Methods

Ethics Approval. Approval for this study was obtained from the Institutional Review Board of the Department of Economics at the University of Zurich (2020-079).

Participants. We have access to educational outcomes of all 6th to 12th graders in public schools directly administered by the São Paulo State Secretariat of Education over the 2018-2020 period.

Data collection. Administrative data shared by the São Paulo State Education Secretariat. We have access to quarterly data on student attendance in math and Portuguese classes, math and Portuguese scorecard test scores, and overall standardized test scores. All statistical analyses were performed within the Secretariat's secure cloud infrastructure. Only summary statistics and regressions results were directly accessible by the researchers, and no data with personal identifiers could be removed from the server. The Secretariat of Education also shared data on which municipalities in the State had issued decrees authorizing schools to resume in-person high-school classes from November onwards.

Measures. *Student dropout risk*. We define high dropout risk equal to 1 if a student had no math and no Portuguese grades on record in that school quarter, and 0 otherwise. *Standardized test scores*. For Q4-2020, only overall standardized test scores are available. For all previous quarters, we average across math and Portuguese standardized test scores. *Attendance*. We use attendance in the analysis of the effects of school reopening in the pandemic. This metric combines online and in-person attendance, and online or offline assignment completion (handing in homework through the app or in-person, at the school gate).

Analysis method. We estimate the impacts of remote learning through a differences-in-differences strategy, contrasting variation in dropout risk and standardized test scores between Q1 and Q4 in 2020 relative to that in 2019. We also present results of naive comparisons between Q4-2020 and Q4-2019, and of differences-in-differences analyses contrasting variation in dropout risk and standardized test scores between Q4-2019 and Q4-2020 relative to its 2018-2019 counterpart – both of which conflate the effects of other changes in 2020, in particular, changes in standardized tests from in-person to remote. We refine our estimates of treatment effects on test scores by matching observations based on their propensity score, the predicted probability of taking the exam within each quarter and grade, based of student and school characteristics. We also re-weight observations by the inverse of their propensity score, to obtain estimates representative for the universe of students within each grade. All analyses absorb grade fixed effects. We cluster standard errors at the school level, allowing random shocks to the outcomes of interest to be arbitrarily correlated with schools.

We also estimate heterogeneous treatment effects by municipal-level per capita Covid-19 cases over that period through non-parametric methods, after residualizing variation with respect to student and school characteristics (allowed to influence learning outcomes differentially in Q1 and Q4), to parse out other effects of disease activity on learning outcomes.

Last, we estimate intention-to-treat (ITT) effects of resuming in-person classes on educational outcomes also through a differences-in-differences strategy, but in this case contrasting municipalities which authorized schools to reopen for in-person activities to those that did not, before and after in-person classes returned for high-school students. We also undertake a triple-differences analysis, in which we contrast differences between middle- and high-school students within municipalities that allowed schools to reopen to those within municipalities that did not, before and after school reopening. We cluster standard errors at the municipality level in these analyses.

Data Availability: The dataset that supports the findings of this study cannot be made available as it contains student personal identifiers (*Código do aluno*, which uniquely identifies each student to the Education Secretariat). As such, the data are not accessible outside of the cloud infrastructure of the São Paulo State Secretariat of Education.

Code Availability: Syntax for the central claims of the paper can be found at https://github.com/Carlosalbertobdc/school-reopening.

Acknowledgments: This study was funded by the Inter-American Development Bank (IADB). We are grateful to the IADB and the São Paulo State Education Secretariat for celebrating the partnership that made it possible to access administrative data on student-level educational outcomes. We also acknowledge the São Paulo State Education Secretariat's staff members who contributed to this study, especially Guilherme Corte, Vinicius Georges and Advaldo Mesquita Moreira Junior. Last, we thank Julien Christen for his help organizing the references. The content is solely the responsibility of the authors.

Author contributions: GL, CAB, OLN and JC take responsibility for the integrity of the data and the accuracy of the data analysis. GL and OLN decided to publish the paper. CAB was responsible for analyzing the data. GL, CAB and OLN drafted the manuscript. GL and CAB contributed to statistical analysis. CAB led data management. GL, CAB, OLN and JC critically revised the manuscript. All authors had responsibility for the decision to submit for publication.

Author information: GL and OLN received fees from the Inter-American Development Bank (IADB) for the design of this study. JC is an IADB staff member. CAB declares no competing interests. Correspondence and requests for materials should be addressed to GL.

