Figure 2 that students with siblings in grades 1–11 experienced larger learning losses than only children during school closures. Now, I examine how these effects vary across grades and persist after schools reopened. As shown in Figure 3, results for elementary school students reveal larger and more persistent effects, with those who have two or three siblings experiencing GPA losses of up to 0.04 SD even three years after reopening. Estimating the DiD specification separately by grade (panel (b)) confirms that younger students were more affected than older ones.

How do these effects vary across different types of schools? I examine elementary school students (grades 1 to 6) separately by urban and rural location, and by public and private school status. Figure A.1 shows that the effects are similar across all groups. This consistency is informative, as these schools adopted different approaches to remote learning during closures. Students in rural areas had less access to computers and the internet, and their lessons relied more on radio than those of students in urban areas. However, in both settings, most communication between teachers and families occurred through cellphones and WhatsApp. The similarity of effects between public and private schools is also notable. While public schools primarily relied on asynchronous instruction, private schools might have been quicker to adopt synchronous virtual classes.

GPA and standardized exams capture distinct dimensions of achievement. GPA tends to be more associated with non-cognitive skills such as self-discipline (Duckworth and Seligman, 2005) and broader personality traits (Borghans et al., 2016), whereas standardized tests primarily measure cognitive ability. Consequently, improvements in one do not necessarily translate into gains in the other (Jackson, 2018). The observed decline in GPA could therefore reflect that students in larger families struggled more with completing homework or submitting assignments on time, without corresponding learning losses in standardized exams. To test this possibility, I also examine standardized test scores collected after schools reopened as a more direct measure of academic performance and cognitive skills.

C.2. Standardized Exams

I examine effects on standardized exams, which provide comparable measures across schools and over time. Unlike GPA, standardized exams are not available for every year and every grade. They were not taken during school closures, and only grade 2 and grade 4 students take it in elementary school. In some cases this is done for the full population of students in the grade and in others only for a representative sub-sample.[10] For grade 2 and grade 4 exams, I perform a DiD estimation between 2019 and 2022, using tested students from a representative sample of the population. For grade 4, I can compare the full population of students using 2018 and 2024 national exams.

Table 4 shows the estimate for the effect of having siblings during school closures on standardized test scores. For grade 2, the effect is –0.034 SD for math and –0.023 SD for reading. Losses for grade 4 are larger than those for grade 2, with up to –0.052 SD for math and –0.036 SD for reading for the same sample of years.[11] Results using the national exams of 2018 and 2024 point in the same direction. For grade 4, the results are even higher than –0.1 SD for those who had three siblings. The larger magnitude in Panel C is particularly interesting when noting that Panels A and C are testing the same cohort of students in the post-period: those who were in grade 1 in 2021. That is, learning losses in standardized test scores are not only persistent, but the gap widened

[10] Prior to COVID-19, the test was administered to the full population of grade 2 students from 2007 to 2016, and then to representative subsamples in 2018, 2019, 2022, and 2023. The full population of grade 4 students was tested in 2016, 2018, and 2024, while only representative subsamples were tested in 2019, 2022, and 2023.

[11] A 0.05 SD in standardized exams is associated with 0.4 SD increase in socioeconomic status or a fifth of the gap between the bottom and top quartile of the socioeconomic status

further after schools reopened.

These results highlight two facts. First, although part of the losses in GPA seem to recover after schools reopened, actual performance in standardized exams does not attenuate over time. Second, even students who had not started school yet in 2020 experienced learning losses. This is important because it shows that the mechanism is relevant even for children who were not yet in school in 2020. Those students can also be impacted by the presence of a sibling, which may constrain household resources.

C.3. Expectations

Finally, I examine whether the sibling gap extends to parents' educational expectations, which may influence future investments in their children's education. Along with the standardized exams, parents were surveyed and asked about the maximum level of education they expected their children would obtain. As is common, these expectations typically exceed actual attainment levels: almost 80% of parents expected their children to obtain a higher education degree, and 40% expected them to obtain a master's or PhD degree.

Table 4 shows significant reductions in these expectations among parents of students with siblings. The results are noisier in Panels A and B for effects on expectations for higher education degrees, with generally negative point estimates but not always significant. However, Panel C, which uses a larger sample, shows more consistent results. Both expectations for higher education and for a graduate degree decline as a consequence of having a sibling, by an average of 2.5 and 1.7 percentage points, and by as much as 5.7 and 3.5 percentage points for those with three siblings—almost a 10% decline from baseline levels.

VI Mechanisms

Having established that learning losses during school closures were larger for children with siblings, particularly among younger students and those from larger families, this section explores potential mechanisms underlying these patterns. I test three possibilities. First, I examine whether birth order differences explain the results, as later-born children tend to have lower educational outcomes. Second, I assess whether siblings competed for limited material or technological resources, such as computers, or disrupted one another's learning. Third, I investigate whether the dilution of parental time and attention explains the observed effects, using heterogeneity by socioeconomic status and age, as well as a regression discontinuity design based on school starting-age cutoffs.

A. Birth Order

To assess whether birth-order effects drive these differences, I begin by examining if the results are explained by the advantage typically associated with being first-born. Research has shown positive effects on being first-born on educational outcomes (Behrman and Taubman, 1986; Black, Devereux and Salvanes, 2005; Price, 2008). If these differences make them better prepared to deal with school closures, being later-born, rather than having siblings, might be causing the effects I have identified. To isolate the role of birth order, I restrict the sample to first-born children and compare those who had siblings with those who did not.[12]

Panel (a) of Figure 4 shows the estimated effect of having a sibling, restricting the sample to first-born children. The magnitudes are similar to those obtained using all children in the sibling sample, which suggests that even though birth order may play a role in explaining differences in overall educational attainment, it does not produce differential learning losses due to school closures. In the following sections, I include only the oldest sibling in the sibling sample.

B. Sibling disruptions

During school closures, students not only spent more time at home but also relied on it as the primary environment for studying, reviewing lessons, and taking tests. Because

[12]Because first-born children may also take on increased responsibilities when having siblings during school closures, I also consider the youngest siblings instead and compare them to only children. The results are consistent.

learning took place within the household rather than at school, siblings rather than peers could have influenced this process by disrupting one another or by competing for access to technology and learning materials.

To test whether technological constraints explain the results, I split the sample by household access to computers and the internet. If competition for these resources were driving the results, I would expect households that lack those resources to not exhibit the negative effect of having a sibling. However, Panels C and D of Table 5 show that the negative effects are present even in households with neither a computer nor internet to access remote education easily. This suggests that the results are not driven by siblings competing for material resources.

Students who are closer in age may interact more and be more disruptive. To test if siblings distracted one another, I examine differences in effects by age gap. The results, presented in panel (b) of Figure 4, show that having a sibling one to two years apart has a similar effect as having a sibling three to five years apart, suggesting that age gap is not a meaningful factor in explaining the effect of a sibling. However, for siblings of the same age, this negative effect disappears and even becomes slightly positive. This pattern reflects a combination of sibling and parental mechanisms. Students of the same age attend the same grade and likely share classes and assignments.[13] On the one hand, it is more likely that in this context having a sibling is beneficial as cooperation is more plausible, especially when most of these are twins. On the other hand, it is easier for parents to keep up with two children doing the same work than with two children doing different schoolwork. This is the case for children with siblings, since most do not share homework, teachers, web resources, or even class schedules, all of which increases the amount of time parents must invest in monitoring and supporting their children.

C. Parental Time and Attention

I next examine whether parental time constraints explain the negative effects of siblings during school closures. When schools closed, teachers' roles diminished and

[13] Because I do not have birth dates for the full sample, age gap has been proxied by the difference in grades between siblings.

learning shifted almost entirely to the household, requiring parents to take a more active role and allocate additional time in supporting their children's learning. Because parental time is limited, families with multiple children may have been less able to dedicate sufficient attention to each child. I find this to be the most plausible mechanism driving the observed results.[14]

If parental time is binding and siblings dilute this resource, I expect to see larger effects where parental time investment is both more necessary and more common. When parents invest little time in their children's learning, those with and without siblings should be equally affected by the loss of school-based support. In contrast, parents who usually invest more time in their children face a trade-off when that time must be divided among several children.

Consistent with this hypothesis, panel (b) of Figure 3 shows larger effects for younger students, where parental involvement tends to be greater. Similarly, Table 5 shows stronger sibling effects among families in the upper socioeconomic quartiles and weaker or null effects among those in the lowest quartile, where parental investment is already limited. This pattern aligns with descriptive evidence from 2015 in panel (a) of Figure A.2, which shows that parental time investment is substantially lower for households in the bottom part of the socioeconomic index distribution.[15] Table A.2 also shows larger effects among families whose parents held higher educational expectations and for students in the top quartile of baseline achievement. This pattern may reflect compensatory behavior, where parents allocate additional time to children who are performing worse, a contrast to Giannola (2024), who finds that parents invest more in higher-achieving children, particularly when constrained.

To further this analysis, I exploit the school starting-age cutoff in Peru to identify the spillover effects of younger siblings starting school on older siblings' academic

[14]According to Bergman (2021), providing parents with information about students' missing assignments improve performance in both GPA and standardized tests, partly by increasing parental involvement. In the context of this paper, where parents often mediated communication between students and teachers, parents with multiple children may have been less informed about their children's schoolwork due to time constraints, which could have reinforced the observed disparities.

[15]Parents are asked whether they help their children with homework, discuss what they did in school, or clarify their questions. A standardized index of parental investment is constructed using these responses.

performance. Under this rule, children who turn six years old by March 31 are required to enroll in grade 1 in that same academic year, while those who turn six after the cutoff must wait until the following year. This cutoff creates a sharp discontinuity in school entry timing that is otherwise unrelated to previous family characteristics. When a younger sibling starts school, they presumably require fewer hours of parental care, which frees up overall time for parents and may improve the academic outcomes of older siblings through increased parental attention later in the day. Panel (a) of Figure 5 shows that the distribution of ages around the cutoff is smooth, and panel (b) confirms that the first stage is sharp, indicating the rule is strictly enforced.

I use this discontinuity to estimate the effect of a younger sibling starting school, and therefore spending less time at home, on the GPA of older siblings who are still in school:

$$Y_{if} = \gamma_0 + \beta_1 ABOVE_{jf} + f(AgeCutoff_{jf}, ABOVE_{jf}) + \tau_t + \epsilon_{if} \tag{7}$$

where Y_{if} is the standardized GPA of child i in family f, based on the running variable, $AgeCutoff_{jf}$, which is the age of the youngest sibling j by the cutoff date. $ABOVE_{jf}$ captures the discontinuity around the cutoff, $f(\cdot)$ includes local linear controls for the running variable, and τ_t captures year fixed effects.

