

The effect of school closures on standardised student test outcomes

Citation for published version (APA):

Maldonado, J. E., & De Witte, K. (2022). The effect of school closures on standardised student test outcomes. *British Educational Research Journal*, 48(1), 49-94. <https://doi.org/10.1002/berj.3754>

Document status and date:

Published: 01/02/2022

DOI:

[10.1002/berj.3754](https://doi.org/10.1002/berj.3754)

Document Version:

Publisher's PDF, also known as Version of record

Document license:

Taverne

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

The effect of school closures on standardised student test outcomes

Joana Elisa Maldonado^{a*}  and Kristof De Witte^{a,b} 

^a*KU Leuven, Belgium;* ^b*UNU-MERIT, Maastricht, the Netherlands*

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 article evaluates the effects of school closures based on standardised tests in the last year of primary school in the Dutch-speaking Flemish region of Belgium. Using a 6-year panel, we find that students of the 2020 cohort experienced significant learning losses in three out of five tested subjects, with a decrease in school averages of mathematics scores of 0.17 standard deviations and Dutch scores (reading, writing, language) of 0.19 standard deviations as compared to previous cohorts. This finding holds when accounting for school characteristics, standardised tests in Grade 4 and school fixed effects. Given the large observed effect sizes, the effect of school closures appears to be a combination of lost learning progress and learning loss. Moreover, we observe that inequality both within schools and across schools rises by 7% for mathematics and 8% 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

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. 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 policymakers voiced concern about the lost progress for students and the resulting educational inequality (Armitage & Nellums, 2020; Azevedo *et al.*, 2020). A common critique is that policymakers 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. This article fills this void by estimating the impact of the COVID-

Corresponding author. Leuven Economics of Education Research, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium. Email: 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. The authors received funding from the Flemish Science Organisation (Grant No. S000617N). The authors declare that they have no relevant or material financial interests that relate to the research described in this article.

19 crisis and its school closures on standardised test scores and inequality within and between schools.

Providing policymakers 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. The first studies present exploratory evidence, for example from surveys (Andrew *et al.*, 2020; Champeaux *et al.*, 2020; Di Pietro *et al.*, 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 *et al.*, 2020; Eyles *et al.*, 2020; Frenette *et al.*, 2020; Kuhfeld *et al.*, 2020). 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 *et al.*, 2020).

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 the tune of 63–68% in reading and 37–50% in mathematics (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 school interruptions in other contexts disregard the unique characteristics of the COVID-19 crisis. A major difference from previous situations of school closures is the widespread use of digital tools and online teaching. Clark *et al.* (2020) present the 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. The first experimental evidence from a randomised controlled trial in Botswana shows positive effects of a text messaging intervention during school closures (Angrist *et al.*, 2020).

Recent evidence from the UK compares test scores of students in all years of primary school from 2020 to results from 2019, and documents substantial decreases in attainment, which appear to be larger for disadvantaged students and schools (Blainey *et al.*, 2020). Similarly, Engzell *et al.* (2020) make use of standardised tests in primary schools in the Netherlands, taken before and after the 2020 school lockdown, and find learning losses of 0.08 standard deviations as well as larger negative effects for disadvantaged students.

This article 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 Dutch-speaking Flemish schools in Belgium. The data used in the analysis covers a large number of Flemish schools over a period of 6 years from 2015 to 2020 and hence provides strong statistical power. The unique panel structure of the data allows us to assess the school averages in different subjects over time and to identify the deviation of the 2020 data from previous years. In particular, the exact same test was administered in 2019 and 2020. Since the 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. Next to the country setting, this study differs from the evidence of Blainey *et al.* (2020) and Engzell *et al.* (2020) as we consider a wider range of subjects, compare to a longer time series of test results, estimate the effect of a longer duration of school lockdown and use standardised tests with lower attrition rates.

The results show that the 2020 cohort experienced significant learning losses in three out of five tested subjects, with a decrease in school averages of mathematics scores of 0.17 standard deviations and Dutch scores (reading, writing, language) of 0.19 standard deviations compared to the previous cohort. The findings hold when accounting for school characteristics and standardised tests of Grade 4 of the cohort. The observed effect appears to be a combination of lost learning progress (i.e. lack of learning due to school closures) and learning loss (i.e. loss of previously obtained knowledge, as often discussed in the context of summer learning loss).¹ The results remain stable in a number of robustness checks, such as limiting the sample using the same schools across all years or using school fixed effects.

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 as the share of students with low socioeconomic status (SES) increases. 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.

Data

Setting

In Flanders, all schools were suddenly and unexpectedly closed by the National Security Council on 16 March 2020. From 15 May 2020 onward, the partial reopening of

primary and secondary schools started under strict conditions. This 9-week period of school closures included 2 weeks of Easter holidays from 6 to 19 April.

In the 3 weeks before the Easter holidays, distance teaching took place. Since schools were not prepared for this sudden change to distance teaching and could not necessarily reach all students, teachers were instructed to only repeat and practice previously taught materials. The organisation of the distance learning was at the responsibility of schools. While the funding mechanism is the same for all schools (De Witte *et al.*, 2019), the policymaking capacity of schools is likely to differ, such that the implementation and practice of distance learning tools likely differed widely across schools. In practice, many primary schools referred their students to online platforms with exercises or distributed exercises on paper. Broadband connection coverage is very high in Flanders (Eurostat, 2021), but 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 (Grymonprez, 2020), students lacked laptops at home (Heymans *et al.*, 2018) and about 12% of students did not have a quiet place to work at home (OECD, 2017)).

In the 4 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 4 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 as 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 2 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 taught exactly in this period, it can be assumed that, similar to the Netherlands, most teaching hours were spent on the fundamental subjects, that is language, mathematics and reading (Inspectie van het Onderwijs, 2020).

On 15 May, conditional upon some safety measures, schools were allowed to offer physical classes to the first, second and sixth grades 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 into 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 8 June, all grades of primary school were allowed to reopen full time.

Online as well as presence classes were continued as usual until the end of the school year on 30 June. Unlike other years, schools were allowed to use the time until 30 June for teaching and assessment, giving the possibility for 2 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 work 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.

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 covers a time span of 6 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 online version of a new test is piloted at the end of May, and pilot schools do not take part in the real test. The test results are only shared with the school and are not made public or shared with the central government. Teachers do not have any incentive to teach to the test, as the tests are not used to evaluate the performance of teachers or students, in order to avoid strategic behaviour (Vlaamse Onderwijsraad, 2018). In 2020, the tests were communicated and organised in the same way as in previous years.

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 6-year period under consideration. In all 6 years, the subjects of mathematics and language (Dutch) were tested. From 2016 onward, science and social sciences were introduced into the tests, first as a combined subject (world studies) and from 2018 onwards as separate subjects. In 2019, French (second language) was added as an additional subject.

In addition, the tests of each subject were slightly modified over the years to accommodate the 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. In 2019 and 2020, the exact same test with identical grade boundaries was used.

Based on the different test versions, we split the sample into three time periods for the analysis. First, we compare the year 2019 to 2020, since the exact same test was administered in these 2 years. Second, we compare the period from 2017 to 2020 and third, from 2015 to 2020, by additionally controlling for the test version. The impact of the school closures during the COVID-19 crisis can be identified by the year 2020, since no other changes took place in that year. The curriculum, funding, school organisation and overall setting in 2020 were the same as in the previous year.

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 2 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 2 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 with the respective Grade 6 data of 2 years later. In all years, the Grade 4 tests covered the three subjects of mathematics, Dutch and world studies.

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 on whether the school is a special needs school, that is a school for children with specific education needs for whom the common curriculum with reasonable adjustments in a regular school is not feasible, as well as the share of students with special needs in regular schools, that is the share of children with specific education needs for whom individual adjustments are made to the common curriculum. Regarding teachers, the data contains the number of teachers at the school in absolute terms as well as full-time equivalents (FTEs) by age group. We use this information to compute the share of teachers above the age of 50, which might matter in the context of COVID-19 as older persons may have a higher expected cost of infection, and hence put more pressure on the school management to not reopen.

Furthermore, the administrative data contains a rich set of measures of 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 from 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, defined as students who speak a different language from 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. The descriptive statistics for each year are provided in Tables A1–A7 in Appendix A.

Moreover, we received data about effective school opening days during the weeks of partial and full reopening of schools. This data comprises the daily number of students expected at school for the period from 15 May to 30 June 2020. Not all schools opened at the maximum allowed capacity for the maximum number of days. Figure B2 in Appendix B shows the average share of students expected at school in May and June 2020. Therefore, it is possible that the test results differ by the amount of teaching at school that students effectively received during the reopening period. Using the data about the effective school opening days during the weeks of partial reopening of the schools, we find that the marginal effect of opening one additional week at maximum capacity does not affect the estimated effects.

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 Appendix A 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 having the situation at least sufficiently under control. 97% of interviewed Catholic schools indicated reaching 80–100% of their students with distance teaching.

At the end of April, 88% of Catholic schools were planning to reopen for Grade 6 in May. On 8 May, 22% of interviewed schools indicated that they would not reopen or were undecided whether to reopen about 10 days later. On 20 May, 34% of interviewed schools indicated that they would use the maximum amount of teaching hours allowed to take place at school.

Sample and attrition

The sample comprises 401 schools in 2020, 1,164 schools in 2019, 1,150 schools in 2018, 1,062 schools in 2017, 1,033 schools in 2016 and 1,017 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. In all years, participating schools had, on average, lower shares of students with low SES, with most of these differences being statistically significant in *t*-tests. In addition, participating schools have in most years higher average grades on the Grade 4 tests than

schools that 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 schools in the school network and external validity could be limited.

Regarding 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 schools' characteristics. Table A7 in Appendix A therefore also compares the participating schools of the year 2020 to the participating schools of the year 2019. Relative to the 2019 schools, schools participating in 2020 were smaller, had a lower share of students whose mothers are low-educated, more students with special needs, fewer Grade 6 students who experienced grade retention and higher Grade 4 scores in world studies. There is, however, no overall pattern in the attrition to the 2020 sample, since, for example, most indicators of SES 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.²

Figure B1 in Appendix B shows the distribution of test scores across the years under consideration for each subject. The distributions clearly show the changes in test versions, which occurred in 2016, 2017 and 2019. While the distribution of scores is similar in 2015 compared to 2016 and in 2018 compared to 2017, the distribution of scores in 2020 compared to 2019 is 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 proceed with regression analyses.

