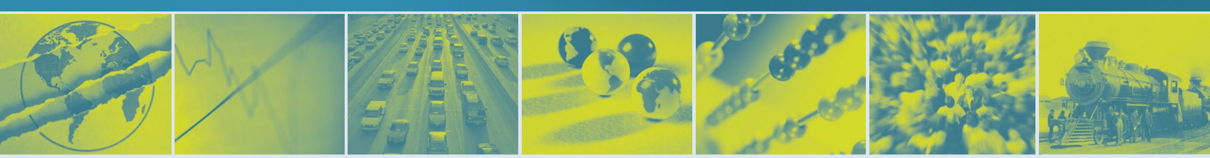
KU LEUVEN

DEPARTMENT OF ECONOMICS

The resiliency of school outcomes after the COVID-19 pandemic. Standardised test scores and inequality one year after long term school closures

Letizia GAMBI and Kristof DE WITTE

FACULTY OF ECONOMICS AND BUSINESS



DISCUSSION PAPER SERIES DPS21.12

OCTOBER 2021

The resiliency of school outcomes after the COVID-19 pandemic. Standardised test scores and inequality one year after long term school closures.*

Letizia Gambi† Kristof De Witte‡

October 15, 2021

Abstract

Almost two years after the largest disruption of education in history, the question remains as to whether, and to what extent, school outcomes are resilient and inequality persists. To answer these questions, this paper exploits a unique panel data-set with standardised test scores and administrative data pertaining to the last year of primary education in the Flemish region of Belgium. For the subjects of native language (Dutch), math, science, social science and foreign language (French), exactly the same standardised tests were administered in 2019 (pre-pandemic), 2020 and 2021. Our empirical specification captures (un)observed heterogeneity at school level, a time trend, and time-varying control variables. The resilience in school outcomes differs per subject as we observe additional attainment deficits in the Dutch and French language one year after the pandemic. For math, the impact of the COVID-19 school closures is halted, but not reversed yet. For science, students in the 2021 cohort have started catching up (though insignificantly) with previous cohorts, while the 2021 test scores improved significantly for social sciences. Notwithstanding the halted attainment deficits in math in 2021, a quantile analysis suggests that the math test scores of the best-performing students in

*The data for this study are protected by a confidentiality agreement and we are precluded from sharing the data with others. We would like to thank Marijke De Meyst, Anton Derks, Steven Groenez, Joana Maldonado, Marc-André Chénier, Silvia Palmaccio, André Decoster, Pieter Verachtert and participants to the LEER seminar for assistance with the data and/or useful insights and comments. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

†Letizia Gambi, Leuven Economics of Education Research (LEER), KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium, E-mail: letizia.gambi@kuleuven.be

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

a school (i.e., quantile 70 to 95) have significantly declined, while those of low-performing students seem to have slightly improved (though insignificantly). One year after the COVID-19 school closures, the inequality within schools seems to have increased in the Dutch language and decreased in mathematics. Further, the findings suggest that targeted remedial actions (in particular summer schools), which were mainly focusing on the most vulnerable students, were successful in halting attainment deficits. However, further policy attention should also be given to the best-performing students, who seem to fall behind one year after the pandemic.

Keywords: COVID-19; School closures; Attainment deficits; Educational attainment; Standardised tests
JEL-classification: I21, I24
Declarations of interest: none.

1 Introduction

Almost two years after the outbreak of the COVID-19 pandemic, which has caused the largest disruption to education imparted in school in history (UN, 2020), attainment deficits and increased educational inequality are high on the policy agenda.[1] As a result, there are growing number of studies that have collected and analysed test scores of students in the aftermath of the COVID-19 crisis. Overall, the available evidence suggests that there is an increase in the inequality within and between schools, and attainment deficits in comparison to learning outcomes before the school closures (see review by Patrinos & Donnelly, 2021). In particular, educational attainment has been found to be significantly lower when measured through tests that were administered in June 2020, immediately after the first period of school closures (in Belgium, Maldonado & De Witte (2021); in the Netherlands, Engzell et al. (2021) and Haelermans et al. (2021b); in the UK, Blainey et al. (2020); in the USA, Kogan & Lavertu (2021)). Although some studies did not observe changes in test scores (e.g. in France, for sixth-grade students in French and mathematics (OECD, 2021); in the USA, among grade 3-8 students in reading outcomes (Kuhfeld et al., 2020); in Australia, among primary school students (Gore et al., 2021)), also after the 2020 summer break, most estimated effects were still significantly negative (in the USA, Domingue et al. (2021) and Pier et al. (2021); in the UK, Rose et al. (2021); in Germany, Schult et al. (2021) and Depping et al. (2021)). The achievement deficits in the aftermath of the school closures are worrisome as lower test outcomes

[1]The literature uses multiple terms to indicate the lost progress of student learning due to the COVID-19 school closures. Next to 'attainments deficits', terminology as 'learning loss', 'slow down in student learning', 'lack of student progress', or 'shortfall in education outcomes' is used. Throughout this paper we use 'attainment deficits'.

have been linked to reduced human capital formation which, in turn, can translate into long-term challenges. For example, human capital formation has been associated with annual lifetime earnings (Chetty et al., 2014), future earnings (Psacharopoulos et al., 2020), employment (Currie & Thomas, 2001), and general prosperity (Hanushek & Woessmann, 2020). However, these troubling expectations only hold in the absence of resilience of school outcomes and, hence, to the extent that the attainment deficits are not caught-up. This paper sheds light on this question by examining how the learning outcomes of students and inequality within schools evolved one year after the school closures.

As there are two different mechanisms at work, it is unclear whether and to what extent the lower test scores due to the disruption of schooling will be reversed one year after the start of the pandemic. On the one hand, theoretical models (e.g., Agostinelli et al., 2020) suggest that poorer levels of knowledge and skills might accumulate over time. In the absence of remedial actions, attainment deficits and inequality between students is predicted to increase. Second, empirical evidence suggests that school outcomes are not as resilient as often hoped for. For instance, earlier disruptions of schooling have been shown to have long term impacts (Belot & Webbink, 2010; Jaume & Willén, 2019). Third, the heterogeneous psychological impact on students might also result in hampered learning in the year after the school closures. Although students with personality traits of high conscientiousness and openness might have benefited from the remote learning, students with high extraversion and neuroticism are expected to decrease their school results (Iterbeke & De Witte, 2021). Finally, the year after the initial (long term) school closures was characterized by (national and local) disruptions in schooling; the use of face masks that reduced the ability of students and teachers to communicate, interpret, and mimic the expressions (Spitzer, 2020); and lower well-being of children.

On the other hand, some of the mechanisms resulting in the initial attainment deficits simply vanished once schools reopened. First, the lack of peer effects (i.e. students from different socioeconomic status (SES) groups mixing at school) has been shown as an important channel of attainment deficits (Agostinelli et al., 2020). Second, lower learning outcomes of students were linked to reduced instruction time that was not fully compensated by remote teaching (Di Pietro et al., 2020). Moreover, remote teaching during the pandemic has been shown to have a lower effectiveness (Lichand et al., 2021). These two mechanisms disappeared once schools reopened. Third, acting in the same direction, thanks to significant policy attention and investments (De Witte & Smet, 2021), attainment deficits might have been mitigated. Multiple countries invested significantly in evidence-based interventions such as summer schools, (private or small group) tutoring, focus on the core curriculum, accelerated learning, ICT equipment or mentoring. Fourth, students that took a test in June 2020 were unlikely to have reached the same level of preparation as students that were in the same grade in 2019. This is different for students taking the test in June 2021, who might even be better prepared than normally as many outside school activities were cancelled due to (national) restrictions and, consequently, more instruction time was available in 2021 than in other years. Finally, during the school disruptions in the school year 2020-2021 (i.e. the second and third pandemic waves), teachers and schools were better prepared for remote teaching, and providing online feedback such that these school closures might not have further hampered student learning.

The empirical literature analysing the effects on test outcomes one year after the COVID-19 school closures is still limited. So far, only a few studies have addressed this issue. In the Netherlands, Haelermans et al. (2021a) found that the initial attainment deficits had been partly caught up by March 2021. Nonetheless, there were still observed shortfalls amounting to 0.17 standard deviations (SD) for reading, 0.14 SD for mathematics, and 0.08 SD for spelling as compared to tests in March 2020. In England, analysing data from the summer of 2021, Blainey & Hannay (2021) observed decreasing but remaining shortfalls in mathematics and especially in the English language, and further increases in inequality. Further, Blainey & Hannay (2021) showed that, except for younger students, attainment in reading was close to pre-pandemic levels. In the Czech Republic, test scores in the Czech language and mathematics from the same tests performed in 2020 and 2021 show that pupils in fifth grade lost the equivalent of about 3 months of learning since the start of the pandemic (PAQ Research, 2021). In Italy, after observing a positive trend in test scores until 2019, the National Institute for the Educational Evaluation of Instruction and Training (INVALSI)[2] observed similar results in primary schools for both mathematics and Italian language, in 2021, with test scores that are mostly unchanged in comparison to 2019 (INVALSI, 2021a).[3] Similarly, in Australia, the first evidence from the Naplan data administered in May 2021 suggested that students performed similar to the test conducted in 2019 (ACARA, 2021). Yet, it should be noted that Australian educational attainments had been improving significantly (e.g., 0.4 SD for reading in grade 3, in the period 2008-2019) since the beginning of the Naplan tests (ACARA, 2019). Finally, public schools in Maryland, USA, recorded a drop in students' performances in math and literacy among all grades at the end of the 2020-2021 school year compared to 2018-2019. Among fifth graders, a 23.5 percentage point decrease in the number of students meeting the standards was found for literacy, while a decrease 25.8 percentage points was found for math (MCPS, 2021).

Given the conflicting mechanisms and the contradictory early evidence, from a policy point of view, assessing the impact of the school closures one year later represents a fundamental step to inform the implementation of (further) adequate remedial measures. For example, a recent survey responded

[2]INVALSI yearly assesses students' educational achievement in mathematics and reading at different stages in the educational system. In primary school, tests are administered at the end of the second and fifth grade.

[3]Differently, results in secondary schools for both math and Italian language showed a decline of 10 percentage points compared to 2018-2019 school year (INVALSI, 2021b).

by 142 Ministries of Education around the world found that 1 in every 3 countries have not taken measures to provide the needed support to students to catch up on their educational attainment after the school closures (UNESCO et al., 2021), while other countries already undertook significant policy actions (De Witte & Smet, 2021).

This paper focuses on a unique data-set comprising of administrative data and standardised tests in grade 6 of primary education in the Flemish region of Belgium, which makes it an interesting region to study because of several reasons. First, the comparison of standardised test scores collected in June 2020 with earlier standardised tests suggested significant attainment deficits in multiple subjects (Maldonado & De Witte, 2021). Given the potential accumulation of attainment deficits, it is relevant to see how the test scores changed. Second, students in primary school went back to full-time face-to-face education in the school year 2020/2021, with 'only' two weeks of extra school closures taken throughout the whole year. Given that in most countries there were longer additional schooling disruptions, the 'one-year later effect' observed in Flanders serves as a lower bound on attainment deficits for other education systems. Third, exactly the same standardised tests were administered in 2019, 2020 and 2021. As a result, this increases the internal validity of the analysis. The data are available for multiple subjects, i.e. mathematics, native language (Dutch), sciences, social sciences and foreign language (French). Fourth, even before the pandemic, the Flemish education system was characterised by a significant inequality in test scores (Franck & Nicaise, 2018). Maldonado & De Witte (2021) show that inequality within and between schools further increased due to school closures. In the aftermath of the first wave of school closures, multiple remedial policy interventions targeting the most vulnerable students (e.g. tutoring, summer schools) have been taken, potentially impacting test score inequality. Finally, we have unique panel data for Flanders allowing for an empirical specification that uses a school fixed effects regression analysis, hence controlling for (un)observed time invariant school characteristics. The data are available at both school and pupil level, and are augmented with administrative data as well as standardised tests since the school year 2015/2016. In the empirical specification, we also include a time trend capturing the steady decline of Flemish students' performances in international tests (e.g. PIRLS, PISA, TIMSS).[4] Multiple robustness tests, including coarsened exact matching, show the robustness of the results.

The results indicate significant attainment deficits since the start of the pandemic. In all subjects, except for social science, we observe lower student test scores after the pandemic than before. Moreover, the estimates provide a mixed picture of the resilience of student test scores one year after the long term COVID-19 school closures. The significant attainment deficits stopped for math, while the test scores cautiously started improving for science and significantly improved

[4]In this respect, thanks to the additional data wave, the empirical strategy in the present paper differs from Maldonado & De Witte (2021) as it includes school fixed effects and a time trend.

for social sciences. For the Dutch and French language, the attainment deficit further deepened in 2021. In spite of the overall halting in attainment deficits in math in 2021, using a quantile analysis we find that the math test scores of the best-performing students (i.e., quantile 70 to 95) have significantly declined, while those of low-performing students seem to have slightly improved (though not significantly). For the Dutch language, all students, irrespective of the test score quantile, increased in 2021 their attainment deficits. Further, the results suggested that the test score inequality rose for the Dutch language in 2021, and decreased for math. When considering marginal effects based on socioeconomic indicators, we find that attainment deficits increased for those schools that count more students with low educated mothers, while lower average decreases in attainment (though insignificant) are found for schools with students from a more disadvantaged neighbourhood. Finally, with respect to the impact of remedial actions, our results suggest that, despite evidence that summer schools were established in the most disadvantaged areas, in areas with summer schools the attainment deficits in both the Dutch language and mathematics were halted in 2021 as compared to 2020, contrary to areas without summer schools.