When schools operate normally, being above the cutoff implies that students begin grade 1 and therefore spend less time at home. This reduction in childcare needs for parents may generate positive spillovers by increasing the time they have available for their other children already in school. These positive spillover effects are shown in Figure 5 and Panel A of Table 6 for the optimal bandwidth in column (1) and for different bandwidth sizes in columns (2)–(6).[16] This occurs in a context where families expected to offload responsibilities to schools. However, during school closures, all children remained at home both before and after reaching school starting age. Panel B shows

[16]Based on Calonico et al. (2017).

that, in that case, older children no longer benefit when younger siblings start school. To estimate changes in these spillovers, I use a difference-in-regression-discontinuity design that compares the regression discontinuity estimates during closures with those obtained when schools were open, as specified in the following equation:

$$
\begin{aligned}
(8) \quad Y_{ift} = \delta_0 &+ \beta_2 ABOVE_{jf} \cdot Closures_t \\
&+ \alpha_0 ABOVE_{jf} + \alpha_1 Closures_t + f(AgeCutoff_{jf}, ABOVE_{jf}) + \tau_t + \epsilon_{ift}
\end{aligned}
$$

The effect of interest is captured by β_2. Effects for math and reading are shown in Panels C and F of Table 6, and the negative effects of –0.012 SD and –0.013 SD in column (1) show the estimates for the optimal bandwidth for math and reading, respectively.[17] When schools are closed and children have to stay at home rather than start school when they turn six, parents can no longer benefit from the increased time availability that school entry would normally create. This is reflected by these negative effects, which show the loss of positive spillovers that older siblings would have otherwise benefited from.

To explore the mechanisms behind the positive spillovers from having a sibling at school rather than at home, I use reported measures of parental time investment from the surveys. Although these measures are unavailable during or before school closures, I can examine their spillover effects in 2022 and 2023, when schools reopened. For these years, I construct an index of parental investment based on several questions about the frequency of activities such as explaining topics, helping with schedules, and asking questions about school.

Panel (d) of Figure 5 shows a small positive effect on the parental investment index for grade 4 students when a younger sibling starts school rather than spend more time at home. The results in Table 7 also show increases in specific activities, with parents spending more time explaining topics and helping with schedules. However, although

[17]Because Panels A and B and Panels D and E use different samples, I estimate the optimal bandwidth for each and choose the minimum for each pair.

the point estimates remain positive, most of the effects are no longer statistically significant when using the optimal bandwidth, except for "help with schedule." Because the sample for this measure is relatively small, part of the loss of significance may reflect limited statistical power.

Overall, the evidence points to parental time constraints as the main channel. Birth order and material competition do not explain the observed patterns, and although there is some indication of positive spillovers in specific cases such as twins, there is no consistent evidence that siblings were disruptive otherwise. The heterogeneity analysis and school starting-age discontinuity results are consistent with what would be expected if parental time constraints were the mechanism behind sibling-related learning losses during school closures, when parents faced greater demands for childcare and educational support.

VII External Validity

In this section, I assess the external validity of the findings by examining whether children with siblings also experienced larger learning losses during school closures in other countries. I draw on two complementary sources: data from the Programme for International Student Assessment (PISA) and evidence from Singh, Romero and Muralidharan (2022).

A. PISA

I first use data from PISA, which measures the educational performance of 15-year-old students in math, reading, and science and is conducted primarily in OECD countries (with additional non-OECD participants) every three years. The last four assessments were administered in 2012, 2015, 2018, and 2022. Because the 2015 and 2018 rounds did not include information on sibling status, I compare the gap in learning outcomes between children with and without siblings in 2012 and 2022.

Panel (a) of Figure 6 shows how the score distributions shift leftward for both groups but more markedly for those with siblings. To examine this more closely, I first estimate the learning losses among children with siblings in each country and children without siblings (only children). Then I estimate how different these learning losses are for both groups. My estimate is analogous to a simple 2×2 DID estimate without controls: $DID = \Delta_{2022-2012}Sibs - \Delta_{2022-2012}OC$. As in the previous analysis, if children with siblings experienced larger losses, this difference will take a negative value.

Additionally, using data from UNESCO on length of school closures for each country, I estimate the number of weeks each country had of full or partial closures. The results, shown in panel (b) of Figure 6, reveal two patterns. First, most developed and developing countries are below the 0 line, meaning that children with siblings also experienced larger losses between 2012 and 2022. Second, the size of these larger losses is highly correlated with the duration of school closures. This provides some suggestive evidence that across different lengths of closures and policy measures to deal with them, children with siblings had a harder time adjusting to remote schooling.

B. India

To complement the cross-country evidence, I assess whether a similar sibling gap in learning losses appears in India, which also experienced prolonged school closures. In Peru, standardized exams were not administered until one year after schools reopened, allowing time for learning recovery. In contrast, Singh, Romero and Muralidharan (2022) study a setting in which students were tested after closures ended. Children scored about 0.7 SD lower in math and 0.34 SD lower in Tamil (the local language) in December 2021 compared to those of the same age and gender in the same villages in August 2019. The authors find greater learning losses among children whose mothers had not completed high school (grade 12).

To complement their analysis, I extend their results by adding a dimension of heterogeneity that is the focus of my paper: whether the children had siblings. The results, presented in Table 8, show substantially larger learning losses among children with siblings. The size of these effects is –0.11 SD for math and –0.065 SD for Tamil.

As in the Peruvian case, the estimates are robust when controlling for heterogeneity by the mother's education and are larger relative to those differences.

VIII Conclusions

I find evidence of an overlooked dimension of school closures: larger families struggled more to compensate for the loss of formal schooling. These families faced greater and more persistent learning losses, primarily because parents could not fully substitute for teachers when time had to be divided among multiple children. While siblings may have offered emotional support or social support during the pandemic (Hughes et al., 2023; Lampis et al., 2023), this did not translate into academic benefits. More broadly, the results highlight the essential role that schools and educational services play in mitigating the inherent time constraints of larger households.

International evidence suggests that similar patterns emerged elsewhere, with larger effects in countries that experienced longer school closures. These findings have important implications for education policy during crises. Differences in parental time constraints are a key source of variation in learning losses across family structures. Because even high-resource families cannot overcome this limitation, policies focused solely on technological or financial support may overlook an important source of inequality. Crisis response should therefore consider family structure and include additional support for larger families, such as targeted tutoring or remote-learning structures that reduce parental supervision requirements, to mitigate long-term human capital losses.

REFERENCES

Angrist, Joshua D., and William N. Evans. 1998. "Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size." *The American Economic Review*, 88(3): 450–477. Publisher: American Economic Association.

Angrist, Joshua, Victor Lavy, and Analia Schlosser. 2010. "Multiple Experiments for the Causal Link between the Quantity and Quality of Children." *Journal of Labor Economics*, 28(4): 773–824. Publisher: [The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago].

Becker, Gary S., and Nigel Tomes. 1976. "Child Endowments and the Quantity and Quality of Children." *Journal of Political Economy*, 84(4, Part 2): S143–S162. Publisher: The University of Chicago Press.

Behrman, Jere R., and Paul Taubman. 1986. "Birth Order, Schooling, and Earnings." *Journal of Labor Economics*, 4(3): S121–S145. Publisher: [The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago].

Bergman, Peter. 2021. "Parent-Child Information Frictions and Human Capital Investment: Evidence from a Field Experiment." *Journal of Political Economy*, 129(1): 286–322. Publisher: The University of Chicago Press.

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

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2005. "The More the Merrier? The Effect of Family Size and Birth Order on Children's Education*." *The Quarterly Journal of Economics*, 120(2): 669–700.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes. 2010. "Small Family, Smart Family?: Family Size and the IQ Scores of Young Men." *Journal of Human Resources*, 45(1): 33–58. Publisher: University of Wisconsin Press Section: Articles.

Borghans, Lex, Bart H. H. Golsteyn, James J. Heckman, and John Eric Humphries. 2016. "What grades and achievement tests measure." *Proceedings of the National Academy of Sciences*, 113(47): 13354–13359. Publisher: Proceedings of the National Academy of Sciences.

Calonico, Sebastian, Matias D. Cattaneo, Max H. Farrell, and Rocío Titiunik. 2017. "Rdrobust: Software for Regression-discontinuity Designs." *The Stata Journal*, 17(2): 372–404. Publisher: SAGE Publications.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects*." *The Quarterly Journal of Economics*, 133(3): 1107–1162.

Conley, Dalton, and Rebecca Glauber. 2006. "Parental Educational Investment and Children's Academic Risk: Estimates of the Impact of Sibship Size and Birth Order from Exogenous Variation in Fertility." *Journal of Human Resources*, 41(4): 722–737. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System].

Das, Jishnu, Stefan Dercon, James Habyarimana, Pramila Krishnan, Karthik Muralidharan, and Venkatesh Sundararaman. 2013. "School Inputs, Household Substitution, and Test Scores." *American Economic Journal: Applied Economics*, 5(2): 29–57.

Duckworth, Angela L., and Martin E. P. Seligman. 2005. "Self-discipline outdoes IQ in predicting academic performance of adolescents." *Psychological Science*, 16(12): 939–944.

Fredriksson, Peter, Björn Öckert, and Hessel Oosterbeek. 2016. "Parental Responses to Public Investments in Children: Evidence from a Maximum Class Size Rule." *Journal of Human Resources*, 51(4): 832–868. Publisher: University of Wisconsin Press Section: Articles.

Giannola, Michele. 2024. "Parental Investments and Intra-household Inequality in Child Human Capital: Evidence from a Survey Experiment." *The Economic Journal*, 134(658): 671–727.

Haelermans, Carla, Madelon Jacobs, Rolf van der Velden, Lynn van Vugt, and Sanne van Wetten. 2022. "Inequality in the Effects of Primary School Closures Due to the COVID-19 Pandemic: Evidence from the Netherlands." *AEA Papers and Proceedings*, 112: 303–307.

Houtenville, Andrew J., and Karen Smith Conway. 2008. "Parental Effort, School Resources, and Student Achievement." *Journal of Human Resources*, 43(2): 437–453. Publisher: University of Wisconsin Press Section: Articles.

Hughes, Claire, Luca Ronchi, Sarah Foley, Caoimhe Dempsey, Serena Lecce, and I.-Fam Covid Consortium. 2023. "Siblings in lockdown: International evidence for birth order effects on child adjustment in the Covid19 pandemic." *Social Development*, 32(3): 849–867. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/sode.12668.

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

Jackson, C. Kirabo. 2018. "What Do Test Scores Miss? The Importance of Teacher Effects on Non–Test Score Outcomes." *Journal of Political Economy*, 126(5): 2072–2107. Publisher: The University of Chicago Press.

Jakubowski, Maciej, Tomasz Gajderowicz, and Harry Anthony Patrinos. 2023. "Global learning loss in student achievement: First estimates using comparable reading scores." *Economics Letters*, 232: 111313.

Lampis, Valentina, Sara Mascheretti, Chiara Cantiani, Valentina Riva, Maria Luisa Lorusso, Serena Lecce, Massimo Molteni, Alessandro Antonietti, and Marisa Giorgetti. 2023. "Long-Lasting Effects of Changes in Daily

Routine during the Pandemic-Related Lockdown on Preschoolers' Language and Emotional–Behavioral Development: A Moderation Analysis." *Children*, 10(4): 656. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics*, 37(2): 187–204. Publisher: ASA Website _eprint: https://doi.org/10.1080/07350015.2016.1227711.

