COVID-19 disruptions
Attainment gaps and primary school responses

Research Report

February 2022

Ben Weidmann
Rebecca Allen
Dave Bibby
Rob Coe
Laura James
Natasha Plaister
Dave Thomson

Please cite as:

Weidmann, B., Allen, R., Bibby, D., Coe, R., James, L., Plaister, N. and Thomson, D., 'COVID-19 disruptions: Attainment gaps and primary school responses', Education Endowment Foundation, 2021.

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education.
Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

This project was run by colleagues at FFT, TeacherTapp, and the EEF. The team consisted of Ben Weidmann, Dave Thomson, Natasha Plaister, Becky Allen, Rob Coe, Dave Bibby, and Laura James.

FFT was responsible for the recruitment of schools, collecting pupil assessment data, and teacher survey data. This was largely led by Laura James. Pupil data collection permission was followed up by Ruth Jameson as part of FFT's School Support Team. Dave Thomson, Natasha Plaister, and Dave Bibby from FFT were responsible for analysing data. Ben Weidmann and Rob Coe wrote the report with support from other FFT colleagues. Becky Allen designed and analysed the teacher surveys. The project was supported by EEF colleagues Jennifer Stevenson, Celeste Cheung, and Jamila Boughelaf.

Acknowledgements

We would like to thank the school support team at FFT for their efforts to keep strong relationships with schools in collecting both assessment and teacher survey data, which enabled analysis to go ahead. We also thank Jennifer Stevenson, Celeste Cheung, and Jamila Boughelaf at the Education Endowment Foundation for their support and guidance throughout the project.

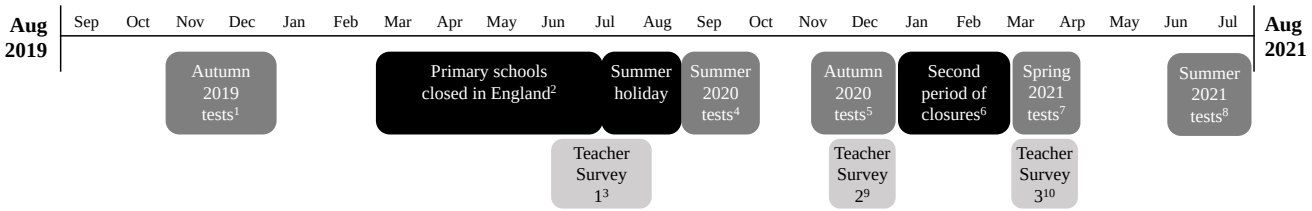
Executive summary

NOTE: this report supersedes an interim report published in May 2021. This report analyses more assessment data (including assessments from spring 2021 and summer 2021) and a third teacher survey.

This descriptive, longitudinal study focuses on primary schools in England. It has three aims:

1. **to quantify changes in attainment gaps since the onset of COVID-19**: our focus is on gaps between economically disadvantaged pupils[1] and their peers in Years 2 to 6;
2. **to describe how teachers and schools responded to the challenges of COVID-19**: we examine three periods—
   a. the 'initial closure period', March to May 2020 when schools were suddenly shifted to remote learning for most pupils;
   b. the 'reopening period' when pupils returned to schools for the start of the 2020/2021 year; and
   c. the 'second closure period' from January to March 2021, when a rise in COVID-19 cases meant that schools were again closed to a majority of students; and
3. **to explore associations between school responses to covid and changes in attainment gaps**.