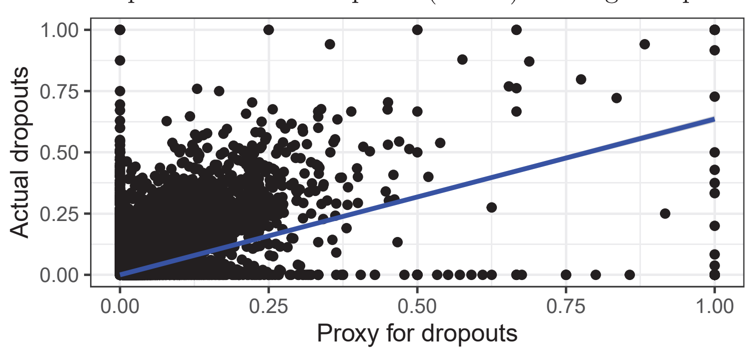
A. Validation of the proxy for student dropouts

This Appendix compiles evidence to validate our proxy for student dropouts (high dropout risk, equal to 1 if a student had no math or Portuguese grades assigned to them in the administrative data for that quarter, and 0 otherwise).

To do so, we use administrative data from the São Paulo State Education Secretariat, which includes information on both math and Portuguese grades and actual dropouts for public high-school students in 2019. Concretely, administrative dropouts equal to 1 if a student was enrolled in a State school in 2019 but not in 2020, and 0 otherwise. We restrict attention to 6th to 11th graders, as we cannot compute administrative dropouts for high-school seniors.

Figure A.1 plots the prevalence of administrative and proxy dropouts at the classroom level, for the universe of 6th-11th graders of São Paulo State. Even though administrative dropouts are measured with error – as students might not re-enroll for alternative reasons, from moving to a different State to switching over to a private school –, the figure showcases that the classroom-level actual and proxy dropouts are highly correlated, with a coefficient of approximately 0.7. While 10% of middle- and high-school students dropped out of the State public schools in 2019, that figure was over 6-fold among those with missing math and Portuguese grades by the end of the school year.

Figure A.1: Scatter plot of student dropouts (actual) and high dropout risk (proxy)



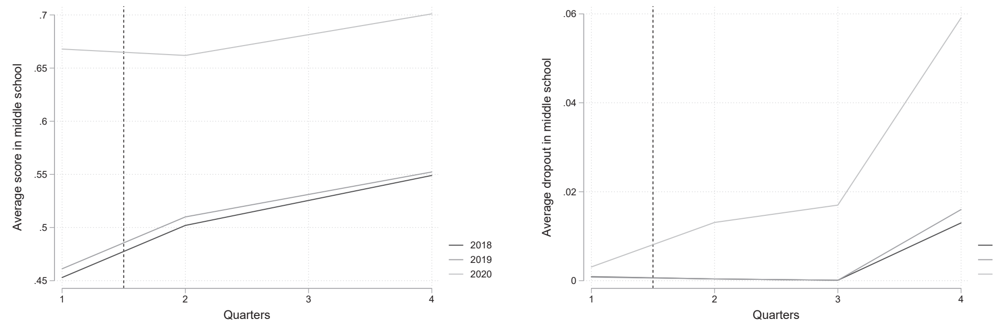
Notes: The figure plots classroom-level dropouts according to administrative data (on the vertical axis) and according to our proxy (on the horizontal axis), for Q4-2019 school year. Administrative dropouts = 1 for students enrolled at a State public school in 2019 but not in 2020, and 0 otherwise. High dropout risk = 1 for students without math and Portuguese grades on record at Q4, and 0 otherwise. The regression line is estimated through OLS.

B. Descriptive statistics

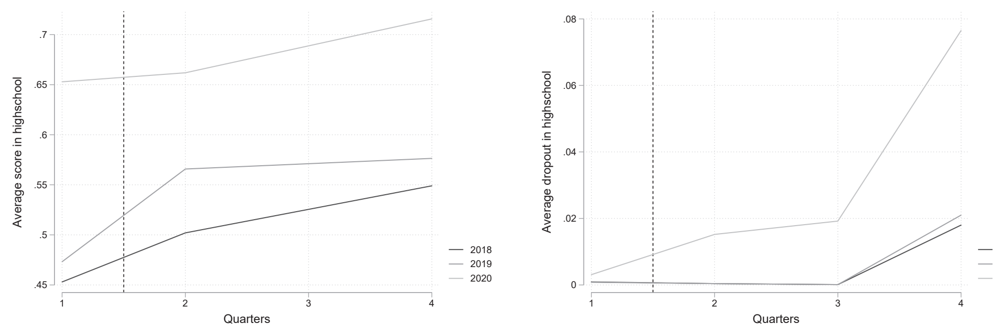
Figure B.1 showcases aggregate trends in educational outcomes across schools quarters for 2018-2020, separately for middle- and high-school students.