The paper unfolds as follows: Section 2 presents the setting, while Section 3 describes the standardised test and administrative data. After outlying the applied empirical specification, we present the main results for education attainments and inequality. Section 6 discusses the marginal effects, and the results from a quantile analysis. We conclude with multiple robustness tests and a conclusion.

2 Setting

2.1 The Flemish education system

This paper focuses on primary education in the Flemish region of Belgium. Primary education lasts for 6 years (from age 6 to 12) after which students move on to secondary education (from age 12 to 18). The language used as a medium of instruction is Dutch, while French is taught as the first foreign language beginning from grade 5 of primary school. In the Flemish community of Belgium, not only do schools enjoy a high degree of autonomy, but also children (and their parents) have full freedom to choose any primary or secondary school (i.e. there are no catchment areas). In this heterogeneous educational landscape, the network of Catholic schools (Katholiek Onderwijs Vlaanderen) is publicly funded, but privately-run, and it is the largest education provider in Flanders. It attracts a slightly more advantaged student population (Cherchye et al., 2010).

Although there are no central examinations organised at the national level, the network of Catholic schools administers standardised test every year to students in grade 6. These tests allow schools to self-evaluate their student's performances and are part of the network's internal quality assessment. As a result, the test outcomes are not shared with the general public or central government (see section 2.2). Since 2016, a downward trend has been observed in the results for reading comprehension in grade 4 of Flemish primary schools, as measured by the international study of reading achievement PIRLS (Dockx et al., 2019). The negative trend in attainment carries on in secondary school as also confirmed by the OECD's Programme for International Student Assessment (PISA), whose 2018 assessment found the test scores of 15-year-old students in Flanders to be declining in all subjects (reading, mathematics and science) (PISA, 2018).

2.2 Education during the COVID-19 pandemic

In an attempt to slow down the spread of the COVID-19 virus, all Flemish primary and secondary schools were closed for a total of nine weeks during the first wave of the pandemic in the Spring of 2020. In particular, schools were closed from March 16, 2020 to May 15, 2020. As the school closure came as a surprise for all stakeholders and as the initial aim was to reopen schools after only a short closure, schools were advised to only review previously taught material during the first three weeks of the school closures. This period of rehearsing materials was followed by two weeks of the Easter holidays. Soon it became clear that schools would be closed for a longer period than initially expected. Therefore, in the next four weeks, schools were asked to start 'pre-teaching', i.e. covering new material in remote teaching that was planned to be taught again as soon as schools reopened. From May 15 onward, grades 1, 2 and 6 of primary education and the last year of secondary education were allowed to partially reopen. Students in grade 6 were allowed to attend school for a maximum of two full days, or four half-days per week, while remote teaching was kept on days when students had to stay at home. Furthermore, the maximum number of students per class was limited to fourteen, which caused most classes to split into two groups. For the other grades, remote teaching continued until June 8, such that for those students schools were closed for a total of 12 weeks. From June 8, all classes of primary education were allowed to reopen full-time until the start of the summer break on July 1 (Chenier et al., 2021; Maldonado & De Witte, 2021).

During the first period of school closures, which corresponds to about a third of a traditional school year, education was severely disrupted as the Flemish education system was poorly prepared for remote teaching. Despite the availability of ICT in schools, teachers were rarely integrating ICT in their pre-pandemic courses. Followed by this, they participated in fewer ICT-related professional development initiatives. Additionally, computer-assisted instruction methods were also uncommon. Before the crisis, the education inspectorate reported after inspection that the majority of the schools had an insufficient ICT strategy (De Witte, 2021). Although broadband coverage is very high in Flanders, particularly students with a vulnerable background could not be reached. Moreover, the same student group also lacked a dedicated quiet area for studying and were also subject to lower levels of parental involvement. Furthermore, during the period of 'pre-teaching', it was asked to limit the instruction time to half hours of a normal school day. This potentially slowed down the learning progress (Di Pietro et al., 2020).

After two months of summer break, the 2020-2021 school year started in relatively normal conditions on September 1. Flemish primary schools were allowed to reopen with full-time face-to-face education which was kept almost unchanged throughout the year. During the second and third wave of the COVID-19 pandemic, schools were closed twice for a short period of time: one week in November (in the week before the traditional Autumn break) and one week in April (in the week before the traditional Easter holidays). Yet, some anomalies compared to the usual teaching conditions were in place, possibly influencing the didactic process. Until Easter 2021, face masks were mandatory in the fifth and sixth grades of primary education once infection was detected. After Easter 2021, face masks became mandatory for all students in the fifth and sixth grades. Once a cluster of infections was detected, some classes and schools were also placed in quarantine. Moreover, all activities outside school were forbidden until June 2021, resulting in more instruction time.

To tackle the potentially negative consequences of the COVID-19 school closures, the Flemish government responded with a series of targeted remedial actions both during the school year and during the summer holidays (a more detailed overview of the policy responses in terms of additional education funding allocated in Belgium and other European countries can be found in De Witte & Smet (2021)). With respect to the former, additional funding was devoted to ICT, hiring and/or training teachers, and counselling and/or assistance for students. With respect to the latter, during the 2020 summer break, 138 fully subsidized summer schools were organised by local authorities and several education providers (schools, school communities and other education stakeholders). Summer schools targeted mainly low SES students, who were taught at 226 different locations for at least 10 full days (or 20 half days) in small classes of maximum 14 students. About 7,500 students participated in a summer school. The summer schools aimed at mitigating the attainment deficits, and preparing pupils for the 2020-2021 school year. The majority of summer schools took place during the last three weeks of August 2020, and were offered to students from primary education (Verachtert et al., 2020).[5]

3 Data

The analysis relies on a panel data-set covering a seven years period from 2015 to 2021.[6] The data-set combines: (1) data on standardised tests that are administered every year in June by the network of catholic schools in Flanders (Katholiek Onderwijs Vlaanderen), in the last year of

[5]Six out of ten summer schools hosted only students from primary schools, while 18% of the summer schools focused on both primary and secondary education students (Verachtert et al., 2020).

[6]In the remainder of the paper, we refer to 2015 to indicate school year 2015-2016, and so on.

primary school (grade 6); and (2) administrative data comprising several school characteristics.

3.1 Standardised tests

We observe student test scores at the individual level. We use both individual-level data, as well as aggregation at the school level. Because of organisational reasons, not all subjects were yearly tested since 2015.[7] The subjects 'mathematics' and (native) language (Dutch) were tested in all seven years under consideration.[8] The subjects science and social sciences were added to the tests in 2016, while the subject 'French' (second language) was introduced for the first time in 2019. In some years, exactly the same test was administered. In particular, four different test versions were administered since 2015: a test version in 2015, a different test version in 2016, the same test in 2017 and 2018, and the same test in 2019 and 2020. With a specific eye on the research question of the present paper, in 2021, both a new test version and the same 2019/2020 test version were administered in 64% of all participating schools. Thanks to schools taking both tests, we could equate the new 2021 test to the test taken in 2019 and 2020 (see section 4.2 for more details).

To allow for a comparison across time and with other literature, the test score variable has been normalized to mean 0 and standard deviation 1, per test version.[9] The standardisation per test allows estimates to be interpreted as the percentage of standard deviation. Given the different test versions, the results compare three time periods for the analysis: 2019-2021 (i.e. using only data from exactly the same or equated test), 2017-2021 (i.e. two different test versions) and 2015-2021 (i.e. four different test versions).[10]

Using the standardised tests as base, we constructed indicators for the inequality within schools. Using the individual-level data, we measure inequality within schools as the Gini coefficient and the spread between the 90 to 10 percentiles of the distribution of students. The Gini coefficient measures the inequality among values of the distribution of test scores, and it ranges between 0 (perfect equality) and 1 (maximal unequal distribution of test scores). While the Gini coefficient summarises the entire test scores distribution, the Ratio of 90/10 takes the ratio of scores at two

[7]The data is available for several subjects, each focusing on a number of sub-contents: Mathematics (focus on rounding; arithmetic; geometry), Dutch language (focus on reading comprehension), Science (focus on technology), Social Science (focus on Time) and French (focus on listening and reading comprehension).

[8]It is impossible to distinguish sub-subjects, such as algebra or geometry for mathematics; grammar, spelling or reading for language.

[9]Standard deviation (SD) is a measure of variability in educational attainments. Approximately 68 % of test scores are expected to fall between minus one and plus one SD around the mean.

[10]Unfortunately, standardised tests in grade 4 were not administered in school years 2019 and 2020, thus not allowing us to account for previous knowledge of the 2021 cohort. Therefore, and contrary to Maldonado & De Witte (2021), we do not include grade 4 tests in the analysis.

points in the distribution, i.e. the ratio of the top 10% of test scores (decile 10) to the lowest 10% of scores (decile 1).

3.2 Administrative data

The test score data is enriched with administrative data on school characteristics, characteristics of students in grade 6, and teachers' characteristics. The former data correspond to variables that earlier literature has found to have an impact on student performance. It includes the number of students enrolled in the school (as a proxy for school size), the share of girls, special needs education, the share of students with special needs, and indicators of socioeconomic status (SES) of the student (neighbourhood, mother's education and language spoken at home). As the standardised tests are administered in grade 6, we also add administrative data for this specific student population. In particular, we include information on the share of girls, the share of grade 6 students with grade repetition, the share of grade 6 students retaking grade 6 (i.e., as a proxy for the rehearsal effect of the test) and SES indicators for grade 6 students. The teachers' characteristics comprise the number of teachers in the school, and the share of teachers older than 50 years (as a proxy for increased risk for COVID-19 related diseases).

3.3 Sample and attrition

As participation in the standardised tests is voluntary for schools, the number of schools that takes the test varies each year. This results in an unbalanced panel structure. As shown in Table 1, with the exception of 2020, on an average there are about 35,000 students and 1,100 schools taking the test every year. This corresponds to about 85% of the catholic schools participating in the standardised tests. Despite the large majority participation of Catholic schools in the tests, the voluntary participation might result in selection effects of schools. Appendix A compares the school characteristics of those schools that participated in the tests for at least one subject and those schools of the same school network that did not participate in the test. The descriptive statistics in Appendix A confirm that the participating schools have a more advantaged student population in terms of neighbourhood of residence and mother education, and have a smaller share of students that are considered to be slow learners. The remainder school characteristics are all balanced. Given that the private Catholic schools attract in general a more advantaged student population, the latter observation suggests that the results might result in lower bound estimates of attainment deficits for the Flemish education system. Similarly, Maldonado & De Witte (2021) established that the 2015-2020 sample is not representative of the entire Flemish education system (nor of the network of the Flemish Catholic schools), as it is based on a more privileged student population on average. Figure B1 in Appendix shows the distribution of raw test scores (not standardised) for each subject across years 2019-2021.

Table 1: Participating schools and students in the different test years

	Schools		Students
	Total	*Participating*	*Participating*
	N	*N*(%)	*N*
2015	1293	1017 (79%)	29,280
2016	1293	1033 (80%)	30,783
2017	1288	1062 (82%)	32,344
2018	1293	1150 (89%)	36,057
2019	1289	1164 (90%)	37,321
2020	1300	401 (31%)	12,406
2021	1313	1158 (88%)	37,324

'Total' refers to the overall number of schools in the school network that offers grade 6. 'Participating' refers to the sub-sample of schools (and students) that participated in the test in each respective year.

4 Methodology

4.1 Empirical specification

We examine the effect on learning outcomes and inequality one year after the first wave of COVID-19 school closures. To distinguish between the overall impact of COVID-19 and the change in 2021, we construct two variables.

First, we construct a dummy variable '$COVID-19$' to indicate if the test scores were measured after the COVID-19 crisis. Hence, the dummy equals 1 in the years 2020 and 2021, and 0 otherwise. This variable measures the overall effect of the COVID-19 school closures. The estimated coefficient can be interpreted in terms of change in standard deviations due to the COVID-19 crisis. Second, we construct a dummy variable '*Change-in-2021*' to indicate if the test score was measured in 2021. This dummy, equal to 0 in all other years, measures the change in test scores one year after the first wave of COVID-19 school closures. Combining the two dummy variables in one regression model provides the net effect of the extent to which the school outcomes are resilient after the closures, ceteris paribus the impact of COVID-19 on test scores (i.e., COVID-19 variable).

The overall effect of the COVID-19 school closures, and the average change of standardised test scores in 2021, is investigated by means of the following regression model:

$$\begin{aligned} y_{i,t} = \alpha + \beta(COVID-19)_{i,t} + \gamma(Change-in-2021)_{i,t} + \\ + \delta X_{i,t} + \lambda Test_{i,t} + \theta Trend + \nu_i + \epsilon_{i,t} \end{aligned} \tag{1}$$

where $y_{i,t}$ denotes the analysed outcome variable (i.e. test score in each subject, inequality measures) of school i at time t. The outcome is regressed on a constant α, the dummy $(COVID-19)$ and the $(Change-in-2021)$ dummy. β and γ represent the two main coefficients of interest as β identifies the effect of the school closures, while γ isolates the sole effect on test scores 2021 relative to 2020 results. The school fixed effects (ν_i) remove any omitted variable bias arising from unobserved time-constant heterogeneity across schools (e.g. school and school board management quality, location of the schools, etc.) that may be correlated with regressors. Next, $Test$ are dummies capturing differences between the test versions. $Trend$ is a continuous-time trend that controls for the overall change in test scores over the years. $X_{i,t}$ denotes a vector of time-varying control variables at school, grade 6 and teacher level. Tables B1-B5 in Appendix indicate that gradually adding them to the regression model does not significantly change the results. Finally, the standard errors $\epsilon_{i,t}$ are clustered at the school level to capture the intra-school correlation of the outcome variable.