This is an observational study. The sample sizes differ across research questions and range between 5,000 to 19,000 pupils (in 35 to 117 schools). Our samples are similar to the broader population in terms of prior achievement and levels of disadvantage. The study uses data from five assessment points for reading and maths. We focus on a constant sample of students, all of whom completed a Rising Stars assessment in 2019 prior to the COVID-19 pandemic. The study also administered three teacher surveys: the first focused on the initial period of lockdown, the second focused on the autumn 2020 term when students returned to face-to-face instruction, the third focused on the second period of school closures from January to March 2021. A timeline of our data sources is presented below:



Notes: [1] Autumn 2019 tests were primarily conducted in November (42%) and December (52%). The remainder were almost all in October. [2] Schools were closed on 25 March. During the period from 25 March to 1 June, most schools were only open to a very small group of children. From 1 June 2020, many primary schools reopened for children in reception, Year 1 and Year 6. [3] Teacher Survey 1 was focused on the initial period of partial school closures (March–July 2020) and was administered in July 2020. [4] Summer 2020 tests were conducted in-person when schools reopened in September for the start of the 2020/2021 academic year. [5] Autumn 2020 tests were almost all in either December (65%) or November (32%). [6] The majority of primary schools were closed from 5 January and were reopened on 8 March; during this period most schools were only open to some children.[7] Spring 2021 tests were conducted in March 2021 (90%). [8] Summer 2021 tests were conducted in June (71%) and July (25%) 2021. [9] Teacher Survey 2 was focused on in-person education in the autumn term of 2020/2021 and was administered in Dec 2020. [10] Teacher Survey 3 was focused on the second period of school closures and was administered in March 2021.

Table 1: Summary of study findings

Research question	Finding
How did 'disadvantage gaps'—attainment gaps between FSM6 pupils and their peers—change between autumn 2019 and summer 2021?	We find evidence that disadvantage gaps for maths widened since the onset of the COVID-19 pandemic. Before the pandemic, the gap in our sample for maths was 0.45 ES units. On average, across Years 2 to 6, the maths gap between disadvantaged pupils and their peers widened by an estimated 0.05 ES units (with a plausible range of 0.02 to 0.08). This represents an increase of between 4% to 17% in the pre-COVID-19 disadvantage gap. Our best estimate suggests an increase in the gap for maths of around one month's progress. This widening primarily happened during the initial lockdown.

[1] Defined by having received free school meals status within the previous 6 years (FSM6).

Research question	Finding
	There was no discernible change in the disadvantage gap for reading.
Were there associations between how schools responded to **remote learning** and changes in disadvantage gaps?	We examined five variables describing primary schools' responses to remote learning: 'phoning students', 'timetabling', 'frequency of work submission', 'use of technology platforms', and providing live lessons. We examined these variables in two periods: the initial lockdown (March to May 2020) and the second period of partial school closures (January 2021 to March 2021). We did not find clear associations in either period between these variables and changes in disadvantage gaps.
Were there associations between how schools responded during the **reopening** period (September to December 2020) and changes in the disadvantage gap?	We examined five variables describing primary school practice in the autumn 2020 term: providing videos/live streams for *absent pupils*, providing videos/live streams for absent *classes*, extra learning time, reducing the curriculum, and more small group interventions. Overall, we did not find clear associations between these variables and changes in disadvantage gaps.

Additional findings

- In maths, the strongest evidence for widening disadvantage gaps came in Years 2 and 3. While this echoes findings from other studies suggesting that younger children have been more affected by COVID-19 disruptions (Curriculum Associates, 2020; Juniper Education, 2020; Rising Stars, 2021; Tomasik et al., 2020) we were unable to draw clear conclusions about whether attainment gaps have grown more for younger children in maths, relative to older children.

- Across the three teacher surveys, we observed considerable variation in how schools responded to the challenges of COVID-19. Take, for example, the practice of phoning students: during the initial lockdown (March to May 2020), 23% of teachers reported phoning students at least once a week while 37% never phoned students. Similar variation was observed during the second period of widespread school closures (January to March 2021). During this period, 23% of teachers reported phoning students at least once a week while 48% never phoned students (and 30% of teachers phoned students once or twice during the half-term).

- We asked teachers very similar questions during two periods of school closures (March to May 2020 and January to March 2021). In the second period of partial school closures, we found that teachers were doing *more* of the activities we asked about. For example, during this second period, 76% of teachers asked that students submit work daily (or more often) compared to 31% in the first lockdown. The second period of closures also saw increased rates of online platform use, live lessons, and timetabling.