Methodology

Data of different school years has been combined to a panel dataset. This panel data structure allows for a comparison of schools over different time periods. 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 2 years, and thus represents the strongest sample. The estimation is based on the following equation:

$$y_{i,j} = \alpha_j + \beta_j \text{COVID} - 19 + \delta_j X_i + \varepsilon_{i,j} \quad (1)$$

where $y_{i,j}$ denotes the average score in the respective subject j at the level of 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 and $\varepsilon_{i,j}$ denotes the residual. This equation is estimated for the subjects of 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 relative 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 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:

$$y_{i,j} = \alpha_j + \beta_j \text{COVID} - 19 + \gamma_j T + \delta_j X_i + \varepsilon_{i,j} \quad (2)$$

where $y_{i,j}$ again denotes the average score in the respective subject j at the level of 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 of mathematics, Dutch, science and social sciences 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 (2). For this time period, covering all 6 years, the regression is only estimated for the two subjects that were tested in all years, that is mathematics and Dutch.

Results

Results for each subject

Table 1 presents the results for all subjects with the standardised test score as outcome variable and all control variables. The three panels present the results for the three different sub-samples.³ In each panel, COVID-19 is identified as the coefficient for the year 2020. In the second and third panel, we account in addition for the test version which changed across years.⁴

The first column of Table 1 shows that the size and significance of the COVID-19 coefficient for the mathematics score appears to be robust across sub-samples, with an effect size of -0.15 to -0.17 standard deviations, significant at the 5% level. Table B1 in Appendix B confirms that the results also remain robust when including different sets of control variables. This means that students of the 2020 cohort have school averages in mathematics of almost one-fifth of a standard deviation lower than students participating in the standardised tests in previous years.

The second column of Table 1 shows the same regression results for the standardised Dutch scores. In all samples, the effect sizes are slightly larger, that is more negative, for Dutch scores than for mathematics scores. In the 2019–2020 sample, the 2020 cohort has a decrease in Dutch scores with 0.19 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 Appendix A, 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 in more depth below. Table B2 in Appendix B shows that the results for Dutch also remain robust when including different sets of control variables.

The third column of Table 1 shows the main results for the social sciences scores for the two available sub-samples of 2019–2020 and 2017–2020. For this subject, the results are less pronounced. The effect is found to be negative, but not significant in the 2019–2020 sample. In the 2017–2020 sample, the effect of COVID-19 becomes significant. However, even then, the effect size is, with 0.13 standard deviations,

Table 1. Main results for all subjects

	Test score				
	Mathematics	Dutch	Social sciences	Science	French
<i>2019–2020</i>					
COVID-19	–0.17** (0.08)	–0.19*** (0.06)	–0.12 (0.07)	–0.15 (0.10)	–0.18*** (0.05)
<i>N</i>	1,287	1,478	1,070	836	1,322
<i>2017–2020</i>					
COVID-19	–0.16** (0.08)	–0.20*** (0.06)	–0.13* (0.07)	–0.16* (0.09)	
<i>N</i>	3,470	3,655	2,403	2,161	
<i>2015–2020</i>					
COVID-19	–0.15* (0.08)	–0.18*** (0.05)			
<i>N</i>	5,511	5,691			

Notes: Test scores are standardised at the test level to have a mean of 0 and a standard deviation of 1. COVID-19 is a dummy variable for the year 2020. Control variables include the test version, school characteristics, year 6 characteristics, teacher characteristics and year 4 scores. The same test was administered in 2017–2018 and 2019–2020.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

smaller than those for Dutch and mathematics and only significant at the 10% level. 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 been covered sufficiently during the distance teaching and reopening periods.

For science, we observe in the fourth column of Table 1 a non-significant learning loss of 0.15 standard deviations in the 2019–2020 sample. However, this estimate becomes significant at the 10% level in the 2017–2020 sample. Moreover, the results in Table B4 in Appendix B indicate a learning loss of 0.18 standard deviations, significant at the 10% level, when we do not control for the year 4 test scores.

Finally, the fifth column of Table 1 shows the main results for French. French has only been tested in the last 2 years and can, therefore, only be compared from 2019 to 2020. As for mathematics, Dutch and science, French scores are also found to be significantly lower in 2020 compared to 2019, with an effect of –0.18 standard deviations, significant at the 1% level. Table B5 in Appendix B confirms that this finding holds when including different sets of control variables. This means that the effect sizes are comparable in Dutch and French, and only slightly smaller in mathematics.⁵ Only in science and social sciences are the decreases in standardised test scores not statistically significant throughout samples and specifications.

Inequality assessment

As the tests are administered at the 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. The first three panels of Table 2 show the results for three inequality measures that capture inequality within schools for the subjects of 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 in 2019 of 0.14 for mathematics and 0.13 for Dutch, the Gini coefficient increases due to the COVID-19 crisis by 0.01 for mathematics and Dutch, significant at the 1% level for mathematics and at the 5% level for Dutch. This corresponds to an increase by 7% and 8% 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 90th percentile to the score of the 10th percentile. With a mean of 2.07 in mathematics and 2.02 in Dutch, inequality as measured by the 90/10 ratio increases significantly by 0.15 for mathematics and 0.09 for Dutch. This suggests that the difference between the top and bottom performers increased significantly due to the school closures.

A similar increase in within-school inequality is observed for the generalised entropy index, which is a relative inequality index that is, with the chosen sensitivity parameter of -1 , more sensitive to changes at the bottom of the distribution (Cowell, 2016). With a mean of 0.05 in mathematics and Dutch, it increases by about 0.01 in both subjects, significant at the 1% level for mathematics and at the 10% level for Dutch. 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 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.⁶

Similarly, inequality is found to increase across schools as well. The bottom panel of Table 2 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 0.13 for Dutch from 2015 to 2019, and increases significantly by 0.01 in both subjects. This equals an increase of 7% for mathematics and 8% for Dutch, amounting to a similar effect size as the increase in within-school inequality in Dutch and a smaller effect in mathematics. The 90/10 ratio has a mean of 2.03 for mathematics and 1.83 for Dutch and increases, respectively, by 0.1 in mathematics. In Dutch, there is a small negative effect of 0.03. Compared to the within-school effects, this is a smaller increase in inequality across schools. The entropy measure increases by 0.01 for mathematics and Dutch, from a previous mean of 0.05 for mathematics and 0.04 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

Table 2. Inequality within and across schools

	Mathematics			Dutch		
	Gini coefficient	Ratio 90/10	Entropy	Gini coefficient	Ratio 90/10	Entropy
<i>Within schools</i>						
2019–2020						
COVID-19	0.01*** (0.00)	0.15*** (0.05)	0.01*** (0.00)	0.01** (0.00)	0.09* (0.04)	0.00* (0.00)
N	1,287	1,287	1,287	1,478	1,476	1,478
Mean	0.14	2.07	0.05	0.13	2.02	0.05
2017–2020						
COVID-19	0.01*** (0.00)	0.15*** (0.05)	0.01*** (0.00)	0.01** (0.00)	0.08* (0.04)	0.00* (0.00)
N	3,470	3,468	3,470	3,655	3,653	3,655
Mean	0.13	1.94	0.04	0.12	1.91	0.04
2015–2020						
COVID-19	0.02*** (0.00)	0.22*** (0.05)	0.01*** (0.00)	0.01*** (0.00)	0.10** (0.04)	0.01*** (0.00)
N	5,511	5,506	5,511	5,691	5,689	5,691
Mean	0.12	1.90	0.04	0.11	1.80	0.03
<i>Across schools</i>						
2015–2020						
COVID-19	0.01*** (0.00)	0.10*** (0.00)	0.01*** (0.00)	0.01*** (0.00)	−0.03*** (0.00)	0.01*** (0.00)
N	5,826	5,826	5,826	5,826	5,826	5,826
Mean	0.14	2.03	0.05	0.13	1.83	0.04

Notes: 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 2017–2018 and 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 10th percentile to the score of the 90th percentile. A higher value of the 90/10 ratio indicates higher inequality. Entropy is based on a generalised entropy index $GE(-1)$, identifying the deviation from perfect equality. The mean is the baseline mean (i.e. computed excluding the 2020 cohort).

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

capture the time trend in inequality. Table B8 in Appendix B shows that the increase in inequality as a result of school closures, both within and across schools, remains for mathematics when including a time trend in the regression. For Dutch, the inclusion of a time trend results in smaller and less significant estimates.

Marginal effects

By quintile of the distribution of scores. To examine whether the COVID-19 school closures influenced the students at the top and 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 B3 in Appendix B 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 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.

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 5 hours or more per day during the school lockdown, while a substantial share of students only studied for 2 hours or less per day (Huber & Helm, 2020). We therefore decompose the effects by SES to identify which groups of students experienced larger learning losses due to school closures.

We estimate the marginal effects based on SES indicators in Grade 6 of the school. Figures 1a and b show the marginal effects based on the four SES indicators in Grade 6 for mathematics and Dutch for the full sample of 2015 to 2020. The figures are based on a linear regression (solid lines) and the dotted lines indicate the confidence intervals.

One measure of SES 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 SES, 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 displays 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 common measure of SES and the educational background of students is the mother's education level. We 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 b confirm this expectation for both mathematics and Dutch. In both subjects, learning losses increase as the share of students with a low-educated mother in the school increases.