Figure B.1: Trends in dropout risk and standardized test scores

Panel A: Middle-school students



Panel B: High-school students



Notes: The figure showcases trends in high dropout risk and standardized test scores across school quarters for 2018, 2019 and 2020, pooling data for all middle-school students (Panel A) and all high-school students (Panel B). High dropout risk = 1 if the student had no math or Portuguese grades on record for that school quarter, and 0 otherwise. Standardized test scores from quarterly standardized tests ($AAPs$), averaging math and Portuguese scores for that school quarter.

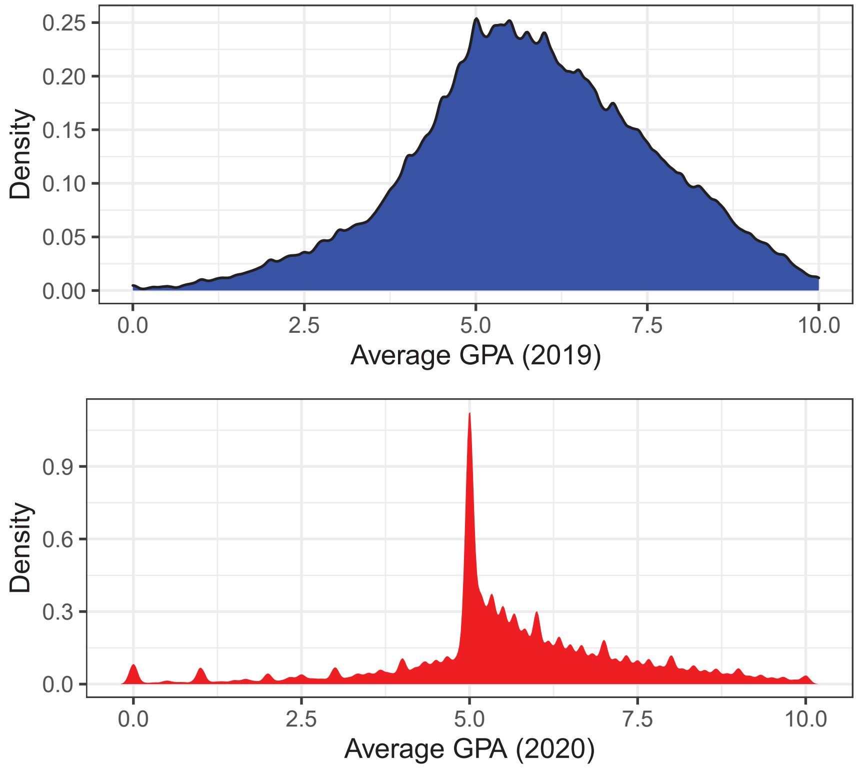
C. Comparability between 2019 and 2020 test scores

A potential concern with our econometric results is that standardized test scores in 2020 might not be comparable to those in 2019. In the main text, we discussed that they are not directly comparable in one key dimension: in 2020, exams were taken remotely. That difference is consequential – average standardized test scores are substantially higher in 2020, relative to 2019 (as Appendix B shows). We account for that difference in our empirical strategy by comparing *changes* in test scores within 2020 (under remote exams) to those within 2019 (under in-person exams). While there were other changes in standardized tests between 2019 and 2020 – in particular, the simplified curriculum recommended for Brazilian schools during the pandemic(*24*) was reflected in 2020 standardized tests(*25*) –, most importantly, such changes were *not* differential across school quarters: the Q1-2020 AAP already reflected the simplified curriculum, benefiting from re-planning efforts that happened early on, as the state of the pandemic worsened in the country.

Having said that, if standardized tests were graded *disproportionately* favourably in Q1-2020, as classes were transitioning from in-person to remote, our strategy would still over-estimate learning losses under remote classes. This Appendix provides evidence against this hypothesis. While we do observe strategic grading in Q1-2020 with respect to student GPA, we do *not* find similar evidence with respect to standardized tests scores.