- We did not find clear evidence of associations between school approaches to remote learning and relative changes in attainment. We found no clear associations between changes in attainment and the variables 'phoning students', 'timetabling', 'live lessons', 'frequency of work submission', or 'use of technology platforms'.

- Results exploring the association between school responses and changes in attainment gaps should be interpreted with caution as the analysis was not supported by any information about *how* different responses were implemented. Moreover, we could only analyse data at the school level. This level captured some, but certainly not all, of the variation in teacher responses.

Introduction

Background and motivation

This section provides the background to, and motivation for, our study, which has three parts.

- In part 1, we **quantify changes in attainment gaps since the onset of COVID-19**. We focus on gaps between economically disadvantaged pupils and their peers. As our measure of economic disadvantage is free school meal status in the past six years ('FSM6'), we refer to these inequalities in educational attainment as the 'disadvantage gap'.
- In part 2, we **describe how teachers and schools responded to COVID-19 disruptions**. There are three periods of interest:
    - the 'initial closure period' from March to May 2020 when schools were suddenly required to shift to remote learning;
    - the 'reopening period' when pupils returned to schools for the start of the 2020/2021 academic year; and
    - the 'second closure period' from January to March 2021, when a rise in COVID-19 cases meant that schools were again closed to a majority of students.
- Part 3 combines parts 1 and 2. We **explore associations between school responses and changes in disadvantage gaps in attainment**.

1. Measuring changes in educational inequalities during COVID-19

COVID-19 caused unprecedented disruptions to schooling. Existing research has shown how disruptions exacerbated inequalities in access to education. During the first period of school closures (March to May 2020) data from England suggests that compared to their more affluent peers, children in low income households spent less time on education, had less parental support in terms of time and expertise, received fewer paid-for educational services, and had more problems with access to devices and the internet (Andrew et al., 2020; Cullinane & Montacute, 2020; Hupkau & Petrongolo, 2020). Research from the United States illustrated how students in economically disadvantaged households engaged less with online educational platforms (Bacher-Hicks et al., 2020; Chetty et al., 2020).

A growing set of studies has focused on how inequalities in educational inputs translate into attainment gaps. We build on this literature by estimating changes in attainment gaps between disadvantaged students and their peers in Years 2 to 6. Existing literature has tended to show that gaps in attainment have widened since the onset of the pandemic. We review specific findings below. First, however, we note that the literature faces several challenges in providing estimates of how COVID-19 has impacted disadvantage gaps in attainment. These challenges include:

- **Making comparisons across time (pre- and post-COVID-19) using a longitudinal sample, or comparable samples**. COVID-19 disruptions have led to large changes in the composition of students who take assessments, compared to those who do not sit tests. Many in-person tests have been cancelled or heavily disrupted. Research has shown that attrition from online assessments has been high, and may differ systematically across socioeconomic groups (Domingue et al., 2021). In making comparisons across time—especially in examining whether (and by how much) disadvantage gaps in attainment are changing—it is important to compare apples with apples. The gap between FSM pupils and their peers can be influenced by the composition of the sample at different time points.
- **Using standardised measures of attainment, administered in a consistent way across time**. Ideally, attainment data should be standardised rather than based on teacher assessments. To get a clear picture of how the magnitude of underlying inequalities may be changing, tests should be conducted in the same way (preferably in person) across different time periods. If, for example, testing conditions after the onset of COVID-19 became systematically more difficult for economically disadvantaged students, which is especially plausible for tests administered at home, then resulting estimates of changes in gaps may not reflect underlying differences in learning.
- **Using individual-level indicators of disadvantage**. Quantifying attainment gaps between disadvantaged pupils and their peers using school-level or regional-level measures of disadvantage

provides indirect estimates. In many cases these estimates will understate the magnitude of gaps. As an example, consider an analysis comparing the attainment of pupils in 'low-FSM' schools (defined as those with a below-average percentage of FSM pupils) with attainment in 'high-FSM' schools (defined as having an above-average percentage of FSM pupils). In the latter, it may be the case that a majority of the pupils may not be from disadvantaged backgrounds. This will tend to mask any gaps.