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 lower SES had a more difficult environment for home schooling than students in families with higher SES. This can be linked to the presence of practical learning facilities, such as a desk to study at and a device to follow online classes, as

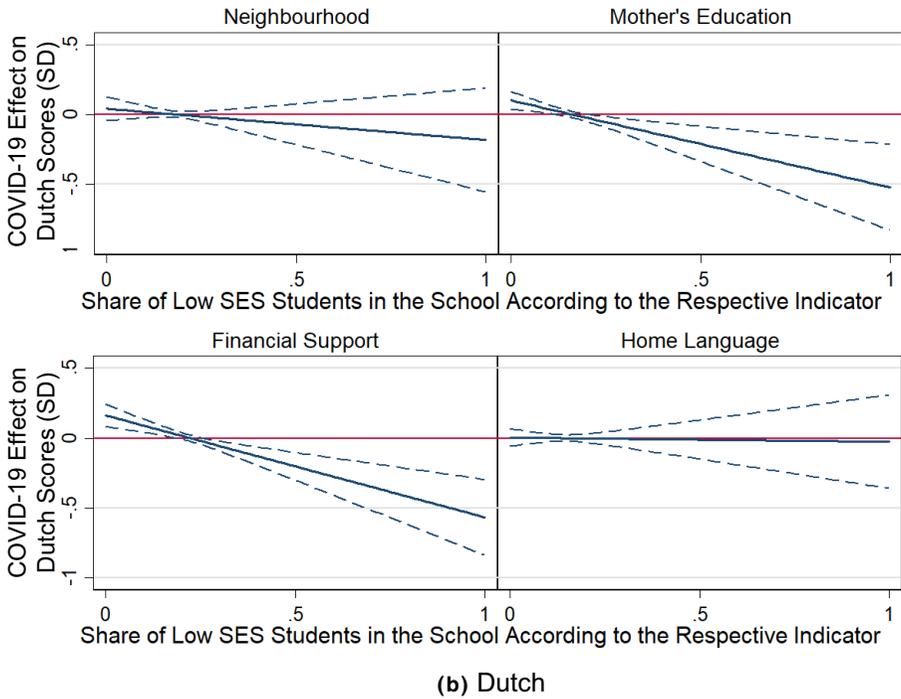
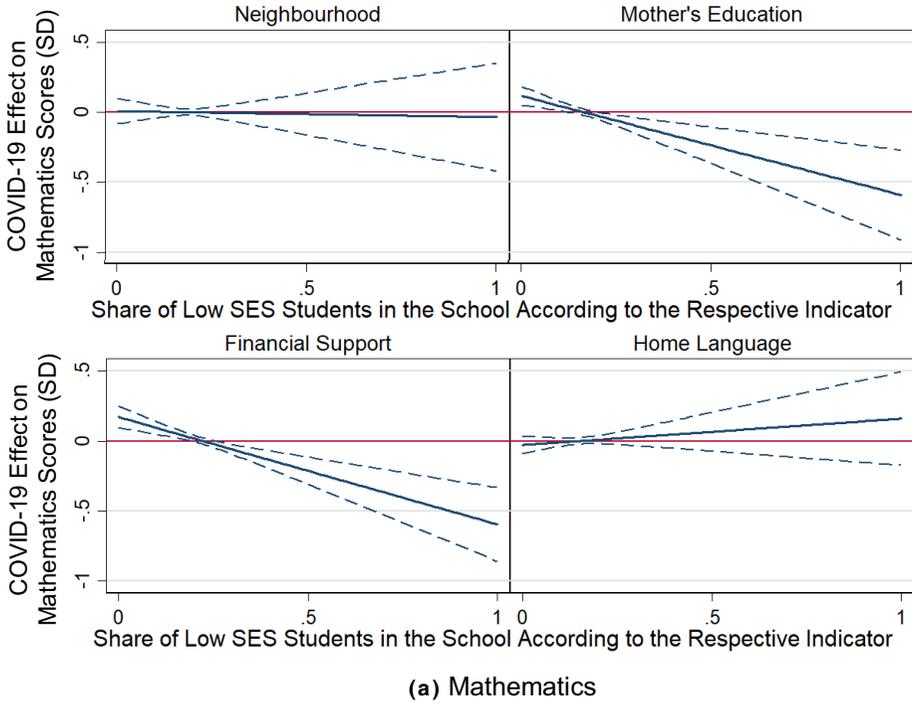


Figure 1. Marginal effects by socioeconomic status: (a) mathematics; (b) Dutch

Notes: Based on a multivariate regression with the 2015–2020 sample, using the full set of control variables for school characteristics, year 6 characteristics and teacher characteristics. [Colour figure can be viewed at wileyonlinelibrary.com]

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 b 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 what extent parents were able to support home schooling and help their children with distant learning. It can therefore be expected that schools with a larger share of students who 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 b suggest that there are no significant differences in learning losses by the share of students with another home language, in either subject.

In summary, we observe that 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 with increasing shares of students with a mother with a low education level and shares 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.⁷

Figures B4 and B5 in Appendix B show that, similarly, inequality within schools, as measured by the Gini coefficient, is increasing as the share of students receiving financial support and the share of students with a mother with a low education level increase. Across schools, the Gini coefficient increases for Dutch, with increasing shares of students with a different home language.

Figure B6 in Appendix B shows that learning losses decrease as the Grade 4 grade point average (GPA) increases, 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 B7 in Appendix B shows that, for Dutch, the learning loss remains constant along the population size. For mathematics, even after accounting for all observed SES characteristics, there appears to be a clear pattern of increasing learning losses as population size increases, with the largest learning losses, and the largest spread of scores, in the biggest cities.

Robustness checks

This section presents 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 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. The first column of Table 3 shows that the

Table 3. Robustness checks

	Same schools	Fixed effects	Matching	Matching + year 4
<i>Mathematics</i>				
2019–2020				
COVID-19	−0.18*	−0.14**	−0.12	−0.12
	(0.11)	(0.07)	(0.09)	(0.09)
<i>N</i>	246	1,287	572	563
2017–2020				
COVID-19	−0.18	−0.15**	−0.15*	−0.16*
	(0.11)	(0.06)	(0.09)	(0.09)
<i>N</i>	492	3,470	1,237	1,141
2015–2020				
COVID-19	−0.17*	−0.12*	−0.15*	−0.16*
	(0.10)	(0.07)	(0.08)	(0.08)
<i>N</i>	738	5,511	1,793	1,654
<i>Dutch</i>				
2019–2020				
COVID-19	−0.31***	−0.32***	−0.15**	−0.16**
	(0.11)	(0.06)	(0.07)	(0.07)
<i>N</i>	246	1,478	772	763
2017–2020				
COVID-19	−0.31***	−0.29***	−0.20***	−0.21***
	(0.10)	(0.05)	(0.07)	(0.07)
<i>N</i>	492	3,655	1,434	1,336
2015–2020				
COVID-19	−0.32***	−0.29***	−0.19***	−0.21***
	(0.10)	(0.05)	(0.06)	(0.06)
<i>N</i>	738	5,691	1,988	1,848

Notes: Test scores are standardised at the 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 2017–2018 and 2019–2020. The regressions in the first column only include schools that participated in the tests in each year. The regressions in the second column include school fixed effects. In the regressions in the last two columns, schools were matched 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. In the last column, matching was done in addition on year 4 scores.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

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 0.18 standard deviations for mathematics and a larger effect size of 0.31 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.

Another approach to account for differences between schools is to include school fixed effects. The second column of Table 3 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 in mathematics, with 0.14 standard deviations, and larger for Dutch, with 0.32 standard deviations.

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 *et al.*, 2009). We sequentially match the participating 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. The third column of Table 3 shows that the effects for the matched sample confirm the main results, with negative effects in both subjects. With a learning loss of 0.12 standard deviations for mathematics and 0.15 standard deviations for Dutch, the effect size is robust to matching schools as well, although the statistical significance decreases for mathematics.

Second, in addition to the school characteristics, we match further based on the average scores in Grade 4. The last column of Table 3 shows the results from an estimation based on a matched 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 for both subjects in all sub-samples. The effect size is similar to the first matching approach, with 0.12 standard deviations for mathematics and 0.16 standard deviations for Dutch, although the statistical significance for mathematics again decreases. These effect sizes are still in line with the main results.

Finally, Table B7 in Appendix B shows that the main results for mathematics and Dutch are also robust to the exclusion of participating special needs schools.

Conclusion and discussion

This article provides 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 dataset with standardised test scores from a large share of Flemish schools over a period of 6 years spanning 2015 to 2020, combined with administrative data.

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 0.17 standard deviations for mathematics and 0.19 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.⁸

The large observed effect sizes appear to be a combination of lost learning progress and learning loss rather than only lost progress. Bloom *et al.* (2008) demonstrate that, by Grade 5, student achievement improves by about 0.4 standard deviations over the course of an academic year, suggesting that the COVID-19 effect is larger than what could be expected from the loss of instruction time at school. A conservative estimate for the lost progress due to a reduction of instruction time is 0.06–0.18 standard deviations (Di Pietro *et al.*, 2020). In addition, learning loss due to forgetting previously obtained knowledge, similar to the summer learning loss, can already account for approximately 0.18 standard deviations (Kuhfeld *et al.*, 2020; Atteberry & McEachin, 2021). Given the far-reaching impact on social interactions during the COVID-19 lockdown, it is also likely that psychological factors could have negatively impacted students' test outcomes (Iterbeke & De Witte, 2020). It is of course possible that factors such as mental health not only affected learning progress, but also performance during testing, and therefore led to larger effects. However, it is unlikely that this accounts for more than a small part of the effect. Finally, teachers often practice with students for tests, while this practice time could have been skipped during the school lockdown and partial reopening of schools in 2020.

To put these observed effects into perspective, Chetty *et al.* (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 standard deviation decrease in standardised test scores could decrease future employment probability by 0.86% (Currie & Thomas, 2001). On the one hand, this effect of reduced human capital accumulation due to learning losses could, at least partially, be mitigated as the complete generation of students was equally affected by the school closures. Hence, in line with the job competition theory, the relative schooling of the current generation of youngsters remains unaffected, such that the impact on earnings and unemployment might be lower than what can be expected based on pre-COVID research. On the other hand, this research shows that not all students appear to be affected equally, as inequality between students and schools has risen. It is therefore likely that the groups of vulnerable students who recorded relatively larger learning losses will suffer disadvantages on the labour market and negative effects on earnings in the long-term.

Regarding the external validity of our findings, it has to be kept in mind that the sample under consideration represents a slightly advantaged population of schools in terms of grades on standardised tests and other background characteristics. As we find larger learning losses in more disadvantaged schools, the learning losses in the general population could potentially be larger than the effects recorded in this study. Moreover, the Grade 6 students were among the first to re-enter school in Flanders. Hence, it can be expected that the observed effects for the Grade 6 cohort are a lower bound for students from other grades who 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 SES: learning losses appear to be increasing depending on most indicators for SES as well as population density, while they

are decreasing as Grade 4 scores increase. 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.

Given the global nature of the pandemic and the ongoing school closures in many countries, these findings are highly relevant for policymakers worldwide. Our results might be an upper bound for countries which experienced a period of shorter school closures or which relied on better distance learning strategies (e.g. as in the Netherlands; Engzell *et al.*, 2020), while they serve as lower bound estimates for countries with longer school closures or less developed remote teaching strategies (e.g. as in the USA; Slavin, 2020). It is possible that, as knowledge accumulates, early learning shortfalls escalate in the long run, resulting in grade repetition and school dropout. Yet, as students are resilient, learning losses could also fade out over time, as students might recoup the learning losses at a faster rate when returning to school, benefitting from the positive effects of maturity on learning. Further research should examine this. Moreover, future research can explore the heterogeneity in learning losses by correlating it with different implementations of distance learning or policymaking capacity of schools.