Figure C.1 showcases that GPA grading changed considerably between Q1-2019 and Q1-2020. In 2019, the distribution of grades was close to a normal distribution. In 2020, in turn, we observe considerable bunching around the minimum passing grade. Besides minimum attendance, GPA is the key variable determining grade progression. Since grading for regular exams is decentralized at the teacher level, teachers might have felt like they had a mandate to try to prevent students from falling through the cracks in such a difficult time. The State later changed the grade progression rules, preventing grade repetition for almost every student in 2020 – rendering such manipulation ultimately unimportant, although revealing of teachers' strategic grading behavior.

Figure C.1 GPA distribution in Q1-2019 and Q1-2020

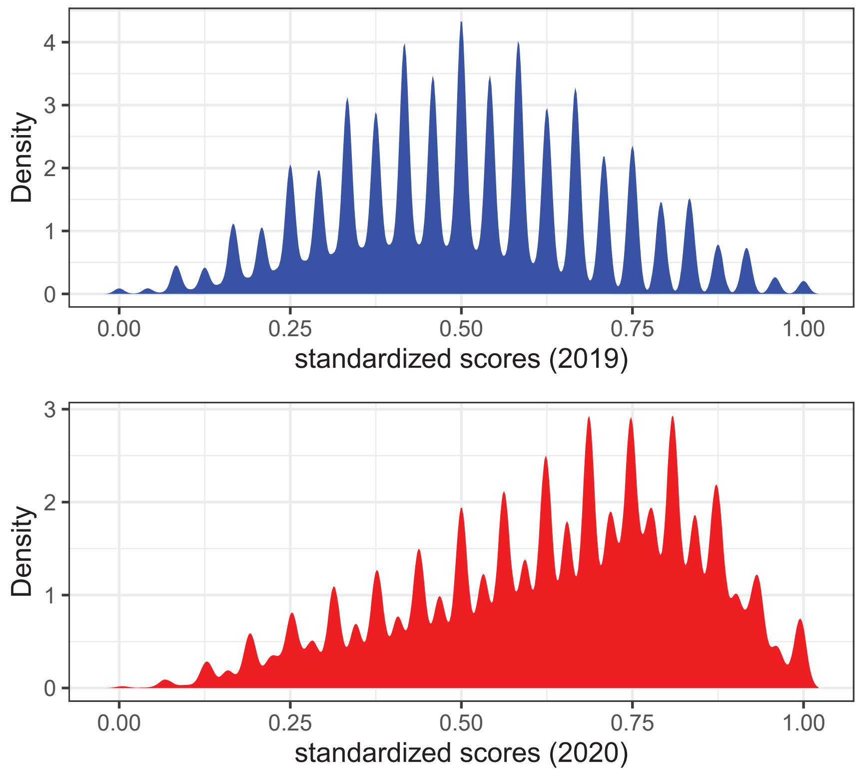


Notes: The figure shows the GPA distribution for all students in Q1-2019 and Q1-2020. Math and Portuguese GPA were averaged at the student-quarter level.

For standardized tests, however, such incentives for strategic manipulation were absent from the get-go. Such tests are optional, with no bearing on students' academic prospects, and they are not graded by the teachers, but rather by external graders.

In fact, Figure C.2 shows that, unlike GPA, the distribution of standardized tests scores in Q1-2019 displays no evidence of bunching relative to that of Q1-2020.

Figure C.2 Distribution of standardized test scores in Q1-2019 and Q1-2020



Notes: The figure shows the distribution of standardized scores for all students in Q1-2019 and Q1-2020. Math and Portuguese GPA were averaged at the student-quarter level.

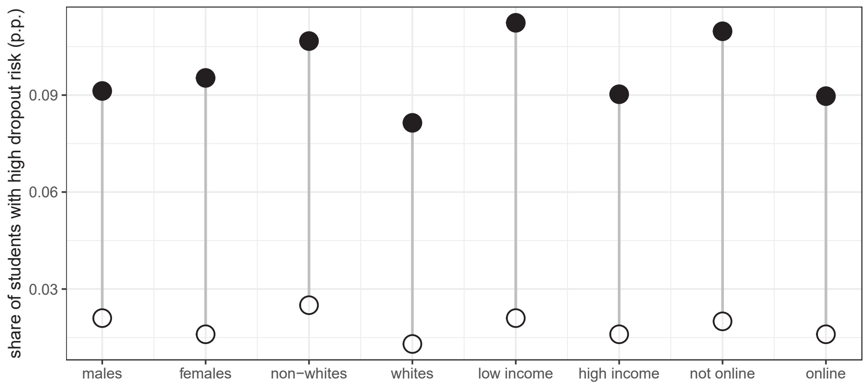
The non-smoothness in the distributions (both in 2019 and 2020) reflect the fact that we only have access to standardized test scores rounded to the closest integer. While the 2020 distribution clearly has more mass on higher scores, Appendix D shows that average standardized test scores are higher throughout 2020, relative to 2020. In fact, such scores do increase between Q1- and Q4-2020; just not as much as in the counterfactual, under in-person classes.

