Durham University

The differential impact of Covid-19 related school closures on English primary school pupils' writing performance

EDO/FR-000023248 Final Report

Jens F. Beckmann, Jochen Einbeck & Alexandra Hunt

November 2024

Executive Summary

Objective:

The present study examined the impact of school closures due to the Coronavirus Disease 2019 (COVID-19) pandemic on the acquisition of writing skills among pupils in English primary schools. A particular focus was placed on the potential differential effects of selected factors that are discussed as moderating responses to disruptions in the provision of education. These include the ‘Pupil Premium’ status, which can be seen as a person-level indicator of economic disadvantage, and ‘Geographic Region’, which is a more indirect, environment-level indicator of economic disadvantage. Other factors considered in the analyses were ‘Sex’ and ‘School Type’. Subsequent analyses were conducted to investigate the impact of the interruption of in-person teaching on the relative importance of pupil characteristics and school (i.e. environmental) characteristics contributing to progress in the acquisition of writing skills during primary school.

Approach:

The study utilised a large-scale, longitudinal data set, which was extracted from a database provided by No-More-Marking[1]. The analyses were based on repeatedly assessed writing scores from N = 189,534 primary school pupils across five age cohorts (year 1 to year 5). For each pupil two assessments of their writing performance were available. The interval between these assessments varied between 8 and 15 months, the second of the two assessments took place in the school year succeeding the first.

The total sample consisted of cohorts of pupils that were either affected or unaffected by COVID-19 related 3-month school closures in 2020 between their first and second assessments of writing performance. This setting made it possible to adopt a quasi-experimental design, which allowed to establish some basic causal inferences regarding the effects of school closures on progress in the development of writing skills at pupil level. The availability of repeated assessments of writing performance over time on an individual pupil level permits a more appropriate modelling of potential effects on learning and development, as it allows for the examination of intra-individual change processes rather than relying on cross-sectional between-group comparisons.

The writing performance scores for each individual pupil were derived from comparative judgments, an innovative approach to assessing a complex concept such as writing quality reliably, validly, and efficiently. Comparative judgment is a performance scoring method based on an iterative process of pairwise comparisons of scripts. Two one-page scripts are presented parallel on a computer screen and evaluators are asked to judge which of the two is the better one. Each script is iteratively paired with other scripts and judged accordingly. As a result, each script accumulates a record of ranking scores that reflect its relative position to other scripts in terms of perceived quality. Comparability of performance scores across age groups (i.e., cross-sectionally) and over time (i.e., longitudinally) is achieved through anchored scaling procedures. Both forms of comparisons were realised in the analyses presented.

The three main research questions addressed in this report refer to (1) the extent to which loss in in-person teaching resulted in learning loss, (2) the extent to which various learner and school characteristics moderate the response to COVID-related school closures, and (3) the extent to which the relative importance of being part of a class (in terms of teaching group) or a particular school shifts as a result of school closures.

[1] No-More-Marking is a company (see https://www.nomoremarking.com/) that offers solutions for schools to assess writing performance on the basis of comparative judgement.

The generalisability of the findings in relation to answering the research questions is determined by two major factors: firstly, the representativeness of the sample for the population of primary school pupils in England; and secondly, the comparability of the two sub-cohorts (affected vs. non-affected by COVID-19-related school closures) per Age Cohort.
Descriptive analyses of the data indicated that overall, the sample can claim representativeness in terms of sex ratio and the proportion of pupils in receipt of pupil premium. In terms of the adopted dichotomous categorisation of school type into independent and state-funded schools, independent schools tend to be proportionally underrepresented in the sample. The distribution of schools across the various geographic regions within England also reflects positively on the representativeness of the sample used in this research.

Results:

RQ1: *To what extent does the loss in in-person teaching opportunities caused by COVID-19 related school closures translate into learning loss?*

The effects of three months interruption in the provision of in-person teaching on primary school pupils' progress in acquiring writing skills appear to be rather small. In more concrete and tangible terms, across all age groups pupils affected by a 3-month loss of in-person teaching due to COVID-19-related school closures make on average 2.3 points less progress than their non-affected peers. For contextualisation, this average 2.3 points difference is observed on a scale with a plausible score range from 200 and 800 points and a mean score of 485.7 and a standard deviation of 66.2. The size and direction of the difference between affected and non-affected sub-cohorts varies across Age Cohorts from 8.8 points *in favour of the non-affected* sub-cohort to 1.4 points *in favour of the affected* sub-cohort.

RQ2: *To what extent do certain learner and school characteristics have an influence on the extent to which loss in in-person teaching opportunities translates into learning loss?*

The potential moderating influence of four selected covariates on the effect of school closures related to COVID-19 on the development of writing performance was analysed. These include 'Pupil Premium' status, pupil's 'Sex', the 'School Type' pupils attend, and the 'Geographic Region' in which their school is situated.
When including 'Pupil Premium' in the analyses, we first confirmed the existence of a "pupil premium gap" of about 17.7 points on average across the five Age Cohorts. The attainment gap in terms of writing skills between pupils in receipt of the pupil premium and those who do not tends to decrease with age to its lowest level of about 13.4 points at year 5. The findings regarding potential differential effects of school closures suggest that pupils in primary school who are in receipt of the pupil premium make, on average, less than one point less progress during the 3-month period than their peers who are not in receipt of pupil premium. Results further indicate that pupils in the two youngest age cohorts (i.e. Y1Y2 and Y2Y3) who are in receipt of the pupil premium demonstrate less progress of approximately 1.5 and 3.3 points, respectively. However, the status of being a pupil premium recipient in older year groups tends to be negligible to inconsequential in terms of contributing to the observed general effect of the 3-month school closures. In other words, learners who receive the pupil premium tend, on average, to demonstrate the same level of relative learning loss as their non-premium peers, thereby maintaining the pre-existing attainment gap.

Analyses considering pupils' sex revealed an average attainment gap of about 21.4 points in favour of female pupils. The sex differences in terms of writing skills were relatively stable across Age Cohorts during primary school. When affected by a 3-month interruption of in-person teaching, female pupils also tend to show on average a 0.6 points smaller effect on their progress in the development in writing skills. In other words, the pre-existing sex related attainment gap remains largely unaffected.

Across the five age cohorts, the average performance of pupils attending independent schools was found to be 32.9 points better than that of their peers attending state-funded schools. This attainment gap between the two categories of pupils was observed to have only a weak tendency to reduce with age. With regard to the differential effects on progress, independent schools appear to successfully compensate, or even in early Age Cohorts, over-compensate for otherwise observed reductions in progress under school closure conditions. Pupils attending independent schools when affected by school closures make 1.9 points more progress than their non-affected peers. This is with the exception of Age Cohort Y4Y5, where a reduction in progress of 5.9 points was found. In contrast, pupils attending state-funded schools tend to make on average 2.9 points less progress during the three months of school closure. In other words, the school-type related attainment gap seems to have widened somewhat as a result of COVID-19 related school closures.

Analyses that differentiate between geographic regions of England revealed high levels of consistency in *intra-regional differences* across age groups. That is, within each of the 10 regions we observed a systematic increase in average performance with age. The pattern of *inter-regional differences* across Age Cohorts, however, was less consistent. That is, with little exceptions, performance rankings of the 10 regions varied across Age Cohorts.
The average range of general performance differences across regions covered 23 points. In comparison to this "region-related" attainment gap, the average range of differences in progress that can be uniquely attributed to 'Geographic Region' was 6.1 points for the 3-months of school closure across Age Cohorts. Although the relative size of the differential effect of 'Geographic Region' on the impact of school closures is small in comparison, it is strongly aligned with the average performance shown in general by pupils attending schools in a geographic region. In other words, regions with higher average performance tend to also show smaller negative effects of the pandemic on education. From an aggregative, i.e. regional, perspective, this suggests that high performance may constitute some form of resilience towards the potential negative effects of school closures related to the pandemic. This may result in a tendency for regional attainment gaps to widen. A similar phenomenon was observed for the inclusion of 'Pupil Premium', 'Sex', and 'School Type' in the respective moderator analyses.

RQ3: *To what extent do COVID-19-related school closures change the relative importance of being in a particular teaching group/class or school.*

The results of the analyses in relation to research question 3 corroborated the reasonable expectation that with a loss of in-person teaching, the relative importance of the school environment diminishes. While the relative importance of class or teaching group remains largely comparable between un-affected and affected sub-cohorts, the reduction of school-related influence was counter-balanced by an increase in the relative importance of individual pupil characteristics. This result may be interpreted as a reminder of the importance of helping pupils across all age groups to develop resilience as a pupil attribute.

Implications:

From a perspective that primarily focusses on the size of the numerical differences in the effects of COVID-19 related school closures on the development of writing performance during primary school in England, the results obtained from the various analyses could be interpreted of having limited practical relevance. In other words, the loss of in-person teaching during the 3 months of school closures in 2020 seem not to have resulted in substantial learning losses in terms of the development of writing skills in primary school pupils. However, the systematicity and consistency of the result pattern across different age groups, and the inclusion of different covariates, in conjunction with the substantial size of the near-representative sample, warrants the findings to be taken seriously. Further research should therefore investigate the potential impact of prolonged or repeated periods of school closures on the development of writing skills, including the relative magnitude of effects and the possibility of a further widening of pre-existing attainment gaps. Additionally, it would be valuable to examine differential trajectories of any form of recovery.

Notwithstanding this, it is crucial to note that while the analyses in this report offer a comprehensive and systematic *description* of the observed outcomes, they do not necessarily provide *explanations* of the causal mechanisms that underpin the emergence of the observed effects. Yet, recommendations or policies for targeted interventions require to be based on sound evidence for *explanatory* causality, which cannot be substituted by *descriptive* causality. Nonetheless, the results presented in this report are important and useful for informing assumptions about potential explanatory causal mechanisms. These assumptions, however, need to be specifically tested, which requires methodological approaches different to the ones realised in this report.

Acknowledgements:

We would like to thank No-More-Marking (visit www.nomoremarking.com) for providing the data used in the analyses which led to this report, and the Nuffield Foundation (www.nuffieldfoundation.org) for providing the funding to carry out this work. We would also like to thank Karen Jones for leading the funding acquisition, as well as Dr Bilal Ashraf for carrying out the initial data cleaning and preparation for analysis.

The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare and Justice. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics, the Ada Lovelace Institute and the Nuffield Family Justice Observatory. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation.

1 Introduction

The Covid-19 pandemic caused an unprecedented impact on education. The disruption to the education system has affected 1.5 billion students across the world (UNESCO, 2023). As national lockdowns were imposed across many countries, in-person teaching was halted and replaced with online learning. Specifically in England, the timeline of events with potential impact on education was as follows:

- Academic year 2018/19: the last uninterrupted academic year prior to the pandemic.
- Academic year 2019/20: the first disruption whereby a national lockdown was imposed in March 2020. Schools remained open to "vulnerable" students and children of "keyworkers" only. Schools then reopened in June 2020 to certain year groups. These included nursery, reception, Year 1, and Year 6 pupils at the primary school level and Year 10 and Year 12 at the secondary school level.
- Academic year 2020/21: Schools reopened to all pupils in September 2020 and remained open until the third and final national lockdown imposed in January 2021. School closures (that exempted "vulnerable" pupils and children of "keyworkers") related to that national lockdown lasted until March 2021.

The first interruption amounts to around 80 days (appr. 3 months) of school closure, the second interruption represents about 45 days of loss in in-person teaching. In the analyses presented here we focus on the first phase of COVID-19-related school closures only.

The lack of in-person teaching is believed to have had a negative impact on students' learning in general. Unsurprisingly, the COVID-19 pandemic has prompted a plethora of analyses, including systematic reviews and meta-analyses (e.g., Betthäuser et al., 2023), aimed at identifying and quantifying the "learning loss" (Engzell et al., 2021) – a lower level of progress than anticipated – experienced by affected learners.

A prominent sub-topic shared across research studying the impact of COVID-19 in education refers to the differential, moderating effects of socio-economic background of pupils. Pupils regarded as disadvantaged were expected to be stronger impacted by COVID-19-related school closures than so-called non-disadvantaged pupils (Rose et al., 2021, through NfER and EEF; Department for Education, DfE, 2021a, 2021b). Disadvantage as a covariate of interest is operationalised in different forms and with different foci. For instance, eligibility to free school meals is often used as a proxy for socio-economic status (e.g., in England), in other study contexts parents' educational level is used as a differentiator in terms of disadvantage (e.g., Engzell et al., 2021). An arguably more indirect approach to operationalising socio-economic disadvantage pertains to geographic region. Another covariate expected to moderate school closure effects is sex, where boys are expected to be more affected than girls.

Popular attainment domains studied in COVID-19 related research include reading and maths. The current study seeks to obtain insights regarding the differential impact of COVID-19 related disruptions to the educational provision on the development of primary school pupils' writing skills. An effective acquisition of sufficient levels of writing skills is an important educational outcome (Graham et al., 2013). With its functional links to reading skills writing skills play an important role in the analysis and interpretation of information, and subsequently in acquiring knowledge. In other words, writing skills are instrumental as an enabler to future learning, be it in the context of schooling or more broadly.

While COVID-19-related interruptions are considered to have had a general negative impact on pupils' learning progress, it is expected that the impact of these interruptions will differ in severity across pupils with different personal and contextual characteristics. It is these differential effects that are in the main focus of this project. The availability of a large-scale longitudinal data set – provided by No-More-Marking (NMM) – offers an interesting opportunity to not only analyse the impact of COVID-19 related interruptions of school-based education in general, but it also allows to learn more about their differential effects on the acquisition of writing skills in particular.

The elucidation of differential effects will further our knowledge and therefore strengthen the evidence base that should underpin decision making processes, be it at the level of the classroom, the school, across local authorities, or education policy in general.

The research conducted in this project is unique in at least four ways:

(1) The analyses presented here utilise an extensive, yet under-utilised database containing information about a large, near representative sample of primary school pupils, which is conducive to valid generalisations of our findings.

(2) The utilised data represent repeated assessments of writing performance over time at the individual pupil level. This allows for a more appropriate modelling of potential effects as intra-individual change processes rather than relying on cross-sectional between group comparisons.

(3) The database includes cohorts that either were or were not affected by COVID-related interruptions of their education. This allows for quasi-experimental comparisons, which in turn creates a more solid foundation for causal inferences.

(4) The writing performance scores for each individual pupil are derived from comparative judgments, an innovative approach to assessing a complex concept such as writing quality reliably, validly, and efficiently (Curcin et al., 2019; McGrane, 2023; Verharvert et al., 2019; Wheadon et al., 2019).

The research objectives and subsequent research questions are outlined in section 2. Section 3 introduces details of the dataset that will be used to address the research questions and provides information regarding some core descriptive statistics.

In section 4 the modelling is outlined that underpins the analyses of the data to address the research questions. Section 5 presents the results; section 6 provides interpretation and contextualisation of the results and offers recommendations.

2 Research Objectives and Research Questions

In the education literature a range of person and context variables are discussed to be associated with educational outcomes. In addition to various proxy markers for socio-economic differences, such as pupil premium, eligibility for free school meals, or geographic regions, other variables, including sex, school type, and class size are considered contributing factors to learning success.

The objective of this research is to elucidate the effects of large-scale interruptions on academic outcomes, with a particular focus on writing performance, in the context of the COVID-19 pandemic. To that end we analyse a comprehensive set of secondary data that includes repeated measures of writing performance for cohorts of primary school students (Y1 to Y6) who have or have not been affected by COVID-19-related school closures.

Three main research questions will be addressed. The first question focusses on the general effect of COVID-19 related school closures on the development of writing skills of pupils in Year 1 to Year 6 in primary schools. In other words,

RQ1: To what extent does the loss in in-person teaching opportunities translate into learning loss?

The second research question refers to potential moderating effects of selected pupil and contextual variables on the impact of COVID-19 related school closures on the development of writing performance. In other words,

RQ2: *To what extent do certain learner and school characteristics have an influence on the extent to which loss in in-person teaching opportunities translates into learning loss?*

In relation to the second research question the following learner and school characteristics are considered: age, socio-economic disadvantage, sex, school type, and geographic region.

Writing performance, like any form of performance, is a multi-determined phenomenon. Changes in writing performance, whether as a result of learning and development or as a result of COVID-19-related school closures, may be associated with variations at pupil, class or school level. Hence, the third research question aims to gain insight into potential shifts in the variance composition of writing scores as a result of COVID-19-related school closures. In other words,

RQ3: To what extent do COVID-19-related school closures change the relative importance of being in a particular teaching group/class or school?

3 Methodology

3.1 The Data

To address the research questions, we access a large secondary dataset provided by No-More-Marking (NMM). This dataset of N = 720,283 observations across 105 variables which in addition to pupils' writing scores also includes various demographic information related to the individual pupils themselves and information regarding the schools they attend. For the purpose of this research a subset of variables has been formed, to only consider the variables relevant to the research questions to be addressed. These variables include:

- Pupil ID
- Class ID
- School ID
- School start year
- Sex
- Pupil premium status based on eligibility for free school meals
- Writing scores and corresponding year group for academic years 2018/19, 2019/20, and 2020/21
- Geographical Region of school
- School census information from the academic year 2018/19 and the academic year 2020/21, these include:
    - Type of school, e.g., state, independent
    - Ofsted rating of school

Pupils' writing scores have been operationalised through a comparative judgement approach employed during the academic year 2018/19 up to and including the academic year 2021/22.

