KU LEUVEN

DEPARTMENT OF ECONOMICS

The effect of school closures on standardised student test outcomes

Joana Elisa MALDONADO and Kristof DE WITTE

FACULTY OF ECONOMICS AND BUSINESS

DISCUSSION PAPER SERIES DPS20.17

SEPTEMBER 2020

The Effect of School Closures on Standardised Student Test Outcomes

Joana Elisa Maldonado* Kristof De Witte†

September 22, 2020

Abstract

The school closures owing to the 2020 COVID-19 crisis resulted in a significant disruption of education provision leading to fears of learning losses and of an increase in educational inequality. This paper evaluates the effects of school closures based on standardised tests in the last year of primary school in Flemish schools in Belgium. The data covers a large sample of Flemish schools over a period of six years from 2015 to 2020. We find that students of the 2020 cohort experienced significant learning losses in all tested subjects, with a decrease in school averages of mathematics scores of 0.19 standard deviations and Dutch scores of 0.29 standard deviations as compared to the previous cohort. This finding holds when accounting for school characteristics, standardised tests in grade 4, and school fixed effects. Moreover, we observe that inequality within schools rises by 17% for math and 20% for Dutch. Inequality between schools rises by 7% for math and 18% for Dutch. The learning losses are correlated with observed school characteristics as schools with a more disadvantaged student population experience larger learning losses.

Keywords: COVID-19; school closures; learning losses; standardised tests
JEL classification: I21, I24

*Joana Maldonado (corresponding author). Leuven Economics of Education Research (LEER), KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium (e-mail: joanaelisa.maldonado@kuleuven.be. The data for this study is protected by a confidentiality agreement and we are precluded from sharing the data with others. We would like to thank Marijke De Meyst, Maarten Penninckx, Jerissa de Bilde, Anton Derks, Steven Groenez, Bieke De Fraine, Pieter Vos, Johan Geets, Teun Pauls and André Decoster for valuable comments and suggestions. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

†Kristof De Witte: Leuven Economics of Education Research (LEER), KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium (e-mail: kristof.dewitte@kuleuven.be); United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT), Boschstraat 24, 6211 AX Maastricht, the Netherlands (e-mail: k.dewitte@maastrichtuniversity.nl).

1 Introduction

In the spring of 2020, the world experienced the largest disruption of education in history which affected 94% of the world's student population (United Nations, 2020). Due to the restrictions owing to COVID-19, schools in more than 190 countries had to close for several weeks or months in order to prevent the spread of COVID-19. Although the school lockdown was initially widely accepted as a necessary measure to deal with the rising pandemic, many researchers, teachers, parents and policy-makers voiced concern about the learning losses for students and resulting educational inequality (Armitage & Nellums, 2020; Azevedo, Hasan, Goldemberg, Aroob Iqbal, & Geven, 2020; United Nations, 2020). A common critique is that policy-makers are well-informed about the benefits of school closures, by a vast amount of research modelling the spread of the pandemic, while information about the costs of school closures is lacking.

Providing policy-makers with the correct information about the costs of school closures is crucial for further management of the pandemic as well as to be able to design and implement appropriate policies to deal with the consequences of school closures. Therefore, studies about the effects of school closures are needed to clarify the extent to which students' learning can be affected and which groups of students are suffering the most when lessons at school are suspended.

Given that a school lockdown of such scope and duration is unprecedented, research about its consequences is still limited. First studies present exploratory evidence, for example from surveys (Andrew et al., 2020; Di Pietro, Biagi, Costa, Karpiński, & Mazza, 2020; Huber & Helm, 2020; Iterbeke & De Witte, 2020). The initial predictions about the effects of school closures are based on previously collected data of school interruptions and loss of instruction time, such as summer learning loss, teacher strikes, reforms or natural disasters (Bao, Qu, Zhang, & Hogan, 2020; Eyles, Gibbons, & Montebruno, 2020; Frenette, Frank, & Deng, 2020; Kuhfeld et al., 2020).[1] Other predictions rely on suggestive extrapolations based on the loss of a share of a year of schooling (Azevedo et al., 2020; Haeck & Lefebvre, 2020; Kaffenberger, 2020; Psacharopoulos, Collis, Patrinos, & Vegas, 2020).

[1] Kuhfeld et al. (2020) provide a comprehensive overview of the effects of school interruption related to summer learning loss and absenteeism found in previous literature.

These first predictions paint a negative picture: The loss in education as a share of a regular school year learning gain could result in future earning losses equivalent to 15% of future gross domestic product (Psacharopoulos et al., 2020). Based on previous studies on summer learning loss, students could suffer a reduction of the learning gains of a regular school year to 63-68% in reading and 37-50% in math (Kuhfeld et al., 2020), and kindergarten children could experience a literacy loss of 67% (Bao et al., 2020). A simulation of a calibrated model using PISA data predicts that today's grade 3 students could lose 1.5 years' worth of learning by the time they reach grade 10 (Kaffenberger, 2020). Globally, the loss of schooling could be between 0.3 and 0.9 years of schooling adjusted for quality, as predicted by simulations using data on 157 countries (Azevedo et al., 2020). Merely considering the reduction in learning time already leads to a predicted learning loss of 0.82-2.3% of a standard deviation per week (Di Pietro et al., 2020). In addition, based on Canadian PISA data, the socioeconomic skills gap could increase by as much as 30% (Haeck & Lefebvre, 2020). Surveys show that students in families with a higher income and better educated parents spent more time studying during school closures, had better studying supplies at home and received more support (Andrew et al., 2020).

While surveys can only report correlations of self-reported measures, predictions based on other contexts of reduced teaching at school disregard the unique characteristics of the COVID-19 crisis that clearly differ from past school interruptions. Clark, Nong, Zhu, and Zhu (2020) present first evidence based on data collected during the 2020 school lockdown, using a difference-in-differences approach with data from three Chinese middle schools. They find that online learning provided during the school lockdown has a positive impact on student achievement compared to the school stopping to provide any support. First experimental evidence from a randomised controlled trial in Botswana shows positive effects of a text messaging intervention during school closures (Angrist, Bergman, Brewster, & Matsheng, 2020).

This paper contributes to the existing literature related to the effects of school closures by evaluating standardised tests administered after the COVID-19 school lockdown in the final year of primary school of Flemish schools in Belgium. The data used in the analysis

covers a large number of Flemish schools over a period of six years from 2015 to 2020 and hence provide strong statistical power. The unique panel structure of the data allows us to assess school averages in different subjects over time and to identify the deviation of the 2020 data from the time trend. Since survey results indicate large differences in study time at home between students (Huber & Helm, 2020), we contribute to the identification of vulnerable groups of students by estimating marginal effects based on a wide range of school characteristics. To the best of our knowledge, this is the first study examining the impact of COVID-19 on standardised test scores.

The results show that the 2020 cohort experienced significant learning losses in all tested subjects, with a decrease in school averages of mathematics scores of 0.19 standard deviations and Dutch scores of 0.29 standard deviations compared to the previous cohort. The findings hold when accounting for school characteristics and standardised tests of grade 4 of the cohort.

Furthermore, we find that inequality both within and across schools increased in 2020, which holds when accounting for the time trend. We also find learning losses to increase in the share of students with a low socioeconomic status. The changes in inequality are hence driven by large learning losses in schools with large shares of disadvantaged students and small learning losses in schools with small shares of disadvantaged students.

This paper is structured as follows. In the next section, the setting of the analysis and the structure of the dataset are introduced. Section 3 discusses the sample and attrition and section 4 establishes the methodology. In section 5, the results are presented and section 6 provides an overview of the robustness checks. Section 7 concludes with a discussion.

2 Data

This section describes the Flemish setting of the school closures. Next, it presents the test data, which constitute the outcome variables of the analysis, and the administrative data, which is added to the test data in order to provide background characteristics of the participating schools.

2.1 Setting

In Flanders, all schools were suddenly and unexpectedly closed by the National Security Council on March 16 2020.[2] From May 18 2020 onward, the partial reopening of primary and secondary schools started under strict conditions. This nine-week period of school closures included two weeks of Easter holidays from April 6 to April 19.

In the three weeks before the Easter holidays, distance teaching took place, but teachers could only repeat and practice previously taught materials. The organisation of the distance learning was at the responsibility of schools and, therefore, the implementation and practice likely differed widely across schools. In practice, many primary schools referred their pupils to online platforms with exercises or distributed exercises on paper. Evidence suggested that not all students could be reached (up to a third of primary school students in the city of Antwerp were not reached), students lacked laptops at home, and about 12% of the students did not have a quiet place to work at home.

In the four weeks after the holidays, teachers were advised to start with so-called 'pre-teaching' in the distance learning, that is previewing new material planned to be taught once schools would reopen. It has to be noted that, due to the high degree of school autonomy in Flanders, pre-teaching was not a strict obligation but only based on guidelines. In an agreement between the Minister of Education and the education providers, this pre-teaching was agreed to be limited to a maximum of four hours per day, that is only half of a regular school day. Although there is no evidence on the precise implementation across Flemish schools, evidence from the Netherlands shows that 4 out of 10 schools estimated that students studied in distance learning even less than half the amount of time than before the school closures (Inspectie van het Onderwijs, 2020). To limit the pressure on parents, the distance learning tasks were agreed to be designed for independent completion by the student with a maximum of two hours of parental involvement per week. In terms of content, it was left to schools and teachers to choose which subjects and topics to cover in the pre-teaching. Although there is no evidence on what schools exactly taught in this period, it can be assumed that, similar to the Netherlands, most teaching hours were

[2] As evidence of the unexpected event, even a few hours before the meeting of the National Security Council, the Flemish Minister of Education declared that schools would not close.

spent on the fundamental subjects, that is language, mathematics and reading.[3]

On May 18, conditional upon some safety measures, schools were allowed to offer physical classes to the first, second and sixth grade of primary school, and the last year of secondary school, for a limited amount of hours. Students in the sixth grade of primary school could come to school for a maximum of two full days or four half days a week. In order to comply with the safety measures, such as a maximum number of 14 students per classroom, most classes needed to be split in two groups. On days that students had to stay home, pre-teaching was continued. However, school managers could independently decide whether their school could be opened safely for all the allowed grades and the maximum allowed number of hours, or less. For example, a fifth of primary schools decided to only open for one or two grades (Baert, 2020). From June 8, all grades of primary school were allowed to reopen full time.

Online as well as presence classes were continued as usually until the end of the school year on June 30. Different than in other years, schools were allowed to use the time until June 30 for teaching and assessment, giving the possibility for two weeks of additional instruction and assessment time. However, not all schools made use of this possibility.

Regarding evaluation, only a few summative and formative assessments were done during the period of school closures and reopening. The majority of schools in Flanders works with online education platforms that allow teachers to see the actions and results of their students. However, the use of such online tools and other forms of assessment differed widely across and within schools, as schools and teachers had a high degree of autonomy in the implementation of instruction and assessment.

Given a school year of 175 days, more than a third of the school year was affected by the school closures and part-time teaching at school.

[3] In the Netherlands, 70% of schools indicated to have spent 75-100% of the teaching hours on language, mathematics and reading, and for 28% of schools this amounted to 50-75% of the teaching hours (Inspectie van het Onderwijs, 2020).

2.2 Standardised Tests

The analysis is based on standardised tests that are administered every year by the network of catholic schools in Flanders (Katholiek Onderwijs Vlaanderen) in grade 6, that is the last year of primary school. Catholic schools are publicly funded, but privately-run schools and constitute the majority of schools in Flanders. The data comprises data over a time span of six years from 2015 to 2020.

The tests are administered in June, at the end of the school year. Teachers choose when to implement the tests in a time period of several weeks. The tests are designed by the network of catholic schools and serve as an internal tool for quality measurement. The test results are only shared with the school, and are not made public or shared with the central government. This ensures that teachers do not have any incentive to teach to the test and students do not have any incentive to study for these tests. At the same time, the test is still high stakes for the students, since it is an important part of the end of grade 6.

The test data is collected at individual level. In each year, the tests of the different subjects can be combined at individual level by the use of anonymous student identification numbers that teachers use when submitting the test results. However, when comparing different years, the data needs to be aggregated at school level, as this implies the comparison of different cohorts, that is different individuals. We therefore conduct the analysis at school level.

The tested subjects were slightly adjusted over the six year period under consideration. In all six years, the subjects mathematics and language (Dutch) were tested. From 2016 onward, science and social sciences were introduced in the tests, first as a combined subject (world studies) and from 2018 onward as separate subjects.[4] In 2019, French (second language) was added as an additional subject.[5]

In addition, the tests of each subject were slightly modified over the years to accommodate

[4]We use the term 'science' for the subject 'science and technology' and the term social sciences for the subject 'people and society'.

[5]The data is based on schools from the Flemish region of Belgium. While Belgium has three official languages (Dutch, French, German), Flemish schools teach exclusively in Dutch. Most students in Flemish schools have Dutch as a first language and French as a second language which they learn in school. Only a minority of students are bilingual at home.

for practical needs of the schools in order to maximise participation. For example, the listening exam in Dutch was removed due to recurring technical issues. In 2019, the sub-parts of the tests were changed in order to shorten the test by including a smaller number of questions to decrease the time investment for participating schools. A smaller number of sub-parts has been tested in 2019 and 2020 in order to increase the accuracy of testing for certain sub-domains of a subject instead of testing all sub-domains more selectively.