To examine the robustness of the results, regression model (1) is estimated on three different samples, corresponding to three separate time periods that follow from the tests' versions. First, the influence of the school closures is analysed on the 2019-2021 sample. Second, we compared the test scores and inequality in the 2017-2021 sample, and, third, in the 2015-2021 sample. Given the fact that not all subjects were administered each year, the regression model (1) is estimated for the subjects of mathematics, native language (Dutch), science, social sciences and foreign language (French) as outcome variable $y_{i,j}$ in the 2019-2021 sample; for all subjects but French in the 2017-2021 sample; and only for mathematics and Dutch in the 2015-2021 sample. The impact on educational inequality is examined in all three samples. Due to multicollinearity issues, the time trend is excluded from the 2019-2021 sample. However, as the mechanisms that underlie the decline in student test scores (e.g. low aspiration and motivation of students, low accountability of schools) are still present after the COVID-19 pandemic, we consider the 2017-2019 sample, with only two test versions and a time trend, as our main specification.

Next to different samples, we analyse the robustness of the results by estimating regression model (1) at both the school and student levels. Estimates at the former level provide evidence on how the average test score of an average school changed due to the educational disruption. On the other hand, estimates at the student level reveal how the average student test score changed due to the school closures. As variation in test scores is larger within schools than across schools, we expect the estimated effects to be smaller in the student-level analysis. However, in comparing the estimated coefficients, caution should be taken as we do not observe individual student characteristics, such that, also in the student level analysis, all control variables are at the school level. The estimated coefficients at pupil level are also likely to hide more heterogeneity than the estimates at the school level, which we test in a marginal effects analysis (Section 6).

4.2 Test equating

Although exactly the same test was administered in 2019 and 2020, a new version of these tests was developed in 2021. In order to achieve direct comparability between the scores of 2019/2020 and 2021, respectively, and also for measuring longitudinal trends of student learning outcomes, a series of equating techniques were performed.

For all subjects but French, a *single group equating design* was used to collect the 2021 test scores.[11] This implies that the same students were administered both the 2019/2020 test version and the new 2021 test version. In particular, 751 schools (64% of the 2021 sample of participating schools) agreed to take both tests for at least one subject. These schools administered first the 2019/2020 test version in May 2021 and administered the novel 2021 test version in June 2021. 421 schools took only the 2021 test version.

On the one hand, the single group design has the advantage of needing only one group of test-takers, and of controlling for differential student proficiency such that any other difference in the test scores will be attributed to testing differences, rather than student differences.[12] On the other hand, order effects (such as tiredness or familiarity with the test) can sometimes affect the students' performance, posing a challenge to the validity of the approach (Kolen & Brennan, 2014). In the present context, students who took both tests did so with about one-month time difference. Hence, they did not take both tests on the same day, such that both tiredness while taking the second test and familiarity with it can be excluded. Yet, it should be noted that the 2020 test could have been considered as an exercise for the students in the preparation of the official 2021 test. Thus, students might have had less incentive and less pressure to perform well on this test. As a result, it is also essential to note that students could have been less prepared for this test compared to when taking the official 2021 test form at the end of June. Nevertheless, this effect might be mitigated as also the test from June 2021 is a low stakes test for the students. Table A2 in Appendix compares the participating schools taking both test versions to the participating schools only taking the 2021 test. Relative to schools that only took the 2021 test version, schools taking both tests are bigger

[11]A similar approach to equating design was followed by the Italian INVALSI tests in 2021, allowing the direct comparison of learning outcomes of cohorts 2019 and 2021 (Desimoni, 2021).

[12]There is no confounding of test difficulty with group (of test takers) differences.

and have a greater number of teachers. All other characteristics are balanced, indicating that participating schools that took both test versions did not differ significantly from those schools who only took the 2021 test.

From a methodological point of view, the test equating was performed using the traditional *linear equating method*. This method adjusts the distribution of the scores. This means that, after equating, the two test versions have a comparable mean and standard deviation. Hence, test scores from the new 2021 test version can be directly compared to the 2019/2020 tests scores. Figure B2 in Appendix shows the distribution of test scores for each subject before equating. While both distributions are negatively skewed, the schools administering the 2020 test version show a lower mean score in every subject, indicating that students performed, on an average, slightly worse in the test version 2020, and better in the new 2021 test version. This difference might be either due to differences in difficulty levels between the two tests, or as students were less prepared for the old 2020 test. However, it should be noted that both tests were low stakes tests for the students, such that the former explanation is more likely than the latter explanation.

For the subject of French, the *common-item nonequivalent groups design* was employed to collect the 2021 test scores. This means that all students were administered the same new test version in June 2021. Nonetheless, this new test version was designed in manner that it encompassed three items in common with the test version administered in 2019/2020. Responses on the common-items were used as a (statistical) anchor to equate the two tests through linear equating, corresponding to the *Tucker observed score method* (Kolen & Brennan, 2014).

5 Results

5.1 Evolution of standardised test scores by subject

Table 2 presents the school and student level results for all test subjects. In the school level analysis (left-hand panel), the outcome variable measures the average standardised score in each subject at the school level. For the analysis at the student level (right-hand panel), the outcome variable is the standardised test score acquired for each student. We first run the analysis on the 2019-2021 sample, and then gradually extend it to the years 2017-2021 (second panel), and finally, to the years 2015-2021 (third panel). The results presented in Table 2 are from the most saturated regression model that includes clustered standard errors at school level, school fixed effects, and controls at school, grade 6 and teacher level. The results are robust to gradually adding these specifications (see Appendix B). In the latter two panels, we account for time trends over the years and for the test version, making the 2017-2021 sample our preferred specification (see section 4). As science and social science were taken the first time in 2017, these subjects are only presented in the 2017-2021 and 2019-2021 samples. The test scores for French are only available for the sample 2019-2021.

School-level analysis. The first column of Table 2 suggests that the standardised test scores for the Dutch language significantly decreased after the COVID-19 pandemic, and continued to decrease in 2021. The overall impact of COVID-19 accounts for an attainment deficit in the Dutch language of 0.24 standard deviations (SD), while the change in scores in 2021 amounts to an additional -0.23 SD (sample 2019-2021). The effect size and significance are robust across the sub-samples, although the coefficients tend to get smaller, that is, the attainment loss is less outspoken, as the sample is extended.[13] This is mainly due to including the time trend in the extended samples. Nevertheless, also in the preferred 2017-2021 sample, compared to before the COVID-19 crisis, attainment in Dutch is 0.17 SD lower after the long term school closures, and an additional 0.12 SD are added in 2021. Hence, compared to before the COVID-19 crisis (and accounting for the overall decline in language test scores as also observed in PIRLS surveys), Dutch attainment is 0.29 SD lower in 2021. The R^2 reveals that the model accounts for a range of 60% to 70% of the variation in the average score of schools across the different sub-samples.

The second column of Table 2 presents the main results for the standardised mathematics scores. The results indicate that the COVID-19 related school closures resulted in a decline in test scores of mathematics accounting for 0.11 SD. This is significant at 10%-level. Adding a trend does not seem to change the sign of the estimates, although the significance disappears in the third panel.[14] Although mathematics attainment was further found to be declining between 0.03 and 0.06 SD in 2021, this is not significantly different from 0. Thus, these estimates aggregately suggest that the impact of the COVID-19 school closures is halted, but not reversed yet.

With regard to science, the third column of Table 2 indicates that the overall effect of COVID-19 is significantly negative. Controlled for the time trend and (un)observed heterogeneity at the school level, science attainment dropped significantly with 0.15 SD. In spite of this overall decline in science test scores, the 2021 coefficient suggests a positive, though insignificant, change. This suggests that students in the 2021 cohort have started catching up with previous cohorts. However, it is essential to note here that this is not yet enough to bring to a significant improvement in their learning outcomes in science.

For social science, the fourth column of Table 2 shows a non-significant overall decrease in test scores in both 2019-2021 and 2017-2021 samples. This decrease amounts to -0.07 and -0.09 SD respectively. At the same time, the 2021 change in social science scores is positive, amounting to 0.09 standard deviations (non-significant) in the 2019-2021 sample, and 0.16 standard deviations

[13] As shown in Table B8 in Appendix, this difference is significantly different from 0.

[14] As shown in Table B8 in Appendix, the estimated coefficients in the different samples are not significantly different from 0.

(significant) in the 2017-2021 sample. This suggests that attainment in social sciences is improving over time.

Finally, a COVID-19 effect size similar to the Dutch language is observed for foreign language (French). However, the negative effect is much less pronounced for the sole change in average scores in 2021. It should be noted that this sample does not allow to include a time trend.

In Table B6 in Appendix, we compare whether the estimated coefficients for the subjects are significantly different from each other. As suggested by the differences in point estimates, the COVID-19 dummy is significantly different for the multiple subjects, while not significantly different between science and social science. Similarly, the 'change-in-2021' dummy is significantly different between the different subjects, but in the case of science versus math, and science versus social science.

Student-level analysis. At the student level, overall, we observe that the estimated coefficients have the same direction and significance as the school level estimates, providing confidence to the earlier discussed estimates. In line with the expectations (see section 4) the estimated coefficients are smaller in the student-level analysis. Controlling for the trend (i.e. the 2017-2021 sample), student attainments decreased significantly by 0.14 SD in the Dutch language, 0.07 SD in mathematics, 0.11 SD in science and 0.06 SD in social sciences since the outbreak of the COVID-19 pandemic. Dutch test scores further declined in 2021, adding up to an attainment deficits of 0.23 SD compared to the pre-COVID cohorts. The significance of the decline was halted in 2021 for mathematics and science, while the 2021 test scores started improving significantly for social sciences.

Comparing the school and student level estimates, despite the differences in estimated coefficients, once controlled for the trend in the 2017-2021 sample and the 2015-2021 sample, the school level and student level estimates do not differ significantly from each other for all subjects except Dutch language (see Table B7 in Appendix). The significantly lower student-level estimates for the Dutch language might have to do with the larger individual variation in Dutch test scores, while the variation in the Dutch language test scores at school level is less outspoken. We explore this further in Section 6.

Summary. Overall, the estimates provide further evidence on the detrimental effect of the COVID-19 pandemic. In all subjects, except social sciences, student test scores are significantly lower right from the beginning of the pandemic. Although the first signs of catching-up are visible for science and social science, and the decline is halted for mathematics, we witness a further deterioration of test scores of the Dutch language in 2021.

Table 2: Main results for all subjects

	School level analysis					Student level analysis				
	Dutch	Math	Science	Social Science	French	Dutch	Math	Science	Social Science	French
Sample 2019-2021										
COVID-19	-0.24***	-0.11*	-0.25***	-0.07	-0.24***	-0.14***	-0.05**	-0.13***	-0.06**	-0.17***
	(0.05)	(0.06)	(0.08)	(0.05)	(0.04)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)
Change in 2021	-0.23***	-0.06	0.04	0.09	-0.08*	-0.10***	-0.04	0.00	0.06**	-0.02
	(0.05)	(0.06)	(0.08)	(0.06)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
N	2272	2107	1190	1433	2003	76696	72127	46201	51413	69536
R^2	0.70	0.78	0.77	0.72	0.78	0.13	0.14	0.15	0.14	0.20
Sample 2017-2021										
COVID-19	-0.17***	-0.12*	-0.15*	-0.09		-0.14***	-0.07***	-0.11***	-0.06***	
	(0.05)	(0.07)	(0.08)	(0.06)		(0.02)	(0.02)	(0.03)	(0.02)	
Change in 2021	-0.12**	-0.03	0.07	0.16**		-0.09***	-0.02	-0.01	0.06***	
	(0.06)	(0.07)	(0.08)	(0.07)		(0.02)	(0.02)	(0.03)	(0.02)	
N	4626	4456	2670	2907		142207	137070	83889	89291	
R^2	0.63	0.71	0.68	0.63		0.12	0.14	0.12	0.11	
Sample 2015-2021										
COVID-19	-0.15***	-0.10				-0.11***	-0.06**			
	(0.05)	(0.07)				(0.02)	(0.03)			
Change in 2021	-0.16***	-0.05				-0.09***	-0.02			
	(0.06)	(0.07)				(0.02)	(0.03)			
N	6668	6504				199924	194725			
R^2	0.60	0.67				0.13	0.14			

Note: Robust standard errors, clustered at school level, between brackets. All outcome variables are standardized (mean 0, standard deviation 1) per test version. COVID-19 is a dummy variable equal to 1 in the years 2020 and 2021, measuring the overall impact of COVID-19 on test scores. 'Change in 2021' is a dummy variable equal to 1 in 2021, measuring the change of standardized test scores in 2021. All regressions include school fixed effects, dummy variables for the test version, a time trend (only for samples 2017-2021 and 2015-2021) and time varying controls at school level, for grade 6 characteristics and characteristics of the teacher. Controls at school level are: the number of students in the school, the share of girls, three SES indicators (neighbourhood, mother's education and home language). Characteristics of year 6 are: the share of girls, three SES indicators, the share of grade repetition and the share of slow learners. Characteristics of the teachers are: the number of teachers and the share of teachers older than 50.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

5.2 Inequality assessment

Earlier literature observed an increase in inequality in learning outcomes due to COVID-19 related school closures (Agostinelli et al., 2020; Engzell et al., 2021; Grewenig et al., 2021; McCoy et al., 2021; Maldonado & De Witte, 2021). Given the targeted policy interventions and the significant attention towards to most vulnerable students (De Witte & Smet, 2021), it is inconclusive how inequality has changed one year after the pandemic.