In order to identify potential effects of COVID-19 related school closures at the pupil level, the analyses will be based on within-person contrasts using repeated measurement of the outcome variable, i.e. writing performance. This required data merging so that individual pupils' writing scores at time 1 (T1) are matched with their respective writing score at time 2 (T2). Pupils' writing score at T2 was obtained in the academic year following the year in which writing score T1 was recorded. For instance, the T1 writing score for a Year 2 pupil obtained in the academic year 2018/19 had to be matched with their writing score now produced in their Year 3 during the academic year 2019/20 (T2). The data in the database can be sub-divided into so-called non-affected cohorts and affected cohorts. So-called non-affected cohorts comprise pupils whose T1 writing score was obtained in the academic year 2018/19 and their subsequent T2 writing score was obtained in the academic year 2019/20 prior to the first COVID-related school closure. So-called affected cohorts comprise pupils who contributed a writing score in the academic year 2019/20 (T1) prior to COVID-19-related school closures, while their subsequent writing score (T2) was obtained in the academic year 2020/21, i.e. after a period of COVID-19-related school closures (Figure 3.1).

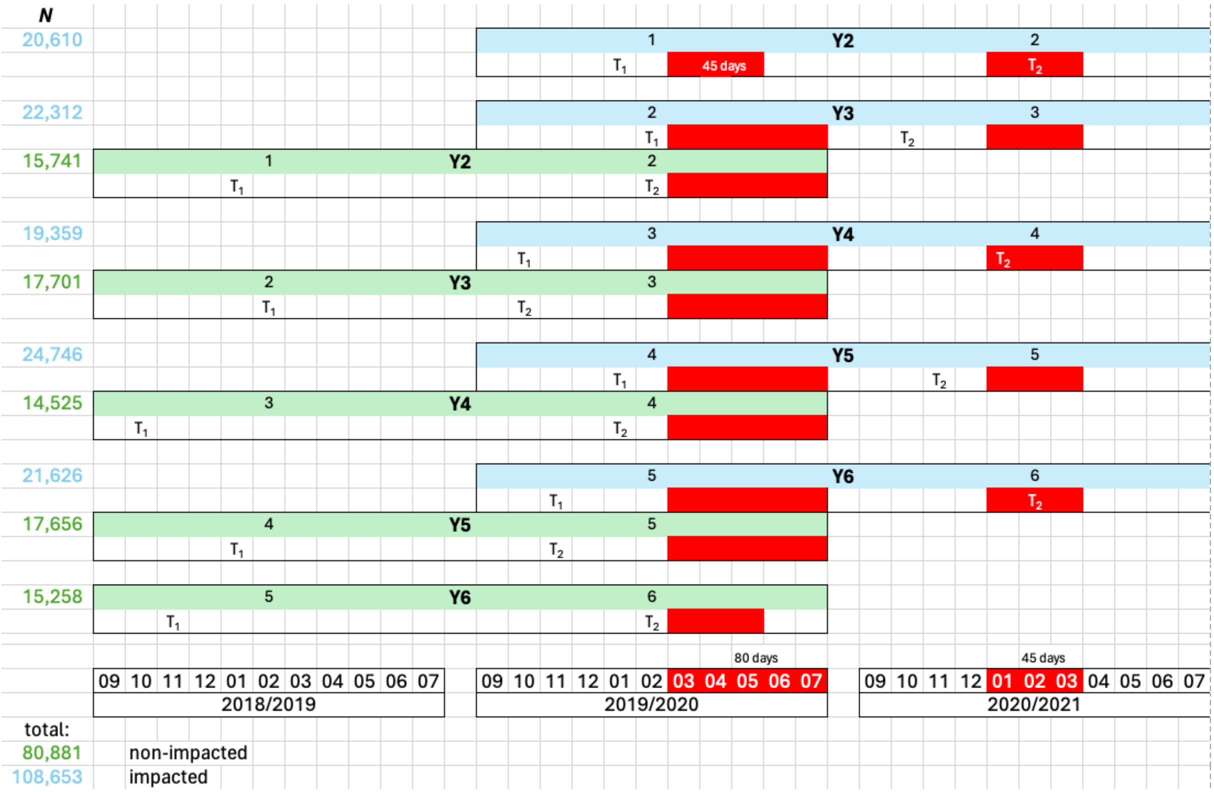

Figure 3.1: Data structure. Size of the different Age Cohorts across the two conditions (non-affected vs affected) and over time (Assessments indicated by T1 and T2, respectively).

The availability of two assessments per pupil provides a unique opportunity to study effects of temporary school closures at a within-person level (in contrast to the prevalent between group, cross-sectional comparisons). The process of compiling datasets with T1 and T2 scores at pupil level resulted in a subset of N = 189,534 datasets across five Age Cohorts which can be subdivided into non-affected sub-cohorts and affected sub-cohorts (Table 3.1).

Table 3.1: Numbers of pupils across Age Cohorts and conditions.

Age Cohorts	non-affected	affected
Y1-Y2	15,741	20,610
Y2-Y3	17,701	22,312
Y3-Y4	14,525	19,359
Y4-Y5	17,656	24,746
Y5-Y6	15,258	21,626
TOTAL	**80,881**	**108,653**

Our sample with complete data for repeated assessments of pupil-specific writing scores represents close to 2 per cent of the entire population of primary school children in England during the period between the academic years 2018/2019 and 2020/2021.

In order to draw meaningful and generalisable conclusions about the potential effects of school closures due to COVID-19, two main prerequisites need to be checked. The first relates to the representativeness of the sample of available records in relation to the population of primary school children in England. The second relates to the comparability of the samples across the two conditions (non-affected vs. affected). To assess the quality of the sample in terms of generalisability and comparability, we analyse the distributions of selected pupil and school variables.
The next set of Tables shows the sex ratio across Age Cohorts and differentiated between affected and non-affected sub-cohorts (Table 3.2 for total, Table 3.3 for Age Cohort-specific).

Table 3.2: Numbers of pupils according to their sex across conditions.

	non-affected	affected
Male	40,469 (50.0%)	54,454 (50.1%)
Female	40,411 (50.0%)	54,177 (49.9%)
Missing	1 (0.0%)	22 (0.0%)

Table 3.3: Age-Cohort specific pupil numbers according to their sex across conditions.

	non-affected	affected
YEAR 1 – YEAR 2		
Male	7,933 (50.4%)	10,275 (49.9%)
Female	7,808 (49.6%)	10,335 (50.1%)
Missing	0 (0.0%)	0 (0.0%)
YEAR 2 – YEAR 3		
Male	8,949 (50.6%)	11,210 (50.2%)
Female	8,752 (49.4%)	11,102 (49.8%)
Missing	0 (0.0%)	0 (0.0%)
YEAR 3 – YEAR 4		
Male	7,112 (49.0%)	9,745 (50.3%)
Female	7,412 (51.0%)	9,614 (49.7%)
Missing	1 (0.0%)	0 (0.0%)

Table 3.3 continued from previous page

	non-affected	affected
YEAR 4 – YEAR 5		
Male	8,845 (50.1%)	12,330 (49.8%)
Female	8,811 (49.9%)	12,416 (50.2%)
Missing	0 (0.0%)	0 (0.0%)
YEAR 5 – YEAR 6		
Male	7,630 (50.0%)	10,894 (50.4%)
Female	7,628 (50.0%)	10,710 (49.5%)
Missing	0 (0.0%)	22 (0.1%)

The sex ratio in the sample studied in this research pretty much mirrors what is to be found in the population of primary school pupils in England (i.e., 49F: 51M). This is also the case for each of the Age Cohorts.

The next tables show how Pupil Premium Status – as proxy indicator for socio-economic disadvantage – is distributed across cohorts (Table 3.4 for total, Table 3.5 separate for each Age Cohort).

Table 3.4: Numbers of pupils with or without Pupil Premium Status differentiated by condition.

status	non-affected	affected
Pupil premium	17,984 (22.2%)	23,085 (21.2%)
Non-pupil premium	62,897 (77.8%)	85,568 (78.8%)
Missing	0 (0.0%)	0 (0.0%)

Table 3.5: Numbers of pupils with or without Pupil Premium Status across Age Cohorts and experienced conditions.

	non-affected	affected
YEAR 1 – YEAR 2		
Pupil premium	2,654 (16.9%)	3,270 (15.9%)
Non-pupil premium	13,087 (83.1%)	17,340 (84.1%)
Missing	0 (0.0%)	0 (0.0%)
YEAR 2 – YEAR 3		
Pupil premium	3,826 (21.6%)	4,241 (19.0%)
Non-pupil premium	13,875 (78.4%)	18,071 (81.0%)
Missing	0 (0.0%)	0 (0.0%)
YEAR 3 – YEAR 4		
Pupil premium	3,548 (24.4%)	4,658 (24.1%)
Non-pupil premium	10,977 (75.6%)	14,701 (75.9%)
Missing	0 (0.0%)	0 (0.0%)
YEAR 4 – YEAR 5		
Pupil premium	4,002 (22.7%)	6,071 (24.5%)
Non-pupil premium	13,654 (77.3%)	18,675 (75.5%)
Missing	0 (0.0%)	0 (0.0%)

Table 3.5 continued from previous page

	non-affected	affected
YEAR 5 – YEAR 6		
Pupil premium	3,954 (25.9%)	4,845 (22.4%)
Non-pupil premium	11,304 (74.1%)	16,781 (77.6%)
Missing	0 (0.0%)	0 (0.0%)

The percentages of pupils with Pupil Premium Status mirror those to be found in the population of English primary school children for the academic years between 2018/19 and 2020/2021 across the five Age Cohorts (i.e., 17.0% in 2018/19, 19.3% in 2019/20, and 23.3% in 2020/21 of all primary school pupils in England were eligible to receive Free School Meals). This and the fact that the differences between cohorts affected or not affected by COVID-19 related school closures across the Age Cohorts are negligible, benefits meaningful, i.e. generalisable comparisons.

Another covariate considered in the research questions is school type. Tables 3.6 and 3.7 indicate how pupil numbers are distributed across different types of schools. The database provided two separate sets of information regarding to the type of school pupils attended. One set offered more specific information, with 12 different school type categories, plus a category for missing information. The other set distinguished only between state-funded and independent schools, in addition to the category for missing information (see Tables 3.8 and 3.9). We report the descriptive statistics for both sets. Reporting the first set of information serves primarily the purpose of providing a as much detailed description of the sample as possible, the other set of information represents the dichotomous categorisation used in later analyses in relation to research question 2.

Table 3.6: Numbers of pupils across different school types and conditions.

School Type	non-affected	affected
Academy converter	27,651 (34.2%)	37,521 (34.5%)
Academy special converter	0 (0.0%)	0 (0.0%)
Academy sponsor led	11,194 (13.8%)	16,292 (15.0%)
Community	21,152 (26.2%)	25,011 (23.0%)
Foundation	1,633 (2.0%)	3,833 (3.5%)
Free	1,364 (1.7%)	1,694 (1.6%)
Free special	0 (0.0%)	0 (0.0%)
Other Independent	812 (1.0%)	1,586 (1.5%)
Other independent special	0 (0.0%)	0 (0.0%)
Service children’s education	0 (0.0%)	0 (0.0%)
Voluntary aided	5,165 (6.4%)	6,739 (6.2%)
Voluntary controlled	1,763 (2.2%)	2,365 (2.2%)
Missing	10,147 (12.5%)	13,612 (12.5%)

Table 3.7: Numbers of pupils across different school types, cohorts, and conditions.

School type per cohort	non-affected	affected
YEAR 1 – YEAR 2		
Academy converter	5,242 (33.3%)	6,891 (33.4%)
Academy special converter	0 (0.0%)	0 (0.0%)
Academy sponsor led	2,163 (13.7%)	2,869 (13.9%)
Community	4,033 (25.6%)	5,131 (24.9%)
Foundation	336 (2.1%)	824 (4.0%)
Free	347 (2.2%)	429 (2.1%)
Free special	0 (0.0%)	0 (0.0%)
Other Independent	133 (0.8%)	185 (0.9%)
Other independent special	0 (0.0%)	0 (0.0%)
Service children's education	0 (0.0%)	0 (0.0%)
Voluntary aided	994 (6.3%)	1,264 (6.1%)
Voluntary controlled	388 (2.5%)	439 (2.1%)
Missing	2,105 (13.4%)	2,578 (12.5%)
YEAR 2 – YEAR 3		
Academy converter	5,851 (33.1%)	7,481 (33.5%)
Academy special converter	0 (0.0%)	0 (0.0%)
Academy sponsor led	2,386 (13.5%)	3,359 (15.1%)
Community	4,731 (26.7%)	5,010 (22.5%)
Foundation	352 (2.0%)	819 (3.7%)
Free	328 (1.9%)	422 (1.9%)
Free special	0 (0.0%)	0 (0.0%)
Other Independent	161 (0.9%)	254 (1.1%)
Other independent special	0 (0.0%)	0 (0.0%)
Service children's education	0 (0.0%)	0 (0.0%)
Voluntary aided	1,198 (6.8%)	1,512 (6.8%)
Voluntary controlled	384 (2.2%)	513 (2.3%)
Missing	23,10 (13.1%)	2,942 (13.2%)
YEAR 3 – YEAR 4		
Academy converter	5,212 (35.9%)	6,819 (35.2%)
Academy special converter	0 (0.0%)	0 (0.0%)
Academy sponsor led	2,213 (15.2%)	2,958 (15.3%)
Community	3,578 (24.6%)	4,461 (23.0%)
Foundation	235 (1.6%)	566 (2.9%)
Free	197 (1.4%)	322 (1.7%)
Free special	0 (0.0%)	0 (0.0%)
Other Independent	154 (1.1%)	291 (1.5%)
Other independent special	0 (0.0%)	0 (0.0%)
Service children's education	0 (0.0%)	0 (0.0%)
Voluntary aided	939 (6.5%)	1,184 (6.1%)
Voluntary controlled	270 (1.9%)	397 (2.1%)
Missing	1,727 (11.9%)	2,361 (12.2%)

Table 3.7 continued from previous page

School type per cohort	non-affected	affected
YEAR 4 – YEAR 5		
Academy converter	5,901 (33.4%)	8,480 (34.3%)
Academy special converter	0 (0.0%)	0 (0.0%)
Academy sponsor led	2,342 (13.3%)	3,911 (15.8%)
Community	4,819 (27.3%)	5,591 (22.6%)
Foundation	374 (2.1%)	869 (3.5%)
Free	257 (1.5%)	257 (1.0%)
Free special	0 (0.0%)	0 (0.0%)
Other Independent	211 (1.2%)	429 (1.7%)
Other independent special	0 (0.0%)	0 (0.0%)
Service children's education	0 (0.0%)	0 (0.0%)
Voluntary aided	1,074 (6.1%)	1,468 (5.9%)
Voluntary controlled	434 (2.5%)	642 (2.6%)
Missing	2,244 (12.7%)	3,099 (12.5%)
YEAR 5 – YEAR 6		
Academy converter	5,445 (35.7%)	7,850 (36.3%)
Academy special converter	0 (0.0%)	0 (0.0%)
Academy sponsor led	2,090 (13.7%)	3,195 (14.8%)
Community	3,991 (26.2%)	4,818 (22.3%)
Foundation	336 (2.2%)	755 (3.5%)
Free	235 (1.5%)	264 (1.2%)
Free special	0 (0.0%)	0 (0.0%)
Other Independent	153 (1.0%)	427 (2.0%)
Other independent special	0 (0.0%)	0 (0.0%)
Service children's education	0 (0.0%)	0 (0.0%)
Voluntary aided	960 (6.3%)	1,311 (6.1%)
Voluntary controlled	287 (1.9%)	374 (1.7%)
Missing	1,761 (11.5%)	2,632 (12.2%)

Table 3.7 indicates that – except for so-called community schools, for which there tends to be a slightly lower proportion of pupils in the affected cohorts – there seem to be no noticeable differences in proportions of pupils across different school types across Age Cohorts and conditions. This is another prerequisite met for meaningful comparisons between conditions and Age Cohorts.
Table 3.8 provides information about the distribution of pupils across two main types of schools (state funded vs independently funded) for both conditions. Table 3.9 lists this information differentiated by Age Cohort.

Table 3.8: Student numbers across different school types and conditions.

School Type	non-affected	affected
State	75,773 (98.9%)	97,534 (98.6.%)
Independent	812 (1.1%)	1,418 (1.4%)
Missing	4,296	9,701

Table 3.9: Student numbers across different school types, Age Cohorts, and conditions.

School type per cohort	non-affected	affected
YEAR 1 – YEAR 2		
State	14,743 (93.7%)	18,652 (90.5%)
Independent	133 (0.8%)	185 (0.9%)
Missing	865 (5.5%)	1,773 (8.6%)
YEAR 2 – YEAR 3		
State	16,399 (92.6%)	20,193 (90.5%)
Independent	161 (0.9%)	242 (1.1%)
Missing	1,141 (6.4%)	1,877 (8.4%)
YEAR 3 – YEAR 4		
State	13,760 (94.7%)	17,314 (89.4%)
Independent	154 (1.1%)	259 (1.3%)
Missing	611 (4.2%)	1,786 (9.2%)
YEAR 4 – YEAR 5		
State	16,422 (93.0%)	22,130 (89.4%)
Independent	211 (1.2%)	367 (1.5%)
Missing	1,023 (5.8%)	2,249 (9.1%)
YEAR 5 – YEAR 6		
State	14,449 (94.7%)	19,245 (89.0%)
Independent	153 (1.0%)	365 (1.7%)
Missing	656 (4.3%)	2,016 (9.3%)

The figures presented in tables 3.5 to 3.9 indicate that pupils attending independent schools are generally under-represented in our sample (about 5% in the population of primary school pupils in England attend independently funded schools). This, together with the small but consistent difference in the proportion of independent schools in the affected cohorts (i.e., slightly higher proportion), warrants some caution when it comes to generalisability of findings of how (dichotomously coded) school type moderates potential COVID effects on English primary school pupils' development of writing skills. The inspection of the descriptives indicates a substantially larger proportion of missing school type information for the affected sub-cohorts. By ways of "logical induction" it seems plausible that these missing pieces of information can be primarily attributed to state-funded schools.

Another co-variate considered in the analyses of potentially differential effects of COVID-19 related school closures was geographic region. Tables 3.10 and 3.11 list how pupil numbers in our sample are distributed across 10 geographic regions in England.