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 article clearly emphasises that school closures are associated with very high costs for students and do not affect all students equally.

Acknowledgements

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, André Decoster, Geraint Johnes, Tommaso Agasisti, Jan Van Damme, Rianne Janssen, Dirk Van Damme, Luka Boeskens, William Thorn, Bastian Betthausen, Per Engzell, Arun Frey, Mark Verhagen, Dieter Verhaest and Diego Gambetta, participants of the OECD Education and Skills Forum, participants of the Flemish Ministry of Education Seminar, the Flemish Education Inspectorate, Flemish Parliament, Dutch Education Inspectorate, participants of the ICFES Seminar 2020, participants of the KU Leuven Centre of Educational Effectiveness and Evaluation Seminar, participants of the University of South Florida Economics Seminar, participants of the ECSR Spring School on The Impact of Covid-19 on Social Inequality and participants of the ULB Workshop on Political Economics for their valuable comments and suggestions.

NOTES

¹ Based on average monthly summer learning loss, Kuhfeld *et al.* (2020) predict 63–68% of learning gains of a regular school year in reading and 37–50% of learning gains in mathematics, assuming an interruption of 3 months.

- ² Within schools, we observe limited attrition in 2020. In 2020, the attrition within schools is lowest for Dutch (90% of students participated in the test) and highest for mathematics (40% of students participated) and science (18% of students participated). While in previous years, the share of students participating in Dutch and mathematics tests was close to 100%, this share was already lower for science, social sciences and French. The attrition in 2020 is likely due to the tests being administered only in some classes, not others.
- ³ 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 (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. To avoid a loss in power, missing year 4 scores are imputed using multivariate normal regression.
- ⁴ 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.
- ⁵ Based on seemingly unrelated regression using the specification with all control variables in the 2019–2020 sample, the coefficients for Dutch, mathematics, science and French do not differ significantly from each other.
- ⁶ As the cardinal properties of inequality measures such as the Gini coefficient and entropy are not well known, the comparison of their percentage changes should be interpreted with sufficient caution.
- ⁷ The marginal effects remain robust when tested on the sub-samples of 2017–2020 and 2019–2020, as well as when using nonlinear specifications.
- ⁸ Similar effect sizes are substantial; 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.

References

- Andrew, A., Cattan, S., Costa-Dias, M., Farquharson, C., Kraftman, L., Krutikova, S. *et al.* (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*. Available online at: <https://ssrn.com/abstract=3663098>.
- Armitage, R. & Nellums, L. B. (2020) Considering inequalities in the school closure response to COVID-19, *The Lancet*, 8, e644. [https://doi.org/10.1016/S0140-6736\(20\)30547-X](https://doi.org/10.1016/S0140-6736(20)30547-X).
- Atteberry, A. & McEachin, A. (2021) School's out: The role of summers in understanding achievement disparities, *American Educational Research Journal*, 58(2), 239–282.
- 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?* Available online at: www.vrt.be/vrtnws/nl/2020/05/14/weer-naar-school/#:~:text=Inhetbasisonderwijsmogenleerlingen,volledagennaarschoolkomen (accessed 28 August 2020).
- Bao, X., Qu, H., Zhang, R. & Hogan, T. (2020) Modeling reading ability gain in kindergarten children during COVID-19 school closures. *International Journal of Environmental Research and Public Health*, 17(6371). <https://doi.org/10.3390/ijerph17176371>.
- Blackwell, M., Iacus, S., King, G. & Porro, G. (2009) cem: Coarsened exact matching in Stata, *The Stata Journal*, 9(4), 524–546.
- Blainey, K., Hiorns, C. & Hannay, T. (2020) *The impact of lockdown on children's education: A nationwide analysis*. Technical Report (Banbury, Hodder Education).
- 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.
- Champeaux, H., Mangiacavchi, L., Marchetta, F. & Piccoli, L. (2020) *Learning at home: Distance learning solutions and child development during the COVID-19 lockdown*. IZA Discussion Paper Series 13819.

- 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.
- Cheung, A. C. & Slavin, R. E. (2016) How methodological features affect effect sizes in education, *Educational Researcher*, 45(5), 283–292.
- Clark, A. E., Nong, H., Zhu, H. & Zhu, R. (2020) Compensating for academic loss: Online learning and student performance during the COVID-19 pandemic, *China Economic Review*, 68, 101629.
- Cowell, F. A. (2016) Inequality and poverty measures, in: M. D. Adler & M. Fleurbaey (Eds) *The Oxford handbook of well-being and public policy* (Oxford, Oxford University Press).
- Currie, J. & Thomas, D. (2001) Early test scores, school quality and SES: Long-run effects on wage and employment outcomes, *Research in Labor Economics*, 20, 103–132.
- De Witte, K., Titl, V., Holz, O. & Smet, M. (2019) *Financing quality education for all – the funding methods of compulsory and special needs education* (Leuven, Leuven University Press).
- Di Pietro, G., Biagi, F., Costa, P., Karpinski, Z. & Mazza, J. (2020) *The likely impact of COVID-19 on education: Reflections based on the existing literature and recent international datasets* (Luxembourg, Publications Office of the European Union).
- Engzell, P., Frey, A. & Verhagen, M. (2020) Learning inequality during the COVID-19 pandemic, paper presented at the *Joint IZA & Jacobs Center Workshop: Consequences of COVID-19 for Child and Youth Development* [online], 19 October.
- Eurostat (2021) *Households with broadband access*. Available online at: https://ec.europa.eu/eurostat/databrowser/product/view/ISOC_CI_IT_H (accessed 17 March 2021).
- 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*. Technical Report 11-626-X No. 103 (Ottawa, Statistics Canada).
- Fryer, R. G. (2016) The production of human capital in developed countries: Evidence from 196 randomized field experiments, in: A. V. Banerjee & E. Duflo (Eds) *Handbook of economic field experiments* (vol. 2) (Amsterdam, North-Holland), 95–322.
- Grymonprez, S. (2020) Jinnih Beels: 'Basisscholen bereiken 20 tot 30 procent van leerlingen niet', *De Standaard*. www.standaard.be/cnt/dmf20200417_04926114.
- Haeck, C. & Lefebvre, P. (2020) Pandemic school closures may increase inequality in test scores, *Canadian Public Policy*, 46(S1), S82–S87.
- Heymans, P. J., Godaert, E., Elen, J., Van Braak, J. & Goeman, K. (2018) *Eindrapport MICTIVO2018. Monitor ICT-integratie in het Vlaamse onderwijs. Eindrapport van O&O-opdracht: Meting ICT-integratie in het Vlaamse onderwijs (MICTIVO)* (Leuven/Ghent, KU Leuven/Universiteit Gent).
- 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(2), 237–270.
- 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?* Available online at: www.onderwijsinspectie.nl/onderwerpen/afstandsonderwijs-tijdens-covid-19 (accessed 28 August 2020).
- 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 20.19.
- Kaffenberger, M. (2020) *Modeling the long-run learning impact of the COVID-19 learning shock: Actions to (more than) mitigate loss*. RISE Insight Series 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, *Educational Researcher*, 49(8), 549–565.

OECD (2017) Table III.10.2 – Availability of a quiet place to study and science performance: Results are based on students' self-reports, *PISA 2015 results (vol. III): Students' well-being* (Paris, OECD Publishing).

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.

Slavin, R. E. (2020) *How much have students lost in the COVID-19 shutdowns?* Available online at: <https://robertslavinsblog.wordpress.com/2020/10/01/how-much-have-students-lost-in-the-covid-19-shutdowns/22> (accessed 20 November 2020).

Vlaamse Onderwijsraad (2018) *Gevalideerde toetsen einde basisonderwijs: hefboom voor kwaliteit?* Available online at: https://assets.vlor.be/www.vlor.be/advice_final_attachments/RBO-RBO-ADV-1819-001.pdf (accessed 18 March 2020).

Appendix A Attrition

Table A1. Attrition and descriptive statistics of participating schools: year 2015

	Attrited	Participated	<i>t</i> -Test		
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	<i>p</i> -Value
Number of students	464	171.11 [95.44]	1,017	189.99 [83.58]	0.00
Share of girls	464	0.46 [0.10]	1,017	0.50 [0.05]	0.00
SES—neighbourhood	464	0.27 [0.28]	1,017	0.19 [0.28]	0.00
SES—mother's education	464	0.20 [0.15]	1,017	0.17 [0.15]	0.00
SES—subsidies	464	0.24 [0.15]	1,017	0.20 [0.14]	0.00
SES—home language	464	0.18 [0.20]	1,017	0.14 [0.20]	0.00
Share of newcomers	191	0.05 [0.09]	146	0.04 [0.05]	0.12
Special needs school	464	0.28 [0.45]	1,017	0.00 [0.03]	0.00
Special needs students	464	0.30 [0.44]	1,017	0.02 [0.05]	0.00
Number of teachers	464	18.78 [10.02]	1,017	15.56 [5.68]	0.00

Table A1. (Continued)

	Attrited	Participated	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Number of teachers as FTE	464	15.12 [9.53]	1,017	12.15 [5.08]	0.00
Teachers: share above 50	464	0.28 [0.16]	1,017	0.30 [0.15]	0.00
Teachers: share above 50 as FTE	464	0.31 [0.19]	1,017	0.33 [0.18]	0.00
Year 6: number of students	464	26.41 [16.36]	1,017	28.22 [14.59]	0.04
Year 6: share of girls	464	0.46 [0.13]	1,017	0.50 [0.12]	0.00
Year 6: SES—neighbourhood	464	0.26 [0.28]	1,017	0.18 [0.28]	0.00
Year 6: SES—mother's education	464	0.19 [0.18]	1,017	0.16 [0.15]	0.00
Year 6: SES—subsidies	464	0.25 [0.17]	1,017	0.21 [0.16]	0.00
Year 6: SES—home language	464	0.16 [0.20]	1,017	0.13 [0.21]	0.00
Year 6: grade repetition	464	0.00 [0.01]	1,017	0.00 [0.02]	0.90
Year 6: slow learners	464	0.13 [0.11]	1,017	0.12 [0.11]	0.20
Year 4: Dutch score	79	66.40 [9.25]	727	67.97 [8.52]	0.15
Year 4: math score	79	64.79 [8.84]	728	65.87 [8.18]	0.30
Year 4: world studies score	78	67.05 [8.59]	726	68.95 [8.39]	0.06