Table 3 presents the results of the change in inequality for the subjects of Dutch language and mathematics. The Gini coefficient ranges between 0 (perfect equality) and 1 (perfect inequality). The 90/10 inequality ratio measures how the best-performing (i.e. the 90th percentile of the distribution) students compared to the weakest performing students (i.e. the 10th percentile of the distribution). Hence, a higher value of the 90/10 ratio indicates higher inequality. As a reference for the inequality measures, the 'mean' presents the average inequality in the sample period before the COVID-19 pandemic.

In the 2019-2021 sample, the average pre-pandemic Gini coefficient amounts to 0.14 for both mathematics and the Dutch language. Using regression model (1), we estimate how inequality changed since the pandemic (variable 'COVID-19') and one year after the pandemic (variable 'Change-in-2021'). Table 3 presents the most saturated model that includes all control variables, school fixed effects, standard errors clustered at school level and a time trend (in the 2017-2021 and 2015-2021 samples).

Using the Gini coefficient as an inequality measure, the estimated coefficient for COVID-19 suggests that inequality in Dutch test scores rose by 0.01 (i.e. a 7% increase) after the pandemic. The rise in inequality continued in 2021, when the spread in Dutch test scores was already 14% higher than in 2019. The change in the Gini coefficient is less outspoken and mostly insignificant in the 2017-2021 and the 2015-2021 samples that include the time trend. The alternative inequality measure of the 90/10 ratio confirms this picture, with an increase of inequality of almost 16% in the 2019-2021 period. A significant 4% increase in 2021 is measured in the extended samples that include the time trend.

In mathematics, the change in inequality within schools is less outspoken and even reversed compared to the Dutch test scores. Although the significance levels differ across the different samples, the 2017-2021 sample estimates suggest that the spread in mathematics test scores declined by about 5% in 2021.

In sum, overall, one year after the COVID-19 school closures, the inequality within schools seems to have increased in the Dutch language, and decreased in mathematics.

Table 3: Inequality within schools

	Dutch		**Math**	
	Gini	Ratio 90/10	Gini	Ratio 90/10
Sample 2019-2021				
COVID-19	0.01***	0.16***	0.00*	0.09
	(0.00)	(0.05)	(0.00)	(0.06)
Change in 2021	0.01***	0.18***	-0.00	0.03
	(0.00)	(0.05)	(0.00)	(0.06)
N	2268	2234	2107	2050
Mean	0.13	2.02	0.14	2.07
R^2	0.64	0.57	0.69	0.58
Sample 2017-2021				
COVID-19	0.00	0.03	0.00	0.06
	(0.00)	(0.05)	(0.00)	(0.05)
Change in 2021	0.00**	0.09*	-0.01***	-0.11**
	(0.00)	(0.05)	(0.00)	(0.05)
N	4623	4601	4456	4421
Mean	0.12	1.91	0.13	1.94
R^2	0.56	0.41	0.59	0.47
Sample 2015-2021				
COVID-19	-0.00	0.04	0.01***	0.13***
	(0.00)	(0.04)	(0.00)	(0.05)
Change in 2021	0.00	0.09**	-0.01**	-0.04
	(0.00)	(0.05)	(0.00)	(0.05)
N	6665	6643	6504	6467
Mean	0.11	1.80	0.12	1.90
R^2	0.59	0.38	0.54	0.43

Note: Robust standard errors, clustered at school level, between brackets. COVID-19 is a dummy variable equal to 1 in the years 2020 and 2021. is a dummy variable equal to 1 in 2021. All regressions include school fixed effects, dummy variables for the test version, a time trend (only for samples 2017-2021 and 2015-2021) and time varying controls at school level, for grade 6 characteristics and characteristics of the teacher. A Gini coefficient of 0 indicates perfect equality, while 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. The mean is the baseline mean (i.e. computed excluding the 2020 and 2021 cohorts).

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

6 Heterogeneity analysis

As the test score analysis indicated lower average educational attainments since the COVID-19 pandemic, and a mixed picture on the test score resilience in 2021, it is interesting to look beyond the impact on the mean. First, we examine the marginal effects. In particular, we measure how the change in test scores one year after the first wave of COVID-19 school closures is different for schools that present diverse socioeconomic compositions of their students' population. Second, we study the distribution of test scores more in-depth using a quantile analysis. Third, we look at heterogeneous effects by the targeted remedial actions (in particular summer schools) implemented.

6.1 Marginal effects

Figure 1 pictures the marginal effects for the 'change-in-2021' dummy, separately for Dutch language and mathematics, for the sample 2017-2021. The figures present the marginal effects as estimated by the school fixed effect regression (solid lines) in the regression model (1). They also show the 95% confidence intervals around the estimates (dashed lines). The horizontal axis presents the percentage of disadvantaged grade 6 students in a school, as measured by disadvantaged neighbourhood (top panel), low education of the mother (middle panel) and not speaking the native (Dutch) language at home (bottom panel). We observe similar marginal effects at student level (see Figure B3 in Appendix).

First, consider the marginal effects by disadvantaged neighborhood. On the one hand, earlier literature indicated that the COVID-19 crisis was particularly harmful for more vulnerable students (OECD, 2020). These students were less frequently reached during the crisis, had a less supportive home situation, and missed more instruction time due to lower ICT availability. On the other hand, it suggests that these students might have especially benefited from the reopening of schools, leading to a quicker catching-up. Moreover, policy interventions were mainly targeting the most vulnerable students, such that these students might show more resiliency in their test scores (see Section 6.3). The upper panel of Figure 1 shows that schools with a higher share of students coming from a disadvantaged neighbourhood experienced lower average decreases in attainment (and are even catching-up, though insignificantly) after one year of full-time face-to-face education. This pattern is similar for Dutch (left-hand side) and mathematics (right-hand side).

Another aspect that has likely influenced the home learning environment of pupils during the school closures is parental involvement (Di Pietro et al., 2020). The level of maternal education, which is defined here as the percentage of students whose mothers' education is at primary level, helps us distinguishing between different levels of parental learning support. In the aftermath of the COVID-19 related school closures, attainment deficits were found to have increased with the percentage of students in a school with a low-educated mother (Maldonado & De Witte, 2021). Yet, maternal education still plays an important role: the middle panel in Figure 1 shows that the 2021 educational attainments decrease as the school counts more students with lowly educated mothers. This is observed for both the Dutch language and mathematics.

Finally, we estimate the marginal effect of the language spoken at home. Once schools reopened, non-native speaking students were again exposed to the Dutch language. Thus, the expectation is that these students will have especially taken advantage of the return to full-time face-to-face education. The bottom left panel in Figure 1 shows that, one year after the school closures, the estimated attainment change in Dutch remains roughly constant (and significant) along the share of non-native students. This suggests that all schools, no matter which language is spoken at home, decreased their attainments in the Dutch language in 2021. Although a similar observation is visible in recent PIRLS studies, it should be noted that the presented marginal effect is measured on top of a time trend. The lower panel of Figure 1 shows that schools with a higher share of non-native language speaking students experienced lower average decreases in math attainment (and are even catching-up, though insignificantly) after one year of full-time face-to-face education.

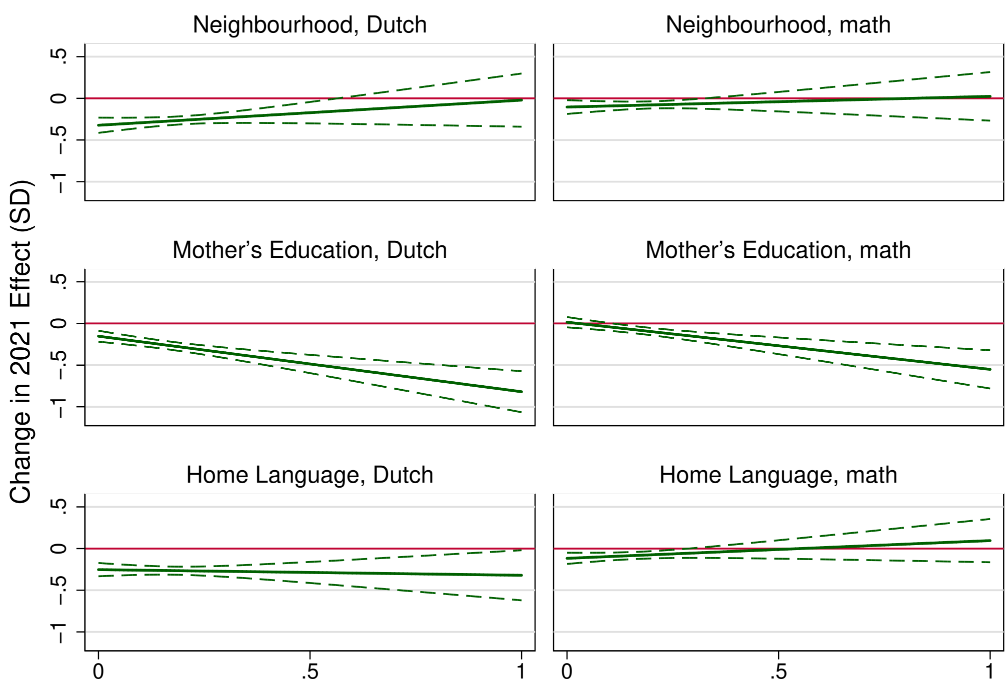


Figure 1: Marginal effects by socioeconomic composition of schools' population
Note: Marginal effects of 'change-in-2021' dummy, as estimated by a school fixed effect regression (solid lines) in regression model (1). 95% confidence intervals around the estimates (dashed lines). Estimates for the 2017-2021 sample at school level (pupil level estimates in Appendix B3).

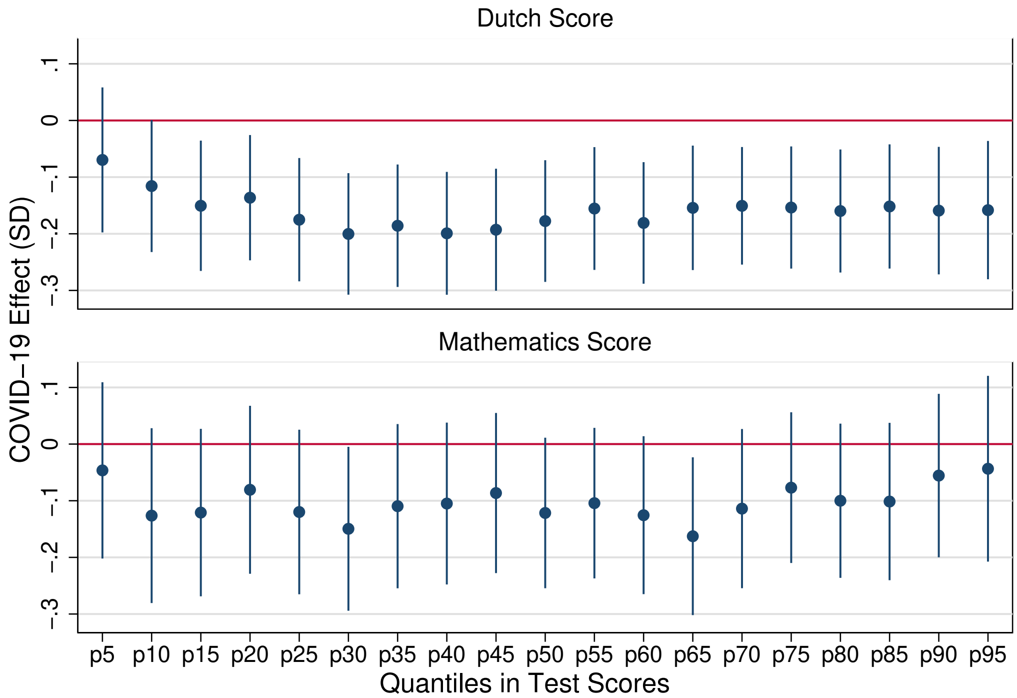
6.2 Quantile analysis

Using a quantile analysis, we analyzed the attainment deficits per quintile of the distribution of students' scores within each school (i.e., per 5%). This approach allows us to investigate whether low- and high-performing students are influenced differently by the COVID-19 school closures, and by the closures one year later.

The results are presented in Figure 2, where the upper figure presents the estimates per quintile for the overall COVID-19 effect and the lower figure shows the estimates per quintile for the 'change-in-2021'. Quintile P95 corresponds to the strongest students (i.e., only 5% of the students in the school obtained better test scores), while quintile P5 corresponds to the weakest students (i.e., only 5% of the students in the school obtained lower test scores). For the Dutch language, all students, except for the weakest 10% of the students, are found to underperform since the pandemic (Figure 2a). One year after the pandemic (Figure 2b), all students decreased their test scores. Only for the strongest students, the attainment deficits are no longer significantly different from zero one year after the pandemic.