Pop-Eleches, Cristian, and Miguel Urquiola. 2013. "Going to a Better School: Effects and Behavioral Responses." *American Economic Review*, 103(4): 1289–1324.

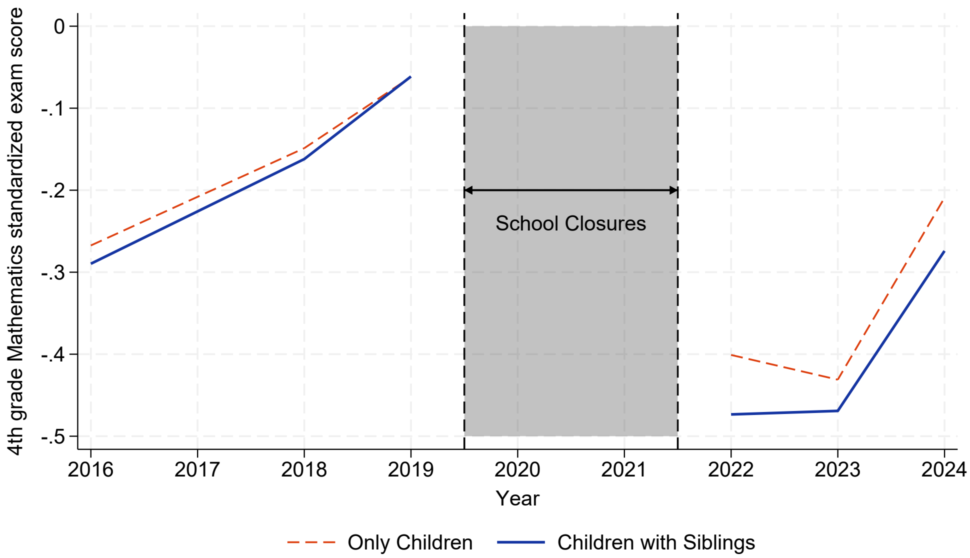
Price, Joseph. 2008. "Parent-Child Quality Time: Does Birth Order Matter?" *Journal of Human Resources*, 43(1): 240–265. Publisher: University of Wisconsin Press Section: Articles.

Singh, Abhijeet, Mauricio Romero, and Karthik Muralidharan. 2022. "Covid-19 Learning Loss and Recovery: Panel Data Evidence from India."
ΩÅslund and Grönqvist

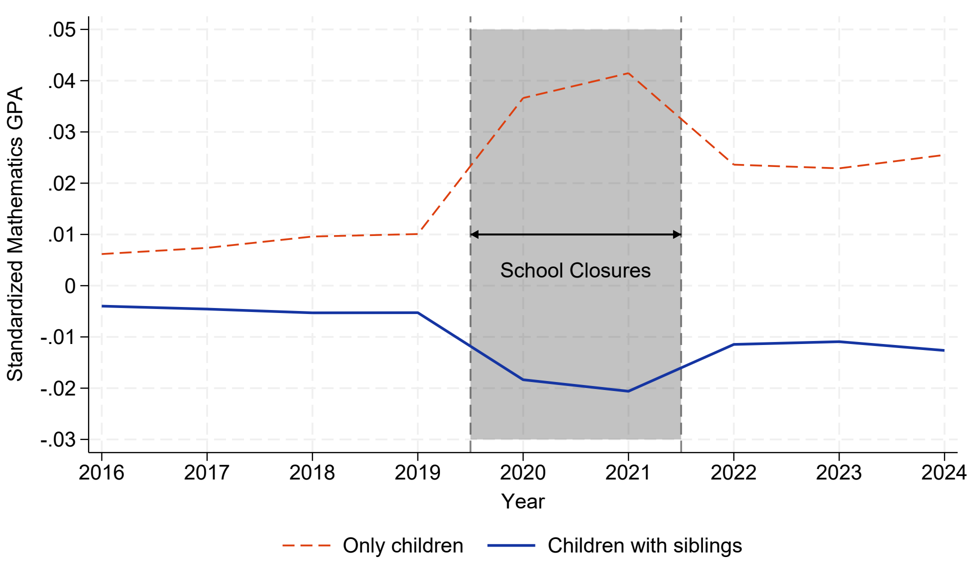
Åslund, Olof, and Hans Grönqvist. 2010. "Family size and child outcomes: Is there really no trade-off?" *Labour Economics*, 17(1): 130–139.

Figure 1. : Trends in education outcomes for only children and children with siblings

(a) Average Standardized Mathematics Exam in 4th grade



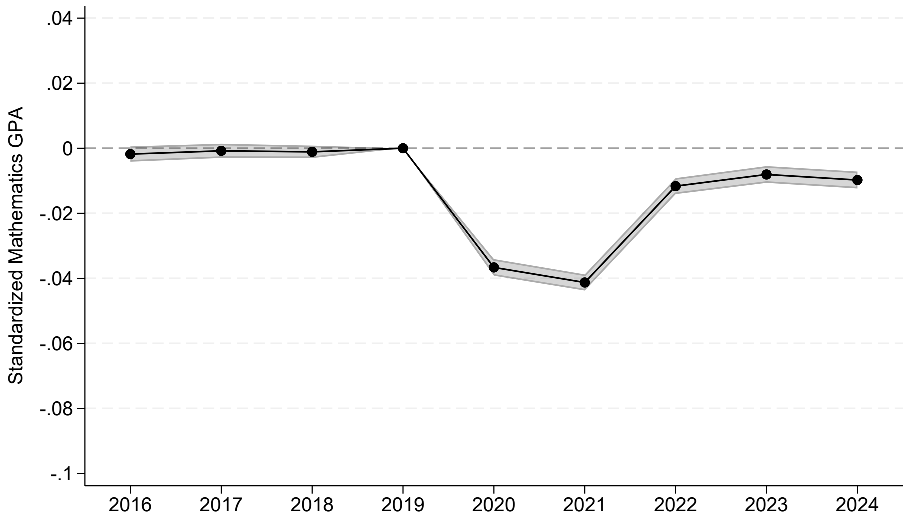
(b) Average GPA standardized within school-grade-year from 1st-6th grade



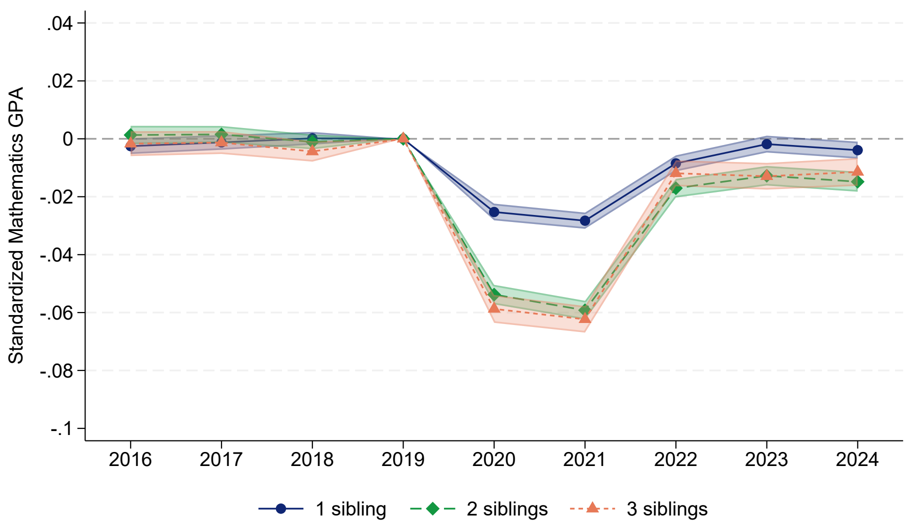
Notes: In *both panels*, the X-axis shows the year in which performance was measured, indicating the years of school closures (2020-2021). In the *top panel*, the Y-axis represents the average standardized exam score for fourth grade mathematics. These scores are comparable across time, with the year 2007 set as reference (mean 0 and standard deviation of 1). These exams are usually taken in November, close to the end of the academic year. The exam was not taken in 2017 or in 2020-2021. In the *bottom panel*, the Y-axis represents average GPA in mathematics from grades 1 through 11, standardized at the school-grade-year level. In both cases, results are show for the sample of children with and without siblings.

Figure 2. : Event Study: Effect of having a sibling

(a) Mathematics GPA: Children with siblings relative to children without siblings



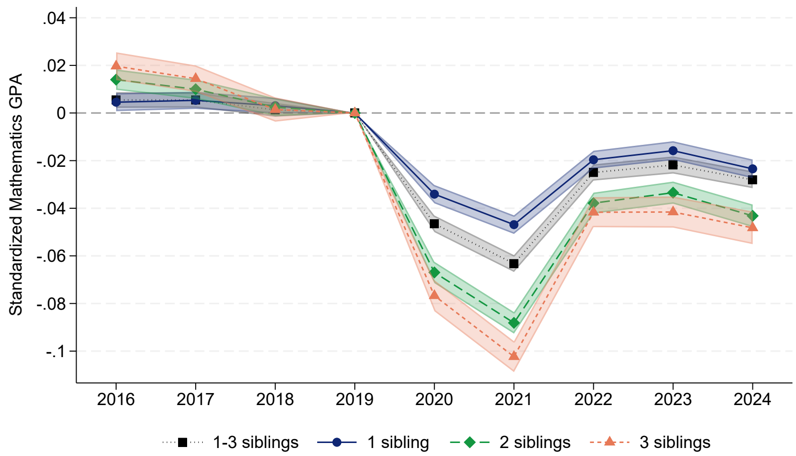
(b) Mathematics GPA by number of siblings relative to children without siblings



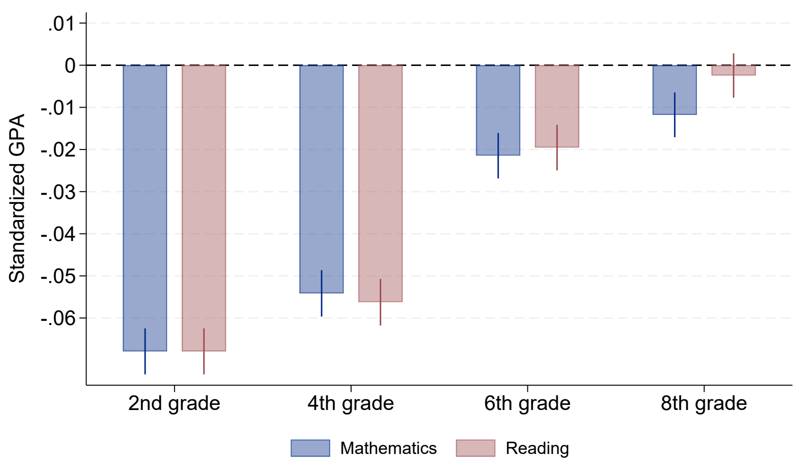
Notes: The *top panel* shows the results from the event study using equation (2) for the GPA in mathematics from grades 1 through 11, standardized at the school-grade-year level. The results are measures for the sample of students with siblings relative to those without siblings. The Sib_i dummy takes the value of one if the student has siblings. In the *bottom panel* I do the same but only considering those with one, two or three siblings respectively compared to those without siblings. That is, estimates for students with one sibling compared to those without siblings are shown in blue circle markers, with two sibling compared to those without siblings are shown in green diamond markers, and with three sibling compared to those without siblings are shown in red with triangle markers. Estimated standard errors, reported in parentheses, are clustered at the primary school level. Shaded areas show the 95% confidence intervals.