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	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Number of students	453	173.79 [103.48]	1,033	192.60 [82.48]	0.00
Share of girls	453	0.46 [0.10]	1,033	0.50 [0.04]	0.00
SES—neighbourhood	453	0.27 [0.28]	1,033	0.19 [0.28]	0.00
SES—mother's education	453	0.20 [0.16]	1,033	0.17 [0.15]	0.00
SES—subsidies	453	0.25 [0.15]	1,033	0.21 [0.15]	0.00
SES—home language	453	0.19 [0.21]	1,033	0.15 [0.21]	0.00
Share of newcomers	215	0.05 [0.11]	256	0.04 [0.05]	0.09
Special needs school	453	0.28 [0.45]	1,033	0.00 [0.04]	0.00
Special needs students	453	0.30 [0.44]	1,033	0.02 [0.05]	0.00
Number of teachers	453	19.04 [10.48]	1,033	15.39 [5.52]	0.00
Number of teachers as FTE	453	15.39 [9.82]	1,033	12.18 [4.97]	0.00
Teachers: share above 50	453	0.29 [0.16]	1,033	0.31 [0.15]	0.00
Teachers: share above 50 as FTE	453	0.31 [0.19]	1,033	0.34 [0.18]	0.01
Year 6: number of students	453	27.17 [17.93]	1,033	28.94 [14.62]	0.07
Year 6: share of girls	453	0.47 [0.14]	1,033	0.50 [0.10]	0.00

Table A2. (Continued)

	Attrited	Participated	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Year 6: SES—neighbourhood	453	0.25 [0.28]	1,033	0.18 [0.28]	0.00
Year 6: SES—mother's education	453	0.19 [0.17]	1,033	0.16 [0.16]	0.00
Year 6: SES—subsidies	453	0.25 [0.17]	1,033	0.22 [0.18]	0.00
Year 6: SES—home language	453	0.17 [0.21]	1,033	0.13 [0.20]	0.00
Year 6: grade repetition	453	0.00 [0.05]	1,033	0.00 [0.01]	0.28
Year 6: slow learners	453	0.12 [0.11]	1,033	0.11 [0.10]	0.25
Year 4: Dutch score	65	62.32 [8.73]	727	63.70 [8.23]	0.22
Year 4: math score	65	64.78 [8.48]	729	67.02 [7.57]	0.04
Year 4: world studies score	65	69.31 [8.13]	726	71.72 [7.66]	0.02

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	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Number of students	425	172.60 [99.53]	1,062	196.64 [86.42]	0.00
Share of girls	425	0.45 [0.10]	1,062	0.50 [0.04]	0.00
SES—neighbourhood	425	0.26 [0.27]	1,062	0.20 [0.29]	0.00

Table A3. (Continued)

	Attrited	Participated	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
SES—mother's education	425	0.19 [0.15]	1,062	0.17 [0.15]	0.01
SES—subsidies	425	0.24 [0.15]	1,062	0.21 [0.16]	0.00
SES—home language	425	0.19 [0.21]	1,062	0.16 [0.21]	0.01
Share of newcomers	228	0.05 [0.09]	321	0.04 [0.05]	0.22
Special needs school	425	0.29 [0.45]	1,062	0.00 [0.06]	0.00
Special needs students	425	0.32 [0.44]	1,062	0.03 [0.07]	0.00
Number of teachers	425	19.29 [10.44]	1,062	15.71 [5.98]	0.00
Number of teachers as FTE	425	15.51 [9.71]	1,062	12.48 [5.36]	0.00
Teachers: share above 50	425	0.29 [0.15]	1,062	0.31 [0.15]	0.01
Teachers: share above 50 as FTE	425	0.32 [0.18]	1,062	0.34 [0.18]	0.03
Year 6: number of students	425	27.32 [17.63]	1,062	29.63 [14.94]	0.02
Year 6: share of girls	425	0.45 [0.13]	1,062	0.50 [0.11]	0.00
Year 6: SES—neighbourhood	425	0.25 [0.27]	1,062	0.19 [0.28]	0.00
Year 6: SES—mother's education	425	0.19 [0.16]	1,062	0.17 [0.16]	0.02
Year 6: SES—subsidies	425	0.25 [0.16]	1,062	0.21 [0.17]	0.00

Table A3. (Continued)

	Attrited	Participated	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Year 6: SES—home language	425	0.18 [0.22]	1,062	0.15 [0.21]	0.02
Year 6: grade repetition	425	0.00 [0.02]	1,062	0.00 [0.02]	0.62
Year 6: slow learners	425	0.11 [0.10]	1,062	0.11 [0.11]	0.98
Year 4: Dutch score	52	67.07 [7.75]	742	69.28 [8.35]	0.05
Year 4: math score	53	63.68 [9.16]	744	66.79 [8.24]	0.02
Year 4: world studies score	52	67.62 [12.32]	741	71.91 [7.34]	0.01

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	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Number of students	339	161.61 [91.92]	1,150	200.22 [89.53]	0.00
Share of girls	339	0.44 [0.12]	1,150	0.50 [0.04]	0.00
SES—neighbourhood	339	0.25 [0.25]	1,150	0.20 [0.29]	0.00
SES—mother's education	339	0.19 [0.14]	1,150	0.17 [0.15]	0.08
SES—subsidies	339	0.25 [0.15]	1,150	0.22 [0.17]	0.01
SES—home language	339	0.20 [0.21]	1,150	0.17 [0.21]	0.03

Table A4. (Continued)

	Attrited	Participated	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Share of newcomers	199	0.05 [0.10]	394	0.04 [0.05]	0.21
Special needs school	339	0.36 [0.48]	1,150	0.00 [0.06]	0.00
Special needs students	339	0.39 [0.46]	1,150	0.03 [0.07]	0.00
Number of teachers	339	20.40 [12.09]	1,150	15.97 [6.17]	0.00
Number of teachers as FTE	339	16.26 [10.87]	1,150	12.75 [5.60]	0.00
Teachers: share above 50	339	0.29 [0.16]	1,150	0.31 [0.15]	0.03
Teachers: share above 50 as FTE	339	0.32 [0.18]	1,150	0.34 [0.18]	0.11
Year 6: number of students	339	25.86 [15.91]	1,150	30.66 [15.90]	0.00
Year 6: share of girls	339	0.46 [0.15]	1,150	0.50 [0.11]	0.00
Year 6: SES—neighbourhood	339	0.24 [0.25]	1,150	0.19 [0.29]	0.00
Year 6: SES—mother's education	339	0.19 [0.15]	1,150	0.17 [0.16]	0.03
Year 6: SES—subsidies	339	0.27 [0.17]	1,150	0.23 [0.17]	0.00
Year 6: SES—home language	339	0.19 [0.21]	1,150	0.16 [0.22]	0.02
Year 6: grade repetition	339	0.00 [0.01]	1,150	0.00 [0.01]	0.44
Year 6: slow learners	339	0.11 [0.08]	1,150	0.11 [0.10]	0.57
Year 4: Dutch score	45	68.19 [9.13]	765	69.50 [7.72]	0.34

Table A4. (Continued)

	Attrited	Participated	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Year 4: math score	47	72.22 [10.16]	767	73.50 [7.15]	0.39
Year 4: world studies score	46	68.68 [9.85]	767	68.90 [8.15]	0.88

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	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Number of students	331	158.09 [92.88]	1,164	201.92 [90.08]	0.00
Share of girls	331	0.44 [0.12]	1,164	0.50 [0.04]	0.00
SES—neighbourhood	331	0.25 [0.24]	1,164	0.20 [0.29]	0.00
SES—mother's education	331	0.18 [0.13]	1,164	0.17 [0.15]	0.41
SES—subsidies	331	0.26 [0.15]	1,164	0.24 [0.17]	0.00
SES—home language	331	0.21 [0.21]	1,164	0.18 [0.21]	0.02
Share of newcomers	201	0.04 [0.04]	433	0.03 [0.04]	0.04
Special needs school	331	0.38 [0.49]	1,164	0.00 [0.04]	0.00
Special needs students	331	0.41 [0.46]	1,164	0.04 [0.06]	0.00
Number of teachers	331	21.29 [12.27]	1,164	16.89 [6.33]	0.00

Table A5. (Continued)

	Attrited	Participated	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Number of teachers as FTE	331	16.53 [11.20]	1,164	12.84 [5.50]	0.00
Teachers: share above 50	331	0.26 [0.15]	1,164	0.29 [0.15]	0.00
Teachers: share above 50 as FTE	331	0.28 [0.18]	1,164	0.31 [0.17]	0.02
Year 6: number of students	331	25.28 [15.99]	1,164	31.49 [16.03]	0.00
Year 6: share of girls	331	0.44 [0.14]	1,164	0.50 [0.11]	0.00
Year 6: SES—neighbourhood	331	0.25 [0.25]	1,164	0.19 [0.28]	0.00
Year 6: SES—mother's education	331	0.18 [0.14]	1,164	0.17 [0.16]	0.25
Year 6: SES—subsidies	331	0.27 [0.16]	1,164	0.24 [0.19]	0.00
Year 6: SES—home language	331	0.20 [0.22]	1,164	0.16 [0.21]	0.01
Year 6: grade repetition	331	0.00 [0.02]	1,164	0.00 [0.01]	0.06
Year 6: slow learners	331	0.10 [0.07]	1,164	0.10 [0.09]	0.45
Year 4: Dutch score	47	65.83 [6.93]	770	65.89 [7.73]	0.96
Year 4: math score	47	71.62 [6.44]	774	71.31 [6.99]	0.75
Year 4: world studies score	46	74.01 [6.44]	769	74.60 [6.25]	0.54

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	<i>t</i> -Test		<i>p</i> -Value
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	
Number of students	1,104	194.05 [94.04]	401	184.44 [84.22]	0.06
Share of girls	1,104	0.48 [0.08]	401	0.50 [0.04]	0.00
SES—neighbourhood	1,094	0.22 [0.27]	400	0.20 [0.29]	0.28
SES—mother's education	1,094	0.18 [0.15]	400	0.16 [0.13]	0.01
SES—subsidies	1,094	0.25 [0.17]	400	0.22 [0.17]	0.01
SES—home language	1,094	0.19 [0.21]	400	0.19 [0.22]	0.77
Share of newcomers	530	0.04 [0.04]	156	0.03 [0.03]	0.08
Special needs school	1,104	0.11 [0.32]	401	0.00 [0.05]	0.00
Special needs students	1,104	0.16 [0.31]	401	0.05 [0.07]	0.00
Number of teachers	1,104	18.94 [9.34]	401	15.99 [6.06]	0.00
Number of teachers as FTE	1,104	14.53 [8.29]	401	11.96 [5.16]	0.00
Teachers: share above 50	1,104	0.29 [0.15]	401	0.30 [0.14]	0.26
Teachers: share above 50 as FTE	1,104	0.31 [0.17]	401	0.31 [0.17]	0.55
Year 6: number of students	1,104	31.65 [16.89]	401	29.44 [15.18]	0.02
Year 6: share of girls	1,104	0.48 [0.13]	401	0.50 [0.13]	0.06