Based on the different test versions, we split the sample in three time periods for the analysis: First, we compare the years 2019 to 2020, since the exact same test was administered in these two years. Second, we compare the period from 2017 to 2020, and third from 2015 to 2020, by additionally controlling for the test version.

A similar standardised test as that of the end of grade 6 is also conducted at the end of grade 4 in the catholic primary schools. A unique feature of the data is that we were able to combine this test data from grade 4 of respectively two years before to the data of grade 6. Given that different anonymous identifiers are used by teachers every year to identify the students, it is not possible to merge this data at student level. Instead, we use the average grade 4 scores of the school as school characteristics in order to account for the value added in the two years before the test at the end of primary school. This use of the data disregards any changes in the student population that may have occurred between grade 4 and grade 6. The data from grade 4 covers the years 2013 to 2018 and is merged to the respective grade 6 data of two years later. In all years, the grade 4 tests covered the three subjects mathematics, Dutch and world studies.[6]

2.3 Administrative Data

In addition to the test data, the analysis makes use of administrative data at the school level. The administrative data comprises general school characteristics, such as the number of students in the school and the share of girls in the school. The data also contains information if the school is a special needs school and the share of students with special needs. Regarding teachers, the data contains the number of teachers at the school in absolute terms as well as full time equivalents (FTE) by age group. We use this information

[6]World studies is a combined subject of the two fields 'people and society' and 'science and technology'.

to compute the share of teachers above the age of 50, which might matter in the context of COVID-19 as older persons might be more risk averse, and hence, put more pressure on the school management to not reopen the schools.

Furthermore, the administrative data contains a rich set of measures of socioeconomic status (SES). These include the share of students coming from a disadvantaged neighbourhood, the share of students with a mother with a low level of education, the share of students who receive financial support from the government and the share of students who speak a different language than the language of instruction at home. This set of measures is used in combination by the government to allocate funds to schools.

As an additional measure of students with an immigration background, the data also comprises the share of newcomers, that is students that speak a different language than the language of instruction at home and only moved to Belgium in the last few years.

The administrative dataset also comprises the number of students, the share of girls and the SES indicators for grade 6. In addition, for grade 6, the data contains the share of grade repetition in grade 6 and the share of slow learners, that is students with a backlog who have repeated at least one grade in the past.

Finally, to provide an overview of the differences in school closures and measures at school level to handle the situation, we add survey data on the COVID-19 crisis to the quantitative analysis. In particular, this data is based on three rounds of telephone surveys that education inspectors conducted with the school management in order to assess the situation during the school closures. Table A8 in the appendix provides an overview of the school inspection data for schools that participated in the standardised tests in 2020, and the schools that did not participate. 95% of participating schools were inspected in at least one of the three inspection rounds. In the first interview round, that is in the last week of April, 95% of catholic schools indicated to have the situation at least sufficiently under control. 97% of interviewed catholic schools indicated to reach 80-100% of their students with distance teaching.

In the end of April, 88% of catholic schools were planning to reopen the school for grade 6 in May. On May 8, 22% of interviewed schools indicated that they would not reopen or

were undecided to reopen about 10 days later. On May 20, 34% of interviewed schools indicated that they would use the maximum amount of teaching hours allowed to take place at school.

3 Sample and Attrition

The sample comprises 402 schools in 2020, 1164 schools in 2019, 1152 schools in 2018, 1062 schools in 2017, 1034 schools in 2016 and 1018 schools in 2015. Due to the voluntary participation of schools, the numbers of observations in the analyses vary. Appendix A shows the comparisons in school characteristics between schools that participated in the tests for at least one subject and schools of the same school network that did not participate in any test for each year under consideration. The t-tests show that, in all years, participating schools had on average lower shares of students with a low socioeconomic status, with most of these differences being statistically significant. In addition, participating schools have in most years higher average grades on the grade 4 tests than schools which participated in the grade 4 tests, but not the grade 6 tests. Yet, this difference is not always statistically significant. Given the population consisting of more advantaged students in the participating schools, this means that the participating schools are not necessarily representative of the overall sample of the schools in the school network and external validity could be limited.

Regarding the internal validity, a concern could be that the participating schools in the year 2020 could differ from the previous years, since the reopening of schools after the school closures could differ across schools based on school's characteristics. Table A7 in the appendix therefore also compares the participating schools of the year 2020 to the participating schools of the year 2019. Relatively to the 2019 schools, schools participating in 2020 were smaller, had a lower share of students whose mothers are lowly educated, more special needs students, less grade 6 students who experienced grade retention, and lower grade 4 scores in mathematics and Dutch. There is no overall pattern in the attrition to the 2020 sample, since, for example, most indicators of socioeconomic status appear to be balanced across the samples. In any case, in order to account for any differences in school

characteristics in the analyses, we control for the complete set of school characteristics from the administrative data.

Figure B1 in the appendix show the distribution of test scores in 2019 and 2019 for each subject. The distribution of scores is in 2020, compared to 2019, slightly more skewed to the left in most subjects with a lower mean score among participating schools. In order to account for background characteristics in the comparison of the mean, we are proceeding with regression analyses.

4 Methodology

The data of the different school years have been combined to a panel data set. This panel data structure allows for a comparison of schools over different time periods using difference-in-differences (DiD) estimation. Based on the content of the tests, we consider three different time periods for analysis: The years 2019 and 2020 can easily be compared in all subjects, since the exact same test was administered in these two years. The estimation is based on the following equation:

Equation I

$$y_{i,j} = \alpha_j + \beta_j \, COVID19 + \delta_j \, X_i + \epsilon_{i,j}$$

where $y_{i,j}$ denotes the average score in the respective subject j at the level of the school i, which is regressed on the COVID-19 dummy, that is a dummy for the year 2020, as well as a vector of school characteristics X_i. We use robust standard errors $\epsilon_{i,j}$. This equation is estimated for the subjects mathematics, Dutch, science, social sciences and French as outcome variable $y_{i,j}$. β_j thus identifies the effect of the school closures due to the COVID-19 pandemic in 2020 relatively to the 2019 data.

To account for time trends over the years and to increase the statistical power, we extend the sample to the years 2017 and 2018. The DiD estimation is repeated for the time period from 2017 to 2020, while taking the change in test versions into account. Accordingly, the estimation is based on the following equation:

Equation II

$$y_{i,j} = \alpha_j + \beta_j \, COVID19 + \gamma_j T + \delta_j \, X_i + \epsilon_{i,j}$$

where $y_{i,j}$ again denotes the average score in the respective subject j at the level of the school i, regressed on the COVID-19 dummy for the year 2020 and the set of school characteristics X_i. T are fixed effects for the test version, which was the same in 2019-2020 and 2017-2018, respectively. This second equation is estimated for the subjects mathematics, Dutch, science and social sciences, as well as an unweighted grade point average (GPA) as outcome variable $y_{i,j}$. French was not tested before 2019.

Finally, we extend the data further to the years 2015 and 2016. This third estimation for the period of 2015 to 2020 is also based on equation II. For this time period, covering all six years, the regression is only estimated for the two subjects that were tested in all years, that is mathematics and Dutch.

5 Results

In the following, the main results are introduced firstly for each subject separately, and secondly for the average of the subjects. In the third subsection, we evaluate the impact of the school closures on inequality based on a range of inequality measures. Subsection 4 presents the effects by quintile of the distribution of scores and marginal effects based on indicators of socioeconomic status.

5.1 Results for Each Subject

In this subsection, the main results are shown for each subject. Table 1 shows the results for the standardised mathematics score as outcome variable. The three panels present the results for the three different sub-samples, respectively without control variables in the first column, with school characteristics as control variables in the second column, and additionally with grade 6 school characteristics in the third column, teacher characteristics in the fourth column and grade 4 school average scores added in the last column.[7] In each

[7]The control variables for school characteristics include the number of students in the school, the share of girls, the school being a special needs school, the share of special needs students and the four SES indicators

panel, COVID-19 identifies the coefficient for the year 2020. In the second and third panel, we account in addition for the test version which changed across years.[8]

The size and significance of the COVID-19 coefficient for the mathematics score appears to be relatively robust across sub-samples and the different sets of control variables. The effect size ranges between minus 0.18 standard deviations and minus 0.25 standard deviations, significant at, at least, the 5%-level. This means that students of the 2020 cohort have school averages in mathematics between one fifth and one fourth of a standard deviation lower than students participating in the standardised tests in the five previous years.

Table 1: Main Result: Mathematics Score

	Mathematics Score				
2019-2020					
COVID-19	-0.190**	-0.231***	-0.252***	-0.249***	-0.186**
	(0.087)	(0.078)	(0.078)	(0.078)	(0.089)
N	1287	1287	1287	1287	856
2017-2020					
COVID-19	-0.190**	-0.226***	-0.248***	-0.243***	-0.203**
	(0.087)	(0.078)	(0.078)	(0.077)	(0.089)
Test Version	0.025	0.074**	0.069**	0.051	0.050
	(0.036)	(0.033)	(0.033)	(0.034)	(0.037)
N	3474	3474	3474	3474	2346
2015-2020					
COVID-19	-0.180**	-0.206***	-0.232***	-0.237***	-0.187**
	(0.085)	(0.077)	(0.076)	(0.076)	(0.088)
Test Version	0.007	0.033***	0.026**	0.023*	0.024*
	(0.013)	(0.012)	(0.012)	(0.012)	(0.014)
N	5518	5518	5518	5518	3792
School Characteristics	No	Yes	Yes	Yes	Yes
Characteristics Year 6	No	No	Yes	Yes	Yes
Teachers	No	No	No	Yes	Yes
Year 4 Scores	No	No	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020).

Table 2 shows the same regression results for the standardised Dutch scores. Again, the three panels show three different samples of years, with the control variables added se-

(financial support, neighbourhood, mother's education and home language). Characteristics of year 6 include the number of students, the share of girls, the four SES indicators, the share of grade repetition and the share of slow learners. The teacher characteristics include the number of teachers and the share of teachers above 50 years of age. The year 4 scores include the school average of grade 4 in mathematics, Dutch and world studies, standardised by year among the schools that participated in the year.

[8]A learning effect when the same test is given in two subsequent years could be possible. However, controlling for 2018 and 2020 being the second year with the same test, leads to a similar result as the main results, meaning that the tests in 2018 and 2020 were similar to the respective previous year.

quentially in each column. In all specifications, the effect sizes are slightly larger, that is more negative, for Dutch scores than for the mathematics scores. In the specification with all control variables and the sample of all six years, the 2020 cohort has a decrease in Dutch scores with 0.26 standard deviations, significant at the 1%-level. This is surprising, since the literature about school interruptions shows learning losses to be larger in mathematics than in reading (Kuhfeld et al., 2020). A possible explanation could be that mathematics is easier to teach in distance learning, as it is simple to provide exercises and tests digitally or as worksheets. As an alternative explanation, from table A6 in the appendix, we observe that about 19% of the students do not speak Dutch at home, such that the loss from these students drives the observed effect. We analyse this more into depth in section 5.4.

Table 2: Main Result: Dutch Score

	Dutch Score				
2019-2020					
COVID-19	-0.237***	-0.258***	-0.262***	-0.274***	-0.286***
	(0.063)	(0.060)	(0.061)	(0.061)	(0.076)
N	1480	1479	1479	1479	982
2017-2020					
COVID-19	-0.237***	-0.248***	-0.257***	-0.268***	-0.280***
	(0.063)	(0.059)	(0.060)	(0.060)	(0.074)
Test Version	0.060*	0.105***	0.099***	0.073**	0.056
	(0.036)	(0.033)	(0.033)	(0.033)	(0.037)
N	3658	3657	3657	3657	2469
2015-2020					
COVID-19	-0.212***	-0.213***	-0.225***	-0.246***	-0.255***
	(0.060)	(0.056)	(0.057)	(0.057)	(0.070)
Test Version	0.017	0.044***	0.038***	0.035***	0.030**
	(0.013)	(0.012)	(0.012)	(0.012)	(0.013)
N	5697	5696	5696	5696	3913
School Characteristics	No	Yes	Yes	Yes	Yes
Characteristics Year 6	No	No	Yes	Yes	Yes
Teachers	No	No	No	Yes	Yes
Year 4 Scores	No	No	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020).

Table 3 shows the main results for the social sciences scores for the two available subsamples of 2019-2020 and 2017-2020. For this subject, the results are less pronounced. In the specification without control variables, the effect is found to be negative, but small and not significant. The same goes for the model with all control variables. Only in the

specifications with the different sets of school characteristics but without grade 4 scores, the effect of COVID-19 becomes significant. However, even in those specifications, the effect sizes are smaller than those for Dutch and mathematics. Therefore, it can be concluded that, in social sciences, the COVID-19 related school closures did not lead to a significant decrease in test scores. As an explanation, it is possible that the topics covered in the text were covered in the part of the school year before the school closures or that all test-relevant topics have sufficiently been covered during the distance teaching and reopening periods.