- **Providing uncertainty estimates for how gaps have changed**. As most system-wide tests were disrupted by COVID-19 (eliminating the possibility of population-level estimates), efforts to quantify the way in which disadvantage gaps have changed are subject to sampling uncertainty. In some analyses, uncertainties are not reported. This is often the case when analyses are focused on 'learning loss' and not on directly estimating attainment gaps.[2] It is also important to distinguish between estimating an attainment gap and estimating a change in that gap before and after COVID-19. Disadvantage gaps will generally have existed before COVID-19 and the change may be small relative to the precision of a sample estimate.

Our study, described below, aims to address these difficulties. We follow a sample of students attending schools that provided in-person tests at five time-points: November/December 2019 (pre-COVID-19 baseline: every student in our sample has a baseline measure of attainment); September 2020 (when schools returned to in-person instruction); December 2020 (after most children had been back at school for a full term); March 2021 (after a second period affected by school closures); and finally July 2021 (at the end of the 2020/2021 school year). Our analytical approach explicitly focuses on estimating gaps and the associated uncertainties. Using this data, our first contribution is to provide direct estimates of changes in the disadvantage gap for primary reading and maths (Years 2 to 6). We now turn to reviewing existing estimates.

Several studies in England use standardised assessments to provide direct evidence about changes in attainment gaps due to economic disadvantage.[3] The two strongest studies are from the Department for Education and Rising Stars. The Department for Education examines attainment in reading and maths for both primary and secondary students (DfE, 2021). Rather than estimating attainment gaps directly, the study models counterfactual outcomes to calculate 'learning loss' due to COVID-19 and then compares levels of lost learning for FSM pupils and their peers. The authors estimate that between autumn 2019 and autumn 2020 the disadvantage gap widened by 0.4 months for reading, and one month for maths.[4] Rising Stars (2020) uses a more straightforward method to estimate disadvantage gaps in reading and maths. For each year group (1 to 6) the authors calculate the gap in mean attainment for autumn 2019 and autumn 2020 and examine the change. Averaging across Years 1 to 6, the authors find that gaps widened by 0.05 ES units for reading and 0.06 for maths.[5] A third study, by NFER, estimated the disadvantage gap in Year 2 for reading and maths in autumn 2020 but was unable to calculate a comparable figure for the pre-COVID-19 period due to lack of information about pupils' FSM status (NFER, 2021).[6] Overall, the most relevant evidence from the U.K. suggests that disadvantage gaps grew after the initial period of school closures in 2020.

Moving beyond the U.K., several U.S. studies provide direct evidence about the extent to which disadvantage gaps have grown. Sass and Goldring (2021), for example, use a similar approach to the DfE (2021) to analyse data in Atlanta for students in Grades 4 to 8. The authors find a widening of disadvantage gaps (defined in terms of free or reduced-price meal eligibility) of 1.7 months for reading and one month for maths.[7] Pier et al. (2021) use the same approach to analyse data in California and also report evidence of widening gaps in

[2] For a systematic review of studies focused on learning loss, see Hammerstein et al. (2021).
[3] Juniper Education (2020) also provides direct evidence of changes in the disadvantage gap using teacher assessments for Years 1 to 6. A clear strength of the Juniper study is that the data covers three periods: autumn 2019, summer 2020, and autumn 2020. The disadvantage gap is defined in terms of the percentage of students who were assessed by their teacher to be achieving at expectation. The gap grew from autumn 2019 to summer 2020 by around 4.8ppt in maths and reading. This widening was reduced by roughly half in the autumn term.
[4] These figures represent averages of the 'autumn 1' and 'autumn 2' tests. To contextualise this change, we note that Thomson (2021) reports that the disadvantage gap is 0.5 ES units. Using the EEF conversion to months' progress, this amounts to six months progress.
[5] We thank Katie Blainey from RS Assessments for providing the underlying data to Figures 7 and 8 so that these numbers could be calculated. Note that the follow-up report from RS Assessments examining the spring 2021 term does not include analysis of disadvantage gaps at the pupil level (Rising Stars, 2021).
[6] While they are unable to provide any direct evidence, the authors compare disadvantage gaps in their autumn 2020 study sample to national estimates of the FSM from Key Stage 1 assessments in 2019, concluding that '[i]t seems that the disadvantage gap is wider than earlier estimates' (NFER, 2021, p. 12).
[7] This figures are averaged across Grades 4 to 8.