Figure 3. : Effect of having a sibling during elementary school

(a) Mathematics GPA for elementary school children



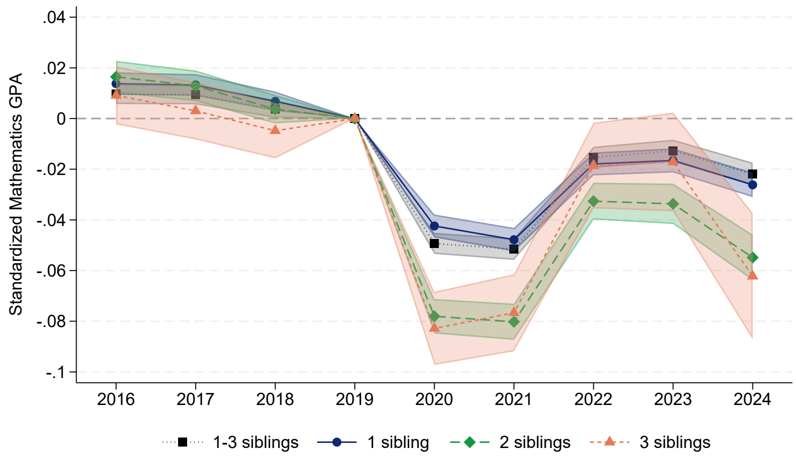
(b) Mathematics GPA by grade



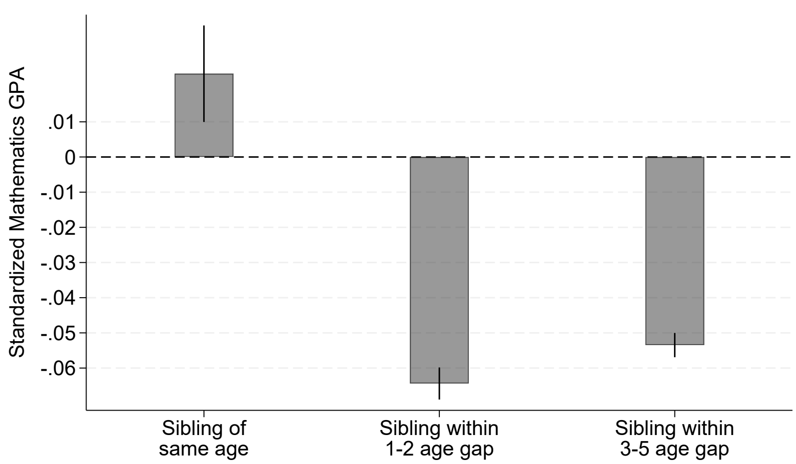
Notes: The *top panel* shows results from the event study using equation (2) for the GPA in mathematics from grades 1 through 6, standardized at the school-grade-year level. The Sib_i dummy takes the value of one if the student has siblings. Estimates for students with one, two or three siblings compared to those without siblings are shown in black square markers, with one sibling compared to those without siblings are shown in blue circle markers, with two sibling compared to those without siblings are shown in green diamond markers, and with three sibling compared to those without siblings are shown in red with triangle markers. Standard errors are clustered at the primary school level. Shaded areas show the 95% confidence intervals. The *bottom panel* shows results from the difference-in-difference estimates using equation (3) for the GPA in mathematics and reading standardized at the school-grade-year level for grades 2, 4, 6, and 8 for the period 2016-2021. The control variables (X_{ist}) used to account for other potential heterogeneity are student's sex, mother's age, mother's age at first birth, and mother's level of education. Standard errors are clustered at the primary school level. Vertical lines in each bar show the 95% confidence intervals.

Figure 4. : Mechanisms: Birth order and age gap

(a) Effect of having a sibling on mathematics GPA for first-born children by number of siblings

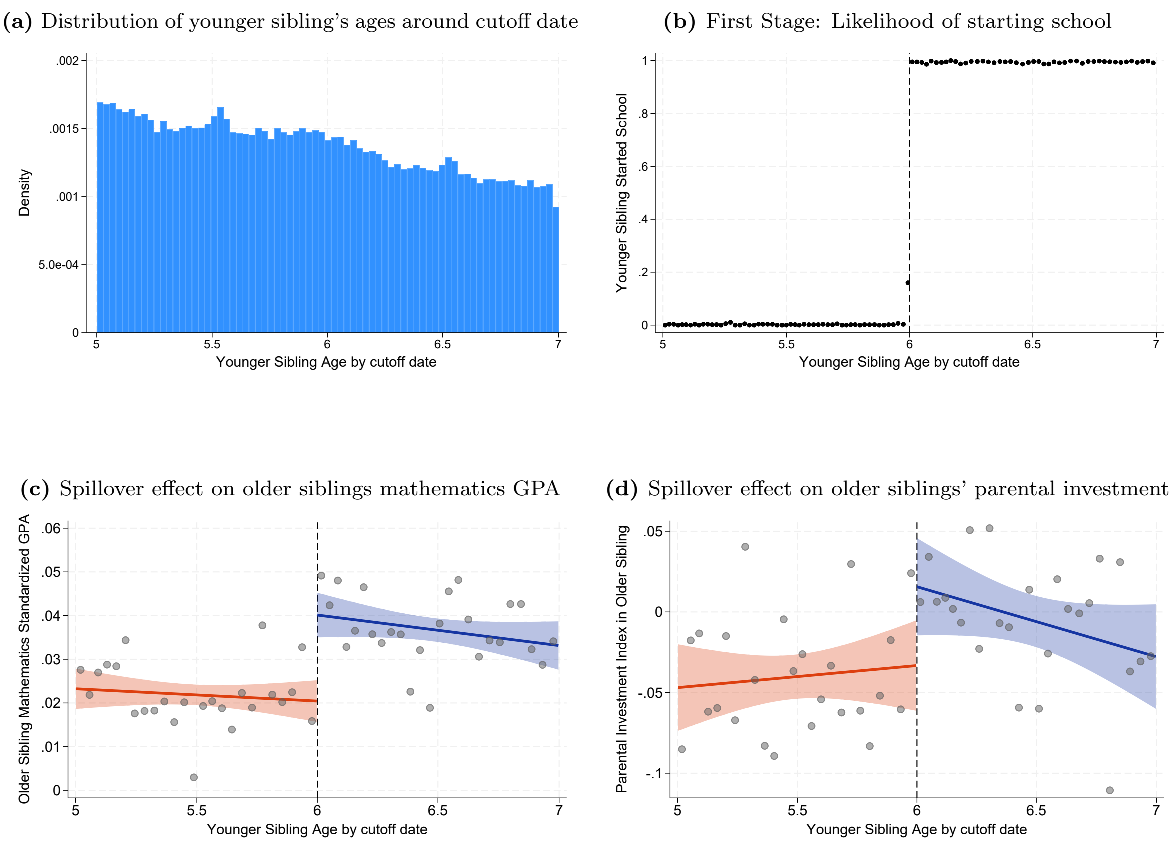


(b) Effect of having a sibling on mathematics GPA by age gap



Notes: The *top panel* shows results from the event study using equation (2) for the GPA in mathematics from grades 1 through 6, standardized at the school-grade-year level. Only the sample of first-born children is consider for both those with or without siblings. The Sib_i dummy takes the value of one if the student has siblings. Estimates for students with one, two or three siblings compared to those without siblings are shown in black square markers, with one sibling compared to those without siblings are shown in blue circle markers, with two sibling compared to those without siblings are shown in green diamond markers, and with three sibling compared to those without siblings are shown in red with triangle markers. Standard errors are clustered at the primary school level. Shaded areas show the 95% confidence intervals. The *bottom panel* shows results from the difference-in-difference estimates using equation (3) for the GPA in mathematics standardized at the school-grade-year level for the period 2016-2021. Three samples were used when comparing to those without siblings: those with a sibling of the same age, with a sibling within 1-2 years of age, and with a sibling within 3-5 years of age. The control variables (X_{ist}) used to account for other potential heterogeneity are student's sex, mother's age, mother's age at first birth, and mother's level of education. Standard errors are clustered at the primary school level. Vertical lines in each bar show the 95% confidence intervals.

Figure 5. : Regression Discontinuity during normal school operation: Spillover effect of having a younger sibling start school



(a) Distribution of younger sibling's ages around cutoff date

(b) First Stage: Likelihood of starting school

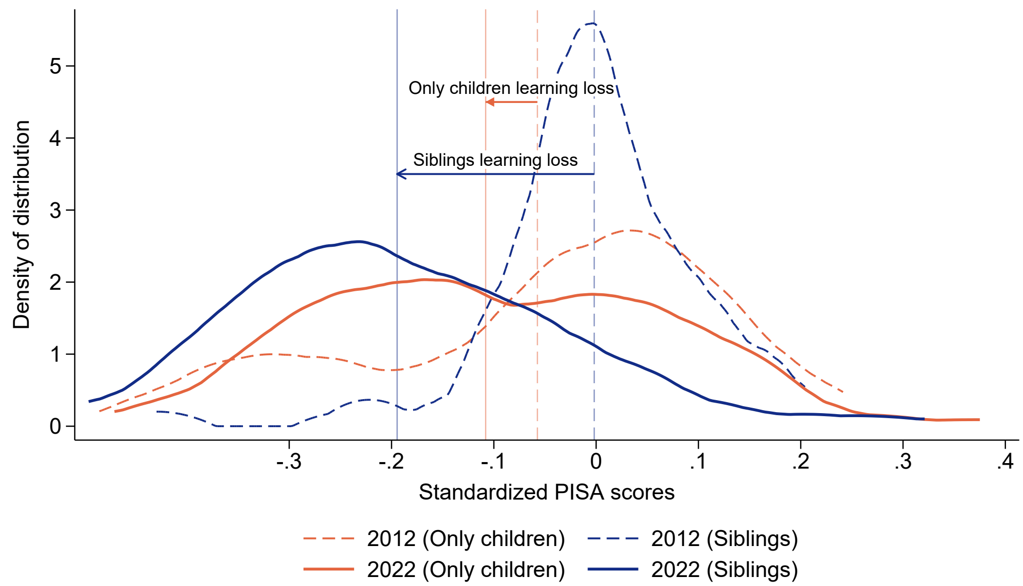
(c) Spillover effect on older siblings mathematics GPA