Table 3: Main Result: Social Sciences Score

	Social Sciences Score				
2019-2020					
COVID-19	-0.076	-0.131**	-0.123*	-0.135**	-0.070
	(0.069)	(0.066)	(0.067)	(0.067)	(0.068)
N	1073	1073	1073	1073	755
2017-2020					
COVID-19	-0.076	-0.143**	-0.142**	-0.155**	-0.086
	(0.069)	(0.066)	(0.065)	(0.065)	(0.068)
Test Version	0.019	0.059	0.050	0.025	0.016
	(0.045)	(0.041)	(0.040)	(0.040)	(0.044)
N	2408	2408	2408	2408	1764
School Characteristics	No	Yes	Yes	Yes	Yes
Characteristics Year 6	No	No	Yes	Yes	Yes
Teachers	No	No	No	Yes	Yes
Year 4 Scores	No	No	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). Social sciences was not tested in 2015 and 2016.

For science, however, the results found for Dutch and mathematics are confirmed. Table 4 shows the main results for the standardised science score for the two available sub-samples of 2019 to 2020 and 2017 to 2020. In all sub-samples and specifications, the COVID-19 coefficient is negative and statistically significant, ranging from a decrease of 0.22 standard deviations to a decrease of 0.33 standard deviations. In the fully saturated specification for the full sample of 2017 to 2020, science scores decreased by 0.32 standard deviations in 2020 compared to previous years.

Finally, Table 5 shows the main results for French. French has only been tested in the last two years and can, therefore, only be compared from 2019 to 2020. In all specifications,

Table 4: Main Result: Science Score

	Science Score				
2019-2020					
COVID-19	-0.238**	-0.220**	-0.239**	-0.232**	-0.333***
	(0.109)	(0.098)	(0.098)	(0.099)	(0.103)
N	836	836	836	836	588
2017-2020					
COVID-19	-0.238**	-0.237**	-0.254***	-0.249**	-0.323***
	(0.109)	(0.096)	(0.097)	(0.097)	(0.103)
Test Version	0.023	0.065	0.051	0.034	0.035
	(0.046)	(0.041)	(0.040)	(0.040)	(0.042)
N	2163	2163	2163	2163	1592
School Characteristics	No	Yes	Yes	Yes	Yes
Characteristics Year 6	No	No	Yes	Yes	Yes
Teachers	No	No	No	Yes	Yes
Year 4 Scores	No	No	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). Science was not tested in 2015 and 2016.

French scores are found to be significantly lower in 2020 as compared to the year 2019, ranging from a decrease of 0.19 standard deviations to a decrease of 0.3 standard deviations, significant at the 1%-level. This means that the effect sizes are comparable in Dutch, science and French, and only slightly smaller in mathematics. Social sciences is thus the only subject in which no significant decrease in standardised test scores is found.

Table 5: Main Result: French Score

	French Score				
2019-2020					
COVID-19	-0.191***	-0.267***	-0.267***	-0.275***	-0.301***
	(0.068)	(0.062)	(0.062)	(0.062)	(0.076)
N	1325	1324	1324	1324	880
School Characteristics	No	Yes	Yes	Yes	Yes
Characteristics Year 6	No	No	Yes	Yes	Yes
Teachers	No	No	No	Yes	Yes
Year 4 Scores	No	No	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). French was not tested in 2015, 2016, 2017 and 2018.

5.2 Results as Grade Point Average (GPA)

In order to consider overall student performance instead of separate subjects, we combine the scores of mathematics, Dutch, social sciences and science to an overall GPA measure. Table 6 presents the results with this combined outcome measure representing the average of the four subjects. The results show a significant decrease in GPA of 0.25 standard deviations in the fully saturated model for the 2017-2020 sample.

Table 6: Main Result: GPA - Mathematics, Dutch, Science And Social Sciences

	GPA: Mathematics, Dutch, Science And Social Sciences				
2019-2020					
COVID-19	-0.225*	-0.297***	-0.314***	-0.289***	-0.228**
	(0.130)	(0.109)	(0.110)	(0.108)	(0.116)
N	719	719	719	719	513
2017-2020					
COVID-19	-0.225*	-0.306***	-0.327***	-0.310***	-0.247**
	(0.130)	(0.108)	(0.107)	(0.105)	(0.112)
Test Version	0.021	0.088**	0.078*	0.044	-0.019
	(0.048)	(0.040)	(0.040)	(0.040)	(0.041)
N	2005	2005	2005	2005	1499
School Characteristics	No	Yes	Yes	Yes	Yes
Characteristics Year 6	No	No	Yes	Yes	Yes
Teachers	No	No	No	Yes	Yes
Year 4 Scores	No	No	No	No	Yes

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). Science and social sciences were not tested in 2015 and 2016.

5.3 Inequality Assessment

Next to the decrease in average test scores, it is relevant to investigate the change in the spread of test scores, since researchers expect an increase in inequality as a result of school closures (Armitage & Nellums, 2020; Haeck & Lefebvre, 2020; Psacharopoulos et al., 2020). We therefore assess the changes in a range of inequality measures both within and across schools.

As the tests are administered at individual level and we only aggregate the data at school level in order to compare the different years, it is possible to make use of the individual level data of each year to compute the inequality within schools. Table 7 shows the re-

sults for three inequality measures that capture inequality within schools for the subjects mathematics and Dutch.

The Gini coefficient takes a value of 0 for perfect equality and 1 for perfect inequality, meaning that higher values are associated with higher levels of inequality. Given a mean within-school Gini coefficient from 2015 to 2019 of 0.12 for mathematics and 0.1 for Dutch, the Gini coefficient increases due to the COVID-19 crisis by 0.02 in both mathematics and Dutch, significant at the 1%-level. This corresponds to an increase by 17% and 20% for mathematics and Dutch, respectively. This suggests that inequality increased significantly within the schools.

The 90/10 ratio is defined as the ratio of the score of the 90^{th} percentile to the score of the 10^{th} percentile. With a mean of 1.9 in mathematics and 1.7 in Dutch from 2015 to 2019, inequality as measured by the 90/10 ratio increases by 0.23 for mathematics and 0.22 for Dutch, both significant at the 1%-level. This suggests that the difference between the top and bottom performers increased due to the school closures.

A similar increase in within-school inequality is observed for the generalised entropy index, which, with a mean of 0.04 in mathematics and 0.03 in Dutch from 2015 to 2019, increases by 0.01 in both subjects, significant at the 1%-level. All inequality measures under consideration thus show an increase in inequality within schools, that is a widening of the spread in scores, in both mathematics and Dutch.

Given that the change in the entropy measure is relatively seen larger than the change in the 90/10 ratio and the Gini coefficient, and the entropy measure being more sensitive to changes at the bottom of the distribution, the rise in inequality is likely driven by a decrease in the bottom of the distribution.[9]

Similarly, inequality is found to increase across schools as well. Table 8 shows the estimations for the changes in the Gini coefficient, the 90/10 ratio and the standard deviation across schools for the full sample. Comparing inequality across schools over the years shows that the 2020 cohort also experienced an increase in inequality across schools.

The Gini coefficient is on average 0.14 for mathematics and 0.11 for Dutch from 2015 to

[9] As the cardinal properties of inequality measures as the Gini coefficient and entropy are not well-known, the comparison of their percentage changes should be interpreted with sufficient caution.

Table 7: Inequality Within Schools

	Mathematics			Dutch		
	Gini Coefficient	Ratio 90/10	Entropy	Gini Coefficient	Ratio 90/10	Entropy
2019-2020						
COVID-19	0.009***	0.162***	0.011***	0.006**	0.088*	0.005*
	(0.003)	(0.048)	(0.004)	(0.002)	(0.045)	(0.003)
N	1287	1265	1287	1478	1437	1478
Mean	0.136	2.068	0.046	0.133	2.021	0.047
2017-2020						
COVID-19	0.009***	0.157***	0.010***	0.005**	0.078*	0.004*
	(0.003)	(0.047)	(0.004)	(0.002)	(0.044)	(0.003)
N	3470	3384	3470	3655	3586	3655
Mean	0.126	1.946	0.039	0.109	1.763	0.030
2015-2020						
COVID-19	0.015***	0.229***	0.014***	0.018***	0.219***	0.014***
	(0.003)	(0.046)	(0.004)	(0.002)	(0.043)	(0.002)
N	5511	5379	5511	5691	5589	5691
Mean	0.122	1.902	0.036	0.100	1.680	0.025

Notes: * p<0.10, ** p<0.05, *** p<0.01. COVID-19 is a dummy variable for the year 2020. In all regressions, the control variables include school characteristics, characteristics of year 6, teacher characteristics and the test version. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). A Gini coefficient of 0 means perfect equality and a value of 1 identifies perfect inequality. The 90/10 ratio is defined as the ratio of the score of the 10þpercentile to the score of the 90þpercentile. A higher value of the 90/10 ratio indicates higher inequality. Entropy is based on a generalized entropy index GE(-1), identifying the deviation from perfect equality. The mean is the baseline mean, i.e. computed excluding the 2020 cohort.

2019, and significantly increases with 0.01 and 0.02, respectively. With an increase by 7% for mathematics and 18% for Dutch, this amounts to a similar effect size as the increase in within-school inequality. The 90/10 ratio has a mean of 2.03 for mathematics and 1.74 for Dutch and increases, respectively, by 0.1 and 0.03. Compared to the within-school effects, this is a smaller increase in inequality across schools. The entropy measure increases by 0.02 for mathematics and Dutch, from a previous mean of 0.05 for mathematics and 0.03 for Dutch.

Nevertheless, it can be argued that inequality is increasing in general over the years, which could mean that the coefficient of the COVID-19 school closures might simply capture the time trend in inequality. Table B2 in the appendix shows that the increase in inequality as a result of the school closures, both within and across schools, remains when including a time trend in the regression.

Table 8: Inequality Across Schools

2015-2020	Mathematics			Dutch		
	Gini Coefficient	Ratio 90/10	Entropy	Gini Coefficient	Ratio 90/10	Entropy
COVID-19	0.012***	0.100***	0.015***	0.020***	0.027***	0.017***
	(0.000)	(0.002)	(0.000)	(0.000)	(0.004)	(0.000)
N	5831	5831	5831	5831	5831	5831
Mean	0.139	2.030	0.046	0.113	1.738	0.031

Notes: * p<0.10, ** p<0.05, *** p<0.01. COVID-19 is a dummy variable for the year 2020. In all regressions, the control variables include school characteristics, characteristics of year 6, teacher characteristics and the test version. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). A Gini coefficient of 0 means perfect equality and a value of 1 identifies perfect inequality. The 90/10 ratio is defined as the ratio of the score of the 10þpercentile to the score of the 90þpercentile. A higher value of the 90/10 ratio indicates higher inequality. Entropy is based on a generalized entropy index GE(-1), identifying the deviation from perfect equality. The mean is the baseline mean, i.e. computed excluding the 2020 cohort.

5.4 Marginal Effects

Given the increases in inequality within and across schools found in the data, not all students have been affected equally by the school closures. It is therefore relevant to identify the groups of vulnerable students who have experienced the largest learning losses. First research predicts that the socioeconomic skills gap could increase by as much as 30% as a result of school closures (Haeck & Lefebvre, 2020). Indeed, surveys indicate that students in families with a higher income and better educated parents had an advantage in terms of material and parental support and spent more time on home learning (Andrew et al., 2020). Therefore, this subsection presents the effects by quintile of the distribution of scores and marginal effects based on indicators of socioeconomic status.

5.4.1 By Quintile of the Distribution of Scores

To examine whether the COVID-19 school closures influenced the students at the top and the bottom of the distribution differently, we assess the learning losses separately by quintile of the distribution of scores. The results do not reveal a clear pattern. Figure B2 in the appendix shows that for mathematics, learning losses are slightly larger in the bottom percentiles, but not significantly different from the top percentiles. In other words, the bottom quintiles performed less in 2020, while there is no significant decrease in the mathematics scores for the top quintiles (nor is there a significant increase in the scores of

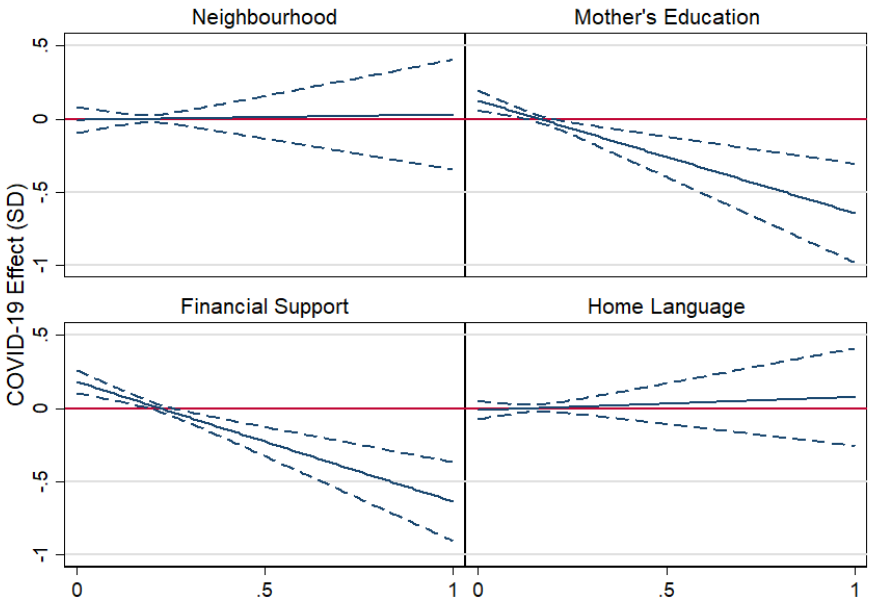
the best performing students). For Dutch, no clear pattern emerges as, irrespective of the quintile, all students underperform in 2020.