For mathematics, as in Maldonado & De Witte (2021), no clear pattern emerges either when focusing on the effect of COVID-19 school closures (Figure 2a) as similar insignificant attainment deficits are observed across all quintiles. Instead, we observe that the students at the top and bottom of the distribution are influenced differently after one year of the school closures (Figure 2b): while the low-performing students in a school seem to have slightly improved their learning outcomes (though not significantly), the best-performing students (i.e., quantile 70 to 95) obtained significantly lower math test scores in 2021. This decline in math performance among the best-performing students is in line with evidence from the OECD PISA tests (PISA, 2018), suggesting that COVID-19 accelerated existing trends in education systems.

a) Marginal effect on 'COVID-19' dummy



b) Marginal effect on 'Change-in-2021' dummy

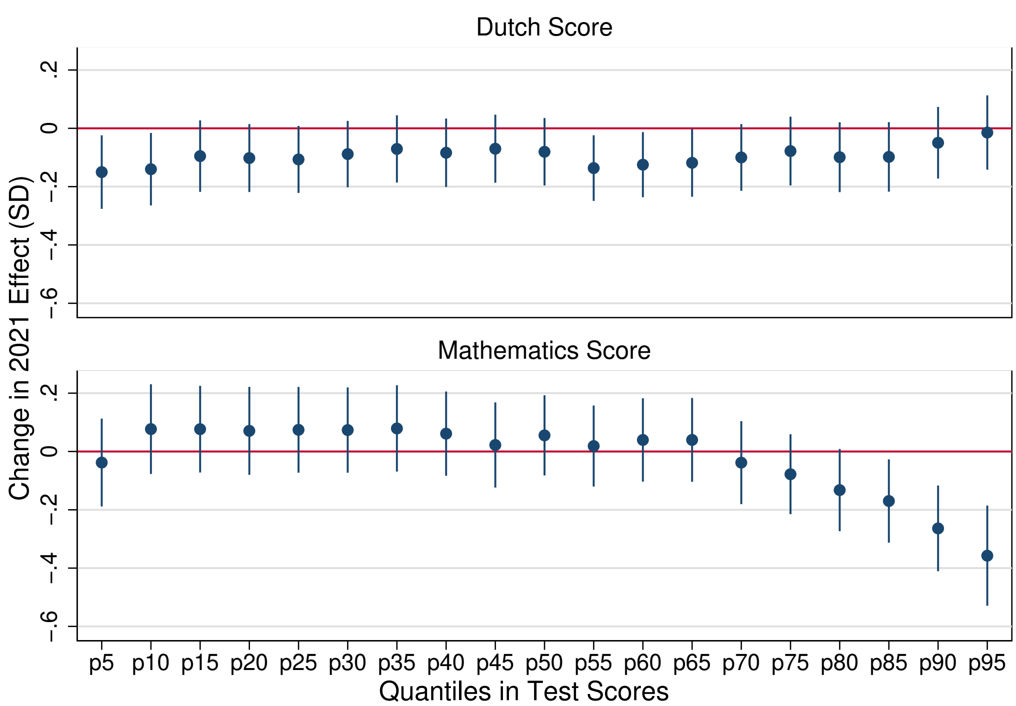


Figure 2: Estimated change of standardised test scores per quintile of test scores within schools
Note: These figures are based on a school fixed effects regression using the 2017–2021 sample which includes dummy variables for the test version, a time trend and time varying controls at school level, for grade 6 characteristics and characteristics of the teacher.

6.3 By targeted remedial actions

Next, we consider the heterogeneity in test scores for the Dutch language and math with respect to the availability of summer schools, which are publicly financed targeted remedial actions aiming at reducing COVID-19 related attainment deficits (see Section 2.2). We split schools into two groups according to whether (or not) the school is located in a postcode area where a summer school was implemented during the 2020 summer break. It should be noted that the remedial action did not necessarily take place at the school location where a student took the standardized tests, but within the same 4-digit postcode area. Summer schools were open to students from any school.[15]

The descriptive statistics in Table A3 in Appendix reveal that the summer schools were established in the postcode areas with a more disadvantaged student population. Confirming earlier findings by Maldonado & De Witte (2021), the results in Table 4 show that the overall COVID-19 effect is more outspoken in postcode areas where a summer school took place for both the Dutch language and math (-0.27 SD and -0.22 SD, respectively). Interestingly, a different picture emerges for the change in test scores in 2021. In 2021, the attainments deficits halted for both the Dutch language and mathematics in the postcode areas where a summer school took place. For the Dutch language, a non-significant decrease in the test scores (-0.02 SD) was found in 2021. For mathematics, we observe a non-significant increase (0.07 SD) in the test scores in 2021. This contrasts to increasing attainment deficits in those postcode areas where no summer school took place, amounting to a significant decrease of -0.17 SD in Dutch, and a non-significant decrease of -0.07 SD for math in the 2017-2021 sample.

As summer schools aim to compensate for the lack of parental support in learning, we might expect that their impact is more outspoken in schools with low-educated mothers. This is confirmed by the marginal effect analysis in Figure B4 in Appendix. The left panel shows that in postcode areas with summer schools the attainment deficits in both the Dutch language and math are halted in 2021 in schools with a larger share of low-educated mothers. In the absence of summer schools, the right panel indicates accumulating attainment deficits in schools with a larger share of low-educated mothers.

Overall, these results suggest that summer schools, which were mainly targeting the most vulnerable students, were successful in keeping the most vulnerable students from falling further behind in learning.

[15]We are unable to link individual student test scores to their potential participation in a summer school.

Table 4: Attainment deficit by targeted remedial actions (i.e. summer school)

	Dutch		**Math**	
	Summer School	No Summer School	Summer School	No Summer School
Sample 2017-2021				
COVID-19	-0.27***	-0.14**	-0.22*	-0.07
	(0.09)	(0.06)	(0.12)	(0.08)
Change in 2021	-0.02	-0.17**	0.07	-0.07
	(0.10)	(0.07)	(0.12)	(0.08)
N	1738	2873	1676	2766
R^2	0.66	0.59	0.73	0.68

Note: Robust standard errors, clustered at school level, between brackets. All outcome variables are standardized (mean 0, standard deviation 1) per test version. COVID-19 is a dummy variable equal to 1 in the years 2020 and 2021, measuring the overall impact of COVID-19 on test scores. 'Change in 2021' is a dummy variable equal to 1 in 2021, measuring the change of standardized test scores in 2021. All regressions include school fixed effects, dummy variables for the test version, a time trend and time varying controls at school level, for grade 6 characteristics and characteristics of the teacher. 'Summer school' refers to schools in the school network that are located in a postcode area where a summer school took place during the 2020 summer break.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

7 Robustness checks

To test the robustness of the findings, we performed a series of additional analyses. First, due to the psychological impact of the COVID-19 pandemic and its school closures, along with less time that was available to students for preparing for the tests administered in June 2020, the 2020 test scores might be downward biased. In a first robustness test, we exclude the 2020 test scores from the sample, mimicking a situation as if the standardised tests were never administered in 2020. In this specification of the model, the estimated coefficient for the dummy 'Change-in-2021' captures the combined effect of COVID-19 on test scores (i.e., 'COVID-19' variable in the main specification) and the additional effect added in 2021. The first column of Table 5 shows that the main findings are robust to this alternative specification. Compared to before the COVID-19 crisis, Dutch language and mathematics attainments are 0.30 SD and 0.14 SD lower in 2021, respectively (in the preferred 2017-2021 sample).

In a second robustness test, we apply a different empirical method. In particular, we use coarsened Exact Matching (CEM; Iacus et al., 2012; Blackwell et al., 2009) in order to control for sample imbalances in covariates. CEM enables the comparison of test outcomes of schools with similar characteristics in the different years. The schools participating in each year are sequentially matched (based on the school characteristics available from the administrative data) to the schools that took part in the test in 2021. As shown in the third column of Table 5, the results for the Dutch language prove to be robust in this alternative econometric method, with a significant attainment deficit of 0.18 SD since the pandemic (COVID-19 dummy), and further deficits of 0.11 SD in 2021 ('Change-in-2021' dummy). For math, the matching analysis suggests a significant attainment deficit of 0.12 SD since the pandemic and an insignificant deficit in 2021. The direction of the effects and their sizes are similar to the main results in all sub-samples.

Finally, the third column of Table 5 shows that the results for Dutch language and mathematics are also robust to the exclusion of special needs schools that participated in the test, as both sign and effect sizes are similar to the main results across all sub-samples.

Table 5: Robustness checks

	Dutch			Math		
	Exclude 2020 test	Matching	Exclude special needs	Exclude 2020 test	Matching	Exclude special needs
Sample 2019-2021						
COVID-19		-0.28***	-0.28***		-0.13**	-0.13**
		(0.05)	(0.05)		(0.06)	(0.06)
Change in 2021	-0.49***	-0.22***	-0.22***	-0.18***	-0.04	-0.04
	(0.04)	(0.05)	(0.05)	(0.03)	(0.06)	(0.06)
N	1856	2256	2266	1924	2087	2103
R^2	0.74	0.70	0.70	0.80	0.79	0.79
Sample 2017-2021						
COVID-19		-0.18***	-0.19***		-0.12*	-0.13**
		(0.05)	(0.05)		(0.06)	(0.06)
Change in 2021	-0.30***	-0.11**	-0.11*	-0.14***	-0.01	-0.02
	(0.07)	(0.06)	(0.06)	(0.06)	(0.07)	(0.07)
N	4241	4582	4611	4286	4410	4443
R^2	0.65	0.63	0.63	0.72	0.71	0.71
Sample 2015-2021						
COVID-19		-0.16***	-0.16***		-0.10	-0.11
		(0.05)	(0.05)		(0.07)	(0.07)
Change in 2021	-0.31***	-0.15***	-0.15***	-0.14***	-0.03	-0.04
	(0.06)	(0.06)	(0.06)	(0.05)	(0.07)	(0.07)
N	6286	6587	6650	6334	6423	6488
R^2	0.61	0.60	0.60	0.68	0.67	0.67

Note: Robust standard errors between brackets. All outcome variables are standardized (mean 0, standard deviation 1) per test version. COVID-19 is a dummy variable equal to 1 in the years 2020 and 2021, measuring the overall impact of COVID-19 on test scores. 'Change in 2021' is a dummy variable equal to 1 in 2021, measuring the change of standardized test scores in 2021. All regressions include school fixed effects, dummy variables for the test version, a time trend (only for samples 2017-2021 and 2015-2021) and time varying controls at school level, for grade 6 characteristics and characteristics of the teacher.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

8 Conclusion and discussion

Almost two years after the first COVID-19 induced school closures, the question remains as to whether, and to what extent, school outcomes are resilient and inequality persists. Although most studies suggested some level of attainment deficits immediately after the first wave of school closures, the early evidence and underlying mechanisms on the school outcomes one year after the pandemic outline a mixed picture.

This paper focuses on the Flemish region of Belgium, where Maldonado & De Witte (2021) found significant attainment deficits in multiple subjects in June 2020, as compared to before the COVID-19 school closures. As a response, the attainment deficits received significant policy attention resulting in evidence-based remedial interventions, such as tutoring and summer schools, targeting mainly disadvantaged students. Moreover, despite different teaching conditions (e.g. the use of face masks), primary schools were able to offer full-time face-to-face education. This paper provides evidence on the effects on learning outcomes and inequality at the end of primary school one year after the first wave of COVID-19 school closures, using a unique data-set that combines standardised test scores from a large share of private Flemish primary schools over the period 2015-2021, and administrative data comprising several school characteristics. From a methodological perspective, we are using a school fixed effects regression that accounts for a time trend and time-varying control variables. Multiple robustness tests, including coarsened exact matching, different samples and different levels of analysis (school level and student level) show the robustness of the results.

The results indicate a significant decrease in test outcomes since the start of the pandemic. In all subjects, except for social science, we observe significant attainment deficits since the start of the pandemic. The point estimates vary from -0.24 SD for foreign language French, -0.17 SD for the native Dutch language, to -0.15 SD in sciences and -0.12 SD in math. Furthermore, the estimates provide a mixed picture on the resilience of student test scores one year after the COVID-19 related school closures. The significant attainment deficits halted for mathematics, while the test scores cautiously started improving for science and significantly improved for social sciences (+ 0.16 SD). For the Dutch language, the attainment deficit that occurred in 2020 further deepened with -0.12 SD in 2021, indicating an additional decline of about 70% as compared to 2020 decline of -0.17 SD. For the foreign language French, school level estimates reveal significant additional attainment deficits (-0.08 SD) in 2021. At the student level, given the larger variation in test scores within schools than between schools, we observe smaller point estimates, although overall the estimated coefficients have the same direction and significance as the school level estimates. In particular, student attainments decreased significantly by 0.17 SD in French language, 0.14 SD in the Dutch language, 0.07 SD in mathematics, 0.11 SD in science and 0.06 SD in social sciences in comparison to the pre-pandemic trend. Estimated at pupil level, Dutch test scores further declined in 2021 with 0.10 SD. With respect to inequality, the results suggest that test score inequality within schools rose for the Dutch language in 2021, and decreased for mathematics.

However, not all students and schools are affected in the same way. Notwithstanding the halted attainment deficits in math in 2021, a quantile analysis suggests that the math test scores of the best-performing students in a school (i.e., quantile 70 to 95) have significantly declined, while those of low-performing students seem to have slightly improved (though not significantly). For the Dutch language, the quantile analysis suggests that all students obtained attainment deficits, irrespective of the quantile. When considering marginal effects of the 'change-in-2021' dummy based on socioeconomic indicators, we find that attainment deficits increase in 2021 for those schools that count more students with lowly educated mothers. Lower average decreases in 2021 attainment (though insignificant) are found for schools with students from a more disadvantaged neighbourhood. At the same time, schools with a more advantaged socioeconomic status composition in terms of neighbourhood of residence, and home language, seem to experience higher (significant) attainment deficits one year after the pandemic. Furthermore, we observe heterogeneity in test scores with respect to the availability of summer schools, which are publicly financed targeted remedial actions aiming at reducing COVID-19 related attainment deficits. In 2021, the attainment deficits observed in the postcode areas with a summer school have halted for both the Dutch language and mathematics. At the same time, the attainment deficits in Dutch increased in those postcode areas where no summer school took place.