Table 3.10: Pupil numbers across 10 geographic regions differentiated by condition.

Region	non-affected	affected
East Midlands	14,174 (17.5%)	13,392 (12.3%)
East of England	11,414 (14.1%)	16,160 (14.9%)
Inner London	8,890 (11.0%)	10,198 (9.4%)
North East	2,007 (2.5%)	2,536 (2.3%)
North West	6,125 (7.6%)	9,846 (9.1%)
Outer London	3,099 (3.8%)	5,892 (5.4%)
South East	8,517 (10.5%)	14,865 (13.7%)
South West	8,668 (10.7%)	9,786 (9.0%)
West Midlands	9,294 (11.5%)	9,931 (9.1%)
Yorkshire and the Humber	4,526 (5.6%)	7,324 (6.7%)
Missing	4,167 (5.2%)	8,723 (8.0%)

Table 3.11: Pupil numbers across 10 geographic regions differentiated by condition and Age Cohort.

Region per cohort	non-affected	affected
YEAR 1 – YEAR 2		
East Midlands	2,593 (16.5%)	2,320 (11.3%)
East of England	2,126 (13.5%)	3,107 (15.1%)
Inner London	1,815 (11.5%)	2,095 (10.2%)
North East	459 (2.9%)	435 (2.1%)
North West	1,240 (7.9%)	1,963 (9.5%)
Outer London	684 (4.3%)	1,219 (5.9%)
South East	1,483 (9.4%)	2,530 (12.3%)
South West	1,901 (12.1%)	1,891 (9.2%)
West Midlands	1,746 (11.1%)	1,891 (9.2%)
Yorkshire and the Humber	847 (5.4%)	1,501 (7.3%)
Missing	847 (5.4%)	1,658 (8.0%)
YEAR 2 – YEAR 3		
East Midlands	2,710 (15.3%)	2,545 (11.4%)
East of England	2,325 (13.1%)	3,388 (15.2%)
Inner London	2,084 (11.8%)	2,033 (9.1%)
North East	445 (2.5%)	566 (2.5%)
North West	1,545 (8.7%)	2,224 (10.0%)
Outer London	705 (4.0%)	1,157 (5.2%)
South East	1,864 (10.5%)	2,790 (12.5%)
South West	1,893 (10.7%)	2,087 (9.4%)
West Midlands	1,993 (11.3%)	2,225 (10.0%)
Yorkshire and the Humber	1,028 (5.8%)	1,608 (7.2%)
Missing	1,109 (6.3%)	1,689 (7.6%)

Table 3.11 continued from previous page

Region per cohort	non-affected	affected
YEAR 3 – YEAR 4		
East Midlands	2,714 (18.7%)	2,460 (12.7%)
East of England	2,002 (13.8%)	3,173 (16.4%)
Inner London	1,473 (10.1%)	1,743 (9.0%)
North East	287 (2.0%)	514 (2.7%)
North West	1068 (7.4%)	1,568 (8.1%)
Outer London	485 (3.3%)	958 (4.9%)
South East	1,592 (11.0%)	2,843 (14.7%)
South West	1,568 (10.8%)	1,644 (8.5%)
West Midlands	1,887 (13.0%)	1,689 (8.7%)
Yorkshire and the Humber	859 (5.9%)	1,146 (5.9%)
Missing	590 (4.1%)	1,621 (8.4%)
YEAR 4 – YEAR 5		
East Midlands	3,486 (19.7%)	3,166 (12.8%)
East of England	2,510 (14.2%)	3,306 (13.4%)
Inner London	2,082 (11.8%)	2,327 (9.4%)
North East	430 (2.4%)	487 (2.0%)
North West	1,233 (7.0%)	2,194 (8.9%)
Outer London	584 (3.3%)	1,403 (5.7%)
South East	1,886 (10.7%)	3,576 (14.5%)
South West	1,756 (9.9%)	2,249 (9.1%)
West Midlands	1,775 (10.1%)	2,348 (9.5%)
Yorkshire and the Humber	923 (5.2%)	1,692 (6.8%)
Missing	991 (5.6%)	1,998 (8.1%)
YEAR 5 – YEAR 6		
East Midlands	2,671 (17.5%)	2,901 (13.4%)
East of England	2,451 (16.1%)	3,186 (14.7%)
Inner London	1,436 (9.4%)	2,000 (9.2%)
North East	386 (2.5%)	534 (2.5%)
North West	1,039 (6.8%)	1,897 (8.8%)
Outer London	641 (4.2%)	1,155 (5.3%)
South East	1,692 (11.1%)	3,126 (14.5%)
South West	1,550 (10.2%)	1,915 (8.9%)
West Midlands	1,893 (12.4%)	1,778 (8.2%)
Yorkshire and the Humber	869 (5.7%)	1,377 (6.4%)
Missing	630 (4.1%)	1,757 (8.1%)

The information presented in Tables 3.10 and 3.11 suggests that the proportional composition of pupil numbers across regions differs between the affected and non-affected cohorts. For instance, for the affected sub-cohorts there are consistently proportionally fewer pupils in the East Midlands region compared to the non-affected cohorts from that region. Conversely, the relative number of pupils from the South-East region tends to be higher for the affected sub-cohorts. The confounding of region (as supposed proxy for socio-economic strength) and condition (affected vs non-affected) creates challenges to the estimation and appropriate interpretation of potential COVID effects and their association with the geographical region.

Another context variable refers to school sex segregation, i.e. whether pupils are being taught in a single-sex or co-educational context. Tables 3.12. and 3.13 present information regarding pupil numbers across these categories.

Table 3.12: Pupil numbers across schools of different categories of sex segregation by condition.

School Type	non-affected	affected
Mixed	70,255 (86.9%)	94,267 (86.8%)
Single (boys)	0 (0.0%)	0 (0.0%)
Single (girls)	479 (0.6%)	774 (0.7%)
Missing	10,147 (12.5%)	13,612 (12.5%)

Table 3.13: Pupil numbers across schools of different categories of sex segregation by condition and Age Cohorts.

School Type per Cohort	non-affected	affected
YEAR 1 – YEAR 2		
Mixed	13,558 (86.1%)	17,936 (87.0%)
Single (boys)	0 (0.0%)	0 (0.0%)
Single (girls)	78 (0.5%)	96 (0.5%)
Missing	2,105 (13.4%)	2,578 (12.5%)
YEAR 2 – YEAR 3		
Mixed	15,316 (86.5%)	19,213 (86.1%)
Single (boys)	0 (0.0%)	0 (0.0%)
Single (girls)	75 (0.4%)	157 (0.7%)
Missing	2,310 (13.1%)	2,942 (13.2%)
YEAR 3 – YEAR 4		
Mixed	12,689 (87.4%)	16,827 (86.9%)
Single (boys)	0 (0.0%)	0 (0.0%)
Single (girls)	109 (0.8%)	171 (0.9%)
Missing	1,727 (11.9%)	2,361 (12.2%)
YEAR 4 – YEAR 5		
Mixed	15,302 (86.7%)	21,448 (86.7%)
Single (boys)	0 (0.0%)	0 (0.0%)
Single (girls)	110 (0.6%)	199 (0.8%)
Missing	2,244 (12.7%)	3,099 (12.5%)
YEAR 5 – YEAR 6		
Mixed	13,390 (87.8%)	18,843 (87.1%)
Single (boys)	0 (0.0%)	0 (0.0%)
Single (girls)	107 (0.7%)	151 (0.7%)
Missing	1,761 (11.5%)	2,632 (12.2%)

As to be expected the proportion of single sex schools is consistently extremely low across Age Cohorts and conditions. While this supports the notion of representativeness of the sample for the population of English primary school children, it also signifies the limitations of comparisons between these categories.

Another context variable refers to schools' Ofsted rating. As it tends to be used as proxy for teaching quality a meaningful and fair comparison of pupils' performances to identify potential effects of COVID-19-related school closures needs to be based on comparable profiles of schools' Ofsted ratings between the two conditions (Table 3.14 and Table 3.15).

Table 3.14: Ofsted ratings across schools in both conditions.

Ofsted Rating	non-affected	affected
Outstanding	47 (9.4%)	62 (9.2%)
Good	279 (55.6%)	362 (53.5%)
Requires improvement	45 (9.0%)	60 (8.9%)
Serious weaknesses	2 (0.4%)	4 (0.6%)
Special measures	6 (1.2%)	6 (0.9%)
Missing	123 (24.5%)	183 (27.0%)

Table 3.15: Ofsted ratings across schools, differentiated by Age Cohorts, separate for conditions.

Ofsted Rating	non-affected	affected
YEAR 1 – YEAR 2		
Outstanding	40 (10.0%)	54 (9.9%)
Good	224 (55.9%)	285 (52.4%)
Requires improvement	34 (8.5%)	47 (8.6%)
Serious weaknesses	1 (0.2%)	4 (0.7%)
Special measures	3 (0.7%)	4 (0.7%)
Missing	99 (24.7%)	150 (27.6%)
YEAR 2 – YEAR 3		
Outstanding	41 (9.5%)	52 (8.9%)
Good	236 (54.5%)	303 (52.0%)
Requires improvement	40 (9.2%)	56 (9.6%)
Serious weaknesses	1 (0.2%)	4 (0.7%)
Special measures	5 (1.2%)	4 (0.7%)
Missing	110 (25.4%)	164 (28.1%)
YEAR 3 – YEAR 4		
Outstanding	32 (8.8%)	43 (8.5%)
Good	210 (57.7%)	279 (55.4%)
Requires improvement	33 (9.1%)	45 (8.9%)
Serious weaknesses	1 (0.3%)	4 (0.8%)
Special measures	2 (0.5%)	4 (0.8%)
Missing	86 (23.6%)	129 (25.6%)
YEAR 4 – YEAR 5		
Outstanding	37 (8.9%)	49 (8.0%)
Good	233 (56.3%)	330 (54.0%)
Requires improvement	36 (8.7%)	57 (9.3%)
Serious weaknesses	2 (0.5%)	4 (0.7%)
Special measures	4 (1.0%)	5 (0.8%)
Missing	102 (24.6%)	166 (27.2%)

Table 3.15 continued from previous page …

Ofsted Rating	non-affected	affected
YEAR 5 – YEAR 6		
Outstanding	38 (10.2%)	42 (7.9%)
Good	210 (56.1%)	285 (53.7%)
Requires improvement	32 (8.6%)	44 (8.3%)
Serious weaknesses	1 (0.3%)	4 (0.8%)
Special measures	3 (0.8%)	5 (0.9%)
Missing	90 (24.1%)	151 (28.4%)

At a purely descriptive level it becomes apparent that the data for the affected sub-cohorts show lower proportions in the categories “outstanding” and “good” across all Age Cohorts. This “shift” cannot necessarily be interpreted as a drop in school quality (from the academic year 2018/2019 – the so called “non-affected” to the academic years 2019/2020 and 2020/2021) as it is not numerically counter-balanced by an increase in the proportions for the categories “requires improvement”, “serious weaknesses”, and “special measures”. It is, however, curious that the higher proportion in the category capturing missing information seems to mirror the lower proportion in the two favourable categories. This signifies that the missingness of data is not at random, which, again, creates a challenge to making sufficiently strong inferences regarding school closure effects.

The final set of descriptive information refers to the outcome variable central in this research project, i.e. writing scores. An appropriate, useful, and meaningful interpretation of empirical findings obtained through sound analyses requires a solid understanding of the information contained in the data used for these analyses.
Writing scores were derived through a process called comparative judgments (Thurstone, 1927a; 1927b; Pollitt, 2012), which is a performance scoring method based on an iterative process of pairwise comparisons of scripts. In this process, in its simplest form, two scripts are presented parallel on a computer screen and evaluators are asked to judge which of the two is the better one. Each script is iteratively paired with other scripts and judged accordingly. As a result, each script accumulates a record of ranking scores reflecting its relative level of superiority over other scripts in terms of perceived quality. For instance, the best in the pool of to be judged scripts will have the highest relative number of "wins". Scripts of perceived average quality in this pool will have close to equal relative numbers of “wins” and “losses”. These raw scores will then be standardised and scaled to facilitate comparisons across time (e.g. Age Cohorts), schools, and condition (e.g., affected vs non-affected). For an appropriate interpretation of scores or results of statistical analyses that utilise these scores it is important to keep in mind that scores reflect relative evaluations and are neither measurements against an absolute standard, nor a direct indicator of an ability. For comparisons of performance scores longitudinally, but also cross-sectionally across Age Cohorts and schools, or geographic regions etc. to be meaningful requires that scores are uniformly scaled. To achieve consistency in the allocation of scores, No-More-Marking (NMM) set up an anchoring process that comprised four components (see also Wheadon et al., 2020, p. 51).

(1) Participating schools had to upload their to-be-assessed scripts within a pre-determined time window through the No-More-Marking online portal.
(2) Of all the uploaded scripts across schools a random sample of 20% was selected by No-More-Marking. These scripts constituted the so-called anchor scripts which were systematically distributed across the pools of scripts that were send back to the schools (except the schools these scripts origin from) for them to be marked.

(3) This was followed by a pre-determined time window of one week in which all schools had to mark their scripts (including the anchor scripts) by employing the procedure of comparative judgment using the specifically designed computer interface of No-More-Marking. Participating schools were advised to complete at least 10 comparisons per script.
(4) As the set of anchoring scripts was systematically interspersed in the set of comparisons for each school, every fifth judgement conducted by teachers was performed on a random pair of scripts from pupils from other schools than their own. The records of judgements for these 'anchor scripts' were than subjected to an estimation procedure according to the Bradley-Terry-Luce (BTL) model (see Bradley & Terry, 1952; Luce, 1959). Serving as anchor values, the resulting scores of this sub-set were than incorporated in the separate estimation procedure related to the judgment records of the scripts from the teachers' respective schools. This way the judgments from different schools could be linked and the resulting scores projected onto a consistent scale.

Table 3.16: Descriptive statistics for writing performance scores across conditions.

Writing scores	non-affected		affected	
	T1	T2	T1	T2
N	80,881		108,653	
Mean (STDEV)	484.48 (70.43)	513.94 (50.48)	486.55 (63.08)	506.70 (52.56)
corr	.67		.68	
IQR	83.60	67.00	83.00	69.00
Range, min – max	0.00 – 680.07	-60.00 – 698.00	-60.00 – 698.00	-82.00 – 1052.00

Table 3.17: Descriptive statistics for writing performance scores across conditions and Age Cohorts.

Writing scores	non-affected		affected	
	T1	T2	T1	T2
YEAR 1 – YEAR 2				
N	15,741		20,610	
Mean (STDEV)	413.19 (56.34)	479.13 (47.67)	411.22 (58.15)	474.13 (56.72)
r(T1,T2)	.70		.69	
IQR	76.36	62.00	79.00	77.00
Range, min – max	225.07 – 643.02	281.00 – 640.00	226.00 – 590.00	-82.00 – 663.00
YEAR 2 – YEAR 3				
N	17,701		22,312	
Mean (STDEV)	471.11 (77.79)	488.22 (44.69)	476.84 (48.63)	485.74 (48.53)
r(T1,T2)	.44		.75	
IQR	64.99	56.00	64.00	65.00
Range, min – max	0.00 – 644.98	-60.00 – 698.00	281.00 – 644.00	297.00 – 633.00
YEAR 3 – YEAR 4				
N	14,525		19,359	
Mean (STDEV)	484.79 (48.62)	518.97 (42.72)	487.17 (45.06)	528.25 (42.69)
r(T1,T2)	.59		.59	
IQR	56.97	52.00	56.00	55.00
Range, min – max	0.00 – 623.60	0.00 – 656.00	-60.00 – 698.00	346.08 – 666.00

Table 3.17 continued from previous page …

Writing scores	non-affected		affected	
	T1	T2	T1	T2
YEAR 4 – YEAR 5				
N	17,656		24,746	
Mean (STDEV)	517.80 (49.26)	533.60 (40.59)	517.93 (43.13)	504.26 (42.98)
r(T1,T2)	.54		.63	
IQR	53.04	47.00	52.00	54.00
Range, min – max	0.00 – 656.07	0.00 – 680.00	0.00 – 650.00	-82.00 – 1052.00
YEAR 5 – YEAR 6				
N	15,258		21,626	
Mean (STDEV)	534.68 (41.05)	552.15 (35.46)	531.88 (41.17)	542.87 (37.97)
r(T1,T2)	.59		.57	
IQR	51.59	43.00	49.00	45.00
Range, min – max	366.76 – 680.07	350.00 – 687.00	0.00 – 680.00	-82.00 – 1052.00

The inspection of the information presented in Tables 3.16 and 3.17 reveals the presence of negative scores (see range information) throughout. Negative performance scores cannot be meaningfully interpreted. They are likely to be the result of too few comparisons performed per the respective scripts. The occurrence of negative performance scores, however, is extremely infrequent (especially in the context of a sample of the given size), so that they – one average – are expected to have little to no impact on the effect estimates in the analyses conducted to address the research questions posed for the current study.

The further inspection of Table 3.16 indicates that both the affected and the non-affected cohorts tend not to differ in their average T1 performance score. This constitutes a favourable prerequisite for later analyses to investigate general and differential effects, especially as it applies across all Age Cohorts (Table 3.17).

Another such generalised observation is the tendency of an increase in homogeneity (i.e., reduction of dispersion) in performance scores from T1 to T2 with one exception. The standard deviation for the performance score at T1 for the non-affected Age Cohort Y2Y3 is considerably larger than the other remaining 19 standard deviations across Age Cohorts and conditions, which indicates a wider dispersion of performance scores in this sub-cohort. The fact that the interquartile range (IQR) for the score distribution tends not to be noticeably different from all the others indicates the presence (of a low number) of extreme scores in the first and/or fourth quartile of the distribution. This, in conjunction with the comparatively low correlation between T1 and T2 (r = .44) could be interpreted as symptoms of compromised measurement quality (i.e., measurement error and reliability), which will have to be taken into account for the interpretation of estimates in relation to answering the research questions.