D. Heterogeneous treatment effects by characteristics

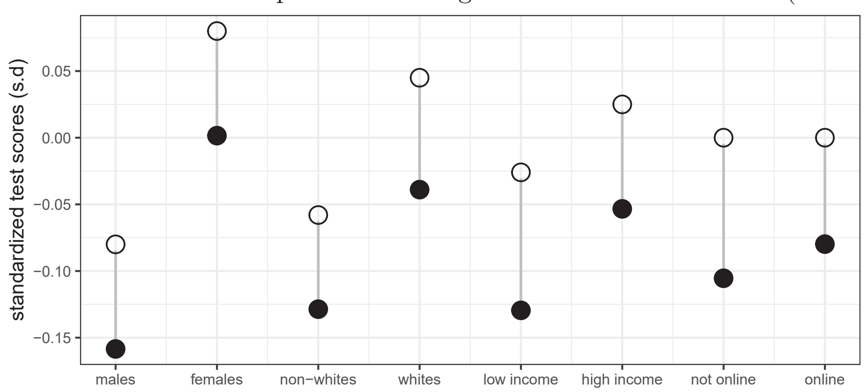
Figure D.1 showcases heterogeneous treatment effects of remote learning on high dropout risk (Panel A) and standardized test scores (Panel B) by student and school characteristics. In the figure, white dots display the Q4-2019 mean for each sub-group, and black dots display its Q4-2020 *predicted* mean, based on our regression estimates.

Figure D.1: Heterogeneous treatment effects by student characteristics

Panel A: Baseline and predicted share of students with high dropout risk (in p.p.)



Panel B: Baseline and predicted average standardized test scores (in s.d.)



Notes: The figure displays Q4-2019 sub-group means (white dots) as well as their predicted Q4-2020 means (black dots), based on regressions following the specification in Column (5) of Table 1, only restricting observations to each sub-group. Sub-groups included are (1) male and female students; (2) white (comprising also students who declare race as yellow or Asian) and non-white (black, brown and indigenous students) students, (3) schools located in below- and above-median per capita income neighborhoods, and (4) schools with and without online academic activities prior to the pandemic. In Panel A, high dropout risk = 1 if the student had no math or Portuguese grades on record for that school quarter, and 0 otherwise. In Panel B, standardized test scores from quarterly standardized tests (*AAPs*), averaging math and Portuguese scores for that school quarter. All estimates absorb grade fixed-effects, parse out the effects of school reopening, control for a third-degree polynomial of the propensity score, and re-weight observations by the inverse of their propensity score. Within each sub-group, we reject the hypothesis of equality of treatment effects on both outcomes (p-value of the joint-tests < 0.01 for each sub-group).

E. Propensity score estimation

Table E.1 presents the marginal probability changes associated with selection into non-null standardized test scores in Q4-2020, relative to Q4-2019. For illustration purposes, the table estimates across all grades, and only displays selected variables); in turn, the propensity scores that we use for both matching and re-weighting observations in the main text are estimated separately for each grade and quarter.

As the table shows, the profile of students who take the standardized test in the last quarters of 2019 and 2020 changes significantly between the two years. In particular, girls, non-white, and under-performing students are under-represented in 2020, as those in schools located in poorer neighborhoods – highlighting that matching and re-weighting are critical for unbiased estimates, especially in face of heterogeneity of treatment effects.

Table E.1: Selection into Q4 standardized test (across all grades)

	Marginal probability change
White x 2020	0.051***
	(0.001)
Male x 2020	0.005***
	(0.001)
Scorecard grade x 2020 (10 scale)	0.045***
	(0.001)
Scorecard frequency x 2020 (100 scale)	-0.001
	(0.001)
Income x 2020 (thousand R$)	0.011***
	(0.001)