Our findings for Dutch language and math confirm earlier theoretical predictions by Agostinelli et al. (2020), arguing that attainment deficits accumulate over time. In fact, even though the targeted policy actions have partly mitigated the negative effects on learning outcomes of more vulnerable students, we still observe (on top of a time trend and school fixed effects) significant attainment deficits in the Dutch language and mathematics, though the latter not significant. For science and social science, the attainment deficits seem to have not accumulated over time, as they both show a recovery. Thus, the results presented in this paper support the findings of Belot & Webbink (2010) and Jaume & Willén (2019) that school outcomes are not as resilient as often hoped for, and that schooling disruptions might have negative impacts on the longer term.

The differences between the subjects can be explained from two angles. First, the interactionist approach to language acquisition and learning (Vygotsky, 1978; Nor & Ab Rashid, 2018) suggests that language acquisition in children is a cyclical and incremental process, which mainly happens through social interaction - not only with adults, but also with peers. Hence, young students that were affected by the COVID-19 long term school closures are likely to have accumulated language delays due to the missed interactions. These effects could be long-lasting. Moreover, the use of face masks during the instruction time likely had a greater impact on the learning process of languages, than on other subjects, as it is more reliant on communication - which has been found to be largely impaired by the use of face coverings (Caniato et al., 2021; Saunders et al., 2021). This is reflected in the further deepening of 2021 attainment deficits in Dutch language compared to 2020, and the additional declines that were observed in French test scores in the sample 2019-2021. Second, early evidence on the use of computer-assisted learning during the lockdown suggested that teaching aided by an adaptive practicing software mitigated, or even reversed, the negative effects of school closures on math learning (Meeter, 2021). Similar software is more common in math education.

The findings of this paper have important policy implications considering the unprecedented challenge posed by the school closures on the educational system and the policy actions undertaken so far (De Witte & Smet, 2021; UNESCO et al., 2021). First, it seems that disadvantaged students have partially recouped the attainment deficits upon returning to school, or at least at a faster rate than the best-performing students. Our finding suggests that the policy actions (in particular summer schools), which were mainly targeting the most vulnerable students, were relatively successful. Nevertheless, our findings suggest that significant policy attention should also be given to the best-performing students, who seem to fall behind one year after the pandemic. Thus, this current set of policy actions should be integrated with the implementation of remedial measures aimed at all students, including the best-performing students. Second, the findings of this paper reinforce the need for continuous close monitoring of the evolution of the test scores and other school outcomes (e.g., school dropout, participation in higher education), in order to avoid the further accumulation of the attainment deficits. Third, the COVID-19 pandemic seems to accentuate existing trends in school outcomes. Specifically in the Flemish education system, the COVID-19 pandemic reinforced the already downward trend of Flemish students' performances in international tests (e.g. PIRLS, PISA, TIMSS). Therefore, next to a 'Build back' strategy, also a 'Build better' strategy should be implemented (see also De Witte, 2021).

References

ACARA (2019). *NAPLAN Achievement in Reading, Writing, Language Conventions and Numeracy: National Report for 2019.* National report. Australian Curriculum Assessment and Reporting Authority, Sydney.

ACARA (2021). *NAPLAN 2021 summary results*. Available online at: `https://reports.acara.edu.au/NAP` [Accessed 24 September 2021].

Agostinelli, F., Doepke, M., Sorrenti, G., & Zilibotti, F. (2020). *When the Great Equalizer Shuts Down: Schools, Peers, and Parents in Pandemic Times.* Working Paper 28264, National Bureau of Economic Research. Series: Working Paper Series.

Belot, M. & Webbink, D. (2010). Do teacher strikes harm educational attainment of students? *Labour*, 24(4), 391–406.

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

Blainey, K. & Hannay, T. (2021). *The impact of school closures on spring 2021 attainment*. Report.

Blainey, K., Hiorns, C., & Hannay, T. (2020). *The Impact of Lockdown on Children's Education: A Nationwide Analysis*. Report.

Caniato, M., Marzi, A., & Gasparella, A. (2021). How much COVID-19 face protections influence speech intelligibility in classrooms? *Applied Acoustics*, 178, 108051.

Chenier, M., Maldonado, J., & De Witte, K. (2021). Student test scores during the covid-19 crisis. *In Dohmen, D. Hurrelmann, K. (Eds.). Generation Corona (149-164). Weinheim/Basel: Beltz Juventa.*

Cherchye, L., De Witte, K., Ooghe, E., & Nicaise, I. (2010). Efficiency and equity in private and public education: A nonparametric comparison. *European Journal of Operational Research*, 202(2), 563–573.

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.

Currie, J. & Thomas, D. (2001). Early test scores, school quality and SES: Longrun effects on wage and employment outcomes. *Worker Wellbeing in a Changing Labor Market*, (pp. 103–132). Publisher: JAI Press.

De Witte, K. (2021). Hoe kunnen we de leerachterstand ombuigen? naar een vlaamse 'build-back' 'build-better' strategie. Available online at: `https://feb.kuleuven.be/drc/LEER/downloads/vlaams-parlement-leerachterstand-12-03-2021.pdf` [Accessed 18 September 2021].

De Witte, K. & Smet, M. (2021). *Financing Education in the Context of COVID-19*. Ad hoc report 3/2021, European Expert Network on Economics of Education (EENEE).

Depping, D., Lücken, M., Musekamp, F., & Thonke, F. (2021). Kompetenzstände Hamburger Schüler*innen vor und während der Corona-Pandemie [Alternative pupils' competence measurement in Hamburg during the Corona pandemic]. In *Schule während der Corona-Pandemie. Neue Ergebnisse und Überblick über ein dynamisches Forschungsfeld* (pp. 51–79). Münster ; New York: DDS – Die Deutsche Schule, Beiheft 17, waxmann edition.

Desimoni, M. (2021). *Il confronto tra le coorti 2019 e 2021: le basi psicometriche. [The comparison between 2019 and 2021 cohorts: the psychometric bases.]*. Nota Metodologica [methodological aspects], INVALSI.

Di Pietro, G., Biagi, F., Dinis Mota Da 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*. JRC121071. Luxembourg: Publications Office of the European Union.

Dockx, J., Van Landeghem, G., Aesaert, K., Van Damme, J., & De Fraine, B. (2019). *Begrijpend lezen van het vierde naar het zesde leerjaar. Herhalingsmeting van PIRLS in 2018 vergeleken met PIRLS 2016. [Reading comprehension from fourth to sixth grade. PIRLS 2018 compared to PIRLS 2016.]*. Technical report, Centrum voor Onderwijseffectiviteit en -evaluatie KU Leuven.

Domingue, B. W., Hough, H. J., Lang, D., & Yeatman, J. D. (2021). *Changing Patterns of Growth in Oral Reading Fluency During the COVID-19 Pandemic*. Policy Analysis for California Education.

Engzell, P., Frey, A., & Verhagen, M. D. (2021). Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, 118(17).

Franck, E. & Nicaise, I. (28-03-2018). Social inequalities in the flemish education system. comparing pisa 2003 - pisa 2015. Leuven Economics of Education Research Conference [PowerPoint slides].

Gore, J., Fray, L., Miller, A., Harris, J., & Taggart, W. (2021). The impact of COVID-19 on student learning in New South Wales primary schools: an empirical study. *The Australian Educational Researcher*, 48(4), 605–637.

Grewenig, E., Lergetporer, P., Werner, K., Woessmann, L., & Zierow, L. (2021). COVID-19 and educational inequality: How school closures affect low- and high-achieving students. *European Economic Review*, 140, 103920.

Haelermans, C., Aarts, B., Abbink, H., Jacobs, M., van Vugt, L., van Wetten, S., & van der Velden, R. (2021a). *A full year COVID-19-crisis with interrupted learning and two school closures: The effects on learning gains and inequality in primary education*. Working paper, Mimeo.

Haelermans, C., van der Velden, R., Aarts, B., Bijlsma, L., Jacobs, M., Smeets, C., van Vugt, L., & van Wetten, S. (2021b). Factsheets eerste schoolsluiting n.a.v. covid-19-crisis. Available online at: `https://www.nationaalcohortonderzoek.nl/factsheets/`.

Hanushek, E. A. & Woessmann, L. (2020). *The economic impacts of learning losses*. OECD Education Working Papers 225.

Iacus, S. M., King, G., & Porro, G. (2012). Causal Inference without Balance Checking: Coarsened Exact Matching. *Political Analysis*, 20(1), 1–24. Publisher: Cambridge University Press.

INVALSI (2021a). I risultati INVALSI 2021 della Primaria in Italiano e Matematica. Section: Prove INVALSI.

INVALSI (2021b). *Rilevazioni nazionali degli apprendimenti 2020-21. I risultati in breve delle prove INVALSI 2021*. Report, INVALSI.

Iterbeke, K. & De Witte, K. (2021). Helpful or harmful? The role of personality traits in student experiences of the COVID-19 crisis and school closure. *Personality and Social Psychology Bulletin*. In Press.

Jaume, D. & Willén, A. (2019). The long-run effects of teacher strikes: evidence from argentina. *Journal of Labor Economics*, 37(4), 1097–1139.

Kogan, V. & Lavertu, S. (2021). *The COVID-19 pandemic and student achievement on Ohio's third-grade English language arts assessment*. Technical report, Ohio State University.

Kolen, M. J. & Brennan, R. L. (2014). *Test Equating, Scaling, and Linking: Methods and Practices*. Statistics for Social and Behavioral Sciences. New York: Springer-Verlag, 3 edition.

Kuhfeld, M., Tarasawa, B., Johnson, A., Ruzek, E., & Lewis, K. (2020). Initial findings on students' reading and math achievement and growth. *NWEA Research*, (pp.$\tilde{1}$2).

Lichand, G., Dória, C. A., Neto, O. L., & Cossi, J. (2021). The impacts of remote learning in secondary education: Evidence from brazil during the pandemic. DOI: `http://dx.doi.org/10.18235/0003344`.

Maldonado, J. E. & De Witte, K. (2021). The effect of school closures on standardised student test outcomes. *British Educational Research Journal*. DOI: `https://doi.org/10.1002/berj.3754`.

McCoy, D. C., Cuartas, J., Behrman, J., Cappa, C., Heymann, J., López Bóo, F., Lu, C., Raikes, A., Richter, L., Stein, A., & Fink, G. (2021). Global estimates of the implications of COVID-19-related preprimary school closures for children's instructional access, development, learning, and economic wellbeing. *Child Development*. DOI: https://doi.org/10.1111/cdev.13658.

MCPS (2021). MCPS students' math proficiency, literacy plummet after year of virtual classes. Section: Schools. Available online at: https://bethesdamagazine.com/bethesda-beat/schools/mcps-students-math-proficiency-literacy-plummet-after-year-of-virtual-classes/ [Accessed 27 September 2021].

Meeter, M. (2021). Primary school mathematics during Covid-19: No evidence of learning gaps in adaptive practicing results. *Trends in Neuroscience and Education*, (pp. 100163).

Nor, M. N. & Ab Rashid, R. (2018). A review of theoretical perspectives on language learning and acquisition. *Kasetsart Journal of Social Sciences*, 39(1), 161–167.

OECD (2020). *The impact of COVID-19 on student equity and inclusion: Supporting vulnerable students during school closures and school re-openings*. Technical report.

OECD (2021). *French national evaluations contribution to the analysis of the impact of the health crisis.* Country education responses to the coronavirus (COVID-19) pandemic.

PAQ Research (2021). Dopady pandemie covid-19 na žáky [impacts of the covid-19 pandemic on pupils]. Section: Schools. Available online at: https://www.paqresearch.cz/ [Accessed 30 September 2021].

Patrinos, H. & Donnelly, R. (2021). *Learning Loss During COVID-19: An Early Systematic Review*. Report. ISSN: 2693-5015.

Pier, L., Hough, H. J., Christian, M., Bookman, N., Wilkenfeld, B., & Miller, R. (2021). *COVID-19 and the educational equity crisis: Evidence on learning loss from the CORE Data Collaborative.* Policy Analysis for California Education.

PISA (2018). *Leesvaardigheid van 15-jarigen in Vlaanderen. Overzicht van de eerste resultaten van PISA 2018. [Reading skills of 15-year-olds in Flanders. Overview of the first results of PISA 2018.]*. Technical report, Universiteit Gent.

Psacharopoulos, G., Collis, V., Patrinos, H. A., & Vegas, E. (2020). Lost Wages: The COVID-19 Cost of School Closures. *SSRN Electronic Journal*.

Rose, S., Twist, L., Lord, P., Rutt, S., Badr, K., Hope, C., & Styles, B. (2021). *Impact of school closures and subsequent support strategies on attainment and socio-emotional wellbeing in Key Stage 1.* Interim Paper 1, National Foundation for Educational Research (NFER), London.

Saunders, G. H., Jackson, I. R., & Visram, A. S. (2021). Impacts of face coverings on communication: an indirect impact of COVID-19. *International Journal of Audiology*, 60(7), 495–506. Publisher: Taylor & Francis. DOI: `https://doi.org/10.1080/14992027.2020.1851401`.

Schult, J., Mahler, N., Fauth, B., & Lindner, M. A. (2021). *Did Students Learn Less During the COVID-19 Pandemic? Reading and Mathematics Competencies Before and After the First Pandemic Wave*. Psyarxiv preprints.

Spitzer, M. (2020). Masked education? the benefits and burdens of wearing face masks in schools during the current corona pandemic. *Trends in neuroscience and education*, 20, 100138.

UN (2020). *Policy Brief: Education during COVID-19 and beyond*. UN Policy Brief.

UNESCO, UNICEF, World Bank, & OECD (2021). *What's Next? Lessons on Education Recovery: Findings from a Survey of Ministries of Education amid the COVID-19 Pandemic*. Report, UNESCO, UNICEF, World Bank, OECD, Paris, New York, Washington D.C.