As would be expected, the average performance scores for older Age Cohorts tend to be higher compared to their younger counterparts. This trend is observable across the non-affected (see Figure 3.2) as well as affected cohorts. There is one exception. The average performance score at T2 for the Age Cohort Y4Y5 (504.26) appears to be surprisingly low in comparison to the T2 score of the Age Cohort Y3Y4 (between cohort difference). In terms of a within-cohort comparison (i.e., the change from T1 and T2), we would expect an increase in performance scores, which would indicate a to-be-expected increase in writing skill levels as a result of maturation and learning. This tends to be the case across cohorts and conditions, with, again, one exception. The average performance score registered for the affected Age Cohort Y4Y5 at T2 signifies a *decline* in performance within this cohort compared to its performance registered at T1. This is unique amongst the 10 different sub-cohorts included in this research. At this

stage it is difficult to speculate about the potential reasons for this anomaly. Nonetheless, it must be taken into account in later analyses and result interpretations.
Before turning to the various analysis steps related to addressing the research questions, we have a look at the overall trajectory of writing skill scores over the course of primary school (see Figure 3.2) using the average scores per Age Cohort as an approximation for the quality of writing performance typical for the respective Age Cohort (see Figure 3.2). This is to see whether the scaling of scores (incl. anchoring) has resulted in a to be expected increase in performance scores over time, indicating learning and development.

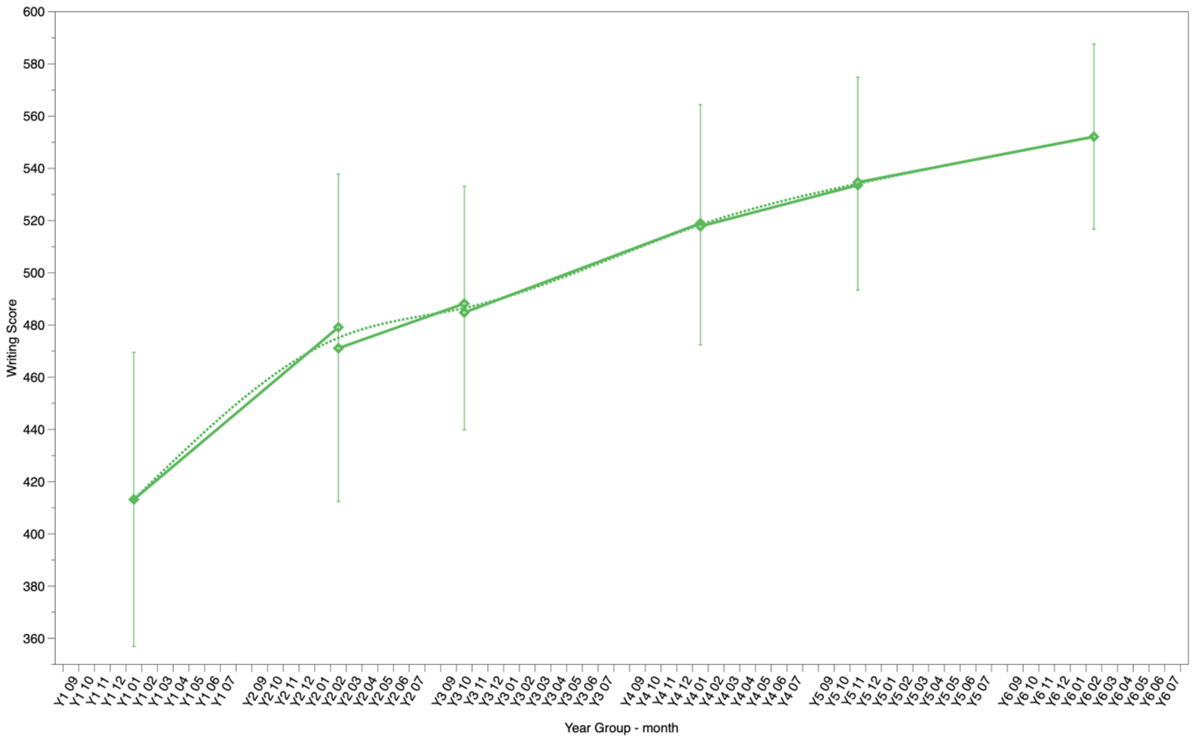

Figure 3.2: Approximation of the developmental trajectory for writing performance scores based on repeated measurements across Age Cohorts (note: only un-affected sub-cohorts are considered).

This trajectory represents the benchmark of learning and maturation-related changes in skill levels over time against which performance trajectories across Age Cohorts who were affected by COVID-19-related school closures have to be mapped.

Figure 3.3 provides a diagrammatic representation of the intra-cohort changes in writing performance for both conditions.

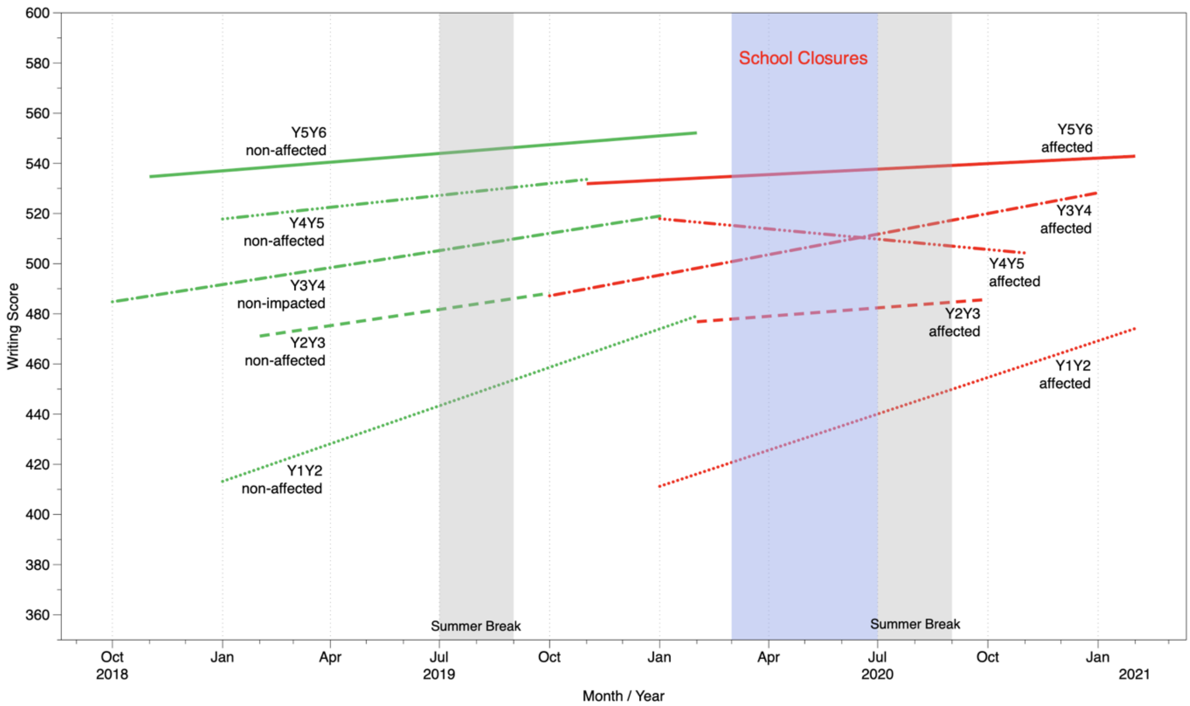

Figure 3.3: Intra-cohort changes in writing performance across Age Cohorts by condition.

Upon inspection of the data presented in Table 3.17 and the information depicted in Figure 3.3 it becomes apparent that the developmental trajectories, especially those for the affected cohorts vary considerably. As already indicated in Figure 3.1, the assessment intervals (i.e. T1 to T2) vary across Age Cohorts (ranging from 8 months to 15 months). This renders an aggregated level of analysis (e.g. treating learning progress between T1 and T2 across Age Cohorts as comparable) inappropriate. We therefore will conduct the effect analyses for each Age Cohort separately.

Prior to modelling, we glean some descriptive data regarding differences (between cohorts exposed to different conditions) in differences (within cohorts from T1 to T2). Notably, within the non-affected and within the affected sub-cohorts the pupils are matched, but between the sub-cohorts they are not (and cannot be) matched. The results of these analyses are provided in Table 3.18.

By ways of summary, so far, we presented information regarding the extent and quality of the database that forms the basis for the analyses to be conducted to address the research questions outlined in section 2.

The database comprises scores of writing performance assessed twice (repeated measurement) for a total of 189,534 English primary school pupils across five Age Cohorts. About 57% of the available datasets stem from pupils that were affected by temporary school closures during the first national COVID-19 lockdown. Meaningful analyses of potential general (RQ1) and differential (RQ2, RQ3) impacts of such school closures on the development of writing performance and the generalisability of findings requires (a) comparability of cohorts that were and were not affected, and (b) representativeness of the sample studied.

The available database is representative of the population of English primary school pupils in terms of the sex-ratio. This is the case for both the affected and non-affected cohorts. The database is also roughly representative in terms of how relative socio-economic disadvantage is distributed across the population of English primary school pupils.

The proportion of pupils attending private (or independent) schools is lower in the sample than is to be found in the reference population. This renders attempts to analyse effects attributable to this school characteristic and the generalisability of respective findings potentially problematic. The distribution of pupils in the sample across geographic regions differs between affected and non-affected cohorts. This imbalance needs to be considered when trying to establish geographic region-related differential effects of school closures on writing performances.

The descriptive statistics related to sex-segregation of schools, while comparable in terms of what is to be found in the population of English primary schools, highlights that contrasting of school types along this dimension is not meaningful.

Statistical information regarding schools' Ofsted ratings indicate a consistently lower proportion of schools in the categories "outstanding" and "good" amongst schools affected by COVID-19 related school closures. If one were to see this variable as a characteristic of a school that *contributes* to how well it is able to cope with school closures the fairness of a comparison between affected and not affected schools would need to be considered questionable. Alternatively, Ofsted ratings could be conceptualised as a school-level *outcome* variable that in itself reflects COVID-19 related impacts on the perceived quality of schooling.

All in all, the available data constitute a promising basis for the main analyses that will address the research questions.

Table 3.18: Mean Differences (and standard deviations) between T1 and T2 per Age Cohort.

	Mean differences between T2 and T1		Difference in Differences (DiD)	T1 – T2 interval in months	DiD per month
cohort	non-affected	affected			
Y1Y2	65.95 (40.96)	62.91 (45.41)	-3.04	13	-0.23
Y2Y3	17.11 (70.90)	8.91 (34.19)	-8.20	8	-1.03
Y3Y4	34.18 (41.93)	41.07 (39.58)	6.89	15	0.46
Y4Y5	15.80 (43.89)	-13.67 (37.05)	-29.47	10	-2.95
Y5Y6	17.46 (34.89)	10.99 (36.74)	-6.48	15	-0.43

A fair comparison of the differences in differences (DiDs) across Age Cohorts requires the varying intervals between the assessments to be taken into account (see penultimate and last column in Table 3.18). Based on a purely descriptive level it becomes apparent that the differences in learning progress between non-affected and affected sub-cohorts vary considerably across Age Cohorts. These differences range from almost 3 points less progress per month for the affected sub-cohort in the Age Cohort Y4Y5 to half a point *benefit* per month for the affected sub-cohort in the Age Cohort Y3Y4. This result pattern underlines the necessity to analyse differential effects for each Age Cohort separately.

In the next section we describe the statistical modelling approach taken to address the research questions.

4 Modelling

Multilevel statistical modelling is carried out to facilitate addressing the research questions. This approach will provide robust estimates of the impact of Covid-19 related school closures on the development of writing attainment for pupils aged five to 12 years. In accordance with the research questions regarding potential differential effects of COVID-19 related school closures variables such as economic deprivation, sex, school type, and geographic region will be included in the respective Difference-in Difference (DiD) regression analytic model.

The structure of the available database allows a longitudinal design to be adopted, such that the response variable is a repeated measurement of writing performance for each pupil (corresponding to T1 and T2). To illustrate, the writing score obtained by a Year 2 student in the 2019/20 academic year (T1) is supplemented by their writing score from the subsequent academic year, 2020/21, when they were in Year 3. In this particular example, this pupil's T2 score would be obtained after the period of approximately 3 months of COVID-19 related school closures, which places this pupil in the so-called affected sub-cohort for Age Cohort Y2Y3 (see also Figure 3.3 above).

4.1 Statistical model for Difference-in-Difference analysis related to RQ1

In technical terms, pupils who experienced a COVID-19 related school closure (March 2020 to July 2020) represent the 'treatment' group, while those who were not exposed to that experience between their T1 and T2 assessment of their writing performance constitute the 'control' group. Given the natural progression of the events in question the impact of school closures will be modelled and analysed in the context of this quasi-experimental design.

With this in view, for each Age Cohort separately, we create a 'base' model to estimate the potential impact of school closures on all pupils in that given Age Cohort. Consider the T1 scores and T2 scores to be associated with time indexes $t = 1$ and $t = 2$, respectively. The model corresponding to a DiD analysis can be formulated as:

$$y_{ikt} = B_0 + B_1(time_{ikt}) + B_2(cohort_{ik}) + \boldsymbol{B_3}(\boldsymbol{time_{ikt}} \times \boldsymbol{cohort_{ik}}) + \mu_k + \tau_i + \varepsilon_{ikt} \quad [4.1]$$

whereby (for all observations in a particular Age Cohort),

- y_{ikt} refers to the observed test scores for pupil i in school k at time $t = 1, 2$ (corresponding to T1 or T2, respectively);
- B_0 is the regression intercept;
- $time_{ikt}$ is a binary categorical variable with T1 as reference category, that is $time_{ikt} = 0$ if the measurement for pupil i in school k is taken at $t = 1$ (T1) and, $time_{ikt} = 1$ if it is taken at $t = 2$ (T2);
- $cohort_{ik}$ is a binary categorical variable for pupil i in school k taking the value 0 if that pupil is in the non-affected sub-cohort and the value 1 if the pupil is in the affected sub-cohort;
- $time_{ikt} \times cohort_{ik}$ is an interaction term, reflecting differences between being affected by school closures or not in the difference between writing scores at T1 and T2 (progress);
- $\mu_k \sim N(0, \sigma_\mu^2)$ is a school-level random effect;
- $\tau_i \sim N(0, \sigma_\tau^2)$ is a pupil-level random effect;
- $\varepsilon_{ikt} \sim N(0, \sigma^2)$ is an error term.

Based on equation [4.1], the coefficient of interest is B_3, which represents the difference in differences estimator (DiD). The Age Cohort specific estimate B_3 represents the net-additional effect attributable to COVID-19-related school closures during the period March and July 2020. This coefficient across the five Age Cohorts informs the answer to research question 1 regarding the general effect of COVID-19 on the development of writing skills.

4.2 Statistical models for Difference-in-Difference-in-Difference analysis related to RQ2

Addressing research question 2, i.e. the question whether there are differential effects related to selective pupil or context variables, requires the systematic inclusion of covariates in the model. This extends equation [4.1] by the respective covariate of interest (signified by the term *covariate* in equation [4.2]) and their respective interaction terms:

$$\begin{aligned} y_{ikt} = {} & B_0 + B_1(time_{ikt}) + B_2(cohort_{ik}) + B_3(covariate_{ik}) + \\ & B_4(time_{ikt} \times covariate_{ik}) + B_5(cohort_{ik} \times covariate_{ik}) + B_6(time_{ikt} \times cohort_{ik}) + \\ & \boldsymbol{B_7(time_{ikt} \times cohort_{ik} \times covariate_{ik})} + \mu_{ik} + \tau_i + \varepsilon_{ik} \qquad [4.2] \end{aligned}$$

In this model coefficient B_6 signifies the effect of COVID-19-related school closures on pupils' progress from T1 to T2 (comparable to coefficient B_3 in equation [4.1]), while B_7 reflects the added, i.e. differential effect of COVID-19-related school closures that can uniquely be associated with the respective covariate. Separate analyses will be conducted for all binary covariates including learners' pupil premium status, pupils' sex, and the type of school they attend using model equation [4.2]. Please note, as none of these covariates depends on time, no index t is incorporated in the equation's notation. While the first pupil premium status and sex are indexed by both i (pupil) and k (school), the covariate school type only depends on k (school). Hence, in this case the covariate is codified as $covariate_k$.

Additionally, we investigate possible differential effects in terms of the geographic region in which their school is situated. The database contains information pertaining to 10 different geographical regions. To accommodate a variable with 10 categories, model equation [4.2] was adjusted, resulting in model equation [4.3]:

$$\begin{aligned} y_{ikt} = {} & B_0 + B_1(time_{ikt}) + B_2(cohort_{ik}) + \textstyle\sum_{j=1}^{9} B_{3j}(region_{kj}) + \\ & \textstyle\sum_{j=1}^{9} B_{4j}(time_{ikt} \times region_{kj}) + \sum_{j=1}^{9} B_{5j}(cohort_{ik} \times region_{kj}) + \\ & \textstyle\sum_{j=1}^{9} B_{6j}(time_{ikt} \times cohort_{ik}) + \sum_{j=1}^{9} \boldsymbol{B_{7j}(time_{ikt} \times cohort_{ik} \times region_{kj})} + \\ & \mu_{ik} + \tau_i + \varepsilon_{ik} \qquad [4.3] \end{aligned}$$

where $region_{kj} = 1$ if *school k* is in *region* $= j$ and 0 otherwise, and, just as for school type, there is no dependence of this covariate on i. Clearly this model has very many parameters, but the nine region-specific interaction parameters B_{7j} in [4.3] can be interpreted just as the parameter B_7 in [4.2], namely the differential effect of COVID-19 related school closures on pupils' writing performance progress in region j. One of the ten regions needs to serve as reference category; this choice is arbitrary, and in our analyses, this will be *East Midlands*. Incidentally, this region is the proportionally strongest in terms of the number of pupils considered in the analyses.