(d) Spillover effect on older siblings' parental investment

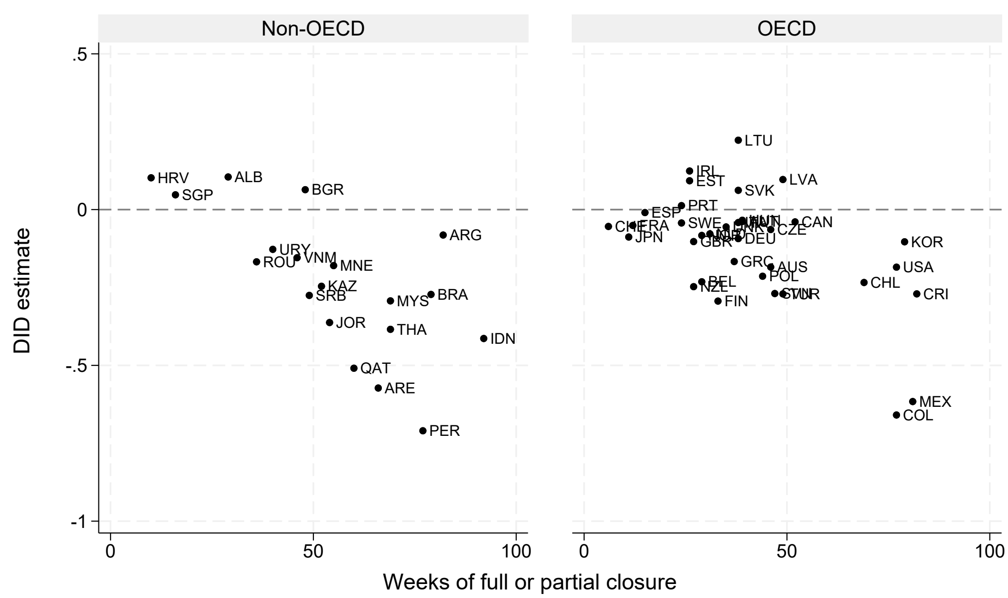
Notes: This figure shows results from the regression discontinuity described in equation . The *top left panel (a)* shows the distribution of the running variable: age of the younger sibling around the starting age cutoff by March 31. The distribution is smooth around the cutoff which provides evidence of no manipulation. The *top right panel (b)* shows the first stage. The Y-axis is the percentage of students who enroll in first grade during that year. The sharp discontinuity shows the almost perfect compliance of the school-starting age rule. The *bottom left panel (c)* shows the estimate of the regression discontinuity equation for the standardized mathematics GPA of the older sibling for years when schools were open (2016-2019, 2022-2023). The *bottom right panel (d)* shows the estimate of the regression discontinuity equation for the investment index reported by parents of fourth grade older siblings for the years 2022-2023. The variable was not available in years 2016-2019. For both panels c and d, the bandwidth used was 365 days. In Table 6, I show results for smaller bandwidths, for reading and for the *difference-in-RD* estimate comparing years of closures with years when schools were open. In Table 7, I show results for smaller bandwidths for the investment index and three of its components. Shaded areas show the 95% confidence intervals.

Figure 6. : International evidence of larger learning losses for children with siblings

(a) Learning gaps in Mathematics between 2012 and 2022 for only children and children with siblings



(b) Change in learning gaps by duration of school closure for OECD and Non-OECD countries.



Notes: The *top panel* shows the distribution of average country PISA scores in mathematics for the 'only children' (red) and 'sibling' (blue) sample for 2012 and 2022 in dashed and solid lines respectively. Individual scores were first standardized using the mean a standard deviation from 2012 and then averaged at the country-year level for both children with and without siblings. The vertical lines show the mean country averages for each of the distributions and the horizontal arrows show the learning loss from 2012 to 2022. The *bottom panel* the Y-axis shows the estimated learning loss for each country as $DID = \Delta_{2022-2012}Sibs - \Delta_{2022-2012}OC$. and the X-axis shows the number of weeks the country had full or partial school closures.

Table 1—: Descriptive Statistics

	Only children	1 sibling	2 siblings	3 siblings
	(1)	(2)	(3)	(4)
Sample share	0.38	0.32	0.20	0.10
Panel A: School and household characteristics				
Urban	0.79	0.79	0.71	0.58
Public School	0.67	0.70	0.80	0.89
Class Size	24.44	24.61	23.46	21.34
Mother with complete secondary	0.39	0.39	0.36	0.28
Panel B: Academic characteristics				
Grade promotion	0.96	0.97	0.96	0.94
Standardized GPA - Mathematics	0.02	0.05	-0.02	-0.10
Standardized GPA - Reading	0.02	0.05	-0.02	-0.10
Observations	1,238,826	1,079,501	683,613	316,529
Panel C: Academic Performance (2nd grade)				
Standardized Exam - Mathematics	0.65	0.75	0.63	0.43
Standardized Exam - Reading	0.77	0.81	0.65	0.44
Observations	819,485	702,368	432,584	193,909
Panel D: Household Characteristics (2nd grade)				
Socio-Economic Index	0.11	0.08	-0.15	-0.44
Internet	0.34	0.32	0.25	0.16
PC	0.35	0.35	0.28	0.19
Education expectation: 5-year college	0.80	0.81	0.75	0.66
Observations	90,998	93,439	58,021	24,544

Notes: This table shows descriptive statistics for four samples of the student population: students without siblings (only children) and students with one, two or three siblings. The first row shows the sample share. The statistics correspond to students from grades 1 through 6 for the year 2016. The first row shows the distribution of students between these four groups. Then, *panel A* shows statistics using on school and mother's education using administrative enrollment data. *Panel B* uses the same data to show academic characteristics of the students. *Panel C* reports baseline academic performance from standardized exams students took in 2nd grade from 2007-2016. *Panel D* reports information from baseline survey responses of second grade students in 2015-2016

Table 2—: DID estimates of having a sibling controlling for confounding heterogeneity

	GPA				
	(1)	(2)	(3)	(4)	(5)
Panel A: 2019–2020 6th grade DID with 2nd grade baseline (Mathematics GPA)					
Sibling x Post	-0.051***	-0.048***	-0.048***	-0.047***	-0.047***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Observations	193,960	193,960	193,960	193,960	193,960
Panel B: 2019–2020 6th grade DID with 2nd grade baseline (Reading GPA)					
Sibling x Post	-0.034***	-0.031***	-0.031***	-0.031***	-0.031***
	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)
Observations	192,367	192,367	192,367	192,367	192,367
Mother's Ed		X	X	X	X
Mother's age			X	X	X
SES				X	X
Score					X

Notes: This table shows results from the *difference-in-difference* estimates using equation (3) for sixth grade students in 2019 and 2020 using second grade baseline characteristics. *Panel A* shows results for GPA in mathematics and *panel B* for reading, both standardized at the school-grade-year level. Column 1 shows the estimates without any controls for heterogeneous effects, analogous to what is done in equation (1). Column 2 adds a dummy for the mother having higher education. Column 3 adds dummies for both mother's age and age at first birth being above 30. Column 4 adds a dummy for baseline socioeconomic status being above the mean. Column 5 adds a dummy for baseline standardized test scores being above the mean. Estimated standard errors, reported in parentheses, are clustered at the primary school level. ***p<0.01, **p<0.05, *p<0.1.

Table 3—: Effect of Family Size on mathematics GPA

	OLS (no controls)	OLS (controls)	First Stage	Second Stage	Observations
Panel A: Pre-Covid (2018-2019)					
Instrument: first two children same sex			0.050*		3,300,349
(Sample: first and second children in families			(0.001)		
with two or more births)					
Number of children in family	−0.081*	−0.070*		0.063*	
	(0.001)	(0.001)		(0.022)	
Panel B: Covid (2020-2021)					
Instrument: first two children same sex			0.047*		2,809,126
(Sample: first and second children in families			(0.001)		
with two or more births)					
Number of children in family	−0.119*	−0.101*		−0.031	
	(0.001)	(0.001)		(0.025)	
Panel C: Difference between COVID and Pre-COVID estimate					
				-.094***	
				(0.033)	

Notes: This table shows results from the *same-sex IV* estimates using equations (5) and (6) using the sample of first and second born children in families with two or more births. *Panel A* shows estimates during years when schools were open and *panel B* shows estimates during years of school closures. Column 1 and 2 show the simple OLS estimates with no controls and controlling for demographics. Column 2 adds a dummy for the mother having higher education. Column 3 adds dummies for both mother's age and age at first birth being above 30. Column 4 adds a dummy for baseline socioeconomic status being above the mean. Column 5 adds a dummy for baseline standardized test scores being above the mean. Estimated standard errors, reported in parentheses, are clustered at the primary school level. ***p<0.01, **p<0.05, *p<0.1.

Table 4—: DID estimates of having a sibling on Standardized Exams and Expectations

	Has a sibling	1 sibling	2 siblings	3 siblings
	(1)	(2)	(3)	(4)
Panel A: 2nd grade students (2019, 2022)				
Mathematics	-0.034***	-0.024**	-0.072***	-0.079***
	(0.009)	(0.009)	(0.012)	(0.019)
Reading	-0.023***	-0.018**	-0.061***	-0.064***
	(0.008)	(0.009)	(0.011)	(0.018)
Max Expectation: 5-year college	-0.007**	-0.015***	-0.007	0.012
	(0.003)	(0.004)	(0.005)	(0.009)
Max Expectation: Graduate level	-0.015***	-0.015***	-0.029***	-0.011
	(0.005)	(0.005)	(0.007)	(0.011)
Observations	226,592	153,784	110,671	77,517
Panel B: 4th grade students (2019, 2022)				
Mathematics	-0.052***	-0.028***	-0.081***	-0.149***
	(0.009)	(0.010)	(0.012)	(0.018)
Reading	-0.036***	-0.010	-0.066***	-0.131***
	(0.009)	(0.010)	(0.012)	(0.017)
Max Expectation: 5-year college	-0.005	-0.009**	0.004	-0.024***
	(0.004)	(0.004)	(0.006)	(0.009)
Max Expectation: Graduate level	-0.012**	-0.010*	-0.009	-0.027***
	(0.005)	(0.006)	(0.007)	(0.010)
Observations	183,473	121,339	90,833	63,819
Panel C: 4th grade students (2018,2024)				
Mathematics	-0.040***	-0.029***	-0.065***	-0.102***
	(0.004)	(0.005)	(0.006)	(0.009)
Reading	-0.062***	-0.050***	-0.096***	-0.132***
	(0.004)	(0.005)	(0.006)	(0.009)
Max Expectation: 5-year college	-0.025***	-0.019***	-0.041***	-0.057***
	(0.002)	(0.002)	(0.003)	(0.004)
Max Expectation: Graduate level	-0.017***	-0.015***	-0.032***	-0.035***
	(0.002)	(0.003)	(0.003)	(0.005)
Observations	758,901	533,684	399,912	299,759

Notes: This table shows results from the *difference-in-difference* estimates using equation (3) for second and fourth grade students controlling for the mother having higher education and for both the mother's age and age at first birth being above 30. The first column shows the estimate of having a sibling when compared to those without a sibling. The second to fourth column show the estimate of having one, two a three siblings respectively when compared to those without a sibling. Each panel shows results for standardized exam scores in mathematics and reading as well as for reported parental expectations on the student obtaining a 5-year college or a graduate degree. *Panel A* shows results for representative sample of second grade students who take the exam in 2019 (pre) and 2022 (post). *panel B* shows results for representative sample of fourth grade students who take the exam in 2019 (pre) and 2022 (post). *panel C* shows results for the universe of fourth grade students who take the exam in 2018 (pre) and 2024 (post). Estimated standard errors, reported in parentheses, are clustered at the primary school level. ***p<0.01, **p<0.05, *p<0.1.