Table A6. (Continued)

	Attrited		Participated		<i>t</i> -Test
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	<i>p</i> -Value
Year 6: SES—neighbourhood	1,094	0.21 [0.27]	400	0.19 [0.29]	0.33
Year 6: SES—mother's education	1,094	0.18 [0.16]	400	0.15 [0.14]	0.00
Year 6: SES—subsidies	1,094	0.26 [0.19]	400	0.23 [0.18]	0.01
Year 6: SES—home language	1,094	0.17 [0.22]	400	0.17 [0.22]	0.94
Year 6: grade repetition	1,104	0.00 [0.01]	401	0.00 [0.02]	0.54
Year 6: slow learners	1,104	0.10 [0.09]	401	0.09 [0.09]	0.16
Year 4: Dutch score	578	63.98 [8.92]	288	65.28 [8.55]	0.04
Year 4: math score	579	70.53 [7.50]	289	72.16 [7.29]	0.00
Year 4: world studies score	566	73.92 [8.34]	284	75.35 [5.93]	0.00

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		<i>t</i> -Test
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	<i>p</i> -Value
Number of students	1,164	201.92 [90.08]	401	184.44 [84.22]	0.00
Share of girls	1,164	0.50 [0.04]	401	0.50 [0.04]	0.44

Table A7. (Continued)

	Participated in 2019		Participated in 2020		<i>t</i> -Test
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	<i>p</i> -Value
SES—neighbourhood	1,164	0.20 [0.29]	400	0.20 [0.29]	0.93
SES—mother's education	1,164	0.17 [0.15]	400	0.16 [0.13]	0.07
SES—subsidies	1,164	0.24 [0.17]	400	0.22 [0.17]	0.22
SE—home language	1,164	0.18 [0.21]	400	0.19 [0.22]	0.46
Share of newcomers	433	0.03 [0.04]	156	0.03 [0.03]	0.13
Special needs school	1,164	0.00 [0.04]	401	0.00 [0.05]	0.78
Special needs students	1,164	0.04 [0.06]	401	0.05 [0.07]	0.00
Number of teachers	1,164	16.89 [6.33]	401	15.99 [6.06]	0.01
Number of teachers as FTE	1,164	12.84 [5.50]	401	11.96 [5.16]	0.00
Teachers: share above 50	1,164	0.29 [0.15]	401	0.30 [0.14]	0.56
Teachers: share above 50 as FTE	1,164	0.31 [0.17]	401	0.31 [0.17]	0.53
Year 6: number of students	1,164	31.49 [16.03]	401	29.44 [15.18]	0.02
Year 6: share of girls	1,164	0.50 [0.11]	401	0.50 [0.13]	0.89
Year 6: SES—neighbourhood	1,164	0.19 [0.28]	400	0.19 [0.29]	0.98
Year 6: SES—mother's education	1,164	0.17 [0.16]	400	0.15 [0.14]	0.02

Table A7. (Continued)

	Participated in 2019		Participated in 2020		<i>t</i> -Test
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	<i>p</i> -Value
Year 6: SES—subsidies	1,164	0.24 [0.19]	400	0.23 [0.18]	0.25
Year 6: SES—home language	1,164	0.16 [0.21]	400	0.17 [0.22]	0.55
Year 6: grade repetition	1,164	0.00 [0.01]	401	0.00 [0.02]	0.21
Year 6: slow learners	1,164	0.10 [0.09]	401	0.09 [0.09]	0.05
Year 4: Dutch score	770	65.89 [7.73]	288	65.28 [8.55]	0.29
Year 4: math score	774	71.31 [6.99]	289	72.16 [7.29]	0.09
Year 4: world studies score	769	74.60 [6.25]	284	75.35 [5.93]	0.07

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		<i>t</i> -Test
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	<i>p</i> -Value
Inspected	1,104	0.77 [0.42]	401	0.95 [0.22]	0.00
Inspected in Round 1A (i.e. 24 April 2020)	1,104	0.47 [0.50]	401	0.56 [0.50]	0.00
Inspected in Round 1B (i.e. 30 April 2020)	1,104	0.29 [0.45]	401	0.37 [0.48]	0.00
Inspected in Round 2 (i.e. 8 May 2020)	1,104	0.10 [0.31]	401	0.14 [0.35]	0.07
Inspected in Round 3 (i.e. 20 May 2020)	1,104	0.10 [0.30]	401	0.13 [0.34]	0.14

Table A8. (Continued)

	Attrited		Participated		<i>t</i> -Test
	<i>N</i>	Mean [SD]	<i>N</i>	Mean [SD]	<i>p</i> -Value
School has situation under control	836	2.56 [0.59]	373	2.51 [0.60]	0.28
Number of students reached with classes	510	2.97 [0.19]	220	2.97 [0.16]	0.66
Pre-teaching taking place	519	0.56 [0.50]	225	0.54 [0.50]	0.64
Teaching all subjects	519	0.33 [0.47]	225	0.28 [0.45]	0.16
Teaching Dutch	837	0.45 [0.50]	375	0.46 [0.50]	0.92
Teaching French	837	0.39 [0.49]	375	0.37 [0.48]	0.49
Teaching maths	837	0.45 [0.50]	375	0.46 [0.50]	0.86
Teaching world studies	519	0.32 [0.47]	225	0.32 [0.47]	0.90
Round 1: reopening in class year 6	837	0.34 [0.47]	375	0.36 [0.47]	0.54
Round 2: reopening school	115	1.26 [0.47]	56	1.25 [0.48]	0.90
Round 3: max hours allowed at school	111	0.33 [0.50]	52	0.33 [0.55]	0.94

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

Table B1. Main result: mathematics score

Mathematics score					
<i>2019–2020</i>					
COVID-19	–0.11 (0.09)	–0.15* (0.08)	–0.17** (0.08)	–0.17** (0.08)	–0.17** (0.08)
<i>N</i>	1,287	1,287	1,287	1,287	1,287
<i>2017–2020</i>					
COVID-19	–0.11 (0.09)	–0.15* (0.08)	–0.17** (0.08)	–0.17** (0.08)	–0.16** (0.08)
<i>N</i>	3,470	3,470	3,470	3,470	3,470
<i>2015–2020</i>					
COVID-19	–0.11 (0.09)	–0.13 (0.08)	–0.16** (0.08)	–0.16** (0.08)	–0.15* (0.08)
<i>N</i>	5,511	5,511	5,511	5,511	5,511
School characteristics		✓	✓	✓	✓
Characteristics year 6			✓	✓	✓
Teachers				✓	✓
Year 4 scores					✓

Notes: 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. We control for the test version, as the same test was administered in 2017–2018 and 2019–2020.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

Table B2. Main result: Dutch score

Dutch score					
<i>2019–2020</i>					
COVID-19	–0.16*** (0.06)	–0.18*** (0.06)	–0.19*** (0.06)	–0.20*** (0.06)	–0.19*** (0.06)
<i>N</i>	1,479	1,478	1,478	1,478	1,478
<i>2017–2020</i>					
COVID-19	–0.16*** (0.06)	–0.18*** (0.06)	–0.20*** (0.06)	–0.21*** (0.06)	–0.20*** (0.06)
<i>N</i>	3,656	3,655	3,655	3,655	3,655
<i>2015–2020</i>					
COVID-19	–0.15** (0.06)	–0.16*** (0.05)	–0.17*** (0.05)	–0.19*** (0.05)	–0.18*** (0.05)
<i>N</i>	5,692	5,691	5,691	5,691	5,691
School characteristics		✓	✓	✓	✓
Characteristics year 6			✓	✓	✓
Teachers				✓	✓
Year 4 scores					✓

Notes: 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. We control for the test version, as the same test was administered in 2017–2018 and 2019–2020.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

Table B3. Main result: social sciences score

Social sciences score					
<i>2019–2020</i>					
COVID-19	–0.05 (0.07)	–0.10 (0.07)	–0.09 (0.07)	–0.11 (0.07)	–0.12 (0.07)
<i>N</i>	1,070	1,070	1,070	1,070	1,070
<i>2017–2020</i>					
COVID-19	–0.05 (0.07)	–0.11* (0.07)	–0.11 (0.07)	–0.13* (0.07)	–0.13* (0.07)
<i>N</i>	2,403	2,403	2,403	2,403	2,403
School characteristics		✓	✓	✓	✓
Characteristics year 6			✓	✓	✓
Teachers				✓	✓
Year 4 scores					✓

Notes: 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. We control for the test version, as the same test was administered in 2017–2018 and 2019–2020.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

Table B4. Main result: science score

		Science score				
<i>2019–2020</i>						
COVID-19	–0.18*	–0.17*	–0.19*	–0.18*	–0.15	
	(0.11)	(0.10)	(0.10)	(0.10)	(0.10)	
<i>N</i>	836	836	836	836	836	
<i>2017–2020</i>						
COVID-19	–0.18*	–0.18*	–0.20**	–0.19**	–0.16*	
	(0.11)	(0.10)	(0.10)	(0.10)	(0.09)	
<i>N</i>	2,161	2,161	2,161	2,161	2,161	
School characteristics		✓	✓	✓	✓	
Characteristics year 6			✓	✓	✓	
Teachers				✓	✓	
Year 4 scores					✓	

Notes: 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. We control for the test version, as the same test was administered in 2017–2018 and 2019–2020.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

Table B5. Main result: French score

		French score				
<i>2019–2020</i>						
COVID-19	–0.08	–0.16***	–0.16***	–0.17***	–0.18***	
	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)	
<i>N</i>	1,323	1,322	1,322	1,322	1,322	
School characteristics		✓	✓	✓	✓	
Characteristics year 6			✓	✓	✓	
Teachers				✓	✓	
Year 4 scores					✓	

Notes: 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.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

Table B6. Main regressions without special needs schools

	Mathematics	Dutch
<i>2019–2020</i>		
COVID-19	–0.17** (0.08)	–0.20*** (0.06)
<i>N</i>	1,285	1,475
<i>2017–2020</i>		
COVID-19	–0.17** (0.08)	–0.21*** (0.06)
<i>N</i>	3,460	3,644
<i>2015–2020</i>		
COVID-19	–0.16** (0.08)	–0.19*** (0.05)
<i>N</i>	5,498	5,677

Notes: 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 2017–2018 and 2019–2020.