5.4.2 By Socioeconomic Status

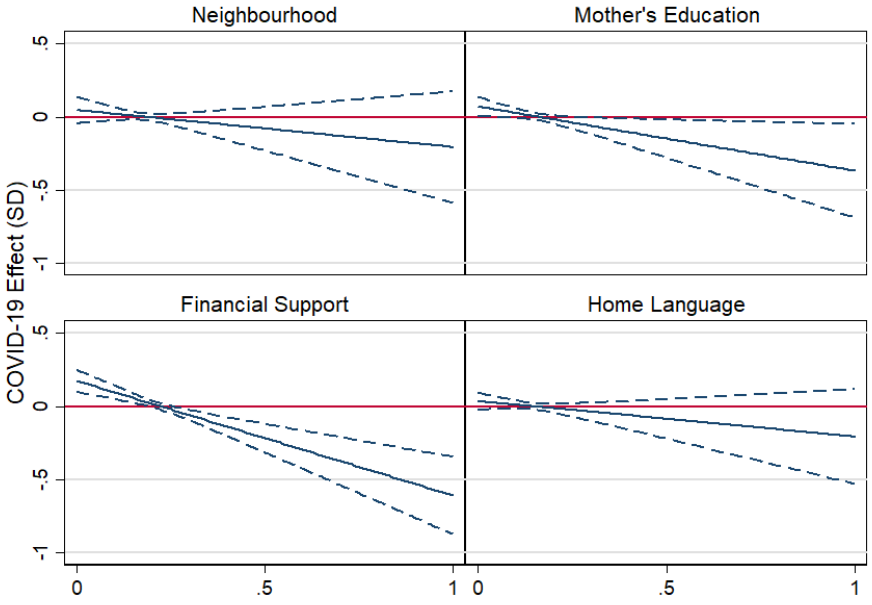
According to a survey conducted in Germany, Switzerland and Austria, less than a third of students had a high level of learning commitment with a study time of five hours or more per day during the school lockdown, while a substantial share of students only studied for two hours or less per day (Huber & Helm, 2020). We therefore decompose the effects by socioeconomic status to identify which groups of students experienced larger learning losses due to school closures.

We estimate the marginal effects based on socioeconomic status indicators in grade 6 of the school. Figure 1a and figure 1b show the marginal effects based on the four socioeconomic status indicators in grade 6 for mathematics and Dutch for the full sample of 2015 to 2020.

One measure of socioeconomic status used in the Flemish administrative school data is based on the neighbourhood where students live. Living in a disadvantaged neighbourhood, as a proxy for low socioeconomic status, could be linked to a less supportive home environment for home schooling. Schools with a higher share of students living in a disadvantaged neighbourhood can therefore be expected to have larger average learning losses in a period of school closures than schools with a lower share of students from disadvantaged neighbourhoods. The top left graph in figure 1a shows that for mathematics, the estimated effect remains constant along the share of students from a disadvantaged neighbourhood. For Dutch, however, the top left graph in figure 1b suggests larger estimated learning losses for schools with higher shares of students from a disadvantaged neighbourhood, although the observed effects are not significantly different from 0.



(a) Mathematics



(b) Dutch

Figure 1: Marginal Effects by Socioeconomic Status.

Based on the 2015-2020 sample, with the full set of control variables for school characteristics, year 6 characteristics and teacher characteristics.

A common measure of socioeconomic status and the educational background of students is the mother's education level. We therefore consider the share of students from families with a mother who obtained at best a primary education degree. During school closures,

students with a more educated mother might have received more support during home schooling in terms of parental tutoring and having been exposed to a better learning environment to stimulate learning at home. The expectation is thus that schools with a lower share of students from families with a low educated mother would have experienced smaller learning losses than schools with a higher share of low levels of mother's education. The top right graphs in figures 1a and 1b confirm this expectation for both mathematics and Dutch. In both subjects, learning losses increase in the share of students with a low educated mother.

Parental support is likely to matter for learning during the school closures in both financial and non-financial terms (Di Pietro et al., 2020). It is possible that students in families with a lower socioeconomic status had a more difficult environment for homeschooling than students in families with a higher socioeconomic status. This can be linked to the presence of practical learning facilities, such as a desk to study and a device to follow online classes, as well as the provision of educational resources, such as books, applications for learning, etc. For example, in Canadian survey data, the number of internet-enabled devices per household member has been shown to be lower in low-income families and lower income households were more likely to rely on mobile devices to access the internet (Frenette et al., 2020). We would therefore expect that schools with a higher share of students who receive financial support would experience larger average learning losses than schools with a lower share of students who receive financial support. The bottom left graphs of figures 1a and 1b show that this expectation is confirmed by the data for both mathematics and Dutch. In both subjects, the estimated effect of the school closures becomes increasingly negative as the share of students receiving financial support rises.

Finally, the language students speak at home could influence to which extent parents were able to support home schooling and to help their children with distant learning. It can therefore be expected that schools with a larger share of students that speak another language than the language of instruction at home incurred larger learning losses than schools with a smaller share of students with another home language. The bottom right graphs of figures 1a and 1b suggests that this is the case for Dutch, but not for mathematics. Since mathematics relies less on language skills, it is an intuitive result that the

learning losses do not change in the share of students who speak another language at home. Conversely, it is evident that speaking another language at home could lead to difficulties in the subject Dutch. This is confirmed, as the learning losses in Dutch seem to (although not significantly different) increase in the share of students who speak another language at home.

In summary, we observe for all four proxies of SES a decrease in performance and hence a stronger effect of the COVID-19 school closures for Dutch if the share of low SES students increases. For mathematics, the average effect remains similar along the share of students from disadvantaged neighbourhoods and students with another home language, but the estimated learning loss increases considerably in the share of students with a mother with a low education level and in the share of students who receive financial support. On all four indicators, a higher level of the SES indicator is associated with a larger confidence interval, indicating a wider spread of scores.[10]

Figure B3 and figure B4 in the appendix show that, similarly, inequality within schools and across schools, as measured by the Gini coefficient, is increasing in the share of students receiving financial support for both Dutch and mathematics, and for mathematics within schools also in the share of students with a mother with a low education level.

Figure B5 in the appendix shows that learning losses decrease in grade 4 GPA for both mathematics and Dutch, meaning that schools with higher average test results in grade 4 suffer lower average learning losses as a result of the 2020 school closures.

In addition, to see whether the learning loss is different for urban and rural areas, we make use of population data provided by Statistics Belgium to compute the marginal effect sizes based on urbanity. Figure B6 in the appendix shows that, for Dutch, there are only small differences in learning loss by population size, with slightly larger learning losses and larger confidence intervals in larger cities. For mathematics, even after accounting for all observed SES characteristics, there appears to be a clear pattern of increasing learning losses in population size, with the largest learning losses, and the largest spread of scores, in the biggest cities.

[10]The marginal effects remain robust when tested on the sub-samples of 2017-2020 and 2019-2020, as well as when using nonlinear specifications.

6 Robustness Checks

In this section, we present different robustness checks that show that the main results discussed in the previous section hold when accounting for various additional aspects. We demonstrate that the results are robust to limiting the sample using the same schools across all years or matching schools based on background characteristics, as well as using school fixed effects.

Restricting the sample to those schools that participated in the tests every year is a simple way to define a sample which is constant over time and thus holds the different school characteristics constant over time. Table 9 shows that the results of an estimation for mathematics and Dutch with schools that participated in all years is similar to the main result for all sub-samples of time periods. As observed before, there are significant negative effects for both mathematics and Dutch, of similar effect sizes of 0.22 standard deviations for mathematics and 0.35 standard deviations for Dutch. The results thus prove to be robust to changes over the years in the composition of the sample in terms of school characteristics.

Table 9: Main Regressions for Schools That Participated Every Year

	Mathematics	Dutch
2019-2020		
COVID-19	-0.283***	-0.361***
	(0.107)	(0.103)
N	246	246
2017-2020		
COVID-19	-0.272**	-0.365***
	(0.106)	(0.101)
N	492	492
2015-2020		
COVID-19	-0.223**	-0.346***
	(0.102)	(0.092)
N	738	738

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. In all regressions, the control variables include school characteristics, characteristics of year 6, teacher characteristics and the test version. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). The regressions only include schools that participated in the tests in each year.

Another approach to account for differences between schools is to include school fixed

effects. Table 10 shows the replication of the main results for mathematics and Dutch with added school fixed effects. Again, the results prove to be robust with significant negative effects in both subjects. Effect sizes are similar to the main results, with 0.17 standard deviations for mathematics, and 0.34 standard deviations for Dutch.

Table 10: Main Regressions With School Fixed Effects

	Mathematics	Dutch
2019-2020		
COVID-19	-0.240***	-0.375***
	(0.067)	(0.056)
N	1287	1479
2017-2020		
COVID-19	-0.166**	-0.352***
	(0.069)	(0.051)
N	3474	3657
2015-2020		
COVID-19	-0.165**	-0.337***
	(0.067)	(0.051)
N	5518	5696

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. All regressions include school fixed effects. In all regressions, the control variables include school characteristics, characteristics of year 6, teacher characteristics and the test version. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020).

Similarly, matching schools based on their characteristics allows for a comparison of groups of schools with similar characteristics in the different years. Matching was done using coarsened exact matching, that is using coarsened variables of characteristics in order to increase the number of matches to maximise the sample size and statistical power (Blackwell, Iacus, King, & Porro, 2009). We sequentially match schools of each year to the schools that participated in 2020.

First, we matched schools based on the school characteristics from the administrative data by matching each year's cohort to the 2020 cohort. Table 11 shows that the effects for the matched sample confirm the main results with significant negative effects in both subjects. With a learning loss of 0.22 standard deviations for mathematics and 0.25 standard deviations for Dutch, the effect size is robust to matching schools as well.

Secondly, we match, in addition to the school characteristics, further based on the average scores in grade 4. Table 12 shows the results from an estimation based on a matched

Table 11: Coarsened Exact Matching Based on School Characteristics

	Mathematics	Dutch
2019-2020		
COVID-19	-0.197**	-0.243***
	(0.088)	(0.072)
N	572	773
2017-2020		
COVID-19	-0.187**	-0.271***
	(0.085)	(0.070)
N	1238	1435
2015-2020		
COVID-19	-0.215***	-0.251***
	(0.082)	(0.065)
N	1794	1991

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. In all regressions, the control variables include school characteristics, characteristics of year 6, teacher characteristics and the test version. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). Matching of schools was done based on all school characteristics as coarsened variables of each year compared to 2020. The 2020 cohort was kept completely, while from the other cohorts only matched observations were kept, in order to maximise matching as well as statistical power.

Table 12: Coarsened Exact Matching Based on Grade Year 4 Scores

	Mathematics	Dutch
2019-2020		
COVID-19	-0.175*	-0.216***
	(0.095)	(0.079)
N	454	660
2017-2020		
COVID-19	-0.158*	-0.237***
	(0.090)	(0.077)
N	874	1078
2015-2020		
COVID-19	-0.183**	-0.231***
	(0.086)	(0.071)
N	1289	1493

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. In all regressions, the control variables include school characteristics, characteristics of year 6, teacher characteristics and the test version. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). Matching of schools was done based on school characteristics as well as grade year 4 scores in mathematics, Dutch and world studies as coarsened variables of each year compared to 2020. The 2020 cohort was kept completely, while from the other cohorts only matched observations were kept, in order to maximise matching as well as statistical power

sample using the coarsened mathematics, Dutch and world studies score averages of grade 4 in addition to the previous matching on school characteristics. Again, the main results are confirmed, the effects are negative and significant for both subjects in all sub-samples.

The effect size is decreased slightly compared to the first matching approach, with 0.19 standard deviations for mathematics and 0.23 standard deviations for Dutch. This effect size is still in line with the main results.

Finally, Table B1 in the appendix shows that the main results for mathematics and Dutch are also robust to the exclusion of participating special needs schools.

7 Conclusion and Discussion

This paper provides first evidence on the effects of school closures during the 2020 COVID-19 crisis on standardised student test scores at the end of primary school. We use a rich data set with standardised test scores from a large share of Flemish schools over a period of six years spanning from 2015 to 2020 as well as administrative data and survey data from the school inspectorate.