Table 5—: DID estimates of having a sibling on mathematics GPA by baseline resources

	1 sibling	2 siblings	3 siblings
	(1)	(2)	(3)
Panel A: Low SES Households (Q1)			
	-0.001	-0.024***	-0.058***
	(0.008)	(0.009)	(0.014)
Observations	282,258	238,260	201,756
Panel B: High SES Households (Q4)			
	-0.037***	-0.065***	-0.134***
	(0.008)	(0.014)	(0.034)
Observations	249,817	193,842	173,621
Panel C: Households with neither PC nor Internet			
	-0.019***	-0.049***	-0.083***
	(0.006)	(0.008)	(0.012)
Observations	468,143	381,437	322,583
Panel D: Households with both PC and Internet			
	-0.028***	-0.056***	-0.076***
	(0.007)	(0.011)	(0.023)
Observations	371,603	294,298	261,885

Notes: This table shows results from the *difference-in-difference* estimates using equation (3) for the mathematics GPA standardized at the school-grade-year level for students in sixth, seventh and ninth grade for which baseline characteristics from standardized test scores and socioeconomic characteristics was available. Column 1 through 3 show the estimate of having one, two a three siblings respectively when compared to those without a sibling. I identified four subsamples of students for grades that were tested before and during closures that also had baseline characteristics: sixth grade students in 2019 and 2020 with baseline characteristics from second grade standardized exams in 2015 and 2016, sixth grade students in 2018 and 2020 with baseline characteristics from fourth grade standardized exams in 2016 and 2018, seventh grade students in 2019 and 2021 with baseline characteristics from fourth grade standardized exams in 2016 and 2018, and ninth grade students in 2019 and 2020 with baseline characteristics from second grade standardized exams in 2018 and 2019. *Panel A* shows results for students in the first (lowest) quartile of socioeconomic status. *Panel B* shows results for students in the fourth (highest) quartile of socioeconomic status. *Panel C* shows results for students with no computer or internet at home. *Panel D* shows results for students with both computer and internet at home. Estimated standard errors, reported in parentheses, are clustered at the primary school level. ***p<0.01, **p<0.05, *p<0.1.

Table 6—: Effects of younger sibling starting school on older sibling GPA

	Standardized mathematics GPA					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Spillover effects on mathematics when schools are open						
Younger sibling	0.025***	0.025**	0.021**	0.024***	0.020***	0.019***
goes to school	(0.007)	(0.012)	(0.009)	(0.007)	(0.005)	(0.004)
Observations	316,576	103,214	206,349	307,449	601,655	1,203,142
Bandwidth (days)	93	30	60	90	180	365
Panel B: Spillover effects on mathematics when schools are closed						
Younger sibling	0.006	0.001	0.011	0.003	-0.001	0.002
goes to school	(0.012)	(0.022)	(0.015)	(0.013)	(0.009)	(0.006)
Observations	101,002	32,857	65,643	97,896	191,203	378,757
Bandwidth (days)	93	30	60	90	180	365
Panel C: Diff-in-RD estimate of spillover on mathematics (closed - open)						
Younger sibling	-0.012*	-0.004	-0.018**	-0.012	-0.012**	-0.014***
goes to school	(0.007)	(0.013)	(0.009)	(0.007)	(0.005)	(0.004)
Observations	417,578	136,071	271,992	405,345	792,858	1,581,899
Bandwidth (days)	93	30	60	90	180	365
Panel D: Spillover effects on reading when schools are open						
Younger sibling	0.016**	0.018	0.016*	0.016**	0.020***	0.019***
goes to school	(0.007)	(0.013)	(0.009)	(0.007)	(0.005)	(0.004)
Observations	334,542	101,752	203,526	303,286	593,439	1,186,652
Bandwidth (days)	100	30	60	90	180	365
Panel E: Spillover effects on reading when schools are closed						
Younger sibling	0.005	0.019	0.012	0.002	0.004	-0.002
goes to school	(0.012)	(0.022)	(0.016)	(0.013)	(0.009)	(0.006)
Observations	107,418	32,492	65,143	97,185	189,831	376,132
Bandwidth (days)	100	30	60	90	180	365
Panel F: Diff-in-RD estimate of spillover on reading (closed - open)						
Younger sibling	-0.013*	-0.004	-0.015*	-0.011	-0.016***	-0.012***
goes to school	(0.007)	(0.013)	(0.009)	(0.007)	(0.005)	(0.004)
Observations	441,960	134,244	268,669	400,471	783,270	1,562,784
Bandwidth (days)	100	30	60	90	180	365

Notes: This table shows the estimates from the regression discontinuities using equations (5) and (8). The running variable is the younger sibling age at the school-starting age date cutoff (March 31st). Results show the effect of the younger sibling starting school on the older sibling's mathematics and reading GPA standardized at the school-grade-year level. *Panel A and D* show results for mathematics and reading respectively for the years in which schools were open (2016-2019 and 2022-2023). *Panel B and E* show results for mathematics and reading respectively for the years in which schools were closed (2020-2021). *Panel C and F* show the *Diff-in-RD* estimate using equation (8). Column 1 shows results for the optimal bandwidth and columns 2 through 6 for different bandwidths between 30 and 365 days. The optimal bandwidth used in *panels A-C* or *panels D-F* the minimum optimal bandwidth of between *panels A-B* or *panels D-E*. ***p<0.01, **p<0.05, *p<0.1.

Table 7—: Effects of younger sibling on parental time investment in older sibling

	Parental time investments					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Spillover effects on parental time investment index						
Younger sibling goes to school	0.019	-0.039	-0.019	0.001	0.024	0.032**
	(0.022)	(0.043)	(0.031)	(0.025)	(0.018)	(0.013)
Observations	34,524	8,675	17,315	25,774	50,183	102,383
Bandwidth (days)	123	30	60	90	180	365
Panel B: Spillover effects on parental time investment (explains topics)						
Younger sibling goes to school	0.035	-0.071	-0.029	0.008	0.038**	0.035***
	(0.022)	(0.044)	(0.032)	(0.026)	(0.018)	(0.013)
Observations	31,752	7,996	15,984	23,743	46,152	94,166
Bandwidth (days)	123	30	60	90	180	365
Panel C: Spillover effects on parental time investment (helps with schedule)						
Younger sibling goes to school	0.040*	-0.040	-0.010	0.017	0.055***	0.048***
	(0.022)	(0.044)	(0.031)	(0.026)	(0.018)	(0.013)
Observations	31,458	7,924	15,819	23,523	45,730	93,197
Bandwidth (days)	123	30	60	90	180	365
Panel D: Spillover effects on parental time investment (search for information)						
Younger sibling goes to school	0.002	-0.004	-0.021	-0.007	0.002	0.015
	(0.023)	(0.045)	(0.032)	(0.026)	(0.019)	(0.013)
Observations	31,605	7,980	15,905	23,626	45,955	93,620
Bandwidth (days)	123	30	60	90	180	365

Notes: This table shows the estimates from the regression discontinuities using equation (5). The running variable is the younger sibling age at the school-starting age date cutoff (March 31st). Results show the effect of the younger sibling starting school on the parental time investment index for the older sibling in fourth grade in the year 2022 and 2023, which is when they were surveyed. *Panel A* shows results for the overall parental time investment index standardized with mean 0 and standard deviation of 1. *Panel B* shows results for wether parents explain topics the student does not understand. *Panel C* shows results for wether parents help students make a schedule or set a time for studying. *Panel D* shows results for wether parents help students search for material or additional information for their homework. Column 1 shows results for the optimal bandwidth and columns 2 through 6 for different bandwidths between 30 and 365 days. ***p<0.01, **p<0.05, *p<0.1.

Table 8—: Learning loss between August 2019 and December 2021 in India - Based on Singh, Romero and Muralidharan (2022)

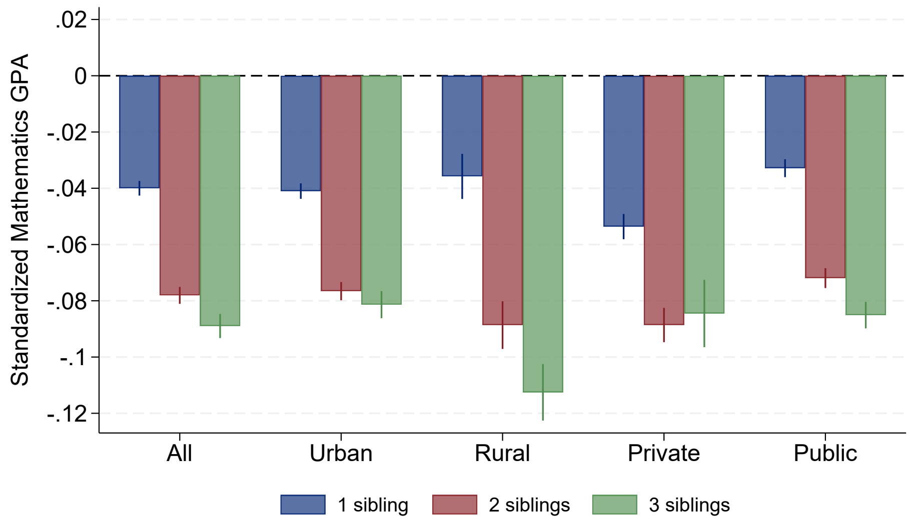
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Math score (in SD)						Tamil score (in SD)					
Wave 1 (Dec 2021)	-.73***	-.74***	-.76***	-.75***	-.68***	-.71***	-.35***	-.35***	-.37***	-.38***	-.32***	-.34***
	(.031)	(.038)	(.042)	(.049)	(.037)	(.046)	(.02)	(.023)	(.027)	(.029)	(.024)	(.031)
Male × Dec 21		.023						-.0074				
		(.041)						(.022)				
Mother Edu: Gr. 9-11 × Dec 21			.019			.021			.0015			.004
			(.053)			(.053)			(.03)			(.03)
Mother Edu: Gr. 12+ × Dec 21			.09*			.084*			.06**			.057**
			(.049)			(.049)			(.025)			(.025)
SES Decile × Dec 21				.0046						.0061		
				(.0075)						(.0039)		
Has Siblings (2-10 yrs old) × Dec 21					-.11**	-.10**					-.065**	-.062**
					(.041)	(.041)					(.025)	(.026)
N. of obs.	13,083	13,083	13,083	13,083	13,083	13,083	13,083	13,083	13,083	13,083	13,083	13,083
R-squared	.33	.33	.33	.33	.33	.33	.31	.31	.31	.31	.31	.31

Notes: This table is based on Table 1 of Singh, Romero and Muralidharan (2022), but adding two columns per panel to include the heterogeneity by sibling. Observations across panels have been adjusted to information in this new variable. The original table estimates learning losses between August 2019 and December 2021 from COVID-19 for mathematics and Tamil ((the local language) using equation $Y_{it} = \alpha_v + \beta_1 Dec2021_t + \beta_2 X_{it} + \epsilon_{it}$ where X_{it} includes demographic characteristics such as sex, mother's education, and socioeconomic status interacted with $Dec2021$. Their results are shown in columns 1 through 4 and 7 through 10. In columns 5, 6, 11, and 12 I include the estimate on learning losses of having a sibling (2-10 years old) and similarly find heterogeneity that is 15%-20% of the total losses. ***p<0.01, **p<0.05, *p<0.1..