elementary/middle school for reading (two months) and maths (0.7 months).[8] Kogan and Lavertu (2021) analyse reading attainment in a state-wide assessment in Ohio and find that economic attainment gaps grew by 0.05 ES units from autumn 2019 to autumn 2020.[9] As the study authors note, however, this is likely an underestimate given the high levels of measurement error in the pupil-level disadvantage indicator, whereby some schools/districts are known to report all students as disadvantaged.

Engzell et al. (2021) examine attainment gaps in the Netherlands between students whose parents have differing levels of education.[10] This study examines attainment from national assessments immediately before COVID-19-related closures (January/February 2020) and soon after schools returned (May/June 2020). Focusing on a composite measure of attainment in Years 4 to 7, the paper finds an increase in gaps of around 1.2 percentile points (roughly 0.03 ES units, ± 0.007).[11]

Finally, a broader set of studies provides indirect evidence about the changes in economic attainment gaps by comparing how COVID-19 has affected average levels of attainment in schools (or neighbourhoods) that have high or low levels of disadvantaged pupils. In the U.K., Rising Stars (2021) finds that schools with a higher percentage of FSM children experienced higher average levels of 'learning loss'. In the U.S., Renaissance (2020) found that schools with higher levels of disadvantaged pupils had greater levels of 'learning loss' compared to other schools (two percentile points greater loss in maths, one percentile point in reading).[12] Similarly, Curriculum Associates (2020) compares estimates of average learning loss in schools from low- or higher-income neighbourhoods. Across Grades 1 to 8 they find a small widening of economic attainment gaps for maths, but a small narrowing of gaps for reading.[13] In Belgium, Maldonado and De Witte (2020) report that schools with higher percentages of students receiving financial support had greater learning losses in maths and Dutch.[14] In Australia, Gore et al. (2021) analyse the effect of COVID-19 on maths results in schools with 'low', 'medium' and 'high' average levels of disadvantage, for Years 3 and 4. The authors found no change in disadvantage gaps in Year 4, but a widening of gaps (two months) between 'low' and 'medium' disadvantage schools.

To summarise, evidence suggests that COVID-19 led to widening inequalities across a range of educational inputs. Schools and neighbourhoods with higher percentages of economically disadvantaged students have tended to be disproportionately affected by COVID-19 in terms of average estimated learning loss. At the pupil level, direct estimates generally show that gaps have widened for primary school students—although the magnitude of gap widening has been somewhat less than our median expectation at the onset of the pandemic (EEF, 2020). The most relevant evidence for this study (DfE, 2021; Rising Stars, 2020) suggests that gaps have grown by around a half to one month's progress.

2. Documenting school responses and their impacts

Our second research focus is to document how schools and teachers responded to the huge disruptions caused by COVID-19. Here, we build on several studies that have surveyed teachers and asked them about their experiences of lockdown and the unexpected closure of schools. Of particular relevance for our report is evidence about how variation in school-level practices differed in terms of the economic circumstances of their pupils. For example, Cullinane and Montacute (2020) reported early in the initial lockdown that pupils from middle class homes were almost twice as likely to be taking part in live or recorded lessons, compared to working class peers (30% compared to 16%, see also Allen et al., 2020 for evidence from later in the initial lockdown). Similarly, schools with more affluent students were more likely to report using an online platform to help set and receive work (Cullinane & Montacute, 2020).