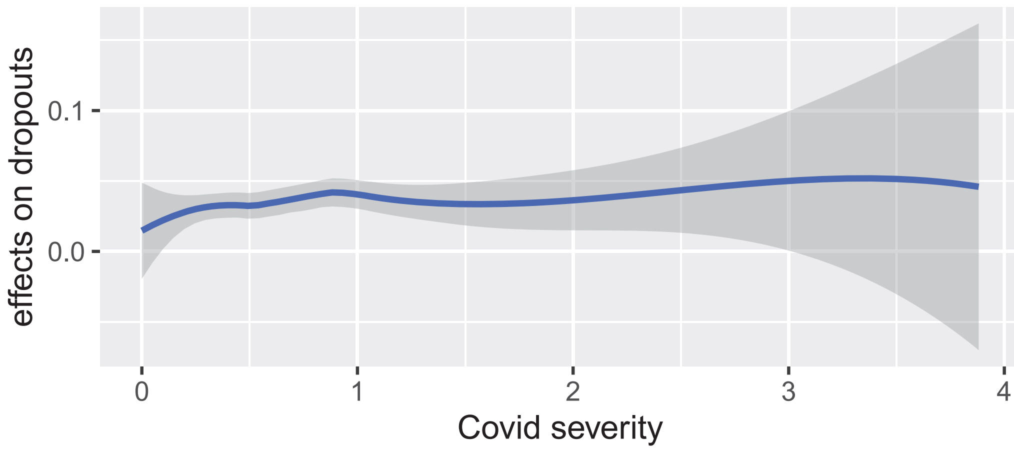
Notes: The table shows marginal probability changes associated with selected variables in a Probit model. The dependent variable is a dummy for taking at least one standardized test over the course of the school year. Additional variables not shown are indicator variables for high school (and its interaction with the 2020 indicator), whites, and males; school attendance; scorecard grades; and school neighborhood's per capita income. Standard errors clustered at the municipality level in parenthesis. * if $p<0.1$, ** $p<0.05$ and *** if $p<0.01$.

F. Non-parametric treatment effects by local disease activity

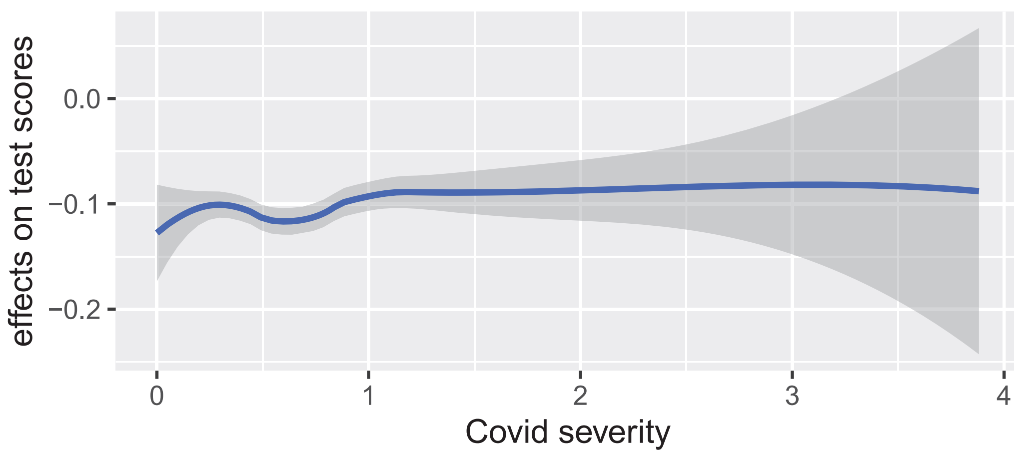
Figure F.1 estimates heterogeneous treatment effects of remote learning on high dropout risk (Panel A) and standardized test scores (Panel B) by variation in per capita Covid-19 cases between Q1- and Q4-2020. In each panel, both variables are residualized with respect to all covariates that we observe and their interactions with a Q4 indicator (= 1 in the last school quarter, and 0 otherwise).

Figure F.1: Non-parametric heterogeneous treatment effects on educational outcomes

Panel A: High dropout risk



Panel B: Standardized test scores



Notes: The figure shows local polynomial regressions of treatment effects on high dropout risk (estimated according to Column 5 of Table 1; Panel A) and treatment effects on standardized test scores (estimated according to Column 5 of Table 1; Panel B) on municipal-level per capita Covid-19 cases x Q4 (labeled as *Covid severity*). Both variables were residualized with respect to a Q4-2020 indicator, all student and school characteristics, municipal-level per capita Covid-19 cases and deaths in each of the previous quarters, as well as interactions between student and school characteristics and per capita Covid-19 cases in each quarter. Estimates are local linear regressions with bandwidth = 0.8. 95% confidence intervals displayed in grey.