Appendix

Appendix A: Additional Tables and Figures

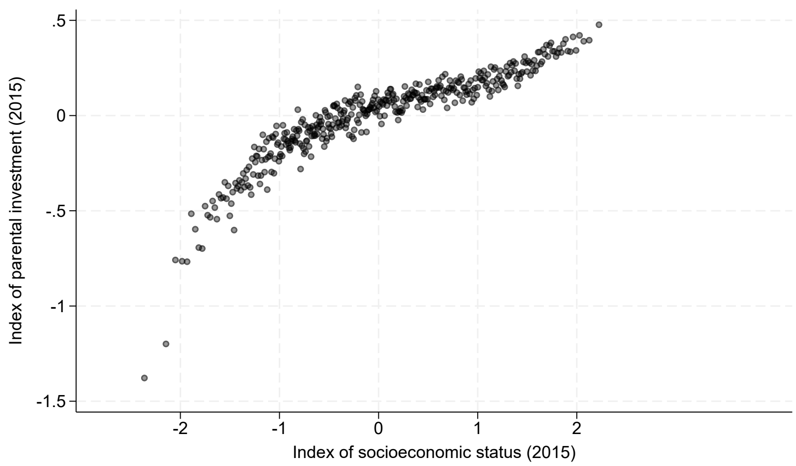
Figure A.1. : Effect of having a sibling on mathematics GPA by school characteristics



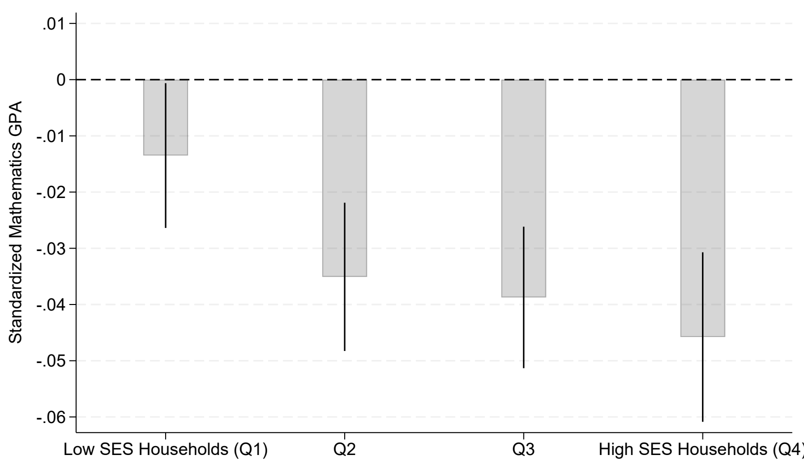
Notes: This figure shows results from the *difference-in-difference* estimates using equation (3) for the mathematics GPA standardized at the school-grade-year level for different samples of the population: all students and student in urban schools, rural schools, private schools and public schools. Each column shows results for the effect of having one, two or three siblings respectively compared to those without siblings. Standard errors are clustered at the primary school level. Vertical lines in each bar show the 95% confidence intervals.

Figure A.2. : Parental investment and socio-economic status

(a) Relationship between parental SES and time investment in 2015

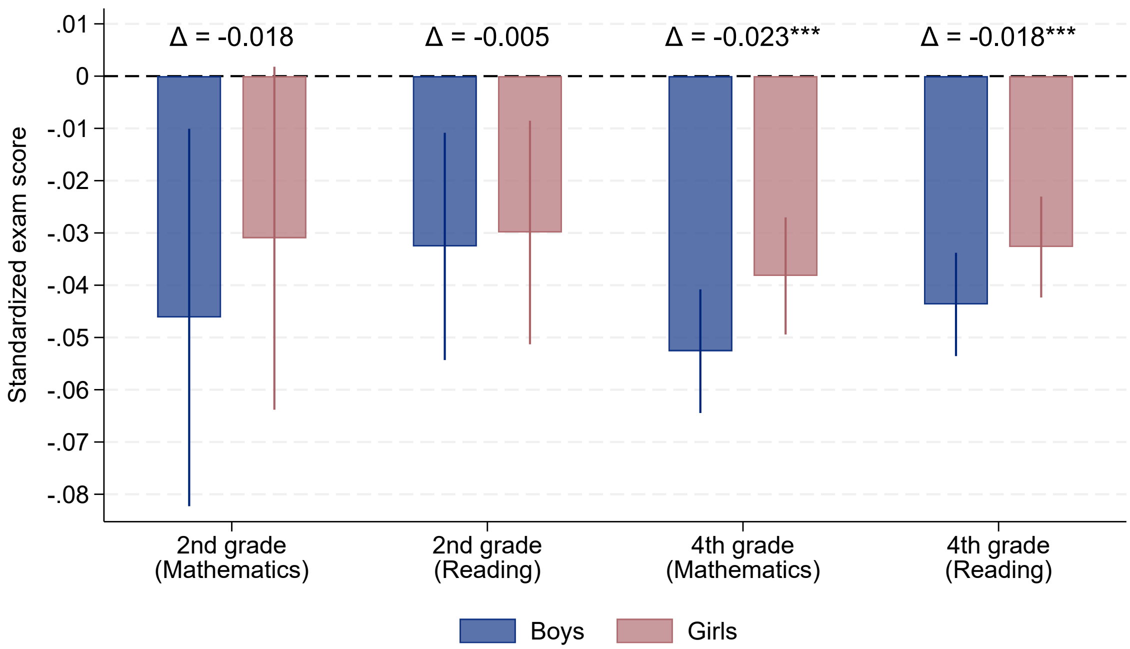


(b) Heterogeneity of sibling effect by socio-economic status



Notes: The *top panel* shows the relationship between parental time investment index (Y-axis) and socioeconomic status index (X-axis) for the year 2015 in which the parental investment index was observed. The *bottom panel* shows results from the *difference-in-difference* estimates using equation (3) for the mathematics GPA standardized at the school-grade-year level for students in sixth, seventh and ninth grade for which baseline characteristics from standardized test scores and socioeconomic characteristics were available as described in the footnote of Table 5.

Figure A.3. : Heterogeneous effects on standardized test scores by gender



Notes: This figure shows results from the *difference-in-difference* estimates of having a sibling using equation (3) for standardized exam scores in mathematics and reading for second and fourth grade students. In each case I estimate the effect for the sample of boys and girls and show the difference on top of the bars. Standard errors are clustered at the primary school level. Vertical lines in each bar show the 95% confidence intervals. ***p<0.01, **p<0.05, *p<0.1

Table A.1—: DID estimates of having a sibling controlling for confounding heterogeneity

	Mathematics GPA					Reading GPA				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: 2019-2020 6th grade DID with 2nd grade baseline										
Sibling x Post	-0.051***	-0.048***	-0.048***	-0.047***	-0.047***	-0.034***	-0.031***	-0.031***	-0.031***	-0.031***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.011)	(0.011)
Mother Ed. x Post		-0.045***	-0.044***	-0.024*	-0.026**		-0.049***	-0.048***	-0.036***	-0.038***
		(0.012)	(0.012)	(0.013)	(0.013)		(0.012)	(0.012)	(0.013)	(0.013)
Age x Post			-0.005	-0.005	-0.003			-0.006	-0.005	-0.003
			(0.017)	(0.017)	(0.017)			(0.017)	(0.017)	(0.017)
Age 1st x Post			0.002	-0.002	-0.002			0.003	0.001	0.000
			(0.010)	(0.010)	(0.010)			(0.011)	(0.011)	(0.011)
2nd grade SES x Post				-0.047***	-0.052***				-0.022**	-0.029***
				(0.011)	(0.011)				(0.011)	(0.011)
2nd grade Score x Post					0.146***					0.151***
					(0.014)					(0.014)
Observations	193,960	193,960	193,960	193,960	193,960	192,367	192,367	192,367	192,367	192,367
Panel B: 2018-2020 6th grade DID with 2nd & 4th grade baseline										
Sibling x Post	-0.042***	-0.040***	-0.036***	-0.035***	-0.033***	-0.044***	-0.042***	-0.038***	-0.038***	-0.035***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Mother Ed. x Post		-0.015***	-0.011*	-0.004	-0.058***		-0.009	-0.004	0.001	-0.050***
		(0.006)	(0.006)	(0.007)	(0.007)		(0.006)	(0.006)	(0.007)	(0.007)
Age x Post			-0.004	-0.005	0.006			0.009	0.009	0.017*
			(0.010)	(0.010)	(0.010)			(0.010)	(0.010)	(0.010)
Age 1st x Post			0.021***	0.019***	0.016***			0.019***	0.017***	0.017***
			(0.006)	(0.006)	(0.006)			(0.006)	(0.006)	(0.006)
4th grade SES x Post				-0.016***	-0.060***				-0.010*	-0.051***
				(0.006)	(0.006)				(0.006)	(0.006)
4th grade Score x Post					0.231***					0.212***
					(0.006)					(0.006)
2nd grade Score x Post					0.104***					0.106***
					(0.008)					(0.008)
Observations	572,522	572,522	572,522	572,522	529,994	565,811	565,811	565,811	565,811	523,853
Panel C: 2019-2020 5th grade DID with 4th grade baseline										
Sibling x Post	-0.038***	-0.037***	-0.041***	-0.040***	-0.039***	-0.032***	-0.032***	-0.037***	-0.037***	-0.037***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)
Mother Ed. x Post		-0.053***	-0.056***	-0.043***	-0.074***		-0.051***	-0.055***	-0.047***	-0.074***
		(0.010)	(0.010)	(0.011)	(0.011)		(0.011)	(0.011)	(0.011)	(0.011)
Age x Post			-0.029**	-0.031**	-0.019			-0.017	-0.018	-0.008
			(0.014)	(0.014)	(0.013)			(0.014)	(0.014)	(0.013)
Age 1st x Post			-0.012	-0.013	-0.007			-0.021**	-0.022**	-0.016
			(0.010)	(0.010)	(0.010)			(0.010)	(0.010)	(0.010)
4th grade SES x Post				-0.030***	-0.062***				-0.015	-0.044***
				(0.010)	(0.010)				(0.010)	(0.010)
4th grade Score x Post					0.296***					0.269***
					(0.010)					(0.010)
Observations	384,813	384,813	384,813	384,813	384,813	380,309	380,309	380,309	380,309	380,309

Notes: This table shows results from the *difference-in-difference* estimates using equation (3) and extends those in Table 2. *Panel A* shows results for GPA in mathematics and reading standardized at the school-grade-year level for sixth grade students in 2019 and 2020 controlling for second grade baseline characteristics. *Panel B* shows results for sixth grade students in 2018 and 2020 controlling for second and fourth grade baseline characteristics. *Panel C* shows results for fifth grade students in 2019 and 2020 controlling for fourth grade baseline characteristics. Column 1 and 6 show the estimates without any controls for heterogeneous effects, analogous to what is done in equation (1). Column 2 and 7 add a dummy for the mother having higher education. Column 3 and 8 add dummies for both mother's age and age at first birth being above 30. Column 4 and 9 add a dummy for baseline socioeconomic status being above the mean. Column 5 and 10 add a dummy for baseline standardized test scores being above the mean. Estimated standard errors, reported in parentheses, are clustered at the primary school level. ***p<0.01, **p<0.05, *p<0.1.