Verachtert, P., Bellens, K., Sneyers, E., Surma, T., & Muijs, D. (2020). *Zomerscholen in Vlaanderen 2020: evalueren en inspireren [Summer schools in Flanders 2020: evaluate and inspire]*. Onderzoeksrapport, Thomas More-hogeschool, Antwerpen.

Vygotsky, L. (1978). Interaction between learning and development. In *Mind in society: The development of higher psychological processes* (pp. 79–91). Cambridge, MA: Harvard University Press.

Appendix A: Attrition

Table A1: Attrition and descriptive statistics of participating schools in the year 2021

	Attrited		**Participated**		t-Test
	N	Mean [SD]	N	Mean [SD]	p-Value
Number of Students	174	207.12 [99.96]	1139	198.69 [88.10]	0.25
Share of Girls	174	0.50 [0.36]	1139	0.50 [0.40]	0.34
SES - Neighbourhood	171	0.25 [0.31]	1131	0.19 [0.28]	0.01
SES - Mother's Education	171	0.20 [0.16]	1131	0.17 [0.15]	0.04
SES - Home Language	171	0.21 [0.23]	1131	0.19 [0.21]	0.20
Newcomers	174	0.01 [0.02]	1139	0.01 [0.03]	0.38
Number of Teachers as FTE	174	13.45 [6.18]	1139	12.85 [5.42]	0.18
Teachers: Share Above 50	174	0.30 [0.13]	1139	0.29 [0.15]	0.61
Year 6: Number of Students	174	33.61 [17.42]	1139	32.21 [16.10]	0.29
Year 6: Share of Girls	174	0.49 [0.09]	1139	0.50 [0.11]	0.24
Year 6: SES - Neighbourhood	171	0.24 [0.30]	1131	0.18 [0.28]	0.02
Year 6: SES - Mother's Education	171	0.19 [0.16]	1131	0.16 [0.16]	0.04
Year 6: SES - Home Language	171	0.20 [0.23]	1131	0.17 [0.21]	0.10
Year 6: Grade Repetition	174	0.00 [0.00]	1139	0.00 [0.01]	0.70
Year 6: Slow Learners	174	0.11 [0.12]	1139	0.09 [0.09]	0.01

Notes: Standard deviations between brackets. 'Attrited' refers to schools in the private school network that offer grade 6, and did not administer the test. 'Participated' refers to the schools that offer grade 6, and administered the test for at least one subject. The p-value is derived from the t-test comparing the attrited with the participating schools. Schools that only administered the old test version 2020 in school year 2021 (14 schools), and special needs schools (5 schools) were excluded from the attrition analysis.

Table A2: Comparison of schools taking both 2020 and 2021 test versions: year 2021

	Test 2021 only		**Both tests**		t-Test
	N	Mean [SD]	N	Mean [SD]	p-Value
Number of Students	421	189.05 [83.75]	737	202.19 [90.63]	0.01
Share of Girls	421	0.50 [0.05]	737	0.50 [0.04]	0.54
SES - Neighbourhood	417	0.19 [0.28]	733	0.20 [0.29]	0.51
SES - Mother's Education	417	0.18 [0.15]	733	0.17 [0.15]	0.75
SES - Home Language	417	0.19 [0.22]	733	0.19 [0.22]	0.80
Newcomers	421	0.01 [0.03]	737	0.01 [0.03]	0.54
Number of Teachers as FTE	421	12.45 [5.39]	737	13.14 [5.82]	0.04
Teachers: Share Above 50	421	0.29 [0.15]	737	0.29 [0.15]	0.82
Year 6: Number of Students	421	30.37 [15.03]	737	32.93 [16.65]	0.01
Year 6: Share of Girls	421	0.50 [0.10]	737	0.50 [0.11]	0.88
Year 6: SES - Neighbourhood	417	0.18 [0.28]	733	0.19 [0.29]	0.47
Year 6: SES - Mother's Education	417	0.16 [0.17]	733	0.17 [0.16]	0.73
Year 6: SES - Home Language	417	0.17 [0.22]	733	0.17 [0.22]	0.69
Year 6: Grade Repetition	421	0.00 [0.01]	737	0.00 [0.01]	0.34
Year 6: Slow Learners	421	0.09 [0.09]	737	0.10 [0.09]	0.85

Notes: 'Test 2021 only' refers to schools in the school network that only administered the test version 2021. 'Both tests' refers to those schools that administered both the old and the new test for at least one subject. The t-test values are p-values that compare the two group of schools. Standard deviations are robust.

Table A3: Comparison of schools located in an area with a summer school: year 2021

	Summer school		**No Summer school**		t-Test
	N	Mean [SD]	N	Mean [SD]	p-Value
Number of Students	433	203.86 [87.36]	720	193.67 [88.86]	0.06
Share of Girls	433	0.50 [0.03]	720	0.50 [0.04]	0.09
SES - Neighbourhood	431	0.31 [0.33]	714	0.13 [0.23]	0.00
SES - Mother's Education	431	0.23 [0.18]	714	0.14 [0.12]	0.00
SES - Home Language	431	0.25 [0.24]	714	0.15 [0.19]	0.00
Newcomers	433	0.02 [0.03]	720	0.01 [0.02]	0.00
Number of Teachers as FTE	433	13.62 [5.43]	720	12.29 [5.37]	0.00
Teachers: Share Above 50	433	0.28 [0.15]	720	0.30 [0.15]	0.05
Year 6: Number of Students	433	32.75 [16.14]	720	31.56 [16.12]	0.22
Year 6: Share of Girls	433	0.51 [0.10]	720	0.50 [0.11]	0.18
Year 6: SES - Neighbourhood	431	0.29 [0.33]	714	0.12 [0.23]	0.00
Year 6: SES - Mother's Education	431	0.22 [0.19]	714	0.13 [0.13]	0.00
Year 6: SES - Home Language	431	0.23 [0.21]	714	0.13 [0.12]	0.00
Year 6: Grade Repetition	433	0.00 [0.00]	720	0.00 [0.00]	0.06
Year 6: Slow Learners	433	0.11 [0.11]	720	0.08 [0.08]	0.00

Notes: 'Summer school' refers to schools in the school network that were located in a postcode area where a summer school took place during the 2020 summer break. The t-test values are p-values that compare the two group of schools. Standard deviations are robust. Special needs schools (5 schools) were excluded from these descriptive statistics.

Appendix B: Additional tables and figures

Table B1: Main results: Dutch score (school level analysis)

	Dutch score			
Sample 2019-2021				
COVID-19	-0.27***	-0.24***	-0.24***	-0.24***
	(0.05)	(0.05)	(0.05)	(0.05)
Change in 2021	-0.23***	-0.23***	-0.24***	-0.23***
	(0.05)	(0.05)	(0.05)	(0.05)
N	2272	2272	2272	2272
R^2	0.69	0.69	0.69	0.70
Sample 2017-2021				
COVID-19	-0.18***	-0.17***	-0.17***	-0.17***
	(0.05)	(0.05)	(0.05)	(0.05)
Change in 2021	-0.12**	-0.13**	-0.12**	-0.12**
	(0.06)	(0.06)	(0.06)	(0.06)
N	4626	4626	4626	4626
R^2	0.62	0.62	0.63	0.63
Sample 2015-2021				
COVID-19	-0.14***	-0.14***	-0.14***	-0.15***
	(0.05)	(0.05)	(0.05)	(0.05)
Change in 2021	-0.16***	-0.17***	-0.17***	-0.16***
	(0.06)	(0.06)	(0.06)	(0.06)
N	6668	6668	6668	6668
R^2	0.59	0.59	0.60	0.60
School fixed effects	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
Test version	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers	No	No	No	Yes

Note: Robust standard errors, clustered at school level, between brackets. All outcome variables are standardized (mean 0, standard deviation 1) per test version. COVID-19 is a dummy variable equal to 1 in the years 2020 and 2021, measuring the overall impact of COVID-19 on test scores. 'Change in 2021' is a dummy variable equal to 1 in 2021, measuring the change of standardized test scores in 2021.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B2: Main results: mathematics score (school level analysis)

	Math score			
Sample 2019-2021				
COVID-19	-0.13**	-0.10*	-0.11*	-0.11*
	(0.06)	(0.06)	(0.06)	(0.06)
Change in 2021	-0.06	-0.06	-0.06	-0.06
	(0.06)	(0.06)	(0.06)	(0.06)
N	2107	2107	2107	2107
R^2	0.78	0.78	0.78	0.78
Sample 2017-2021				
COVID-19	-0.11*	-0.11*	-0.11*	-0.12*
	(0.07)	(0.07)	(0.07)	(0.07)
Change in 2021	-0.02	-0.03	-0.03	-0.03
	(0.07)	(0.07)	(0.07)	(0.07)
N	4456	4456	4456	4456
R^2	0.70	0.70	0.71	0.71
Sample 2015-2021				
COVID-19	-0.09	-0.09	-0.09	-0.10
	(0.07)	(0.07)	(0.07)	(0.07)
Change in 2021	-0.04	-0.05	-0.05	-0.05
	(0.07)	(0.07)	(0.07)	(0.07)
N	6504	6504	6504	6504
R^2	0.66	0.66	0.67	0.67
School fixed effects	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
Test version	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers	No	No	No	Yes

Note: Robust standard errors, clustered at school level, between brackets. All outcome variables are standardized (mean 0, standard deviation 1) per test version. COVID-19 is a dummy variable equal to 1 in the years 2020 and 2021, measuring the overall impact of COVID-19 on test scores. 'Change in 2021' is a dummy variable equal to 1 in 2021, measuring the change of standardized test scores in 2021.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B3: Main results: Science score (school level analysis)

	Science score			
Sample 2019-2021				
COVID-19	-0.27***	-0.26***	-0.25***	-0.25***
	(0.08)	(0.08)	(0.08)	(0.08)
Change in 2021	0.05	0.04	0.04	0.04
	(0.08)	(0.08)	(0.08)	(0.08)
N	1190	1190	1190	1190
R^2	0.76	0.76	0.77	0.77
Sample 2017-2021				
COVID-19	-0.19**	-0.17**	-0.15*	-0.15*
	(0.08)	(0.08)	(0.08)	(0.08)
Change in 2021	0.12	0.08	0.07	0.07
	(0.08)	(0.08)	(0.08)	(0.08)
N	2670	2670	2670	2670
R^2	0.67	0.67	0.68	0.68
School fixed effects	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
Test version	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers	No	No	No	Yes

Note: Robust standard errors, clustered at school level, between brackets. All outcome variables are standardized (mean 0, standard deviation 1) per test version. COVID-19 is a dummy variable equal to 1 in the years 2020 and 2021, measuring the overall impact of COVID-19 on test scores. 'Change in 2021' is a dummy variable equal to 1 in 2021, measuring the change of standardized test scores in 2021.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B4: Main results: Social Science score (school level analysis)

	Social Science score			
Sample 2019-2021				
COVID-19	-0.11**	-0.08	-0.08	-0.07
	(0.05)	(0.05)	(0.05)	(0.05)
Change in 2021	0.09	0.09	0.09	0.09
	(0.06)	(0.06)	(0.06)	(0.06)
N	1433	1433	1433	1433
R^2	0.71	0.72	0.72	0.72
Sample 2017-2021				
COVID-19	-0.11*	-0.10	-0.09	-0.09
	(0.06)	(0.06)	(0.06)	(0.06)
Change in 2021	0.17**	0.16**	0.16**	0.16**
	(0.07)	(0.07)	(0.07)	(0.07)
N	2907	2907	2907	2907
R^2	0.61	0.62	0.63	0.63
School fixed effects	Yes	Yes	Yes	Yes
Time trend	Yes	Yes	Yes	Yes
Test version	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers	No	No	No	Yes

Note: Robust standard errors, clustered at school level, between brackets. All outcome variables are standardized (mean 0, standard deviation 1) per test version. COVID-19 is a dummy variable equal to 1 in the years 2020 and 2021, measuring the overall impact of COVID-19 on test scores. 'Change in 2021' is a dummy variable equal to 1 in 2021, measuring the change of standardized test scores in 2021.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B5: Main results: French score (school level analysis)

	French score			
Sample 2019-2021				
COVID-19	-0.26***	-0.25***	-0.25***	-0.24***
	(0.04)	(0.04)	(0.04)	(0.04)
Change in 2021	-0.07*	-0.08*	-0.08*	-0.08*
	(0.04)	(0.04)	(0.04)	(0.04)
N	2003	2003	2003	2003
R^2	0.78	0.78	0.78	0.78
School fixed effects	Yes	Yes	Yes	Yes
School characteristics	No	Yes	Yes	Yes
Characteristics year 6	No	No	Yes	Yes
Teachers	No	No	No	Yes

Note: Robust standard errors, clustered at school level, between brackets. All outcome variables are standardized (mean 0, standard deviation 1) per test version. COVID-19 is a dummy variable equal to 1 in the years 2020 and 2021, measuring the overall impact of COVID-19 on test scores. 'Change in 2021' is a dummy variable equal to 1 in 2021, measuring the change of standardized test scores in 2021.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B6: Differences among subjects, p-values

Sample 2017-2021		Dutch	Math	Science
Math	COVID-19	0.01***		
	Change in 2021	0.00***		
Science	COVID-19	0.00***	0.09*	
	Change in 2021	0.00***	0.10	
Social Science	COVID-19	0.00***	0.00***	0.48
	Change in 2021	0.00***	0.00***	0.18

The difference among coefficients are based on the school level data.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B7: Differences between school and student level results, p-values

Sample 2017-2021	Dutch	Math	Science	Social Science
COVID-19	0.02***	0.10*	0.52	0.52
Change in 2021	0.00***	0.15	0.79	0.90

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table B8: Differences between results from different samples, p-values