4.3 Statistical model for variance decomposition analysis related to RQ3

In order to address the additional question of whether the level of influence of being in a particular teaching group (i.e. class) or school changed as a result of school closures (see research question 3), slightly different models will be fitted. These models will not be longitudinal in spirit and will not contain a *time* variable but will instead be fitted for the non-affected and affected sub-cohorts separately. However, the T1 score will now be included as a covariate, with the T2 score serving as outcome. Hence, for a specific Age Cohort, we have the model:

$$y_{ijk} = B_0 + B_1(x_{ijk}) + \mu_k + \rho_j + \tau_i + \varepsilon_{ik} \qquad [4.4]$$

where y_{ijk} corresponds to the T2 measurements of pupil i in class j in school k, and $pretest_{ijk}$ corresponds to the T1 (“x”) measurement of that pupil, $\rho_i \sim N(0, \sigma_\rho^2)$ is a class-level random effect, and μ_k, τ_i, are as before. The interest lies in the proportions of σ^2, σ_μ^2 and σ_ρ^2 among the total variance $T = \sigma^2 + \sigma_\mu^2 + \sigma_\rho^2$ for model [4.4], for each of the affected and non-affected sub-cohorts. This analysis will be carried out separately for all five Age Cohorts.

It is noted that all models mentioned above are three-level models. All models are fitted using R function **lmer**. Where **lmer** reported warning messages due to convergence problems or singularities, the results were verified using **glmmTMB**. This led at all occasions to a very similar result as the one originally returned by **lmer**. The result from **glmmTMB** was presented in the Results section in such cases.

5 Results

5.1 Research Question 1: General Effects

The DiD base model has been computed as per equation [4.1], and the analysis proceeded as outlined in Section 4.1. To facilitate the interpretation of the results we report the unstandardised regression weights (B) instead of the commonly used standardised coefficients (β). Table 5.1 provides an overview of the parameter estimates for each Age Cohort.

Table 5.1: Age Cohort specific estimates of regression coefficients (unstandardised).

Age Cohort	Intercept (B_0)	Time (B_1)	Cohort (B_2)	Time by Cohort (B_3)
Y1Y2	411.66	65.95	-1.28	-3.04
Y2Y3	469.51	17.11	7.69	-8.20
Y3Y4	484.58	34.18	1.83	6.89
Y4Y5	516.92	15.80	1.25	-29.47
Y5Y6	534.50	17.46	-2.95	-6.48

Coefficient B_0 in Table 5.1 represents the intercept for the regression, indicating the average performance of the non-affected sub-cohorts at T1. The fact that this coefficient increases systematically from Age Cohort to Age Cohort signifies a to-be-expected systematic improvement in writing skills over time (see also Figure 3.2 above).

Coefficient B_1 indicates the average performance point difference between T1 and T2 for the sub-cohort that was not affected by COVID-19-related school closures. It can be interpreted as the benchmark for the expected change in performance between T1 and T2 against changes in performance in the affected sub-cohort are to be mapped (see also Figure 3.2 above). For a meaningful comparison of progress rates across Age Cohorts the differences in the respective measurement intervals (i.e. the time between T1 and T2) need to be taken into account (see also Table 3.18). For instance, while the 66-points difference observed in Age Cohort Y1Y2 (Table 5.1) is a result of 13 months of learning, the 17-points difference observed in Age Cohort Y2Y3 (Table 5.1) was realised over the course of 8 months. The simplest way of achieving some form of comparability across Age Cohorts is to express the 'Time' related estimates as 'progress per month'. To aid a like-for-like comparison across Age Cohorts, Table 5.2 shows the estimates for B_1 (and B_3), expressed as points per month.

Table 5.2: Age Cohort specific estimates of unstandardised regression coefficients in terms of points per month.

Age Cohort	Time (B'_1)	Time by Cohort (B'_3)
Y1Y2	5.07	-0.23
Y2Y3	2.14	-1.03
Y3Y4	2.28	0.46
Y4Y5	1.58	-2.95
Y5Y6	1.16	-0.43

This approach results in about 5.1 points difference for Y1Y2 and a little bit over 2 points difference for Y2Y3. For the Age Cohorts Y3Y4, Y4Y5, and Y5Y6 the progression rates (i.e. change in writing performance scores per month) are on average 2.3 points, 1.6. points, and 1.2 points, respectively. This trajectory across Age Cohorts, which roughly reflects a to be expected pattern of skill acquisition with the highest level of improvement in the early years and some plateauing in the later years, would otherwise have been obscured by a direct comparison of the B_1 estimates as they are shown in Table 5.1.

Coefficient B_2 in Table 5.1 provides information that is of relevance for the viability of the subsequent analysis. It represents the points difference between the sub-cohorts per Age Cohort at T1, which is the measurement occasion at which we expect the respective two sub-cohorts not to differ. Therefore, small values suggest comparability of the sub-cohorts at T1, which is the precondition for meaningful comparisons of the performance changes to T2 between the affected and non-affected sub-cohorts. As expected, estimates are small, with one exception. For Age Cohort Y2Y3 the sub-cohort that later was exposed to school closures performed on average almost 8 points better at T1 than the sub-cohort that remained unaffected. These "pre-existing" differences need to be taken into account when interpreting the differences between the affected and non-affected sub-cohorts at T2. Our modelling approach accommodates this.

Coefficient B_3 in Table 5.1 reflects the difference in progress (within-person changes in performance between T1 and T2) between sub-cohorts that were and were not affected by school closures between T1 and T2. In technical terms, B_3 represents a 'difference in differences' estimate (DiD). As mentioned before, a fair comparison of the size of the DiD estimates across Age Cohorts requires the differing measurement intervals between T1 and T2 to be considered. The B'_3 estimates in Table 5.2 show the DiDs in terms of difference (between affected and un-affected sub-cohorts) in differences (points progress per month) across Age Cohorts.

A couple of aspects are to be highlighted. The differences in monthly progress between affected and non-affected sub-cohorts (DiD) are rather small overall, with one exception. This is the almost 3 points difference per month – which amounts to about 8.8 points difference for a 3-

month period (approximately the duration of the COVID-19-related school closures) – for Age Cohort Y4Y5. This is to be mapped against the 1.6 points progress per month made by the unaffected sub-cohort (see B'$_1$ in Table 5.2). Such difference does not simply represent – in relative terms – a lack of progress in the affected sub-cohort, it actually suggests *a regression* to a level of performance that is more typical for Y3 pupils. In other words, the data suggest that the affected sub-cohort in that Age Cohort would have lost skills that were already gained. In somewhat of a contrast, for Age Cohort Y3Y4, we notice an average of 0.5 points *more* progress per month in the affected sub-cohort.
Figure 5.1 provides a graphical representation of the effects discussed in this section.

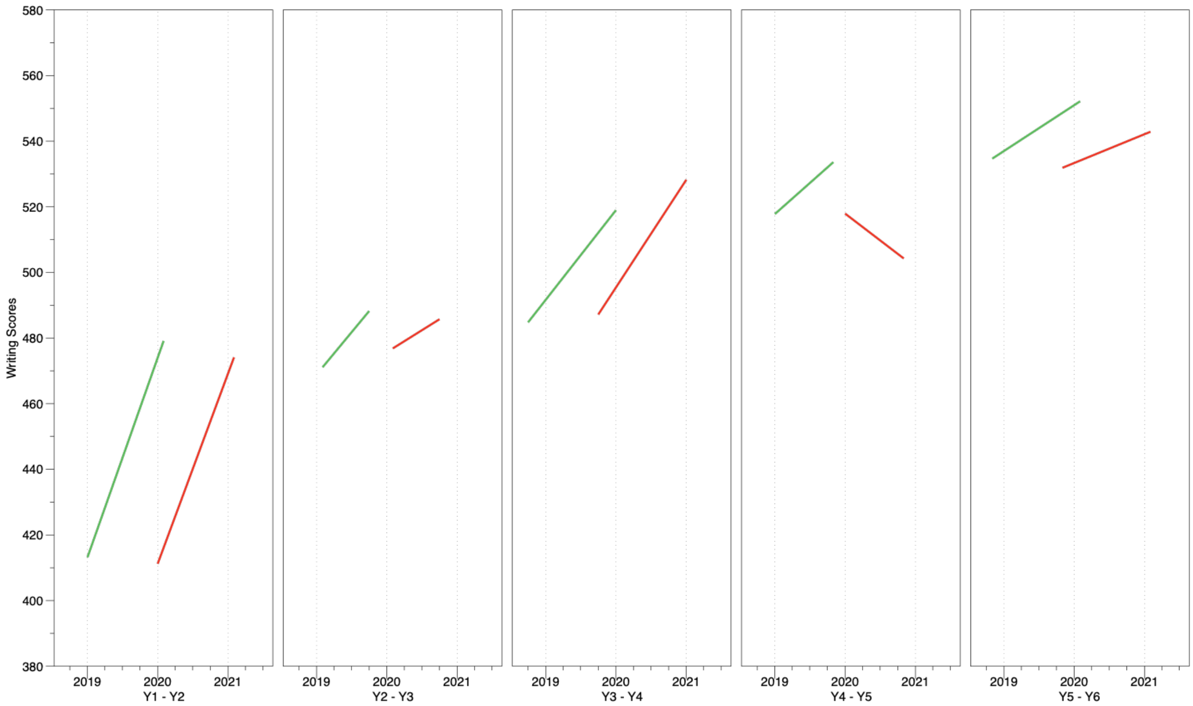

Figure 5.1: Age Cohort specific comparison of performance trajectories of affected (red) and non-affected (green) sub-cohorts.

5.2 Research Question 2: Differential Effects

In addressing research question 2, we exploratively decompose the effects observed under a general DiD perspective. We do this by analysing to what extent selected co-variates, such as pupil premium eligibility, sex, or school type contribute to the effects of COVID-19-related school closures.
The subsequent analyses will enable us to identify, for instance, whether a pupil's sex moderates the extent to which COVID-19-related school closures affected the progress in acquiring writing skills. The same perspective is taken with regard to an indicator of socio-economic depravation (i.e., pupil premium eligibility) or the type of school attended (i.e., state-funded vs independent schools). Analyses based on regression equation [4.2] are conducted separately for each covariate and for each Age Cohort. The respective estimates are presented in Tables 5.3, 5.4, and 5.5 below.

Table 5.3: Unstandardised Regression Weights for Age Cohort specific Models including Pupil Premium (no = 0, yes = 1) as covariate.

Age Cohort	Intercept (B_0)	Time (B_1)	Cohort (B_2)	Premium (B_3)	Time by Premium (B_4)	Cohort by Premium (B_5)	Time by Cohort (B_6)	Time by Cohort by Premium (B_7)
Y1Y2	414.96	65.37	-1.12	-20.44	3.41	-0.81	-1.95	-6.66
Y2Y3	474.53	15.76	6.20	-21.77	6.22	2.52	-6.45	-8.36
Y3Y4	488.98	33.60	1.95	-18.11	2.38	-0.89	6.70	0.84
Y4Y5	520.16	15.83	1.97	-14.69	-0.14	-1.52	-29.50	0.14
Y5Y6	537.93	17.12	-2.92	-13.35	1.32	-1.84	-6.41	-0.10

Table 5.4: Unstandardised Regression Weights for Age Cohort specific Models including Sex (male=0, female=1) as covariate.

Age Cohort	Intercept (B_0)	Time (B_1)	Cohort (B_2)	Sex (B_3)	Time by Sex (B_4)	Cohort by Sex (B_5)	Time by Cohort (B_6)	Time by Cohort by Sex (B_7)
Y1Y2	400.41	66.47	-1.90	22.72	-1.06	1.03	-4.08	2.09
Y2Y3	459.14	17.47	7.16	21.02	-0.74	0.84	-8.82	1.24
Y3Y4	474.32	33.62	2.27	20.19	1.11	-0.36	7.12	-0.43
Y4Y5	506.65	17.17	0.89	20.85	-2.74	0.44	-30.97	2.81
Y5Y6	523.35	20.30	-0.94	22.35	-5.68	-3.86	-9.09	5.23

Table 5.5: Unstandardised Regression Weights for Age Cohort specific Models including School Type (independent = 0, state-funded=1) as covariate.

Age Cohort	Intercept (B_0)	Time (B_1)	Cohort (B_2)	School (B_3)	Time by School (B_4)	Cohort by School (B_5)	Time by Cohort (B_6)	Time by Cohort by School (B_7)
Y1Y2	444.66	65.87	-6.45	-33.17	0.11	5.15	7.81	-11.37
Y2Y3	515.63	-0.33	-4.28	-46.68	17.66	12.36	8.38	-17.34
Y3Y4	518.02	25.82	1.37	-34.34	8.50	0.78	8.22	-1.43
Y4Y5	542.26	13.34	0.46	-25.76	2.34	0.77	-19.81	-9.40
Y5Y6	558.99	7.22	-4.09	-24.58	10.02	0.93	4.52	-10.98

Analogous to the results presentation related to addressing research question 1 (see previous section), we now discuss the results of the analyses conducted in relation to addressing research question 2, which are based on model equation [4.2]. Accordingly, we report the unstandardised regression weights (B) instead of standardised coefficients (β) to aid a more instructive interpretation of the results, especially in terms of their *practical* relevance.

Coefficient B_0 represents the intercept for the regression, indicating the average writing performance of the non-affected sub-cohorts at T1 for pupils not in receipt of pupil premium (Table 5.3), for male students (Table 5.4), or for pupils attending independent schools (Table 5.5), respectively. The size of this coefficient increases systematically from Age Cohort to Age Cohort – regardless of which covariate is included. This signifies a to-be-expected systematic progression in writing skills by age (see also Figure 3.2 above).

Coefficient B_1 reflects the performance increase observed in the non-affected sub-cohorts of learners who are not in receipt of pupil premium (Table 5.3), or are male (Table 5.4), or attend independent schools (Table 5.5). This coefficient can be interpreted as the benchmark for expected change in performance between T1 and T2 against changes in performance in the respective, affected sub-cohort are to be mapped (see also Figure 3.2 above). As discussed in relation to the base model, a comparison of estimates across Age Cohorts requires the differing lengths of the measurement intervals between T1 and T2 to be considered. Table 5.6 displays the transformed 'Time' related estimates as 'progress per month' for the models including different covariates. The second column contains the estimates obtained from the base model analyses discussed earlier for reference.

Table 5.6: Regression estimates for performance change per month across models and Age Cohorts.

	Regression Model without covariate (base model)	Regression Model including Pupil Premium as covariate	Regression Model including Sex as covariate	Regression Model including School Type as covariate
Age Cohort	Time B'_1	Time B'_1	Time B'_1	Time B'_1
Y1Y2	5.07	5.03	5.11	5.07
Y2Y3	2.14	1.97	2.18	-0.04
Y3Y4	2.28	2.24	2.24	1.72
Y4Y5	1.58	1.58	1.72	1.33
Y5Y6	1.16	1.14	1.35	0.48

Two observations are worth mentioning. First, the inclusion of either ‘Pupil Premium’ status or ‘Sex’ into the model tend not to make a difference in the estimates for Age Cohort-specific change rates per month (read horizontally across columns in Table 5.6) in the sub-cohorts that remain un-affected by school closures. In other words, the change rate per month for pupil premium recipients or male pupils tend not to be different from what was observed in the entire sub-cohort of un-affected pupils. This seems to be the case across all Age Cohorts. A differentiation according to ‘School Type’, however, produces systematically smaller estimates (except for Age Cohort Y1Y2). This suggests that pupils un-affected by school closures attending independent schools show, on average, smaller performance increases per months in relative to the entire sub-cohorts (compare column 2 and 5 in Table 5.6). The nominally smaller monthly growth rate in writing performance for pupils in independent schools needs to be interpreted in conjunction with the fact that these pupils tend to achieve substantially higher performance scores throughout (compare B_0 in Table 5.1 with B_0 in Table 5.5). Also, in contrast to ‘Sex’ and, to a degree ‘Pupil Premium’, ‘School Type’ (i.e., in its dichotomous operationalisation ‘independent vs state-funded’) represents an environmental attribute and not a pupil attribute as such. This distinction must be taken into account when interpreting the results as a whole (see later section). The second observation refers to the trajectories of the sizes of the change rates per month across Age Cohorts (read vertically across rows in Table 5.6). Comparing estimates across Age Cohorts reveals a to be expected pattern of skill acquisition with highest levels of improvement in the early years and some plateauing in the later years. In short, the amount of improvement per month declines with age. This phenomenon would remain obscured if the differences in measurement intervals were not taken into account in a direct comparison of the B_1 estimates as they are shown in Tables 5.3 to 5.5.

Coefficient B_2 in Tables 5.3, 5.4, and 5.5 provides information that is of relevance for the viability of the subsequent analyses. B_2 reflects the performance differences at T1 between affected and non-affected sub-cohorts of learners who are not in receipt of pupil premium (Table 5.3), or are male (Table 5.4), or attend independent schools (Table 5.5). As a reminder, the unaffected sub-cohorts at T1 are not expected to differ from the sub-cohorts whose T2 performance will eventually be affected by COVID-19-related school closures. As to be expected, estimates are comparably small, which is a prerequisite for meaningful comparisons of writing performance at T2 between the non-affected and then affected sub-cohorts.
A few exceptions are worth mentioning in this context and deserve further contextualising. For Age Cohort Y2Y3 amongst learners not in receipt of pupil premium the later affected sub-cohort performs on average 6.2 points better at T1 than their peers who remain un-affected (Table 5.3). For the same Age Cohort male learners later affected by COVID-19-related school closures show on average a 7.2 points better performance at T1 than those who remain un-affected (Table 5.4). For pupils educated in independent schools in Age Cohort Y1Y2 the sub-cohort that remains un-affected shows a 6.5 points better performance at T2 (see Table 5.5) Differences of 4.3 points and 4.1 points in the same direction are observed in Age Cohorts Y2Y3 and Y5Y6, respectively. Even if these estimates seem numerically different to the ones obtained across the remaining Age Cohorts (which are close to zero), they are still relatively small in the context of a scale with a mean score of 484.5 and a standard deviation of 70.43 (see Table 3.16). They should therefore not necessarily be interpreted as potential challenges to (a) the representativeness of the sub-cohorts and (b) the comparability of these sub-cohorts for the main analyses.