Table A.2—: DID estimates of having a sibling on mathematics GPA by baseline achievement and expectations

	1 sibling	2 siblings	3 siblings
	(1)	(2)	(3)
Panel A: Low achievement students (Q1)			
	-0.002	-0.041***	-0.103***
	(0.008)	(0.009)	(0.015)
Observations	238,723	201,254	175,478
Panel B: High achievement students (Q4)			
	-0.047***	-0.099***	-0.149***
	(0.008)	(0.012)	(0.023)
Observations	352,745	272,727	236,028
Panel C: Low Education Expectations: Below 5-year college			
	0.008	-0.022	-0.064**
	(0.014)	(0.017)	(0.026)
Observations	81,715	69,691	60,595
Panel D: High Education Expectations: 5-year college+			
	-0.031***	-0.065***	-0.097***
	(0.004)	(0.006)	(0.011)
Observations	980,339	780,019	675,846

Notes: This table shows results from the *difference-in-difference* estimates using equation (3) for the mathematics GPA standardized at the school-grade-year level for students in sixth, seventh and ninth grade for which baseline characteristics from standardized test scores and socioeconomic characteristics was available. Column 1 through 3 show the estimate of having one, two a three siblings respectively when compared to those without a sibling. I identified four subsamples of students for grades that were tested before and during closures that also had baseline characteristics: sixth grade students in 2019 and 2020 with baseline characteristics from second grade standardized exams in 2015 and 2016, sixth grade students in 2018 and 2020 with baseline characteristics from fourth grade standardized exams in 2016 and 2018, seventh grade students in 2019 and 2021 with baseline characteristics from fourth grade standardized exams in 2016 and 2018, and ninth grade students in 2019 and 2020 with baseline characteristics from second grade standardized exams in 2018 and 2019. *Panel A* shows results for students in the first (lowest) quartile of achievement according to baseline standardized exams. *Panel B* shows results for students in the fourth (highest) quartile of achievement. *Panel C* shows results for students whose parents had low expectations for their educational attainment (less than 5-year college). *Panel D* shows results for students whose parents had high expecations (5-year college or more). Estimated standard errors, reported in parentheses, are clustered at the primary school level. ***p<0.01, **p<0.05, *p<0.1.

Appendix B: Other Covid Shocks

If the additional sibling is causing learning losses, is this due to school closures or other related shocks? In this Appendix, I provide evidence that the negative effects of a sibling during this period is in fact due to school closures and not other shocks. Although health or income shocks may be important to explain overall learning losses, they are unrelated to the sibling effects I identify.

A. School Closures

First, I examine what happens when some schools reopened while most remained closed. A small subset of schools resumed operations, either fully or partially, in 2021. I estimate the event study for this group of schools. Figure B.1 shows in Panel (a) that while the effects are similar across most schools in 2020 and 2021, when full school closures were in place, the effects are smaller in 2021 for schools that reopened partially. This attenuation of the effects is consistent with a return to in-person classes.

B. Income Shocks

First, in panel (b) of Figure B.1, I show, using data from household surveys, that average household income per capita declined sharply during the first half of 2020 but had mostly recovered by 2021. The much larger shock in the first year of closures is inconsistent with the roughly equal-sized effects observed in Figure 2.

Second, because I have access to national data, I estimate the effect of having a sibling separately for each region and compare it with regional variation in GDP growth. In panel (c) of Figure B.1, I show that the two are essentially uncorrelated: regions that experienced greater GDP losses do not exhibit larger effects of having a sibling.

Finally, while there is no information on income, there is a socioeconomic index based on household characteristics. In Table B.1, I show results for 2022 and 2023, which are generally not significant, although in a few cases they point in the direction opposite to negative income shocks as a mechanism. That is, the socioeconomic status of larger families is either the same of has slightly improved relative to that of only children. One caveat is that this index is more rigid than income, and families could experience income shocks without an immediate effect on the socioeconomic index, which is based on materials in the household, access to services, assets owned, and parent's education level.

C. Health Shocks

Similar to the analysis of income shocks, I estimate the effect of having a sibling separately for each region and compare it with regional variation in excess deaths from COVID-19. In Panel (d) of Figure B.1, I show that the severity of the pandemic, measured by death rates, was uncorrelated with the magnitude of the estimated sibling effect. Additionally, using information from household surveys in 2021, I find no significant differences in the probability of parents or children reporting COVID-19 symptoms by number of children. On average, 2.34% of parents reported having symptoms in the last month (2.46% among those with one child and 2.09% among those with four), while 0.89% of children reported symptoms (1.11% among only children and 0.77% among those with three siblings).

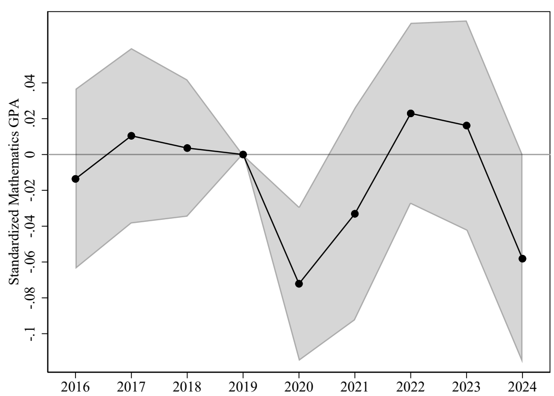
Table B.1—: DID estimates of having a sibling on socioeconomic index

	Has a sibling	1 sibling	2 siblings	3 siblings
	(1)	(2)	(3)	(4)
Panel A: 2nd grade students (2019, 2022)				
Socio Economic Index	0.013	0.007	0.005	0.006
	(0.009)	(0.009)	(0.019)	(0.069)
Observations	96,352	86,837	57,119	49,746
Panel B: 4th grade students (2019, 2022)				
Socio Economic Index	-0.011	-0.011	-0.030*	-0.038
	(0.009)	(0.009)	(0.016)	(0.039)
Observations	86,858	74,479	50,776	42,382
Panel C: 4th grade students (2018,2024)				
Socio Economic Index	0.013***	0.008*	-0.011	-0.034
	(0.004)	(0.004)	(0.009)	(0.032)
Observations	397,895	352,349	260,225	230,647

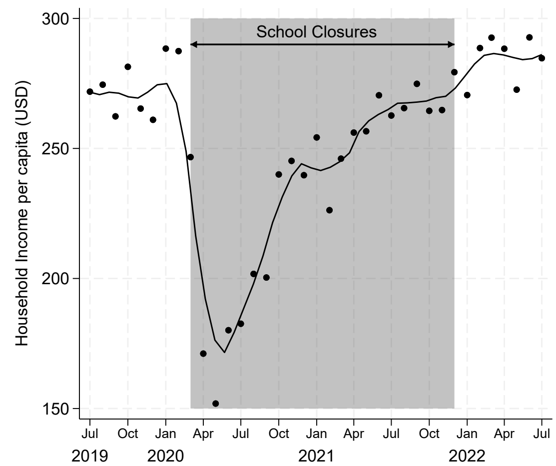
Notes: This table shows results from the *difference-in-difference* estimates of having a sibling on socioeconomic status using equation (3) for second and fourth grade students controlling for the mother having higher education for both mother's age and age at first birth being above 30. The first column shows the estimate of having a sibling when compared to those without a sibling. The second to fourth column show the estimate of having one, two a three siblings respectively when compared to those without a sibling. Each panel shows results for the effect on the socioeconomic status index based on materials in the household, access to services, assets owned, and parent's education level. . *Panel A* shows results for representative sample of second grade students who take the exam in 2019 (pre) and 2022 (post). *panel B* shows results for representative sample of fourth grade students who take the exam in 2019 (pre) and 2022 (post). *panel C* shows results for the universe of fourth grade students who take the exam in 2018 (pre) and 2024 (post). Estimated standard errors, reported in parentheses, are clustered at the primary school level. ***p<0.01, **p<0.05, *p<0.1.

Figure B.1. : Reopened schools, income shocks and health shocks

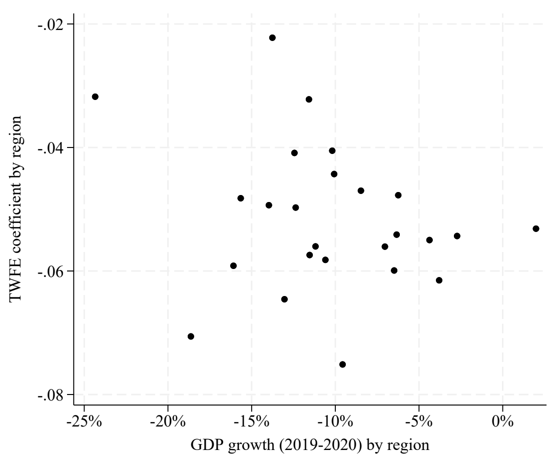
(a) Effects of having a sibling for schools that reopened in 2021



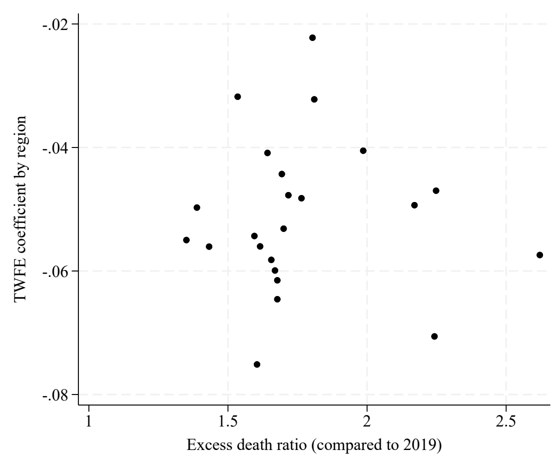
(b) Monthly average household income per capita



(c) Effect of having a sibling and GDP growth



(d) Effect of having a sibling and COVID excess death ratio



Notes: This figure shows evidence that school closures and not other covid related shocks are what is driving the effects. In the *top left panel (a)*, I estimate the event study from equation (2) for schools that reopened early in 2021. As expected, the effect on learning losses of having a sibling is attenuated for those schools in 2021. In the *top right panel (b)* I show the trend in household income per capita estimated using household surveys from 2020-2022 and show that income losses are concentrated in the first part of 2020, and yet, learning losses are equally present in 2020-2021. In the *bottom left panel (c)* I plot the estimate of the effect of having a sibling from equation (1) for each region (Y-axis) against the GDP growth from 2019 to 2020 (X-axis) and show that there is no relationship. In the *bottom right panel (d)* I plot the estimate of the effect of having a sibling from equation (1) for each region (Y-axis) against the excess death rate in 2020 compared to 2019 (X-axis) and show that there is no relationship.