We find that the school closures resulted in significant learning losses and a substantial increase in educational inequality. The 2020 cohort experienced decreases in the school averages of standardised test scores as compared to previous cohorts, amounting to 0.19 standard deviations for mathematics and 0.29 standard deviations for Dutch. This finding holds when accounting for school characteristics and standardised tests in grade 4, as well as when including school fixed effects.[11]

The results thus do not only confirm the fear of significant learning losses, but also show a large effect size. To put these observed effects into perspective, Chetty, Friedman, and Rockoff (2014) observe that raising student achievement by 0.2 standard deviations results, on average, in a 2.6% increase in annual lifetime earnings. Moreover, a 0.2 decrease in standardised test scores could decrease future employment probability by 0.86% (Currie & Thomas, 2001).[12] Alternatively, in the United States, 0.2 standard deviation

[11] Similar effect sizes are substantial as Cheung and Slavin (2016) find in their meta-analysis of 197 RCTs an average effect size on academic achievement of 0.16 standard deviations. Fryer (2016) analyses 105 school-based RCTs and finds an average effect size of 0.05 standard deviations in mathematics and 0.07 standard deviations in reading.

[12] In the study, Currie and Thomas (2001) show that 'A one standard deviation increase in age 16 math scores would translate into a 14% higher wage rate at age 33 for a low or medium-SES person, compared to a return of only 11% for a high-SES person. Similarly, the same increase in age 16 test scores would increase employment probabilities by 7% among low-SES individuals compared to only 3% among high and medium-SES individuals'. Accordingly, we calculate the 0.2 decrease in standardised test scores to equal averaged for

is equivalent to approximately one fourth of the black-white achievement gap (Bloom, Hill, Black, & Lipsey, 2008). Bloom et al. (2008) demonstrate that, by the 5^{th} grade, student achievement improves about 0.4 standard deviations over the course of an academic year[13], suggesting that the COVID-19 effect is larger than what could be expected from the loss of instruction time at school. It is, hence, likely that the learning losses found for the 2020 cohort will, in the long-term, result in disadvantages on the labour market.

Moreover, the observed effect is large considering that grade 6 students could re-enter school among the first. Hence, it can be expected that the observed effects for the grade 6 cohort are a lower bound for students from other grades that returned to school only later.

Furthermore, we find that inequality both within and across schools increased as a result of the COVID-19 crisis. In addition, we find worrying results when considering marginal effects based on the indicators of socioeconomic status. The learning losses appear to be increasing in most indicators for socioeconomic status as well as population density, while they are decreasing in grade 4 scores. This means that schools with a large share of students who were already better-off in terms of their family background or previous grades suffer less learning losses than schools with a larger share of disadvantaged students.

These worrying results call for the immediate implementation of corrective policies that support disadvantaged schools and students in order to maximise the recovery of learning losses. For example, implementing classes on Saturdays and during holidays to help students catch up after the school closures could make up for at least a part of the learning losses. For future policies in further management of the ongoing COVID-19 crisis as well as other potential situations that could require school closures, this paper clearly emphasizes that school closures are associated with very high costs for students.

the three groups of socioeconomic status $\frac{(14+14+11)}{3} * 0.2 = 2.6$.

[13]It should be noted that Chingos, Whitehurst, and Gallaher (2013) argue that schools only account for a fraction of these achievement gains.

References

Andrew, A., Cattan, S., Costa-Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., ... Sevilla, A. (2020). Learning during the lockdown: real-time data on children's experiences during home learning. *IFS Briefing Note*, *BN288*.

Angrist, N., Bergman, P., Brewster, C., & Matsheng, M. (2020). *Stemming Learning Loss During the Pandemic: A Rapid Randomized Trial of a Low-Tech Intervention in Botswana.*

Armitage, R., & Nellums, L. B. (2020). Considering inequalities in the school closure response to COVID-19. *The Lancet*, *8*. doi: 10.1016/S0140-6736(20)30547-X

Azevedo, J. P., Hasan, A., Goldemberg, D., Aroob Iqbal, S., & Geven, K. (2020). Simulating the Potential Impacts of COVID-19 School Closures on Schooling and Learning Outcomes: A Set of Global Estimates. *World Bank Policy Research Working Paper*, *9284*.

Baert, D. (2020). *Weer naar school: welke leerlingen, hoeveel dagen, is er opvang en wat met de veiligheid?* Retrieved from https://www.vrt.be/vrtnws/nl/2020/05/14/weer-naar-school/#:~:text=Inhetbasisonderwijsmogenleerlingen,volledagennaarschoolkomen.

Bao, X., Qu, H., Zhang, R., & Hogan, T. (2020). *Literacy Loss in Kindergarten Children during COVID-19 School Closures.* doi: 10.31235/osf.io/nbv79

Blackwell, M., Iacus, S., King, G., & Porro, G. (2009). cem: Coarsened exact matching in Stata. *The Stata Journal*, *9*(4), 524–546.

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions. *Journal of Research on Educational Effectiveness*, *1*(4), 289–328. doi: 10.1080/19345740802400072

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, *104*(9), 2633–2679. doi: 10.1257/aer.104.9.2633

Cheung, A. C., & Slavin, R. E. (2016). How Methodological Features Affect Effect Sizes in Education. *Educational Researcher*, *45*(5), 283–292. doi: 10.3102/

0013189X16656615

Chingos, M. M., Whitehurst, G. J., & Gallaher, M. R. (2013). School Districts and Student Achievement. *Education Finance and Policy*, *10*(3), 378–398.

Clark, A. E., Nong, H., Zhu, H., & Zhu, R. (2020). *Compensating for Academic Loss: Online Learning and Student Performance during the COVID-19 Pandemic.*

Currie, J., & Thomas, D. (2001). Early test scores, school quality and SES: Longrun effects on wage and employment outcomes. *Research in Labor Economics*, *20*, 103–132.

Di Pietro, G., Biagi, F., Costa, P., Karpiński, Z., & Mazza, J. (2020). The likely impact of COVID-19 on education: Reflections based on the existing literature and recent international datasets. *Publications Office of the European Union, Luxembourg*, *EUR 30275*(JRC121071). doi: 10.2760/126686

Eyles, A., Gibbons, S., & Montebruno, P. (2020). Covid-19 school shutdowns: What will they do to our children's education? *LSE CEP COVID-19 Analysis*, *001*.

Frenette, M., Frank, K., & Deng, Z. (2020). *School Closures and the Online Preparedness of Children during the COVID-19 Pandemic* (Vol. 103; Tech. Rep.). Statistics Canada.

Fryer, R. G. (2016). The production of human capital in developed countries: Evidence from 196 randomized field experiments. In *Handbook of economic field experiments* (Vol. 2 ed., pp. 95–322). North-Holland.

Haeck, C., & Lefebvre, P. (2020). Pandemic School Closures May Increase Inequality in Test Scores. *Research Group on Human Capital Working Paper Series*, *20-03*.

Huber, S. G., & Helm, C. (2020). COVID-19 and schooling: evaluation, assessment and accountability in times of crises-reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational Assessment, Evaluation and Accountability*, *32*, 237–270. doi: 10.1007/s11092-020-09322-y

Inspectie van het Onderwijs. (2020). *COVID-19-monitor Inspectie van het Onderwijs: Wat deden scholen en instellingen, in de periode van schoolsluiting tot aan 23 april, om het onderwijs zo goed mogelijk te continueren?* (Tech. Rep.). Inspectie van het Onderwijs.

Iterbeke, K., & De Witte, K. (2020). Helpful or harmful? The role of personality traits in student experiences of the COVID-19 crisis and school closure. *KU Leuven Department of Economics Discussion Paper Series*, *20.18*.

Kaffenberger, M. (2020). Modeling the Long-Run Learning Impact of the COVID-19 Learn-

ing Shock: Actions to (More Than) Mitigate Loss. *RISE Insight Series*, *2020/017*. doi: doi.org/10.35489/BSG-RISE-RI{_}2020/017

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impacts of COVID-19 school closures on academic achievement. *EdWorkingPaper*, *20-226*. doi: 10.26300/cdrv-yw05

Psacharopoulos, G., Collis, V., Patrinos, H. A., & Vegas, E. (2020). Lost Wages The COVID-19 Cost of School Closures. *World Bank Policy Research Working Paper*, *9246*.

United Nations. (2020). *Education during COVID-19 and beyond* (Tech. Rep.). United Nations.

Appendix A: Attrition

Table A1: Attrition and Descriptive Statistics of Participating Schools: Year 2015

	Attrited		Participated		T-Test
	N	*Mean [SD]*	*N*	*Mean [SD]*	*p-value*
Number of Students	463	171.417 [95.312]	1018	189.829 [83.692]	0.000
Share of Girls	463	0.462 [0.101]	1018	0.497 [0.053]	0.000
SES – Neighbourhood	463	0.270 [0.279]	1018	0.187 [0.279]	0.000
SES - Mother's Education	463	0.198 [0.154]	1018	0.167 [0.146]	0.000
SES – Subsidies	463	0.237 [0.151]	1018	0.200 [0.145]	0.000
SES - Home Language	463	0.182 [0.204]	1018	0.145 [0.203]	0.001
Share of Newcomers	191	0.053 [0.087]	146	0.041 [0.051]	0.117
Special Needs School	463	0.276 [0.448]	1018	0.001 [0.031]	0.000
Special Needs Students	463	0.298 [0.436]	1018	0.024 [0.047]	0.000
Number of Teachers	463	18.816 [10.011]	1018	15.551 [5.689]	0.000
Number of Teachers as FTE	463	15.150 [9.523]	1018	12.141 [5.085]	0.000
Teachers: Share Above 50	463	0.279 [0.155]	1018	0.305 [0.152]	0.004
Teachers: Share Above 50 as FTE	463	0.306 [0.186]	1018	0.334 [0.180]	0.005
Year 6: Number of Students	463	26.464 [16.333]	1018	28.196 [14.609]	0.051
Year 6: Share of Girls	463	0.460 [0.132]	1018	0.500 [0.119]	0.000
Year 6: SES - Neighbourhood	463	0.260 [0.284]	1018	0.176 [0.276]	0.000
Year 6: SES - Mother's Education	463	0.188 [0.177]	1018	0.160 [0.154]	0.004
Year 6: SES - Subsidies	463	0.246 [0.170]	1018	0.211 [0.167]	0.000
Year 6: SES - Home Language	463	0.162 [0.202]	1018	0.129 [0.206]	0.004
Year 6: Grade Repetition	463	0.002 [0.008]	1018	0.002 [0.016]	0.903
Year 6: Slow Learners	463	0.131 [0.114]	1018	0.122 [0.107]	0.178
Year 4: Dutch Score	79	72.559 [8.231]	727	73.302 [7.619]	0.441
Year 4: Math Score	79	64.776 [8.852]	728	65.811 [8.276]	0.319
Year 4: World Studies Score	78	67.050 [8.591]	726	68.951 [8.385]	0.062

Notes: Attrited refers to schools in the school network that did not participate in the test. Participated refers to the schools that participated in the test for at least one subject. The values for the t-test, which compares the attrited and participating schools, are p-values. Standard deviations are robust.

Table A2: Attrition and Descriptive Statistics of Participating Schools: Year 2016

	Attrited		Participated		T-Test
	N	*Mean [SD]*	*N*	*Mean [SD]*	*p-value*
Number of Students	452	173.715 [103.577]	1034	192.616 [82.437]	0.001
Share of Girls	452	0.461 [0.103]	1034	0.497 [0.043]	0.000
SES - Neighbourhood	452	0.268 [0.279]	1034	0.189 [0.281]	0.000
SES - Mother's Education	452	0.198 [0.157]	1034	0.168 [0.145]	0.001
SES - Subsidies	452	0.246 [0.151]	1034	0.211 [0.154]	0.000
SES - Home Language	452	0.190 [0.209]	1034	0.154 [0.205]	0.002
Share of Newcomers	215	0.053 [0.114]	256	0.039 [0.050]	0.087
Special Needs School	452	0.279 [0.449]	1034	0.002 [0.044]	0.000
Special Needs Students	452	0.300 [0.438]	1034	0.025 [0.055]	0.000
Number of Teachers	452	19.051 [10.488]	1034	15.393 [5.520]	0.000
Number of Teachers as FTE	452	15.400 [9.834]	1034	12.178 [4.969]	0.000
Teachers: Share Above 50	452	0.285 [0.157]	1034	0.311 [0.155]	0.004
Teachers: Share Above 50 as FTE	452	0.312 [0.187]	1034	0.340 [0.181]	0.007
Year 6: Number of Students	452	27.162 [17.948]	1034	28.939 [14.612]	0.064
Year 6: Share of Girls	452	0.470 [0.138]	1034	0.501 [0.103]	0.000
Year 6: SES - Neighbourhood	452	0.253 [0.277]	1034	0.181 [0.278]	0.000
Year 6: SES - Mother's Education	452	0.190 [0.174]	1034	0.161 [0.161]	0.003
Year 6: SES - Subsidies	452	0.253 [0.167]	1034	0.221 [0.178]	0.001
Year 6: SES - Home Language	452	0.173 [0.212]	1034	0.134 [0.203]	0.001
Year 6: Grade Repetition	452	0.004 [0.047]	1034	0.002 [0.015]	0.283
Year 6: Slow Learners	452	0.121 [0.108]	1034	0.114 [0.104]	0.243
Year 4: Dutch Score	64	71.026 [9.192]	728	71.899 [7.777]	0.459
Year 4: Math Score	64	64.698 [8.522]	730	67.005 [7.565]	0.035
Year 4: World Studies Score	64	69.198 [8.138]	727	71.722 [7.655]	0.016

Notes: Attrited refers to schools in the school network that did not participate in the test. Participated refers to the schools that participated in the test for at least one subject. The values for the t-test, which compares the attrited and participating schools, are p-values. Standard deviations are robust.