Coefficient B_3 in Tables 5.3, 5.4, and 5.5 reflect the difference in writing scores attributable to the respective covariate at T1 for the un-affected sub-cohorts. With regard to pupil premium (see Table 5.3), we see throughout substantial performance differences at T1 in favour of non-pupil premium learners. These differences range from 13.4 points (Age Cohort Y5Y6) to up to 21.8 (Age Cohort Y2Y3) points. While we can detect a systematic reduction of pupil-premium related differences in writing performance towards higher Age Cohorts, the similarly sized differences associated with sex (Table 5.4) of around 22 points in favour of female learners remain consistent across all Age Cohorts. This can be interpreted as an age-consistent sex effect. The differences in writing performance at T1 between independent and state-funded schools (Table 5.5) are substantial (between 24.6 and 46.7 points in favour of independent schools). There seem to be a slight tendency of a systematic reduction of these differences from Age Cohort to Age Cohort.

Coefficient B_4 in Tables 5.3, 5.4, and 5.5 reflect the extent to which the progress in acquiring writing skills (within-person changes in performance between T1 and T2) is associated with the respective covariate (i.e. pupil premium status, or pupil's sex, or the type of school attended). The estimates refer to the sub-cohort that remained un-affected by COVID-19-relates school closures and therefore represents some form of baseline against which the impact of school closures is to be mapped (see B_7). Positive values for B_4 estimates shown in Table 5.3 indicate that in the un-affected sub-cohort pupil premium recipients tend to make on average *greater* progress than non-pupil premium learners. For the Age Cohort Y2Y3 that remained un-affected by COVID-19-related school closures, we observe, for instance, that pupil-premium recipients make on average 6.2 points *more* progress than non-pupil premium learners.

Negative values for B_4 estimates in Table 5.4 signify that female learners make less progress than male learners. This seems to be the case for the un-affected sub-cohorts for Y5Y6 (5.7 points difference), Y4Y5 (2.7 points difference), and Y1Y2 (1.1 points difference). In light of the substantial and consistent sex differences highlighted in the context of the discussion related to B_3 in Table 5.4, female pupils continue to demonstrate superior performance compared to their male counterparts at T2. The estimates of B_4 indicate, however, that the rate of progress they exhibit between T1 and T2 tends on average to be marginally lower in comparison to their male peers.

Positive values for B_4 estimates presented in Table 5.5 indicate that in the un-affected sub-cohort pupils attending state-funded schools make greater progress than pupils educated in independent schools. In Age Cohorts Y2Y3 and Y5Y6 differences of 17.7 and 10 are observed. As argued earlier, an appropriate interpretation of the size of estimates across Age Cohorts requires the differing lengths of the intervals between T1 and T2 to be considered. To address this issue, we replicated the approach that was taken in discussing the estimates for B_1 across age groups which resulted in progress per months indices (see Table 5.6). Table 5.7 shows the so transformed estimates for B_4 (standardised as change points per month) to aid a fairer comparison across Age Cohorts.

Table 5.7: Regression estimates for the interaction between performance change per month and selected covariates across Age Cohorts.

Age Cohort	Time by Premium B'_4	Time by Sex B'_4	Time by School Type B'_4
Y1Y2	0.26	-0.08	0.01
Y2Y3	0.78	-0.09	2.21
Y3Y4	0.16	0.07	0.57
Y4Y5	-0.01	-0.27	0.23
Y5Y6	0.09	-0.38	0.67

The so derived estimates across Age Cohorts and covariates are generally negligibly small (less than 1 point DiD per month). This suggests that neither pupil premium eligibility, age, sex, nor school type had a moderating impact on the relative progress in the development of writing skills between T1 and T2. As the only marginal deviation from the overall result pattern, for the Age Cohort Y2Y3 we observe that pupils in the un-affected sub-cohorts attending state-funded schools tend to make on average 2.2 points *more* progress per month than their peers who are educated in independent schools.

Coefficient B_5 (in Tables 5.3, 5.4, and 5.4) indicates the association of the differences in average writing performance at T1 between the later affected sub-cohorts and the sub-cohorts that remain un-affected by COVID-19-related school closures with the respective covariate. As this comparison refers to T1, i.e. the measurement occasion prior to any COVID-19-related school closures, these associations are expected to be negligible. This appears to be generally the case, which supports the notion of comparability of sub-cohorts at T1, which is a necessary prerequisite for a meaningful identification of the relative contributions of the respective co-variate to the observed differences in differences (i.e., impact of COVID-19-related school closures on progress in acquiring writing skills).
With regard to pupil premium (Table 5.3), being in receipt of pupil premium in Age Cohort Y2Y3 is associated with an average of 2.5 points better performance at T1. In Age Cohort Y5Y6 the difference between the sub-cohorts appears to be associated with sex. That is, female pupils' performance in the later affected sub-cohort tends to be on average 3.9 points lower than their male peers' at T1 (Table 5.4). Age Cohort Y2Y3 pupils in the later affected sub-cohort who attend state-funded schools tend to show an on average 12.4 points *better* performance at T1 than their peers who attend independent schools.

Coefficient B_6 (Time by Cohort) indicates to what extent the progress in acquiring writing skills (within-person changes in performance between T1 and T2) differs between the sub-cohorts un-affected and affected by COVID-19-related school closures. Estimates for B_6 are derived for non-pupil premium recipients (Table 5.3), for male pupils (Table 5.4), or for pupils attending independent schools (Table 5.5).
A number of observations deserve further attention.

(1) The Age Cohort specific estimates for B'_6 (columns 3, 5 and 7 in Table 5.8) across the various models show high levels of correspondence with the respective B'_3 estimates (column 2), which were derived from a model without considering additional covariates. This suggests that the inclusion of covariates such as 'pupil premium' status, 'Sex', or 'School Type' contributes little to a more differentiated description of the effects of COVID-19-related school closures. This is in addition to the fact that the average DiD estimates are smaller than one point per month.
(2) The B'_6 estimates for the model considering school type in Table 5.8 are – except for Age Cohort Y4Y5 – positive. This indicates that sub-cohorts of pupils attending independent schools that were affected by COVID-19-related school closures show on average somewhat *higher* levels of progress per month than their peers who remained un-affected. While these estimates are also very small (and therefore of no real practical relevance as such), their relative consistency across Age Cohorts (except for Y4Y5) suggests that pupils attending independent schools tend to be able to effectively counter-act the otherwise observed, albeit similarly small, negative effects of COVID-19-related school closures (see B'_3 estimates in the second column of Table 5.8).
(3) In terms of relative size and direction of the estimated effect, Age Cohort Y4Y5 stands out. Regardless of learners' pupil premium status, sex, or school type for this Age Cohort we observe an about 2 to 3 points difference in progress per month between the un-affected and the affected sub-cohorts. Again, the consistency of this pattern across the different models indicates that this anomaly is not associated with pupil premium, sex, or the type of school attended.
(4) For Age Cohort Y3Y4 we also observe consistently (i.e. across all models) a *positive* COVID-19 effect. Again, it is the cross-model consistency rather the size of the effect (i.e., about a half a point per month) that warrants a mentioning. This phenomenon indicates that the affected sub-cohorts made an ever so slight *greater* progress on average than their un-affected peers.

When contrasting the estimates derived from the model that does not include covariates (see second column in Table 5.8, B'$_3$) with the respective estimates derived from the models that include one of the selected covariates (see column 3, 5, and 7, respectively showing B'$_6$ in Table 5.8) then it becomes apparent that the inclusion of the various covariates contribute only little to a more differentiated description the differences in difference (DiD).

Coefficient B_7 presented in Tables 5.3 to 5.5 reflects the unique contribution the selected covariate has in the composition of the difference in difference (DiD) estimate in the respective model for each Age Cohort. Table 5.8 shows the B'$_7$ estimates in terms of point difference per month. In other words, B7 indicates to what extent the respective covariate moderates the difference (between affected and unaffected) in difference (between T1 and T2).
The estimates of B_7 in Table 5.3, as do the B'$_7$ estimates in column 4 in Table 5.8, indicate to what extent the effect of COVID-19-related school closures is moderated by 'Pupil Premium'. For Age Cohort Y1Y2 being in receipt of pupil premium is associated with about half a point less progress per month (6.7 points across a 13-month period); for Age Cohort Y2Y3 pupil premium is associated with one point less progress per months (8.4 points across 8 months). For the remaining Age Cohorts pupil premium is not associated with the difference in progress between affected and non-affected sub-cohorts (see column 4 in Table 5.10 for T1-T2 interval attenuated estimates).
The estimates of B_7 in Table 5.4 indicate whether sex moderates the difference in progress between affected and non-affected sub-cohorts. The consistently low values across Age Cohorts (see column 6 in Table 5.8) suggest that the DiD does hardly differ by 'Sex'. While female pupils tend to generally perform better than male pupils (see B_3 in Table 5.4) they tend on average not to differ in their response to COVID-19-related school closures. Less than a third of a point per month difference in progress from T1 to T2 when affected by school closures can be attributed to pupil's sex (see column 6 in Table 5.8).
The estimates of B_7 in Table 5.5 indicate whether the difference in progress from T1 to T2 between affected and non-affected sub-cohorts is moderated by 'School Type'. As is the case for models including 'Pupil Premium" or pupils' 'Sex' as a covariate, estimates related to school type are also rather small. That is less than one point per month difference in progress between affected and non-affected sub-cohorts that is associated with school type. There seems one minor exception. The estimate for Age Cohort Y2Y3 indicates a negative 2.2 point difference in progress per month that can be attributed to being educated in a state-funded school (see last column in Table 5.8).

Table 5.8: Regression estimates for the interactions between performance change per month and selected covariates across Age Cohorts.

	Model without covariates	Model including Pupil Premium		Model including Sex		Model including School Type	
Age Cohort	Time' by Cohort (B'$_3$)	Time' by Cohort (B'$_6$)	Time' by Cohort by Premium (B'$_7$)	Time' by Cohort (B'$_6$)	Time' by Cohort by Sex (B'$_7$)	Time' by Cohort (B'$_6$)	Time' by Cohort by School Type (B'$_7$)
Y1Y2	-0.23	-0.15	-0.51	-0.31	0.16	0.60	-0.87
Y2Y3	-1.03	-0.81	-1.05	-1.10	0.16	1.05	-2.17
Y3Y4	0.46	0.45	0.06	0.47	-0.03	0.55	-0.10
Y4Y5	-2.95	-2.95	0.01	-3.10	0.28	-1.98	-0.94
Y5Y6	-0.43	-0.43	-0.01	-0.61	0.35	0.30	-0.73

Geographic Region:

Due to the variable's multi-categorical nature the consideration of 'Geographic Region' (see Figure 5.2 for a graphical representation) as a potentially moderating covariate a slightly adjusted modelling approach was required (see section 4.2 above).

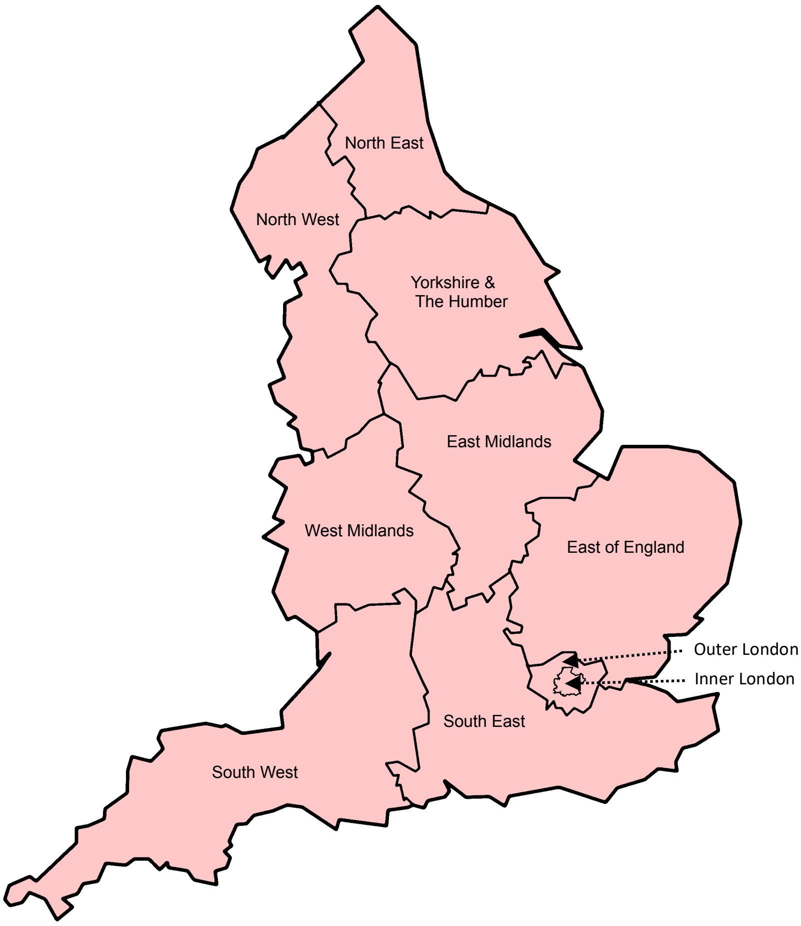

Figure 5.2: Geographic Regions as they are used as potential covariates in this analysis.

To aid a contextualised interpretation of the potentially moderating effects of 'Geographic Region' on effects of COVID-19-related school closures on the acquisition of writing skills during primary school years, we first present some baseline information. These include an overview of how pupils' average performances vary across the 10 different geographic regions. To this effect Table 5.9 informs about the differences in average performances across geographic regions and Age Cohorts relative to the reference Region 'East Midlands'. The top row in Table 5.9 presents the estimates for B_0 from each Age Cohort specific regression indicating the average performance of pupils in the un-affected sub-cohorts at T1. As is to be expected (and has been shown in Tables 5.3 to 5.5 and could be seen in Figures 3.2 and 5.1) the average performance levels increase with age. The B_3 estimates shown in the subsequent rows of Table 5.9 indicate the relative difference in the average performance shown by the unaffected sub-cohorts in each geographical region compared with the average performance of pupils attending schools in the 'East Midlands'. For instance, the average performance of Y1 pupils in schools located in the 'North West' is on average 5.6 points lower than the average performance obtained by their Y1 peers in the East Midlands.

Table 5.9: Unstandardised regression estimates for average T1 performance related to geographic region per Age Cohort (relative to East Midlands, see top row).

Variable	Age Cohort				
(Region)	Y1Y2	Y2Y3	Y3Y4	Y4Y5	Y5Y6
(Reference B_0 - East Midlands)	414.61	481.54	486.41	520.06	534.14
Region (B_3 - East of England)	-12.18	-22.16	-8.42	-8.80	-4.86
Region (B_3 - Inner London)	11.97	3.05	4.90	3.66	3.60
Region (B_3 - North East)	0.34	-4.40	2.54	-5.74	-2.21
Region (B_3 - North West)	-5.58	-14.86	-15.50	-19.77	0.51
Region (B_3 - Outer London)	-4.19	-2.01	1.52	6.13	9.50
Region (B_3 - South East)	-5.76	-27.29	-2.87	-0.67	1.07
Region (B_3 - South West)	-4.37	-19.93	-1.23	-5.04	-1.55
Region (B_3 - West Midlands)	-1.91	-17.74	0.72	4.01	2.04
Region (B_3 - Yorkshire and the	-1.57	-0.95	1.88	-2.64	8.21

The greatest variation in average performance across geographical regions is observed for pupils in year 2, while pupils in year 5 show the least variation. This suggests that there is an overall tendency for these regional differences to become less pronounced with increasing age. The average performance of pupils attending schools in the 'East of England' produces the on average largest negative discrepancies to the reference region (and therefore the lowest average performance overall), followed by the 'North West' and the 'South East'.

To ascertain whether the effects of COVID-19 related school closures on the development of writing skills differ between geographic regions the Age Cohort-specific estimates for the three-way interaction between 'Time', 'Cohort', and 'Geographic Region' based on regression equation [4.3] are to be inspected. This interaction is expressed in the coefficients B_7 for each of the 10 geographic regions considered, which are displayed in Table 5.10. These coefficients represent the point difference in progress between un-affected and affected sub-cohorts that can be attributed to attending a school in the respective geographic region relative to the average school in the 'East Midlands'. The top row in Table 5.10 reports the estimates of B_6 ('Time' by 'Cohort' interaction for pupils attending schools in 'East Midlands') as reference point.
To facilitate a like-for-like comparison of estimates across Age Cohorts, we divided the unstandardised coefficients by the Age Cohort-specific number of months between T1 and T2. The resulting estimates (B'_7 in Table 5.11) reflect the per month point differences in progress between un-affected and affected sub-cohorts attributable to attending a school in the respective geographic region relative to the average school in the 'East Midlands'. For example, Y1Y2 pupils attending schools in the 'North East' if affected by COVID-19-related school closures progress, on average, about 1 point *more* per month than their peers in the same Age Cohort who were not affected by school closures relative to pupils in the 'East Midlands'.

Table 5.10: Unstandardised regression estimates for the region-specific COVID-effects per Age Cohort (relative to East Midlands, see top row).