Sample 2019-2021 vs 2017-2021	Dutch	Math	Science	Social Science
COVID-19	0.03**	0.28	0.68	0.57
Change in 2021	0.02**	0.23	0.37	0.55
Sample 2017-2021 vs 2015-2021				
COVID-19	0.03**	0.72		
Change in 2021	0.02**	0.51		

The difference among coefficients are based on the school level data.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

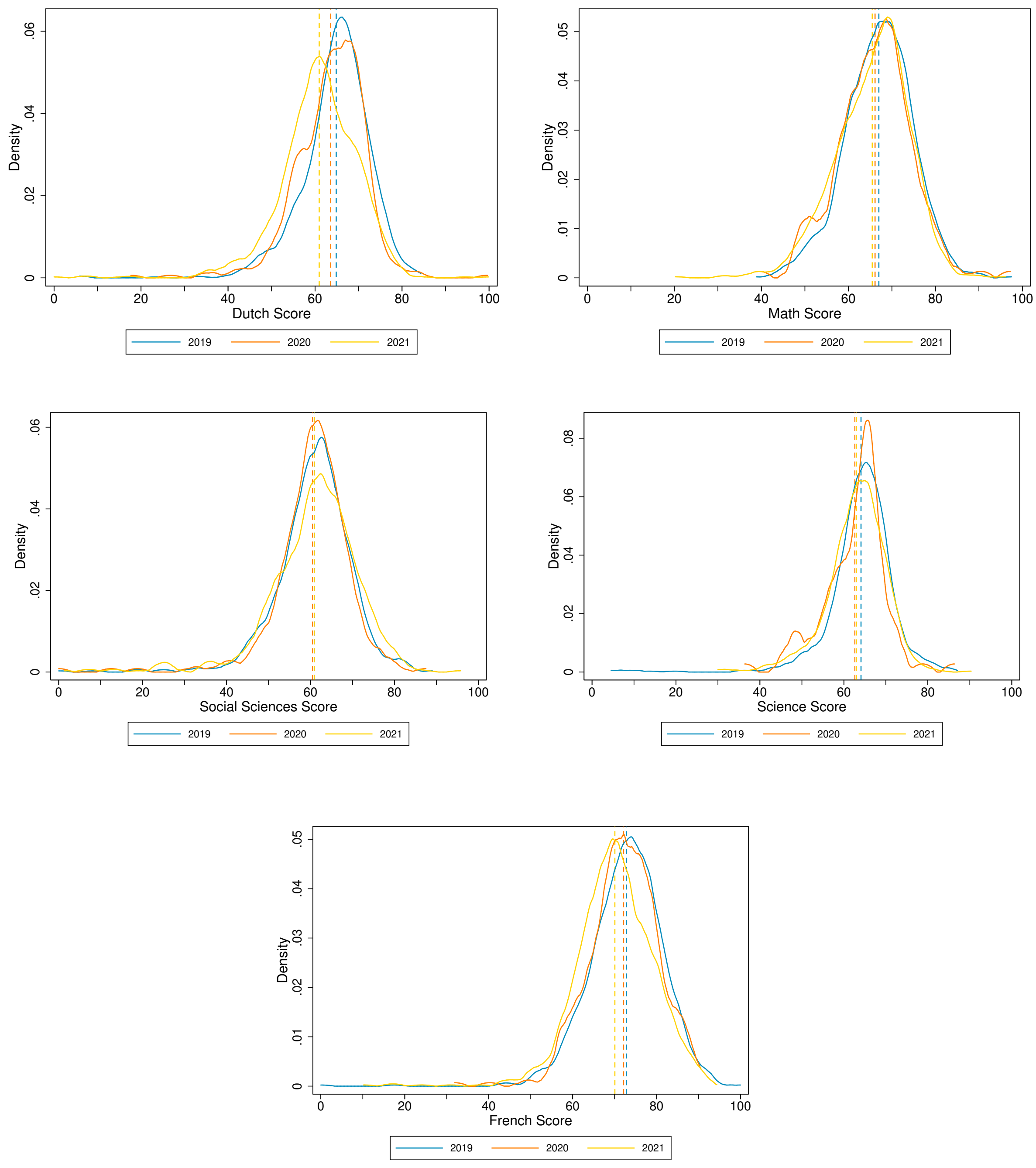


Figure B1: Distribution of scores 2019-2021
Note: Comparison of the density plots of the scores across the years 2019-2021 for each subject (same test). While the density plot is similar across the three years, the distributions of scores in 2020 compared to 2019, and in 2021 compared to 2020, are slightly more skewed to the left in most subjects, showing a gradually smaller mean score among participating schools.

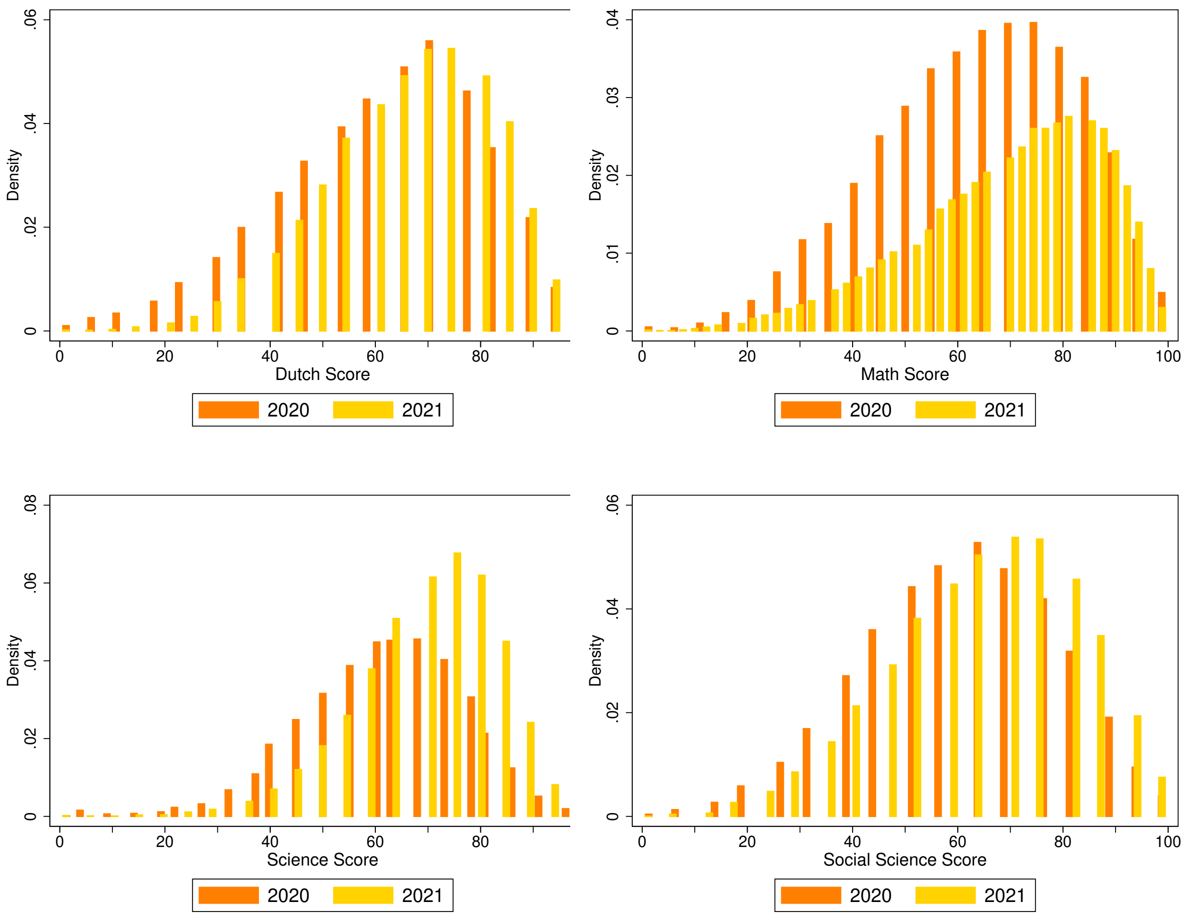


Figure B2: Distribution of scores before linear equating
Note: Comparison of the distribution of the scores for test version 2020 and the new test version 2021, both taken at the end of the academic year 2020/2021. Test version 2020 was administered to 64% of participating schools in order to allow for subsequent equating of scores from test version 2021. While both distribution are negatively skewed, a lower mean score among schools administering the test version 2020 can be observed.

Student level data

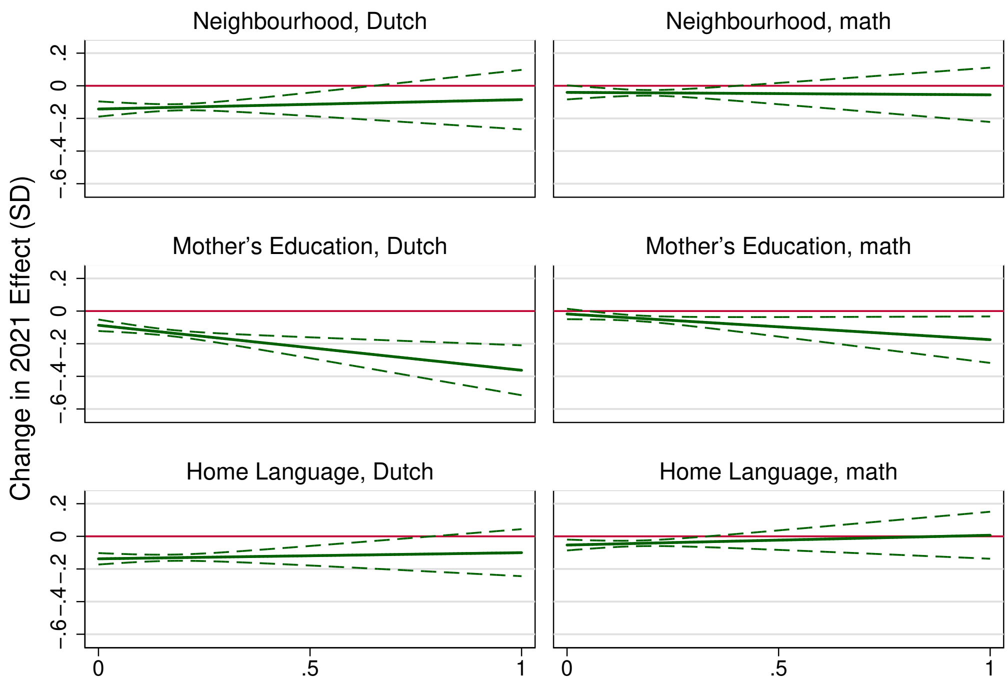


Figure B3: Marginal effects by socioeconomic composition of schools' population
Note: Marginal effects of 'change-in-2021' dummy, as estimated by a school fixed effect regression (solid lines) in regression model (1). 95% confidence intervals around the estimates (dashed lines). Estimates for the 2017-2021 sample at pupil level.

Mother's Education

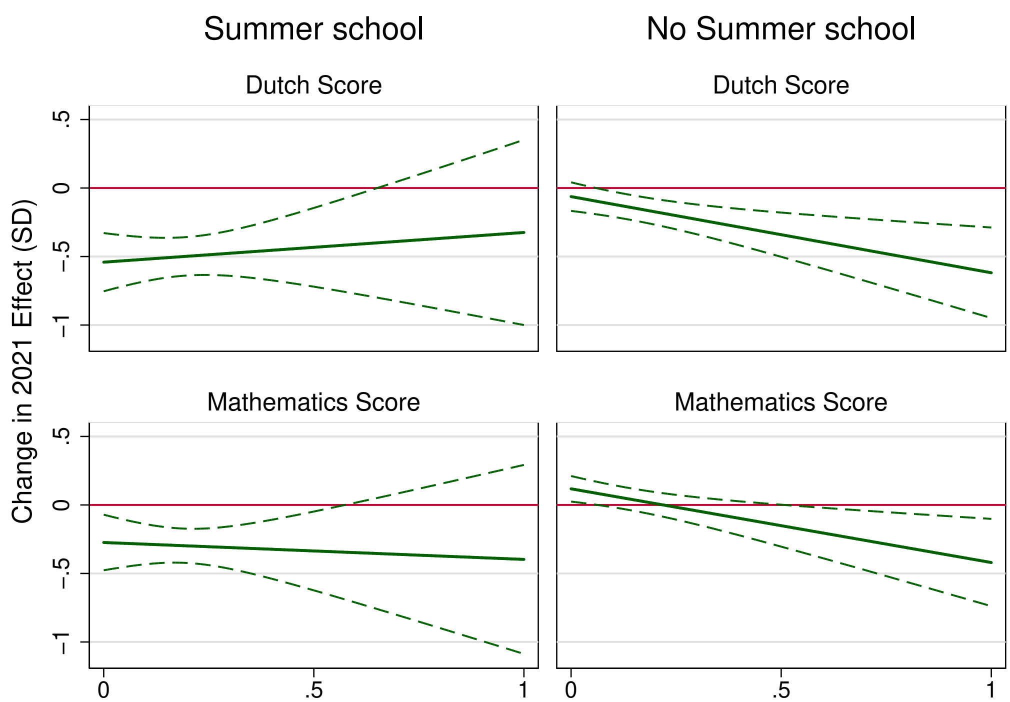


Figure B4: Marginal effects by the share of low-educated mothers in the school population
Note: Marginal effects of 'change-in-2021' dummy ran separately for schools in postcode areas with and without a summer school. Estimated by a school fixed effect regression (solid lines) in regression model (1). 95% confidence intervals around the estimates (dashed lines). Estimates for the 2017-2021 sample at school level.