Table A3: Attrition and Descriptive Statistics of Participating Schools: Year 2017

	Attrited		Participated		T-Test
	N	*Mean [SD]*	*N*	*Mean [SD]*	*p-value*
Number of Students	425	172.605 [99.532]	1062	196.636 [86.417]	0.000
Share of Girls	425	0.455 [0.104]	1062	0.497 [0.044]	0.000
SES - Neighbourhood	425	0.257 [0.274]	1062	0.196 [0.286]	0.000
SES - Mother's Education	425	0.191 [0.147]	1062	0.170 [0.146]	0.013
SES - Subsidies	425	0.240 [0.152]	1062	0.207 [0.155]	0.000
SES - Home Language	425	0.193 [0.208]	1062	0.163 [0.208]	0.013
Share of Newcomers	228	0.046 [0.091]	321	0.038 [0.048]	0.218
Special Needs School	425	0.289 [0.454]	1062	0.004 [0.061]	0.000
Special Needs Students	425	0.315 [0.439]	1062	0.033 [0.072]	0.000
Number of Teachers	425	19.287 [10.437]	1062	15.715 [5.975]	0.000
Number of Teachers as FTE	425	15.512 [9.710]	1062	12.479 [5.358]	0.000
Teachers: Share Above 50	425	0.289 [0.154]	1062	0.312 [0.153]	0.009
Teachers: Share Above 50 as FTE	425	0.316 [0.183]	1062	0.338 [0.176]	0.034
Year 6: Number of Students	425	27.325 [17.632]	1062	29.626 [14.938]	0.018
Year 6: Share of Girls	425	0.453 [0.133]	1062	0.496 [0.111]	0.000
Year 6: SES - Neighbourhood	425	0.247 [0.269]	1062	0.188 [0.283]	0.000
Year 6: SES - Mother's Education	425	0.188 [0.165]	1062	0.167 [0.163]	0.021
Year 6: SES - Subsidies	425	0.248 [0.164]	1062	0.212 [0.169]	0.000
Year 6: SES - Home Language	425	0.179 [0.216]	1062	0.151 [0.212]	0.024
Year 6: Grade Repetition	425	0.003 [0.016]	1062	0.003 [0.018]	0.622
Year 6: Slow Learners	425	0.113 [0.104]	1062	0.113 [0.105]	0.984
Year 4: Dutch Score	52	68.338 [10.125]	742	70.625 [8.423]	0.109
Year 4: Math Score	53	63.680 [9.165]	744	66.774 [8.254]	0.016
Year 4: World Studies Score	52	67.622 [12.324]	741	71.908 [7.338]	0.013

Notes: Attrited refers to schools in the school network that did not participate in the test. Participated refers to the schools that participated in the test for at least one subject. The values for the t-test, which compares the attrited and participating schools, are p-values. Standard deviations are robust.

Table A4: Attrition and Descriptive Statistics of Participating Schools: Year 2018

	Attrited		Participated		T-Test
	N	*Mean [SD]*	*N*	*Mean [SD]*	*p-value*
Number of Students	337	161.098 [91.857]	1152	200.300 [89.503]	0.000
Share of Girls	337	0.442 [0.118]	1152	0.497 [0.043]	0.000
SES - Neighbourhood	337	0.250 [0.251]	1152	0.198 [0.290]	0.001
SES - Mother's Education	337	0.187 [0.137]	1152	0.174 [0.150]	0.114
SES - Subsidies	337	0.246 [0.147]	1152	0.223 [0.167]	0.014
SES - Home Language	337	0.201 [0.209]	1152	0.173 [0.213]	0.033
Share of Newcomers	199	0.045 [0.096]	394	0.036 [0.046]	0.207
Special Needs School	337	0.365 [0.482]	1152	0.003 [0.059]	0.000
Special Needs Students	337	0.390 [0.465]	1152	0.034 [0.068]	0.000
Number of Teachers	337	20.409 [12.119]	1152	15.976 [6.169]	0.000
Number of Teachers as FTE	337	16.267 [10.900]	1152	12.756 [5.598]	0.000
Teachers: Share Above 50	337	0.289 [0.156]	1152	0.312 [0.155]	0.018
Teachers: Share Above 50 as FTE	337	0.321 [0.179]	1152	0.340 [0.180]	0.083
Year 6: Number of Students	337	25.801 [15.932]	1152	30.673 [15.890]	0.000
Year 6: Share of Girls	337	0.456 [0.149]	1152	0.501 [0.109]	0.000
Year 6: SES - Neighbourhood	337	0.241 [0.250]	1152	0.192 [0.288]	0.002
Year 6: SES - Mother's Education	337	0.186 [0.151]	1152	0.167 [0.165]	0.045
Year 6: SES - Subsidies	337	0.266 [0.166]	1152	0.229 [0.174]	0.000
Year 6: SES - Home Language	337	0.189 [0.211]	1152	0.160 [0.215]	0.025
Year 6: Grade Repetition	337	0.002 [0.011]	1152	0.002 [0.011]	0.425
Year 6: Slow Learners	337	0.106 [0.083]	1152	0.109 [0.099]	0.627
Year 4: Dutch Score	45	70.242 [12.709]	765	72.654 [8.069]	0.204
Year 4: Math Score	47	72.093 [10.104]	767	73.463 [7.187]	0.356
Year 4: World Studies Score	46	68.675 [9.847]	767	68.895 [8.147]	0.881

Notes: Attrited refers to schools in the school network that did not participate in the test. Participated refers to the schools that participated in the test for at least one subject. The values for the t-test, which compares the attrited and participating schools, are p-values. Standard deviations are robust.

Table A5: Attrition and Descriptive Statistics of Participating Schools: Year 2019

	Attrited		Participated		T-Test
	N	*Mean [SD]*	*N*	*Mean [SD]*	*p-value*
Number of Students	331	158.088 [92.880]	1164	201.924 [90.080]	0.000
Share of Girls	331	0.437 [0.118]	1164	0.498 [0.042]	0.000
SES - Neighbourhood	331	0.253 [0.243]	1164	0.200 [0.286]	0.001
SES - Mother's Education	331	0.181 [0.126]	1164	0.175 [0.149]	0.410
SES - Subsidies	331	0.264 [0.147]	1164	0.236 [0.174]	0.003
SES - Home Language	331	0.210 [0.213]	1164	0.180 [0.213]	0.024
Share of Newcomers	201	0.042 [0.043]	433	0.035 [0.042]	0.042
Special Needs School	331	0.378 [0.486]	1164	0.002 [0.041]	0.000
Special Needs Students	331	0.408 [0.464]	1164	0.040 [0.060]	0.000
Number of Teachers	331	21.293 [12.268]	1164	16.893 [6.332]	0.000
Number of Teachers as FTE	331	16.531 [11.195]	1164	12.837 [5.500]	0.000
Teachers: Share Above 50	331	0.260 [0.148]	1164	0.292 [0.146]	0.000
Teachers: Share Above 50 as FTE	331	0.281 [0.178]	1164	0.306 [0.168]	0.022
Year 6: Number of Students	331	25.278 [15.988]	1164	31.486 [16.033]	0.000
Year 6: Share of Girls	331	0.445 [0.145]	1164	0.497 [0.108]	0.000
Year 6: SES - Neighbourhood	331	0.250 [0.248]	1164	0.194 [0.284]	0.000
Year 6: SES - Mother's Education	331	0.177 [0.137]	1164	0.167 [0.164]	0.252
Year 6: SES - Subsidies	331	0.273 [0.155]	1164	0.241 [0.194]	0.002
Year 6: SES - Home Language	331	0.199 [0.224]	1164	0.161 [0.214]	0.007
Year 6: Grade Repetition	331	0.004 [0.024]	1164	0.001 [0.009]	0.059
Year 6: Slow Learners	331	0.099 [0.073]	1164	0.102 [0.094]	0.454
Year 4: Dutch Score	47	69.797 [9.345]	770	70.945 [8.023]	0.406
Year 4: Math Score	47	71.617 [6.435]	774	71.295 [6.990]	0.738
Year 4: World Studies Score	46	74.013 [6.439]	769	74.600 [6.249]	0.544

Notes: Attrited refers to schools in the school network that did not participate in the test. Participated refers to the schools that participated in the test for at least one subject. The values for the t-test, which compares the attrited and participating schools, are p-values. Standard deviations are robust.

Table A6: Attrition and Descriptive Statistics of Participating Schools: Year 2020

	Attrited		Participated		T-Test
	N	*Mean [SD]*	*N*	*Mean [SD]*	*p-value*
Number of Students	1103	193.976 [94.058]	402	184.654 [84.225]	0.066
Share of Girls	1103	0.482 [0.077]	402	0.496 [0.043]	0.000
SES - Neighbourhood	1093	0.217 [0.274]	401	0.198 [0.288]	0.263
SES - Mother's Education	1093	0.182 [0.148]	401	0.160 [0.133]	0.005
SES - Subsidies	1093	0.249 [0.169]	401	0.224 [0.167]	0.010
SES - Home Language	1093	0.186 [0.210]	401	0.189 [0.221]	0.808
Share of Newcomers	530	0.035 [0.037]	156	0.030 [0.030]	0.079
Special Needs School	1103	0.114 [0.318]	402	0.002 [0.050]	0.000
Special Needs Students	1103	0.160 [0.306]	402	0.051 [0.069]	0.000
Number of Teachers	1103	18.942 [9.345]	402	15.998 [6.052]	0.000
Number of Teachers as FTE	1103	14.531 [8.296]	402	11.965 [5.153]	0.000
Teachers: Share Above 50	1103	0.287 [0.148]	402	0.297 [0.144]	0.233
Teachers: Share Above 50 as FTE	1103	0.306 [0.171]	402	0.313 [0.174]	0.499
Year 6: Number of Students	1103	31.639 [16.898]	402	29.465 [15.166]	0.017
Year 6: Share of Girls	1103	0.482 [0.126]	402	0.496 [0.125]	0.071
Year 6: SES - Neighbourhood	1093	0.210 [0.274]	401	0.194 [0.285]	0.310
Year 6: SES - Mother's Education	1093	0.177 [0.164]	401	0.147 [0.139]	0.000
Year 6: SES - Subsidies	1093	0.255 [0.187]	401	0.229 [0.184]	0.014
Year 6: SES - Home Language	1093	0.170 [0.215]	401	0.168 [0.222]	0.900
Year 6: Grade Repetition	1103	0.002 [0.010]	402	0.002 [0.019]	0.543
Year 6: Slow Learners	1103	0.100 [0.086]	402	0.092 [0.092]	0.142
Year 4: Dutch Score	577	68.191 [10.515]	290	69.749 [10.291]	0.037
Year 4: Math Score	578	60.188 [6.410]	290	61.520 [6.315]	0.004
Year 4: World Studies Score	565	73.852 [8.514]	285	75.270 [5.999]	0.005

Notes: Attrited refers to schools in the school network that did not participate in the test. Participated refers to the schools that participated in the test for at least one subject. The values for the t-test, which compares the attrited and participating schools, are p-values. Standard deviations are robust.

Table A7: Comparison of Participating Schools in 2019-2020

	Participated in 2019		Participated in 2020		T-Test
	N	*Mean [SD]*	*N*	*Mean [SD]*	*p-value*
Number of Students	1164	201.924 [90.080]	402	184.654 [84.225]	0.001
Share of Girls	1164	0.498 [0.042]	402	0.496 [0.043]	0.414
SES - Neighbourhood	1164	0.200 [0.286]	401	0.198 [0.288]	0.904
SES - Mother's Education	1164	0.175 [0.149]	401	0.160 [0.133]	0.061
SES - Subsidies	1164	0.236 [0.174]	401	0.224 [0.167]	0.212
SES - Home Language	1164	0.180 [0.213]	401	0.189 [0.221]	0.480
Share of Newcomers	433	0.035 [0.042]	156	0.030 [0.030]	0.131
Special Needs School	1164	0.002 [0.041]	402	0.002 [0.050]	0.781
Special Needs Students	1164	0.040 [0.060]	402	0.051 [0.069]	0.004
Number of Teachers	1164	16.893 [6.332]	402	15.998 [6.052]	0.012
Number of Teachers as FTE	1164	12.837 [5.500]	402	11.965 [5.153]	0.004
Teachers: Share Above 50	1164	0.292 [0.146]	402	0.297 [0.144]	0.530
Teachers: Share Above 50 as FTE	1164	0.306 [0.168]	402	0.313 [0.174]	0.496
Year 6: Number of Students	1164	31.486 [16.033]	402	29.465 [15.166]	0.023
Year 6: Share of Girls	1164	0.497 [0.108]	402	0.496 [0.125]	0.859
Year 6: SES - Neighbourhood	1164	0.194 [0.284]	401	0.194 [0.285]	0.997
Year 6: SES - Mother's Education	1164	0.167 [0.164]	401	0.147 [0.139]	0.019
Year 6: SES - Subsidies	1164	0.241 [0.194]	401	0.229 [0.184]	0.245
Year 6: SES - Home Language	1164	0.161 [0.214]	401	0.168 [0.222]	0.575
Year 6: Grade Repetition	1164	0.001 [0.009]	402	0.002 [0.019]	0.207
Year 6: Slow Learners	1164	0.102 [0.094]	402	0.092 [0.092]	0.051
Year 4: Dutch Score	770	70.945 [8.023]	290	69.749 [10.291]	0.074
Year 4: Math Score	774	71.295 [6.990]	290	61.520 [6.315]	0.000
Year 4: World Studies Score	769	74.600 [6.249]	285	75.270 [5.999]	0.111

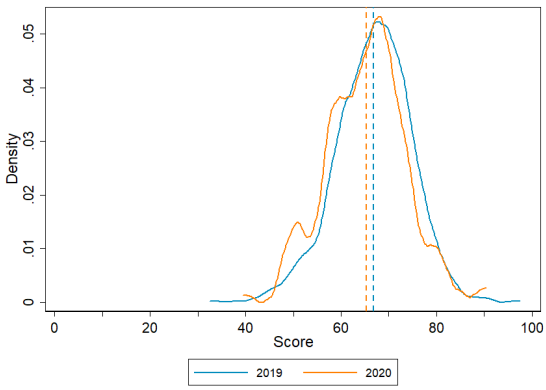
Notes: Attrited refers to schools in the school network that did not participate in the test. Participated refers to the schools that participated in the test for at least one subject. The values for the t-test, which compares the attrited and participating schools, are p-values. Standard deviations are robust.