	Age Cohort				
Unstandardised Regression Weights	Y1Y2	Y2Y3	Y3Y4	Y4Y5	Y5Y6
Time by Cohort for Reference Region East Midlands (B_6)	-2.26	-1.21	2.52	-27.21	-5.89
Time by Cohort by Region (B_7 - East of England)	-5.00	-13.07	1.83	-6.45	-2.91
Time by Cohort by Region (B_7 - Inner London)	3.53	3.28	1.91	4.73	-3.00
Time by Cohort by Region (B_7 - North East)	12.67	5.08	11.38	-3.12	-13.79
Time by Cohort by Region (B_7 - North West)	-6.71	-11.63	-6.56	-17.27	-0.36
Time by Cohort by Region (B_7 - Outer London)	-3.56	12.73	0.15	-4.61	6.77
Time by Cohort by Region (B_7 - South East)	-0.36	-16.25	10.97	2.54	1.25
Time by Cohort by Region (B_7 - South West)	-0.88	-18.79	4.45	1.50	4.26
Time by Cohort by Region (B_7 - West Midlands)	-1.85	-14.29	10.06	-4.63	-1.47
Time by Cohort by Region (B_7 - Yorkshire and the Humber)	-0.59	-2.59	11.03	0.38	2.78

Table 5.11: Interval-standardised regression estimates (points per months) for the region-specific COVID-effects per Age Cohort (relative to East Midlands, see top row).

	Age Cohort				
Interval-standardised Regression Weights	Y1Y2	Y2Y3	Y3Y4	Y4Y5	Y5Y6
Time by Cohort for Reference Region East Midlands (B'_6)	-0.17	-0.15	0.17	-2.72	-0.39
Time by Cohort by Region (B'_7 - East of England)	-0.38	-1.63	0.12	-0.65	-0.19
Time by Cohort by Region (B'_7 - Inner London)	0.27	0.41	0.13	0.47	-0.20
Time by Cohort by Region (B'_7 - North East)	0.97	0.64	0.76	-0.31	-0.92
Time by Cohort by Region (B'_7 - North West)	-0.52	-1.45	-0.44	-1.73	-0.02
Time by Cohort by Region (B'_7 - Outer London)	-0.27	1.59	0.01	-0.46	0.45
Time by Cohort by Region (B'_7 - South East)	-0.03	-2.03	0.73	0.25	0.08
Time by Cohort by Region (B'_7 - South West)	-0.07	-2.35	0.30	0.15	0.28
Time by Cohort by Region (B'_7 - West Midlands)	-0.14	-1.79	0.67	-0.46	-0.10
Time by Cohort by Region (B'_7 - Yorkshire and the Humber)	-0.05	-0.32	0.74	0.04	0.19

While Table 5.11 shows the estimates *relative* to the reference region ('East Midlands'), Table 5.12 presents the *absolute* effects as points of progress per month for each region across Age Cohorts. This perspective allows to explore the results for patterns and helps to better contextualise their interpretation.

Table 5.12: Assessment interval-standardised regression estimates (points per month) for the region-specific COVID-effects on progress per Age Cohort.

	Age Cohort				
Region	Y1Y2	Y2Y3	Y3Y4	Y4Y5	Y5Y6
'East Midlands'	-0.17	-0.15	0.17	-2.72	-0.39
'East of England'	-0.55	-1.78	0.29	-3.37	-0.58
'Inner London'	0.10	0.26	0.30	-2.25	-0.59
'North East'	0.80	0.49	0.93	-3.03	-1.31
'North West'	-0.69	-1.60	-0.27	-4.45	-0.41
'Outer London'	-0.44	1.44	0.18	-3.18	0.06
'South East'	-0.20	-2.18	0.90	-2.47	-0.31
'South West'	-0.24	-2.50	0.47	-2.57	-0.11
'West Midlands'	-0.31	-1.94	0.84	-3.18	-0.49
'Yorkshire and the Humber'	-0.22	-0.47	0.91	-2.68	-0.20

In order to synthesise the insights derived from the analyses pertaining to the covariate 'Geographic Region', we now adopt a perspective in which we aggregate across Age Cohorts. However, it is important to be mindful that this approach inherently overlooks the above-highlighted differences between geographic regions in their dynamics of responding to school closures. Based on such aggregation, the lowest average effect of COVID-19-related school closures across Age Cohorts is observed for 'Outer London', followed by the 'North East' and 'Inner London' (i.e., affected pupils make about 0.4 points less progress per month in these three regions). COVID effects for 'Inner London' also show the highest level of consistency across Age Cohorts. That is, in this region we observe the least amount of variability of the size of a COVID effect across Age Cohorts. The region with the overall strongest negative COVID effect across all Age Cohorts is the 'North West' (about 1.5 points per month less progress, on average), followed by the 'East of England' (about 1.2 points less progress per month, on average, for school closure affected pupils).

For the 'North East' and 'Inner London' the effects of COVID-19-related school closures resulted for the first three Age Cohorts (Y1Y2, Y2Y3, and Y3Y4) in *higher* levels of progress per month. The *positive* COVID effects in these two regions amount to 0.7 points and 0.2. points per month *more* progress for pupils affected by school closures, respectively.

For the Age Cohort Y3Y4, we find positive COVID effects (i.e., affected pupils make around 0.6 points more progress per month) across all regions except the 'North West'. With high levels of consistency across regions, we observed for the Age Cohort Y4Y5 the overall strongest COVID effect (about 3 points per month less progress per month, on average). Again, the pupils attending schools in the 'North West' tend to show the strongest response to COVID-19-related school closures (about 4.5 points less progress per month), not just in this Age Cohort but generally.

Overall, the relative differences attributable to COVID-19-related school closures across geographic regions ranges from 4.5 points less progress (Y4Y5, 'North West'), to 1.4 points more progress per month (Y2Y3, 'Outer London') for the pupils in the affected sub-cohorts. While this suggests that the consideration of 'Geographic Region' as a covariate has very limited practical relevance in terms of a more differentiated description of potential effects of school closures, their consistency over age and place, however, lends credibility to these estimates.

In terms of an overall appraisal of the adequacy of the modelling approach we inspect the conditional R^2 for each of the regressions (see Table 5.13). The conditional R^2 represent the proportion of total variance observed in the outcome variable that is accounted for by the combination of fixed and random predictors in the respective regression model. It can be seen as a proxy of the model quality in terms of effectiveness and efficiency of prediction. By contrasting the conditional R^2 obtained for the baseline models (e.g., no covariates included) with those obtained for models that do include additional covariates one can ascertain whether adopting a more differentiated perspective (e.g., by considering pupil's pupil premium status, their sex, the type of school they attend, or the geographic region in which their school is located) increases the precision of describing potential effects of COVID-19-related school closures on the acquisition of writing performances during primary school. In doing so, two main observations can be made. First, the conditional R^2 for the base model (see second column in Table 5.13) are sufficiently substantial, but they also vary across Age Cohorts (ranging from 54 to 77 per cent of accounted for variance). This variation confirms the necessity to analyse the potential effects of school closures separately for each Age Cohort. Second, there is hardly any difference in the conditional R^2 across the different models for each Age Cohort (see columns 3 to 6 in Table 5.13). This consistency in conjunction with the relatively small values for the regression coefficients signifies that the inclusion of additional covariates as potential moderators such as pupil premium, sex, school type (see B_4 in Table 5.7, B_6 and B_7 in Table 5.8), or geographic region (see Table 5.11) does not increase the effectiveness in accounting for variance in the outcome variable.

Table 5.13: Conditional R^2 as proxies for model fit for regression models across Age Cohorts and selected covariates.

	Conditional R^2				
Age Cohort	Model without covariates	Model incl. Pupil Premium	Model incl. Sex	Model incl. School Type	Model incl. Geographic Region
Y1Y2	.77	.77	.77	.77	.77
Y2Y3	.54	.54	.54	.53	.54
Y3Y4	.66	.66	.66	.66	.66
Y4Y5	.61	.60	.61	.60	.61
Y5Y6	.59	.59	.59	.59	.59

5.3 Research Question 3: Variance components

Research question 3 aims to establish whether and how the level of influence of being in a particular teaching group/class or school changed as a result of COVID-19-related school closures. Through multi-level modelling we achieve a decomposition of the total variance observed in the outcome variable (i.e., writing performance scores) into variance that can be attributed to the school a pupil is attending and the variance that can be attributed to the class (or teaching group) a pupils is part of. The remaining portion of the variance (i.e., the residual in statistical terms) is interpreted as pupil related variance. Differences and potential changes in the composition of variance proportions are to be interpreted as a result of a complex interplay of exogenous factors (e.g., curricula) and endogenous factors (e.g., skill and competency development, incl. growth in self-regulation). Table 5.14 presents the variance components and subsequent percentage values at the school, class, and pupil level for each of the respective Age Cohorts.

Table 5.14: Variance components (absolute and relative) of school, class, and pupil-level across Age Cohorts for affected and non-affected sub-cohorts.

Variance Component	non-affected	affected
YEAR 1 - YEAR 2		
School	118.89 (10.2%)	367.71 (21.3%)
Class	96.73 (8.3%)	50.79 (2.9%)
Pupil	944.90 (81.4%)	1307.37 (75.8%)
Total	1160.52 (100%)	1725.87 (100%)
YEAR 2 - YEAR 3		
School	1546.48 (64.5%)	122.05 (11.9%)
Class	28.11 (1.2%)	49.75 (4.8%)
Pupil	823.17 (34.3%)	854.46 (83.3%)
Total	2397.76 (100%)	1026.26 (100%)
YEAR 3 - YEAR 4		
School	403.20 (30.3%)	251.96 (21.1%)
Class	73.70 (5.5%)	46.02 (3.8%)
Pupil	853.10 (64.1%)	897.47 (75.1%)
Total	1330.00 (100%)	1195.45 (100%)
YEAR 4 - YEAR 5		
School	384.37 (30.6%)	157.73 (14.0%)
Class	72.67 (5.8%)	52.82 (4.7%)
Pupil	797.59 (63.6%)	915.71 (81.3%)
Total	1254.72 (100%)	1126.02 (100%)
YEAR 5 - YEAR 6		
School	97.90 (12.0%)	91.29 (9.4%)
Class	30.42 (3.7%)	45.67 (4.7%)
Pupil	687.74 (84.3%)	834.31 (85.9%)
Total	816.06 (100%)	816.07 (100%)

The variance decomposition conducted in relation to addressing research question 3 allows for a more differentiated description of patterns of potential changes in the relative importance of these sources over time and across conditions. As was noticed when inspecting the descriptive statistics presented in Table 3.17, the variance of writing scores in the un-affected sub-cohort in Age Cohort Y2Y3 appears to be exceptionally high compared to all other Age Cohorts' score distributions. The analyses conducted in relation to research question 3 now indicates that this is mainly due to the variance component that is attributable to School. Its size exceeds the average by a factor of more than three when compared to all other un-affected Age-Cohorts.

Using the un-affected sub-cohorts as benchmark (considering the outlier position of Age Cohort Y2Y3), the proportion of total variance that can be attributed to class or teaching group is rather small (about 5%) while the school related variance component captures on average a bit more than 20%. This consistently observed result pattern confirms the to be expected, namely that classes within a school are more similar (i.e., they vary to a lesser degree) than classes or teaching groups across different schools. The remaining portion of the total variance in the distribution of writing scores rests with the variance at individual pupil level (more than 70%).

The information presented in Table 3.17 also indicated a systematic reduction in the dispersion of score distributions from T1 to T2. No markable difference was observed when comparing the overall variabilities between the un-affected and affected sub-cohorts. The data presented in Table 5.14 in relation to the total variance confirm this result, indicating similarly sized writing score distributions in un-affected and affected sub-cohorts (again, controlling for the distortion imposed by Age Cohort Y2Y3). A more differentiated perspective, as stipulated by research question 3, indicates that there are, however, differences in the composition of the total variances. We observe a reduction of the relative weight of school-related variance in the sub-cohorts affected by COVID-19-related school closures (except for Age Cohort Y1Y2). Subsequently, the shrinkage of the proportion of school-related variance in the composition of the total variance in writing scores is matched by the increase in the relative importance of the variance that is related to pupils' characteristics.

In conclusion, while the overall levels of variance in score distributions remain comparable between un-affected and affected sub-cohorts, the relative weight of its constituting components differs. The relative importance of class or teaching group remains largely comparable between un-affected and affected sub-cohorts; however, the relative contribution of school is reduced, and the importance of pupil characteristics increases subsequently even further (i.e. from 73% to more than 80% averaged across all Age Cohorts) when writing skills development is affected by school closures related to the pandemic.

6 General Discussion

In this section we integrate and synthesise the findings reported in the previous sections. The COVID-19 pandemic has resulted in a multitude of disruptions, including the closure of educational institutions and the cessation of in-person teaching across multiple periods.

The main objective of the present analyses was to ascertain the potential impact of school closures on the acquisition of writing skills during the primary school years. This objective was pursued by addressing three main research questions. Research Question 1 aimed, on a general level, to establish whether, and if so, to what extent the temporary lack of in-person teaching produced a lack of learning (learning loss). Research Question 2 employed a differential approach to investigate the potential moderating effects of selected pupil and contextual variables on the impact of COVID-19 related school closures on the development of writing performance. Research Question 3 aimed to ascertain potential shifts in the importance of

contextual factors, such as school, or class (i.e., teaching group) relative to person factors on writing skills development as it was impacted by COVID-19-related school closures.
These research questions are addressed by analysing the information contained in a database No-More-Marking has made available. The database contains writing skill scores of a total of 189,534 primary school pupils in England, collected between October 2018 and February 2021. The total sample can be subdivided into 5 Age Cohorts, each of which can be further subdivided into sub-cohorts that have or have not experienced COVID-19 related school closures during the period of March to July 2020. Scores for writing skills are derived in the context of comparative judgment. For each pupil two measurements of their writing skills are available. This makes it possible to analyse potential effects of COVID-19 related school closures by employing a so-called difference in difference (DiD) approach; that is, by comparing affected and non-affected sub-cohorts in terms of their progress in writing performance between assessment time 1 and assessment time 2 obtained in the respective subsequent academic year.

The generalisability of the findings in relation to answering the research questions by benefitting from the quasi-experimental data constellation is determined by two major factors, (1) the representativeness of the sample for the population of primary school pupils in England, and (2) the comparability of the two sub-cohorts (affected vs. non-affected) per Age Cohort. Preparatory analyses relying on the information comprised in the database indicate that overall, the sample can claim representativeness in terms of sex ratio and the proportion of pupils in receipt of pupil premium. In terms of the adopted dichotomous categorisation of school type into independent and state-funded schools, the analyses indicate a relative underrepresentation of independent schools in the sample (i.e., 1 to 2% in the sample in contrast to about 5% in the population). As long as the relative under-representation of independently funded schools relative to the proportion of state-funded schools is comparable across the two conditions (see factor (2), which appears to be the case, its consequence with regard to being able to derive trustworthy estimates of the effects of school closures are not as severe. One might even argue that such constellation leads to a rather conservative estimation of the differential effects related to school type. The distribution of schools across the various geographic regions within England also reflects positively on the representativeness of the sample used in this research.
The derivation of meaningful inferences from contrasting writing skill progress under the two conditions also necessitates the establishing of the comparability of sub-cohorts across Age Cohorts. To achieve this, two analysis strategies have been employed. The first of these involves mapping the relative representation of various characteristics (e.g. sex, pupil premium, Ofsted ratings, school type, etc.) across both sub-cohorts. Across the covariates considered in this study the differences in proportional representation between the two sub-cohorts rarely reaches 5%, which lends support to the notion of comparability. The second analysis strategy to check comparability contrasts the mean writing performance scores between the two sub-cohorts at assessment time 1 (T1) across Age Cohorts (see estimates for B_2). Comparability would be indicated by an absence of substantial differences (meaning that the to be compared sub-cohorts start at the same level of writing performance). The regression-analytic approach revealed that the performance differences between the two sub-cohorts ranged from 1.3 points for year 4 to 7.7 points for year 2 (see Table 5.1). In the context of a scale with a mean score of 485.7 and a standard deviation of 66.2, these differences appear to be negligible, thereby not jeopardising the notion of comparability.
The inspection of the descriptive statistics related to the writing scores across time, age, and condition offers further important information. A cross-sectional perspective (i.e. comparing performance scores across Age Cohorts) shows that the average performance scores increase by age. This indicates – as to be expected – a growth in writing skills as pupils progress in their primary school education. Adopting a basic longitudinal perspective (i.e., within-person

change) reveals that pupils' performance increases between assessments. Such increases range from 17.5 points over a period of 15-month (i.e., 1.2 points per month) for Age Cohort Y5Y6 to 65.9 points over a 13-month period (i.e. 5.1 points per month) for Age Cohort Y1Y2. This more directly reflects effects of learning and skill acquisition. In combination, writing performance increases by age, the rate of progress, however, decreases from Age Cohort to Age cohort.
Contrasting the distributions of performance scores obtained at the first and second assessment occasion (T1 and T2) reveals a highly consistent result pattern. The inter-individual variability of performance scores shrinks from the first to the second assessment. This homogenisation tendency could be speculatively interpreted as a result of some form of "re-test" effect. That is to suggest that pupils may have greater familiarity with the writing task itself on the second occasion. At this point, it cannot be ruled out that this familiarisation effect may have been facilitated through repeated exposure to similar tasks as part of the teaching activities between the two measurement occasions.