* $p < 0.10$.

** $p < 0.05$.

*** $p < 0.01$.

Table B7. Inequality within and across schools with time trend

	Mathematics			Dutch		
	Gini coefficient	Ratio 90/10	Entropy	Gini coefficient	Ratio 90/10	Entropy
<i>Within schools</i>						
2015–2020						
COVID-19	0.01***	0.17***	0.01***	−0.00	−0.01	0.00
	(0.00)	(0.05)	(0.00)	(0.00)	(0.04)	(0.00)
Time trend	0.00***	0.04***	0.00***	0.01***	0.10***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
N	5,511	5,506	5,511	5,691	5,689	5,691
Mean	0.12	1.90	0.04	0.11	1.80	0.03
<i>Across schools</i>						
2015–2020						
COVID-19	0.01***	0.03***	0.01***	−0.00***	−0.14***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Time trend	0.00***	0.06***	0.00***	0.01***	0.10***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	5,826	5,826	5,826	5,826	5,826	5,826
Mean	0.14	2.03	0.05	0.13	1.83	0.04

Notes: 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 2017–2018 and 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 10th percentile to the score of the 90th percentile. A higher value of the 90/10 ratio indicates higher inequality. Entropy is based on a generalised entropy index $GE(-1)$, identifying the deviation from perfect equality. The mean is the baseline mean (i.e. computed excluding the 2020 cohort).

* $p < 0.10$.

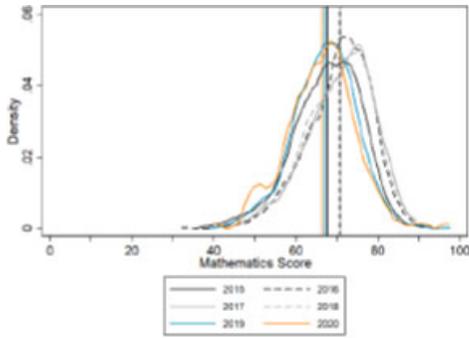
** $p < 0.05$.

*** $p < 0.01$.

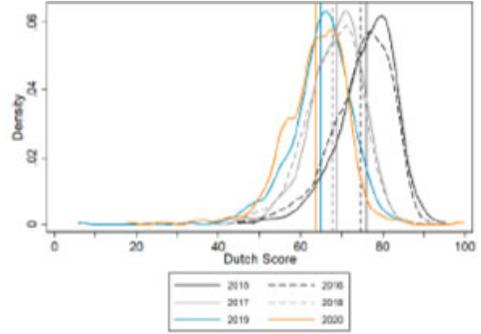
Table B8. Number of questions per test by subject and year

	2015	2016	2017	2018	2019	2020
Mathematics	50	50	21		20	
Dutch	31	31	50		17	
Science		30	17		22	
Social sciences			13		16	
French					20	

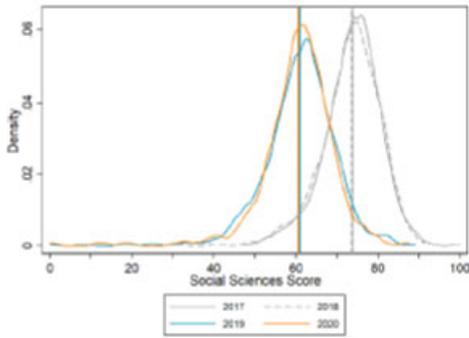
Notes: The same test was administered in 2017–2018 and 2019–2020.



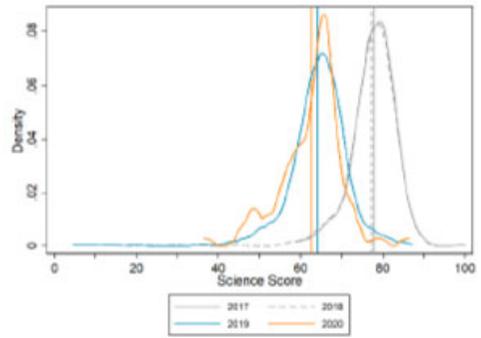
(a) Mathematics



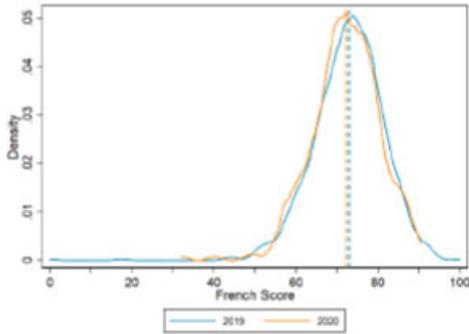
(b) Dutch



(c) Social Sciences



(d) Science



(e) French

Figure B1. Distribution of scores. [Colour figure can be viewed at wileyonlinelibrary.com]

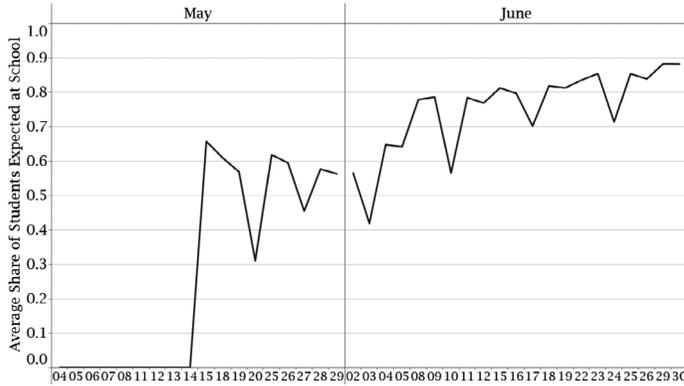
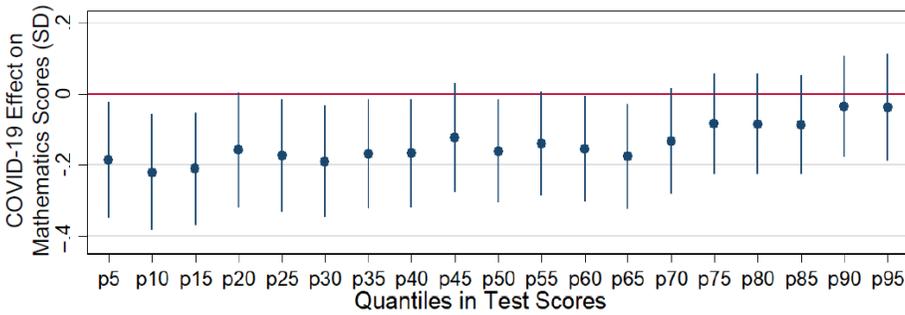
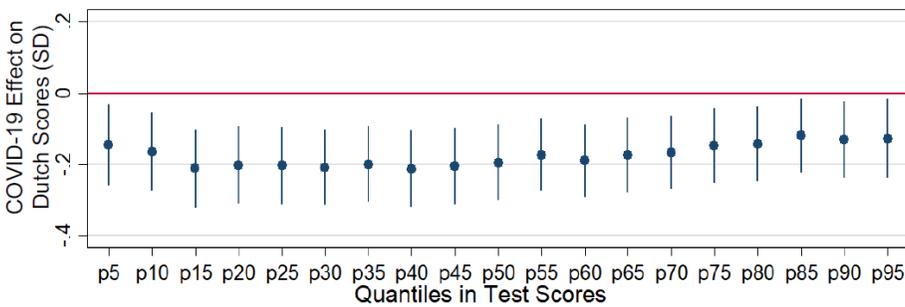


Figure B2. Average daily school opening May and June 2020.



(a) Mathematics



(b) Dutch

Figure B3. Estimated effects per quantile in test scores. This figure and all subsequent figures are based on a multivariate regression with the 2015–2020 sample, using the full set of control variables for school characteristics, year 6 characteristics and teacher characteristics. [Colour figure can be viewed at wileyonlinelibrary.com]