Table A8: School Inspections 2020: Comparison of Participating and Non-Participating Schools

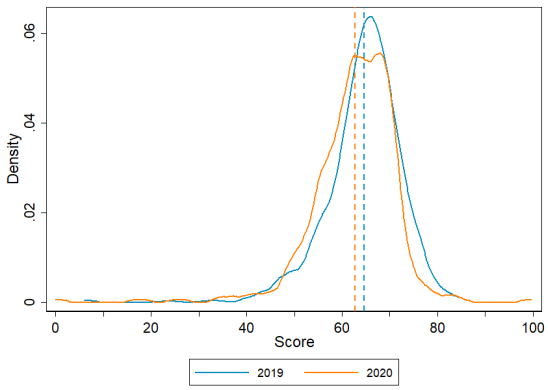
	Attrited		Participated		T-Test
	N	*Mean [SD]*	*N*	*Mean [SD]*	*p-value*
Inspected	1103	0.772 [0.419]	402	0.950 [0.218]	0.000
Inspected in Round 1A *(i.e. 24 April 2020)*	1103	0.470 [0.499]	402	0.562 [0.497]	0.001
Inspected in Round 1B *(i.e. 30 April 2020)*	1103	0.288 [0.453]	402	0.373 [0.484]	0.002
Inspected in Round 2 *(i.e. 8 May 2020)*	1103	0.104 [0.306]	402	0.139 [0.347]	0.074
Inspected in Round 3 *(i.e. 20 May 2020)*	1103	0.102 [0.302]	402	0.129 [0.336]	0.145
School has situation under control	835	2.554 [0.594]	374	2.516 [0.603]	0.303
Number of students reached with classes	509	2.967 [0.190]	221	2.973 [0.163]	0.651
Preteaching taking place	518	0.562 [0.497]	226	0.540 [0.500]	0.581
Teaching all subjects	518	0.332 [0.471]	226	0.279 [0.449]	0.143
Teaching Dutch	836	0.452 [0.498]	376	0.457 [0.499]	0.864
Teaching French	836	0.394 [0.489]	376	0.375 [0.485]	0.539
Teaching Math	836	0.450 [0.498]	376	0.457 [0.499]	0.804
Teaching World Studies	518	0.320 [0.467]	226	0.323 [0.469]	0.946
Round 1: Reopening in class year 6	836	0.340 [0.474]	376	0.356 [0.480]	0.574
Round 2: Reopening school	115	1.261 [0.497]	56	1.250 [0.548]	0.900
Round 3: max allowed hours at school	111	0.667 [0.474]	52	0.673 [0.474]	0.936

Notes: Attrited refers to schools in the school network that did not participate in the test in 2020. Participated refers to the schools that participated in the test for at least one subject in 2020. The values for the t-test, which compares the attrited and participating schools, are p-values. Standard deviations are robust.

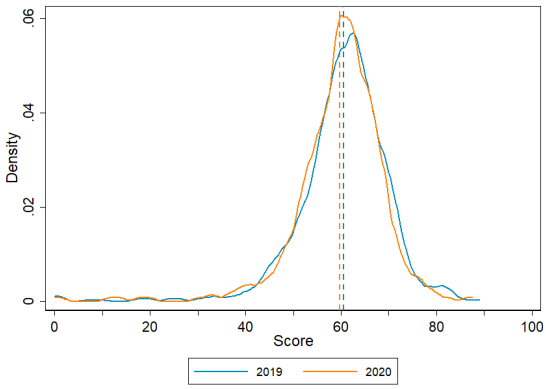
Appendix B: Additional Tables and Figures



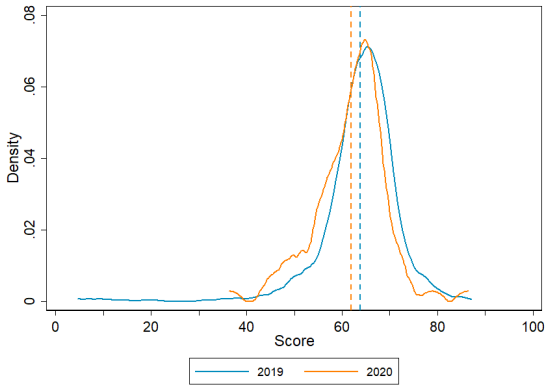
(a) Mathematics



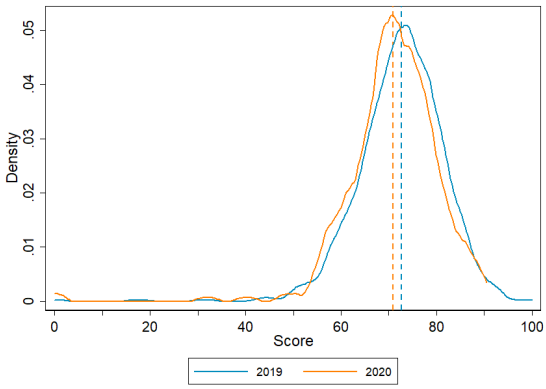
(b) Dutch



(c) Social Sciences



(d) Science



(e) French

Figure B1: Distribution of Scores in 2019-2020

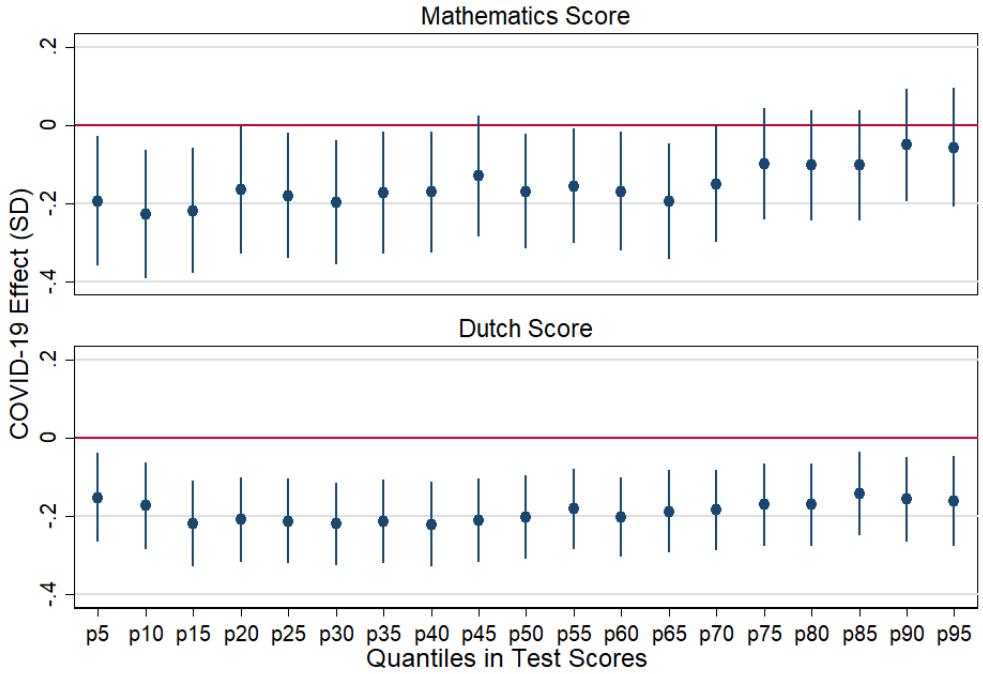


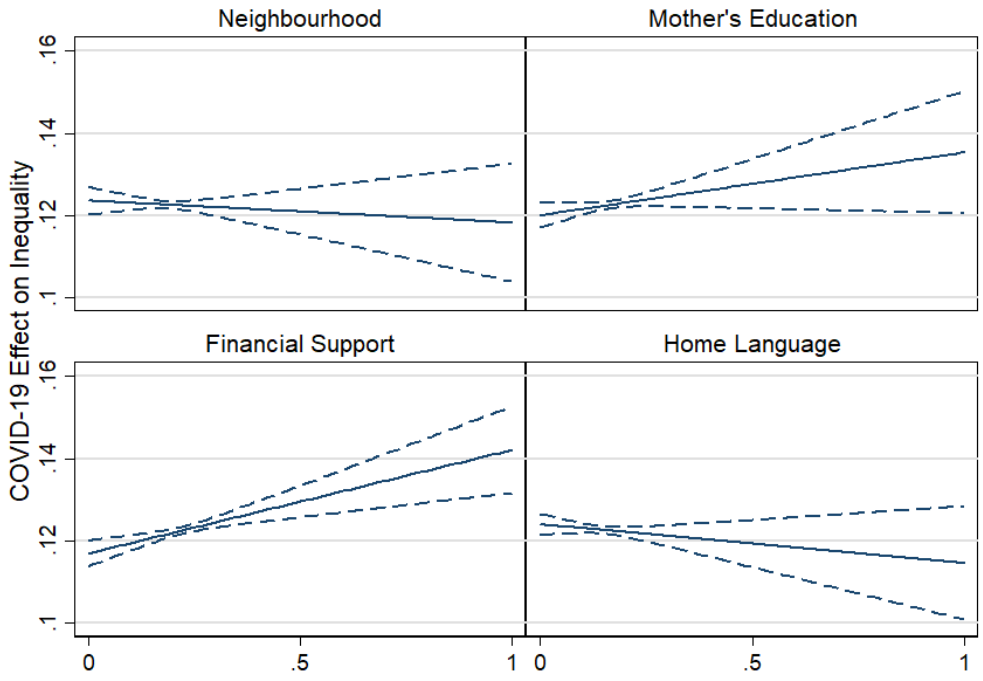
Figure B2: Estimated Effects per Quantile in Test Scores

This figure and all subsequent figures are based on the 2015-2020 sample, with the full set of control variables for school characteristics, year 6 characteristics and teacher characteristics.

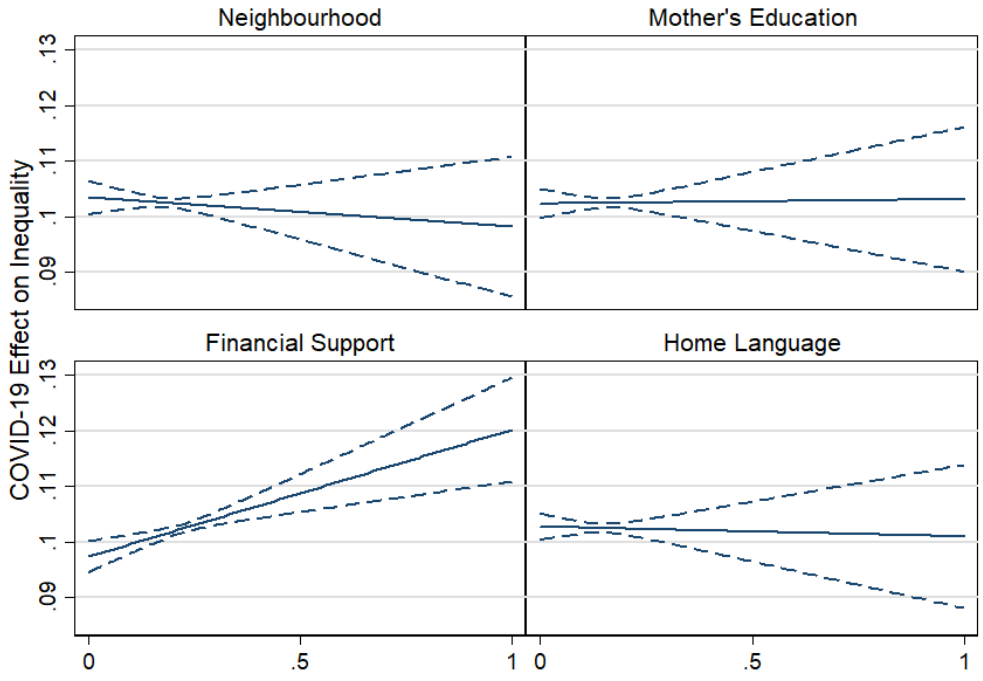
Table B1: Main Regressions Without Special Needs Schools

	Mathematics	Dutch
2019-2020		
COVID-19	-0.250***	-0.278***
	(0.078)	(0.061)
N	1285	1476
2017-2020		
COVID-19	-0.241***	-0.272***
	(0.077)	(0.059)
N	3464	3646
2015-2020		
COVID-19	-0.237***	-0.252***
	(0.076)	(0.057)
N	5505	5682

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. In all regressions, the control variables include school characteristics, characteristics of year 6, teacher characteristics and the test version. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020).