In answering research question 1, the three months disruption of in-person teaching due to COVID-19-related school closures resulted in very small differences in writing skill acquisition. Results indicate that in this 3-month period the affected sub-cohort for Age Cohort Y1Y2 makes 0.7 points in three months less progress than their peers in the un-affected sub-cohort. For Age Cohort Y2Y3 the difference is 3.1 points within this 3-month period. For Age Cohort Y3Y4 the affected sub-cohort makes on average 1.4 points *greater* progress during the period of school closures. Within the time frame of those three months, the affected sub-cohort in Age Cohort Y4Y5 makes 8.9 points less progress in comparison to the 4.7 points progress made by the un-affected sub-cohort during this time. The learning loss for pupils in Age Cohort Y5Y6 during the 3-month period of COVID-19-related school closures amounts to 1.3 points on the writing performance scale. As can also be seen from inspecting Figure 6.1, the effects of COVID-19-related school closures on the acquisition of writing skills during primary school are very small, suggesting limited practical relevance. The result pattern for Age Cohort Y4Y5 suggests more of a substantial loss in learning. As will be later discussed as part of addressing research question 2, both the positive (see Y3Y4) as well as the negative (Y4Y5) effects of COVID-19-related school closures are consistently observed across the 10 geographic regions in England, which somewhat pre-emptively addresses potential concerns that these findings might be outliers. It is important to note, however, that based on a description of a result pattern (as presented here in this report) no prescriptions for interventive actions can (and should) be offered. Effective and meaningful interventions need to be informed by *explanations*, i.e., a clear understanding of the processes and mechanisms that are causal for the emergence of the differences observed. Descriptions – as presented here – are not explanations with causal substance as such and, as a result, cannot form the basis for meaningful recommendations aimed at either non-affected Y3Y4 pupils or affected Y4Y5 pupils. In sum, the loss of 3 months in-person teaching across primary school did not translate into a substantial loss of learning with regard to acquiring writing skills during primary school.

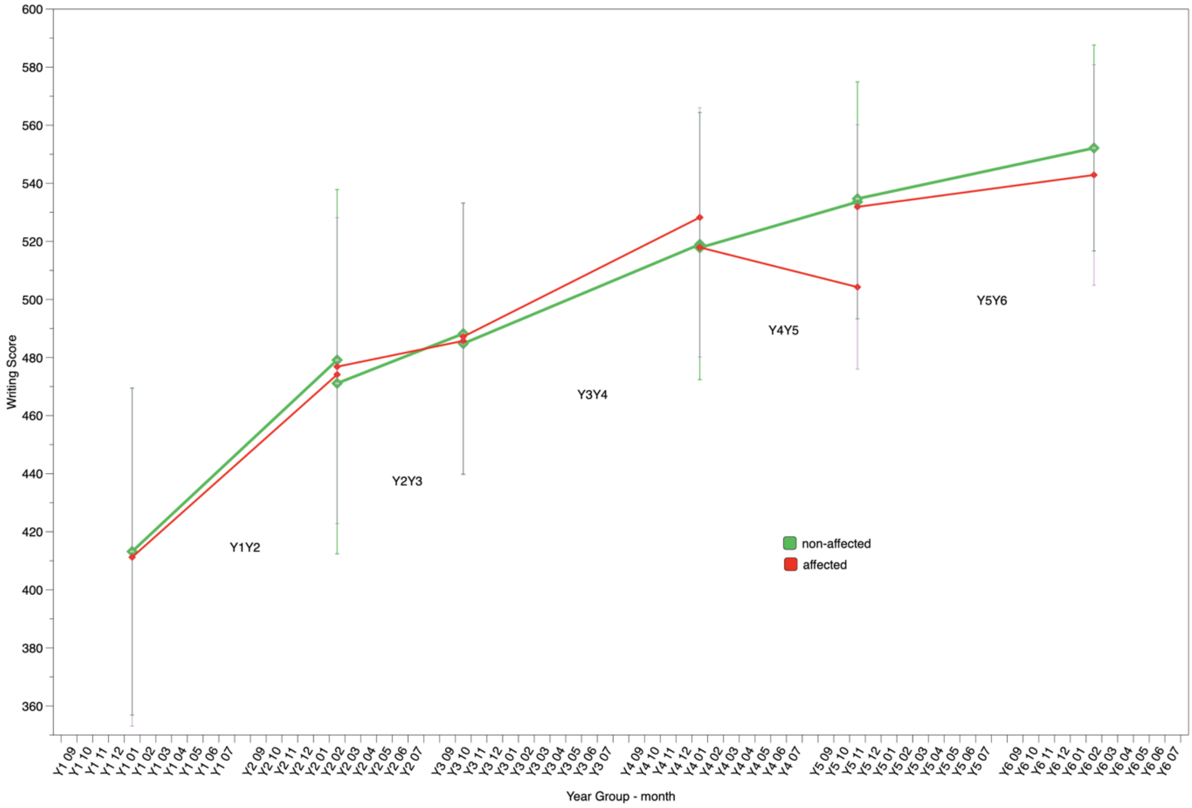

Figure 6.1: Approximation of the developmental trajectories for writing performance scores based on repeated measurements across Age Cohorts and conditions (affected vs non-affected).

In relation to addressing research question 2 (RQ2), we analysed the extent to which the inclusion of selected person- or context-related covariates might contribute to a more differentiated description of the effects of school closures on the development of writing skills. In technical terms, we tested statistically whether systematic variation in the response to school closures (see analyses in relation to RQ1) is moderated by any of the selected covariates. The covariates considered include, pupil premium, sex, school type, and geographic region. The focus of research question 2 is on establishing whether these covariates indicate differential associations to the response to COVID-19-related school closures in terms of progress in writing performance. An appropriate interpretation of potential differential effects has to be based on an appraisal of the extent to which these covariates are systematically linked to inter-individual differences in writing performance in general. This first step reveals an expected result pattern across the analyses related to each of the covariates.

Across Age Cohorts, learners in receipt of pupil premium performed consistently worse than their peers not in receipt of pupil premium. These differences ("pupil premium gap") tend to become smaller with age, starting at 20.4 points difference in year 1 and ending with 13.4 points difference in year 5. In contrast to these general, "pre-existing" differences, the sizes of differential effects that are attributable to pupil premium on the differences in progress between pupils affected vs non-affected by school closures are very small. They range from 1.1 points less progress per month for pupil premium recipients in Y2Y3 to 0.1 points *more* progress per month for pupil premium recipients in Y3Y4. In other words, the impact of by COVID-19 related school closures on progress in writing skills (DiD) does not substantially differ between pupils in receipt of pupil premium and those who are not. Although markedly different in size, the variation of differential effects across Age Cohorts mirror strongly the relative differences of the general effect of the pupil premium disadvantage ($r[B_3, B'_7]_{PupilPremium} = .82$). In other words, the greater the pupil premium gap established at T1 the larger the difference in the relative response to school closures. This association should primarily be seen as an indication of systematicity, which lends the results credibility. But again, these differential pupil premium effects are very small. Across the 3 months of school closures they range from 0.2 points *more* progress for school closure affected pupils in receipt of pupil premium compared to their non-pupil premium peers (Y3Y4) to 3.1 points less progress (Y1Y2).

A similar result pattern was observed for the analyses that included 'Sex' as a covariate. While female pupils across all Age Cohorts performed on average between 20.2 and 22.8 points better than their male peers, the sex-related differences in responding to COVID-19-related school closures on the progress in writing skills, however, range from 0.1 points more progress during these three months for affected male pupils in Y3Y4 (compared to affected female pupils) to 1.1 points differences in favour of females for Y5Y6. Here we also noticed that the differential effects, though minimal in magnitude overall, tend to replicate the sex-related general performance differences across age cohorts ($r[B_3, B'_7]_{Sex} = .55$). Age Cohorts with the largest "sex gap" in general performance tend to also show larger resilience in female pupils towards potential negative effects of COVID-related school closures.

Consistently across Age Cohorts, pupils educated in independent schools showed substantially better performance than pupils educated in state-funded schools. These differences range from 24.6 points (Y5Y6) to 46.7 points (Y2Y3) with a slight tendency for the size of these differences to decrease with age (33.2 points in year 1, reducing to 24.6. points in year 5). As was observed for the other two covariates, the consideration of 'School Type' does add very little to a more differentiated description of the differences in progress between school closure affected and un-affected sub-cohorts. The unique contribution ranges from 0.3 points difference in progress across 3 months (Y3Y4) to 6.5 points difference in progress during the 3 months of school closure (Y2Y3). The differences in size of the differential effects across Age Cohorts mirrors moderately well the differences in size of the general effects ($r[B_3, B'_7]_{SchoolType} = .65$). For Age Cohorts with greater performance gaps between independent and state-funded schools we tend to also observe greater vulnerability towards potential negative effects of COVID-19 related school closures on progress during the 3 months period ranging from 0.3 points to 6.5 points in state-funded schools.

Analyses that differentiate between different geographic regions reveal high levels of consistency in *intra-regional differences* across age groups. That is the increase in average performance with age is consistently observed – albeit at different levels – for each of the 10 regions. The pattern of *inter-regional differences* across Age Cohorts, however, is less consistent. It is important to bear this in mind as inconsistencies in the effect patterns linked to the covariate 'geographic region' make it challenging to pinpoint potential causal mechanisms (i.e., explanations) for the emergence of performance differences to regional differences. Results of our analyses indicate that pupils attending schools in 'Inner London' tend to show very consistently the highest average performance across all Age Cohorts. This is followed by schools in 'Outer London' and 'Yorkshire and The Humber', both, however, with more varied rankings across Age Cohorts. On the opposite side of the performance spectrum, the average performance shown by pupils attending schools in the 'East of England' ranks with highest consistency lowest across Age Cohorts. The two other regions with the poorest average performance in general include the 'North West' and 'South East', both, however, with varied rankings across Age Cohorts. Performance scores range inter-regionally between 14.4 points for year 5 pupils and 30.3 points for year 2 pupils.

As was observed for the other covariates, the unique contribution of 'Geographic Region' to a differential characterisation of effects of COVID-19-related school closures on the progress of writing skill acquisition during primary school is comparatively small. Projected onto a 3-month period (i.e., the duration of school closures), differential effects range from 13.4 points less progress attributable to school closures for Y4Y5 pupils attending schools in the ‘North West’ to 4.3 points *more* progress observed in school closure affected Y2Y3 pupils in ‘Outer London’.

We also observe a strong alignment of the size of differential effects attributable to ‘Geographic Region’ with the size of pre-existing performance differences between regions ($r[B_3, B'_7]_{GeoRegion} = .88$). That is, the lower the average performance shown in general by pupils attending schools in a geographic region (at T1) the higher the proportion of the negative COVID effect that can be attributed to regional differences. Or, in other words, regions with the highest general performance tend to be more resilient towards the potential negative effects of COVID-19-related school closures.

In sum, the analyses in relation to RQ2 indicate that the inclusion of covariates has little to no further “explanatory” potential of contributing to a more differentiated description of the already small effects resulting from the simple difference-in-difference (DiD) analyses conducted in relation to RQ1.

The results of the analyses conducted to address research question 3 confirm the reasonable expectation that inter-individual differences in performance scores when affected by school closures, be it in relation to ‘Pupil Premium’ status, ‘Sex’, ‘School Type’, or ‘Geographic Region’ are associated to a smaller degree with school or teaching group characteristics. The relative importance of pupil characteristics, however, increases under COVID-19 conditions. This result pattern could speculatively be interpreted as a reminder of the importance of helping pupils across all age groups to develop resilience as a pupil attribute.

In conclusion and in an attempt to offer some orientations for how the findings presented in this report could be utilised for policy recommendations, we wish to draw the reader’s attention to a number of points.

When evaluating the overall result pattern, two seemingly contradictory perspectives may be adopted. Firstly, the relatively small score differences obtained across the various analyses could be perceived as an indication of limited practical relevance. Secondly, the consistency of the result pattern across the five age cohorts and different covariates, in conjunction with the substantial size of the near-representative sample, warrants serious consideration of the findings. The relatively small effect sizes could be interpreted as an indication of a positive message, namely that the absence of in-person teaching due to school closures caused by the pandemic did not result in substantial learning losses, on average. At the same time, we need to keep in mind, that the period of school closures analysed was “only” three months. It would be interesting to explore the potential accumulation of effects of further periods of school closures (e.g., January – March 2021) and, maybe even more interestingly, to study differential trajectories of recovery.

Notwithstanding the modest magnitude of the overall and differential COVID effects, it is crucial to be mindful of the fact that the analyses presented here offer a comprehensive and systematic, structural *description* of the observed outcomes. They do not provide *explanations* regarding the reasons or causal mechanisms that led to the observed results. The often-overlooked distinction between description and explanation is particularly important when considering potential interventions or the formulation of policies that inform actions. In other words, devising meaningful interventions requires answers to the *why* question (i.e. explanations) and cannot solely be based on descriptions (which primarily answers the what question). To illustrate, the analyses presented here do not address the question of *why* pupils attending a school located to the west of the Pennines tend to perform at lower levels than pupils in schools situated to the south of the Peak District. If one were to take ‘Geographic

Region' as a proxy of economic disadvantage and were to use this as an implied "explanation" for the observed performance differences, one would need to reconcile that the same level of economic disadvantage (e.g., the 'North East') has detrimental effects on performance in some years (e.g., Y5Y6) but appears to be of benefit in others (e.g., Y1Y2, or Y3Y4). Any implicit assumptions regarding explanatory causal mechanisms for the observed effects (be it intake-, resource- and/or teacher orientated, in the context of 'School Type', for instance) need to be specifically tested using methodological approaches different to the ones realised in the analyses presented in this report.

For an adequate utilisation of the descriptive information provided in this report it is also important to be aware that the selection of covariates included in the current analyses was not conceptually informed. As is the case for most secondary data analyses the selection of covariates was primarily determined by the availability of quantified (or categorised) indicators of attributes of pupils or of the educational environment they are exposed to. In other words, the description offered here cannot claim exhaustiveness either.

In this report, the potential effects of the individual covariates have been analysed in isolation (i.e., separate models for 'Pupil Premium', 'Sex', 'School Type', 'Geographic Region'). Based on the results presented, we cannot conclude, for instance, that a pupil-premium receiving boy in the North-West of England attending a state-funded school is particularly challenged. While these covariates interact (i.e., they are not independent from another) their combined impact is not necessarily simply additive. Take Pupil Premium and Geographic Regions as an example. Methodologically, additivity would require that Pupil Premium was evenly distributed across geographical regions. Conceptually, both covariates ('Pupil Premium' and 'Geographic Region') could be seen as (equally imperfect) proxies for socio-economic disadvantage (one at the pupil level, the other at the school level). Hence, additivity of their effects is not to be expected. A more adequate description of the potential "intersectional" effects of various indicators of economic disadvantage would have required to also include higher order interaction terms in the regression models. This would, however, have resulted in difficult to accommodate levels of complexity in the estimation process.

Finally, in this report, special attention was given to contextualising effects in more concrete and practically tangible terms. For this reason, we refrained from providing "*p*-values" and significance stars, or from discussing standardised β weights in the regressions. Instead, our objective was to facilitate an appreciation of what two scripts might look like that differ by, say, 3 points on a scale with a mean score of 485.7 and a standard deviation of 66.2. As mentioned previously, we consider the systematicity and consistency of the result pattern to be of significant meaning. The intention behind this reporting approach has been to minimise the risk of misinterpretation of findings, which we consider to be particularly important for the utilisation of research evidence informing practice and policy decision-making in education.

7 References

Betthäuser, B.A., Bach-Mortensen, A. M., & Engzell, P. (2023). A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic. *Nature Human Behaviour, 7*, 375-385. https://doi.org/10.1038/s41562-022-01506-4

Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, *39*(3/4), 324-345. https://doi.org/10.2307/2334029

Curcin, M., Howard, E., Sully, K., & Black, B. (2019). Improving awarding: 2018/2019 pilots. Research and analysis (Ofqual). London: Office of Qualifications and Examinations Regulation.

Department for Education (2021). Understanding progress in the 2020/21 academic year: interim findings, January 2021. Renaissance Learning, Education Policy Institute.

Department for Education (2021). Understanding progress in the 2020/21 academic year: Initial findings from the spring term. Renaissance Learning, Education Policy Institute.

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), p.e2022376118.

Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and instruction. *Reading and Writing*, *26*, 1-15. https://doi.org/10.1007/s11145-012-9395-2

Luce, R. D. (1959). *Individual Choice Behaviours: A Theoretical Analysis*. New York: J. Wiley.

McGrane, J. A. (2023). Comparative judgment. In R. J. Tierney, F. Rizvi, & K. Ercikan (Eds.), *International Encyclopedia of Education (Fourth Edition)* (pp. 73-78). Elsevier. https://doi.org/10.1016/B978-0-12-818630-5.09023-0

Pollitt, A. (2012). The method of Adaptive Comparative Judgement. *Assessment in Education: Principles, Policy & Practice*, *19*(3), 281-300. https://doi.org/10.1080/0969594X.2012.665354

Rose, S., Badr, K., Fletcher, L., Paxman, T., Lord, P., Rutt, S., Styles, B., & Twist, L. (2021). *Impact of School Closures and subsequent support strategies on attainment and socio-emotional wellbeing in Key Stage 1*. https://d2tic4wvo1iusb.cloudfront.net/production/documents/pages/projects/Impact-on-KS1-Closures-Report.pdf?v=1706871914

Tadesse, S., & Muluye, W. (2020). The Impact of COVID-19 Pandemic on Education System in Developing Countries: A Review. *Open Journal of Social Sciences, 8*, 159-170. https://doi.org/10.4236/jss.2020.810011

Thurstone, L. L. (1927a). A law of comparative judgment. *Psychological Review, 34*(4), 273-286. https://doi.org/10.1080/0969594X.2012.665354

Thurstone, L. L. (1927b). The method of paired comparisons for social values. *Journal of Abnormal and Social Psychology, 21*, 384-400. https://doi.org/ 10.1037/h0065439

UNESCO. (2023). *UNESCO's education response to COVID-19*. Retrieved January 2024 from https://www.unesco.org/en/covid-19/education-response/initiatives

Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in Education: Principles, Policy & Practice*, *26*(5), 541-562. https://doi.org/10.1080/0969594X.2019.1602027

Wheadon, C., Barmby, P., Christodoulou, D., & Henderson, B. (2019). A comparative judgement approach to the large-scale assessment of primary writing in

England. *Assessment in Education: Principles, Policy & Practice*, *27*(1), 46–64. https://doi.org/10.1080/0969594X.2019.1700212