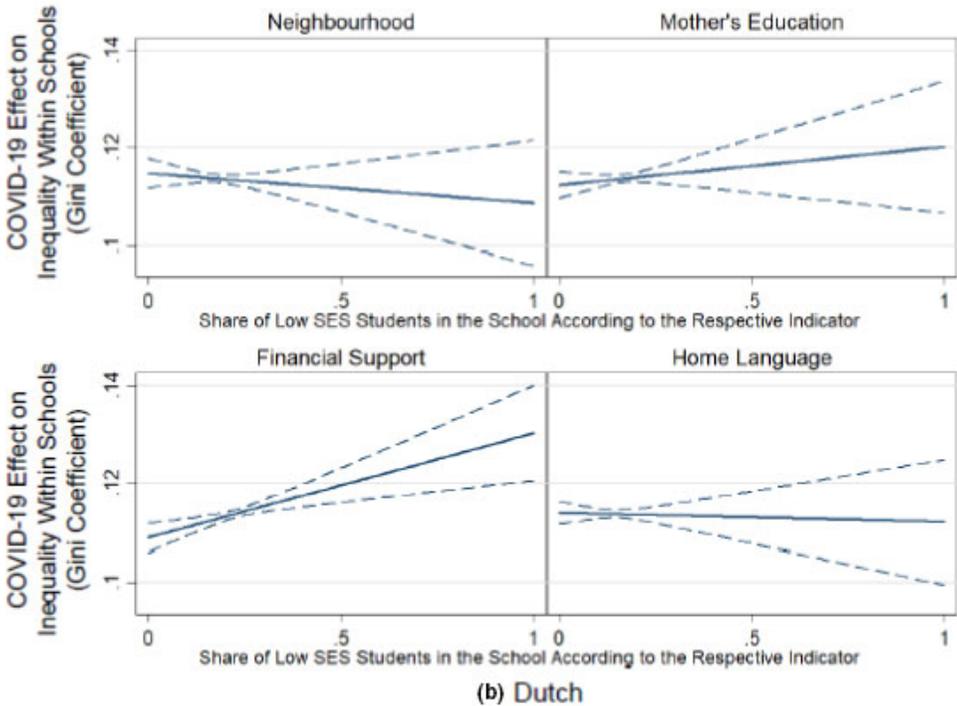
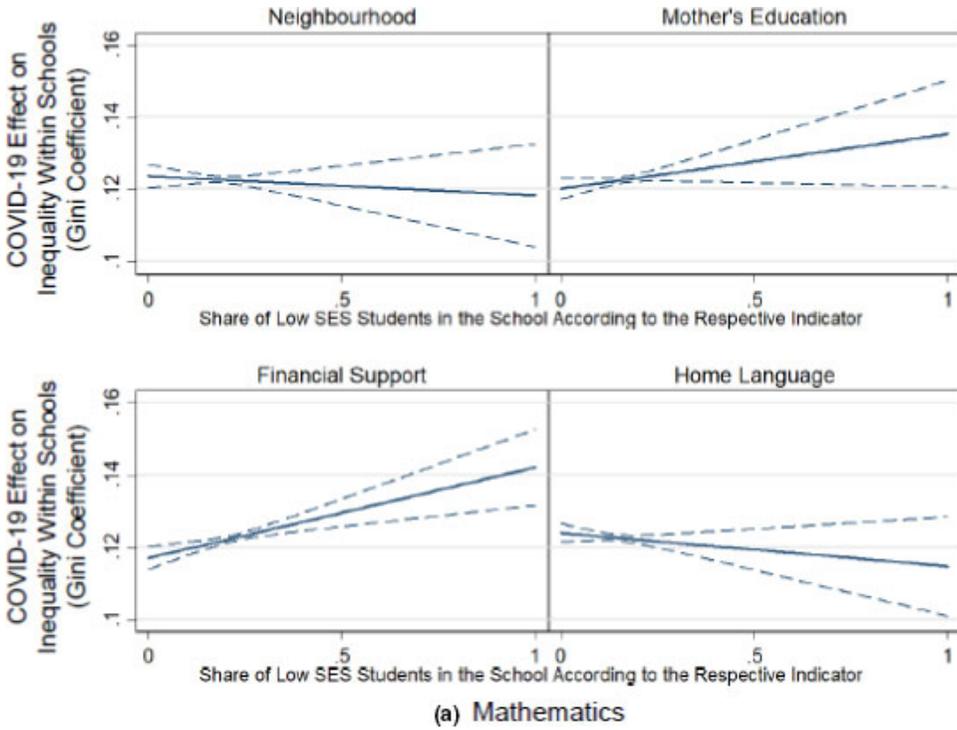
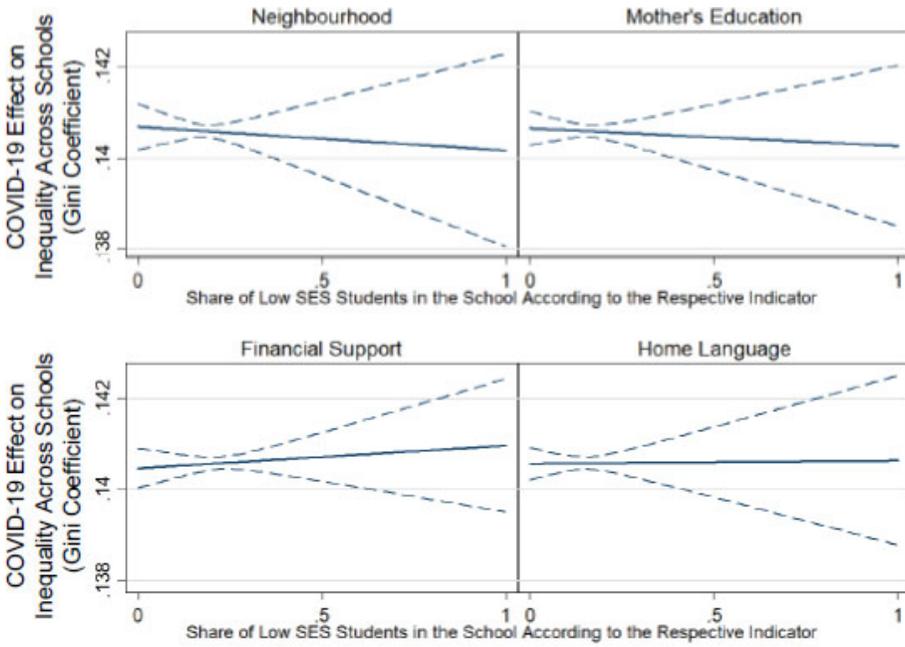
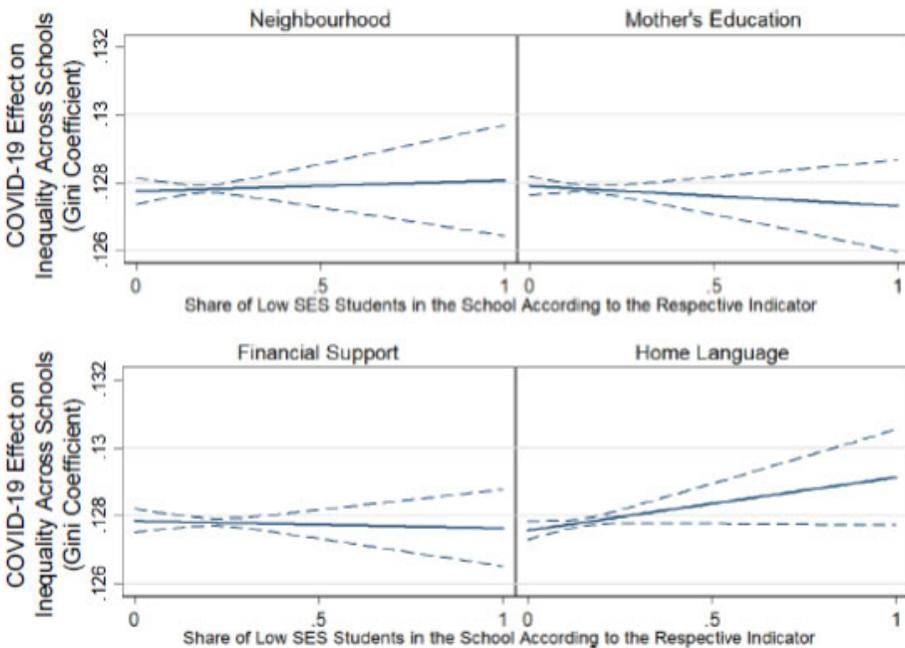


Figure B4. Marginal effects based on SES indicators for the Gini coefficient within schools. [Colour figure can be viewed at wileyonlinelibrary.com]

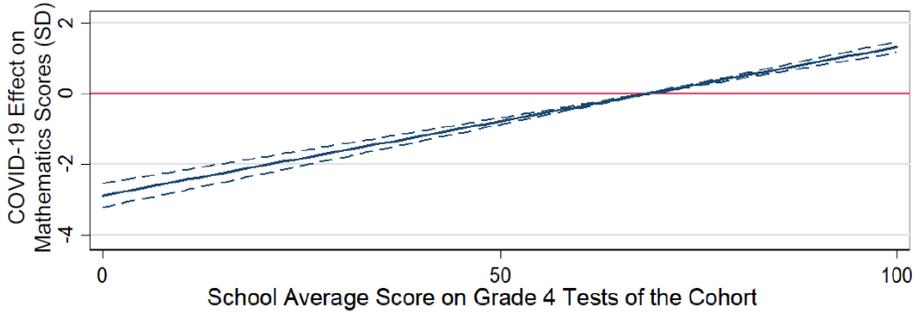


(a) Mathematics

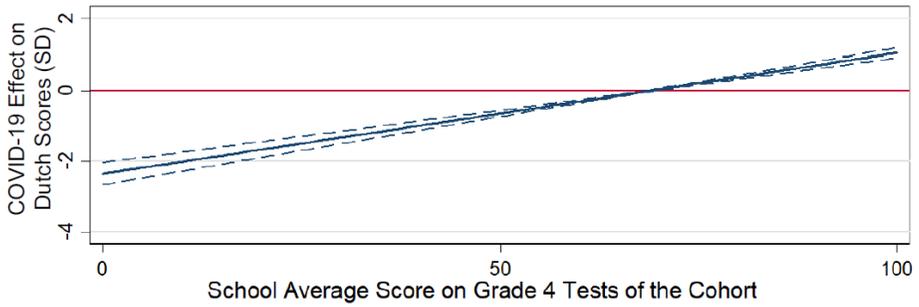


(b) Dutch

Figure B5. Marginal effects based on SES indicators for the Gini coefficient across schools. [Colour figure can be viewed at wileyonlinelibrary.com]



(a) Mathematics



(b) Dutch

Figure B6. Marginal effects based on Grade 4 GPA (mathematics and Dutch). [Colour figure can be viewed at wileyonlinelibrary.com]

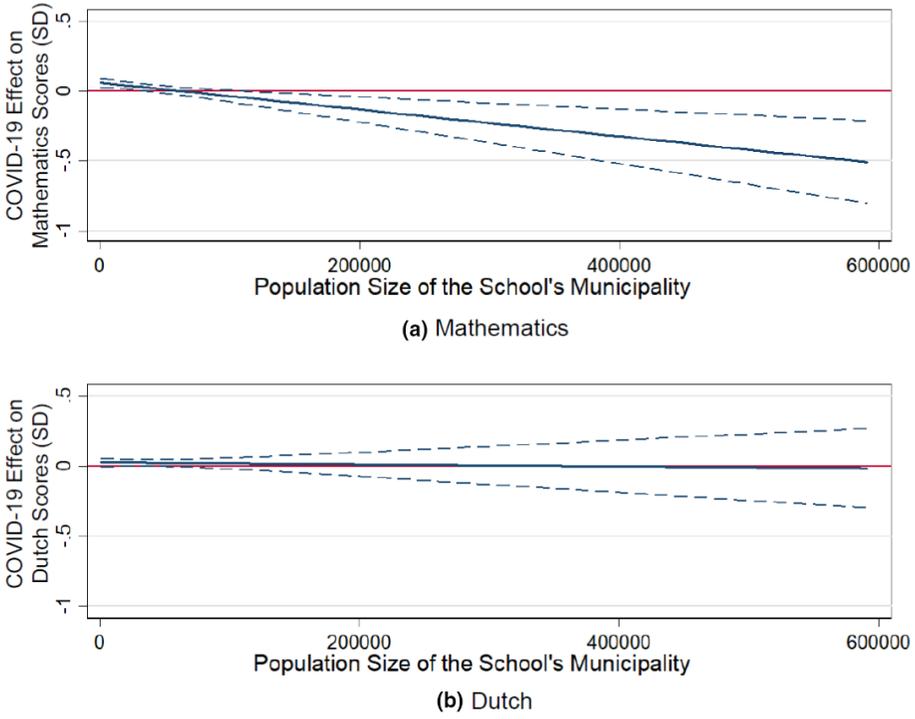


Figure B7. Marginal effects based on urbanity (population): mathematics and Dutch. [Colour figure can be viewed at wileyonlinelibrary.com]