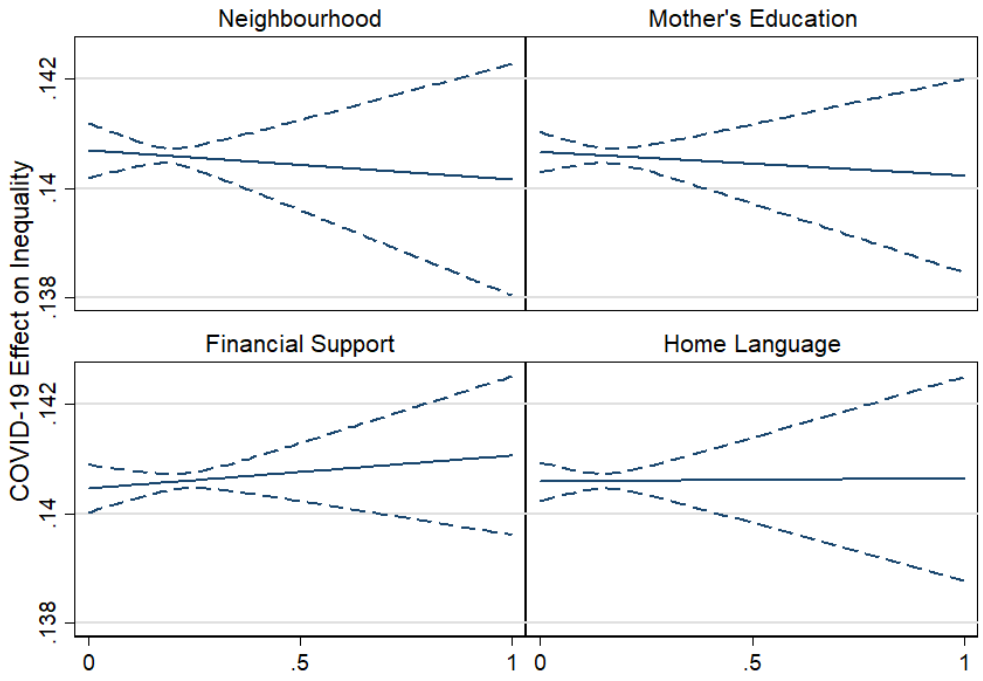


(a) Mathematics

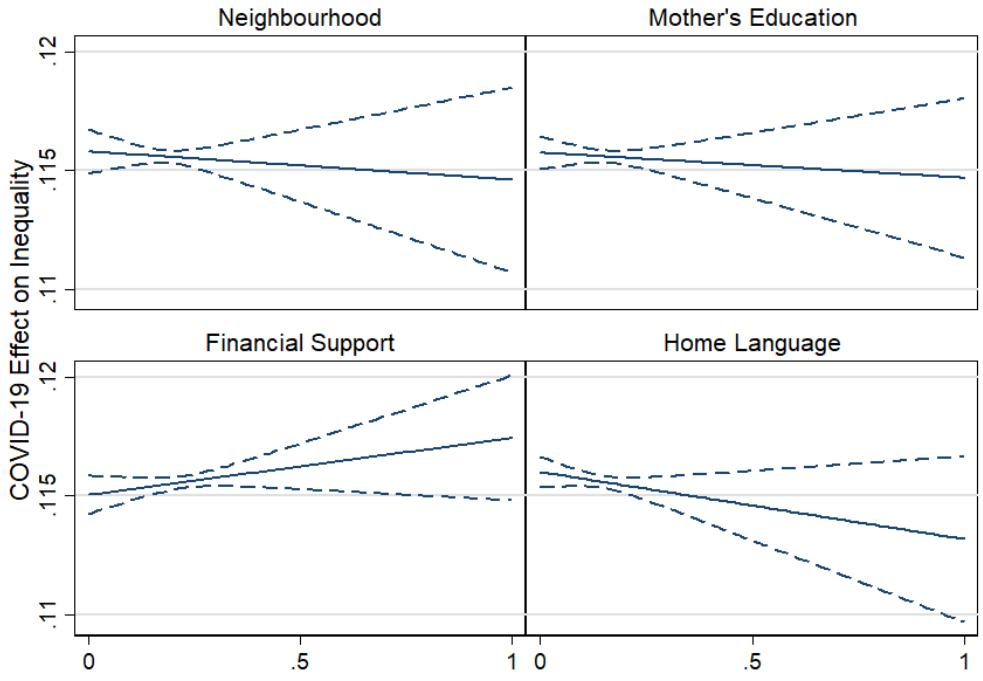


(b) Dutch

Figure B3: Marginal Effects Based on SES Indicators for the Gini Coefficient *Within* Schools



(a) Mathematics



(b) Dutch

Figure B4: Marginal Effects Based on SES Indicators for the Gini Coefficient *Across* Schools

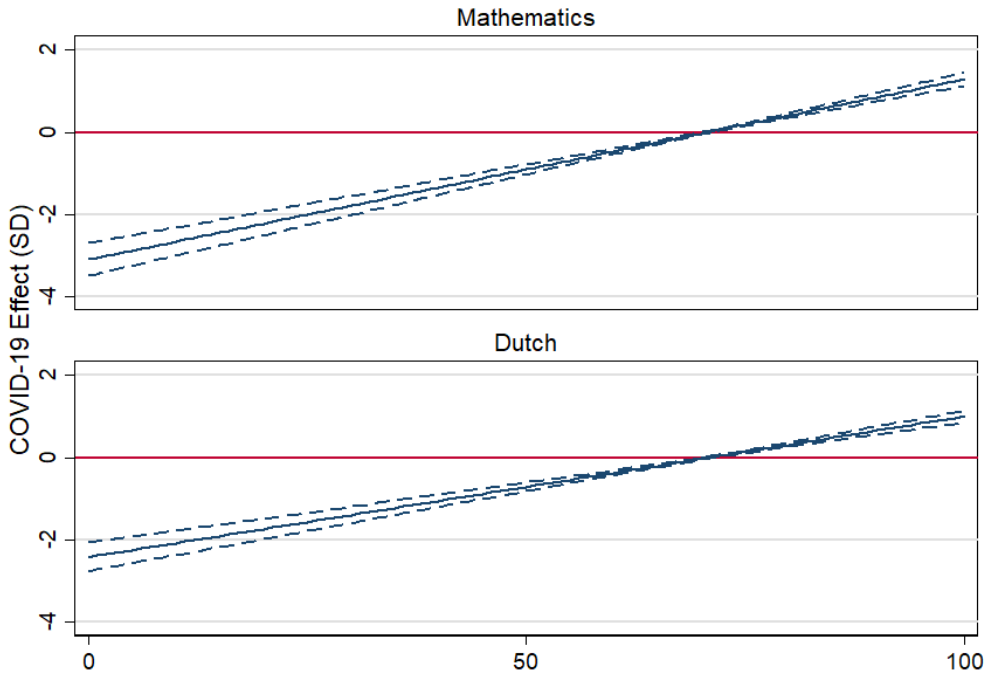


Figure B5: Marginal Effects Based on grade 4 GPA (Mathematics and Dutch)

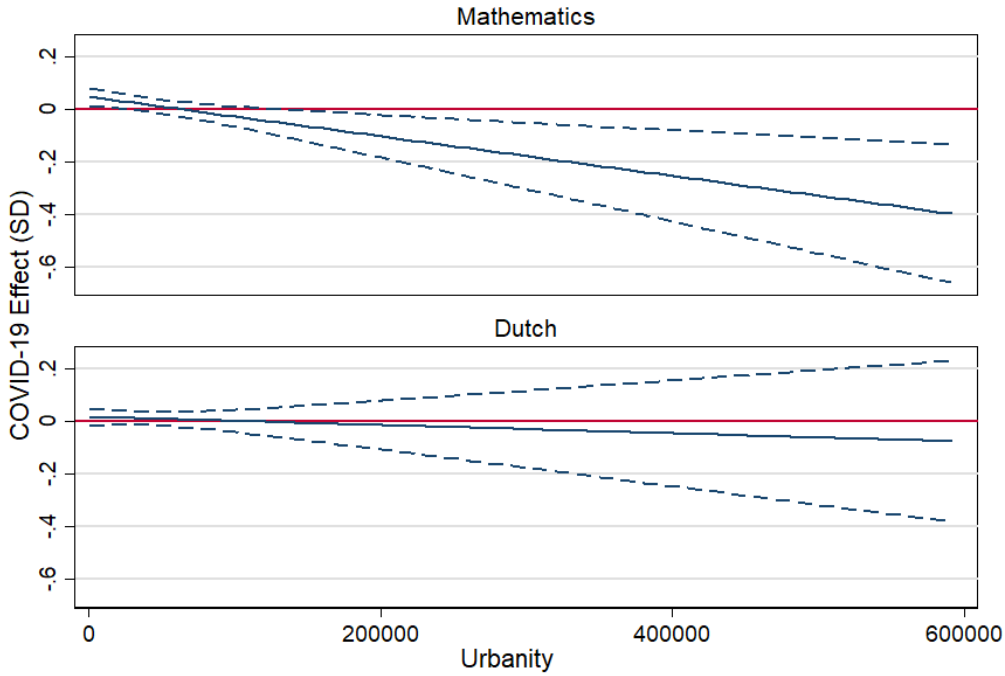


Figure B6: Marginal Effects Based on Urbanity (Population): Mathematics and Dutch

Table B2: Inequality Within and Across Schools With Time Trend

2015-2020	Mathematics			Dutch		
	Gini Coefficient	Ratio 90/10	Entropy	Gini Coefficient	Ratio 90/10	Entropy
COVID-19	0.011***	0.179***	0.011***	0.007***	0.106**	0.006**
	(0.003)	(0.048)	(0.004)	(0.002)	(0.044)	(0.003)
Time trend	0.004***	0.044***	0.003***	0.010***	0.101***	0.007***
	(0.000)	(0.005)	(0.000)	(0.000)	(0.004)	(0.000)
N	5511	5379	5511	5691	5589	5691
Mean	0.122	1.902	0.036	0.100	1.680	0.025
	Gini Coefficient	Ratio 90/10	Entropy	Gini Coefficient	Ratio 90/10	Entropy
COVID-19	0.007***	0.029***	0.011***	0.008***	-0.100***	0.008***
	(0.000)	(0.003)	(0.000)	(0.000)	(0.005)	(0.000)
Time trend	0.004***	0.063***	0.003***	0.011***	0.120***	0.008***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.001)	(0.000)
N	5831	5831	5831	5831	5831	5831
Mean	0.139	2.030	0.046	0.113	1.738	0.031

Notes: * p<0.10, ** p<0.05, *** p<0.01. COVID-19 is a dummy variable for the year 2020. In all regressions, the control variables include school characteristics, characteristics of year 6, teacher characteristics and the time trend. The same test was administered in some years: 1 (2015), 2 (2016), 3 (2017-2018), 4 (2019-2020). A Gini coefficient of 0 means perfect equality and a value of 1 identifies perfect inequality. The 90/10 ratio is defined as the ratio of the score of the 10þpercentile to the score of the 90þpercentile. A higher value of the 90/10 ratio indicates higher inequality. Entropy is based on a generalized entropy index GE(-1), identifying the deviation from perfect equality. The mean is the baseline mean, i.e. excluding the 2020 cohort.

Table B3: Main Regressions for 2019-2020 With Non-Standardised Scores

	Mathematics	Dutch	Science	Social Sciences	French
2019-2020					
COVID-19	-2.039***	-2.210***	-1.921**	-1.315**	-2.489***
	(0.616)	(0.487)	(0.806)	(0.633)	(0.565)
N	1287	1479	836	1073	1324

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. In all regressions, the control variables include school characteristics, characteristics of year 6 and teacher characteristics.

Table B4: Main Regressions With Non-Standardised Scores Without Control Variables

	Mathematics	Dutch	Science	Social Sciences	French
2019-2020					
COVID-19	-1.508**	-1.917***	-1.944**	-0.724	-1.741***
	(0.690)	(0.506)	(0.891)	(0.662)	(0.620)
N	1287	1480	836	1073	1325

Notes: * p<0.10, ** p<0.05, *** p<0.01. Test scores are standardised at test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. All regressions do not include any control variables.

Table B5: Number of Question per Test by Subject and Year

	2015	2016	2017	2018	2019	2020
Mathematics	50	50		21		20
Dutch	31	31		50		17
Science			17		22	
Social Sciences			13		16	
French						20

Notes: The same test was administered in 2017-2018 and 2019-2020.