[8] Elementary/middle school in the data covers Grades 4 to 8. Note that Pier et al. (2021) report a substantial *narrowing* of gaps in high school—although this is beyond the scope of our paper. The outcomes are MAP and STAR assessments.
[9] As the study authors note, this is likely an underestimate given the high levels of measurement error in the pupil-level disadvantage indicator as some schools/districts are known to report all students as disadvantaged.
[10] A limitation of this high quality study is the measure of disadvantage. Household are divided into three categories in terms of parental education: 92% of households fall into the 'high' category, which makes this quite a restrictive measure.
[11] This assumes a normal distribution.
[12] The comparison was between Title 1 schools, that have above-average levels of economically disadvantaged students, and others. This study examined Grades 1 to 8.
[13] The comparison is between autumn 2020 and an average of previous years (autumn 2017, 2018, and 2019). The outcome is 'percentage of students who are 2 or more grades below expectation' on the i-ready diagnostic test.
[14] Other measures of disadvantage such as 'percentage of students who come from low SES neighbourhoods' and 'percentage of students with mothers who have low levels of education' found mixed results.

In May 2020, NFER sent a survey to all primary and secondary schools in England. Around 9% of primary schools responded (Lucas et al. 2020).[15] The survey suggested that schools with higher levels of disadvantaged children had less regular contact with pupils: teachers in the most deprived schools were in regular contact with 50% of their pupils compared to 67% for the least disadvantaged schools. Similarly, senior leaders in the most disadvantaged schools were more likely to report using workbooks or worksheets and significantly less likely to be using live or pre-recorded video lessons.

Our second contribution is to present more detailed evidence about variation in school responses during the initial lockdown (March to May 2020). We also extend existing literature by providing descriptions of how school responses varied when schools reopened to all students in the autumn term (September-December 2020), as well as describing school responses in the second period of widespread school closures (Jan-March 2021). Moreover, by repeating many of the questions from our first survey in the second period of closures, we are able to document how remote learning practices changed as the pandemic evolved.

3. Linking school responses to changes in educational inequalities

The final part of our study examines associations between school responses to COVID-19 disruptions (both during periods of school closures and reopening) and disadvantage gaps in attainment. This builds on a small body of related literature. Lucas et al. (2020) use teacher surveys to report associations between schools' use of a virtual learning environment (VLE) and maintaining a broad curriculum with higher pupil engagement. Andrew et al. (2020), using parent surveys of time-use, report a positive association between the use of 'active school resources' (online classes, video conferencing, and online chat) and more time spent on learning activities after controlling for a range of family socioeconomic factors and home resources. While these studies provide useful evidence, they rely on behavioural proxies for learning (for example, time spent) rather than standardised assessments. In addition to being proxies, these measures are often based on teacher or parent reports, which may be confounded by the fact that the people who are reporting on the behaviours are often responsible for bringing these behaviours about.

Clark et al. (2021) address this issue by analysing exam board results for a sample of ninth grade students in China.[16] The study examined 1,800 students in three schools which pursued two different strategies during a seven-week period of lockdown from February to April 2020. Schools 'B' and 'C' both provided access to an online learning platform that provided students with video lessons, allowed teachers to communicate with students, and gave teachers a means of setting, receiving, and returning assignments. School A provided none of these services and was analysed as a control group. Using a difference-in-difference strategy, the authors estimate that the provision of online services was associated with an increase in exam performance of 0.22 Effect Size units. This study provides valuable initial evidence about the association between school responses and student learning. That said, it has three clear limitations from our perspective. First, it has a very small sample at the school level (three schools in total). Second, the analysis is only able to examine school responses as a set: schools either provided a whole suite of services, or none at all. Last, a lack of information about disadvantage means that the study cannot examine the impact of school response on disadvantage gaps.

Our third contribution is to provide novel estimates of the association between multiple school practices and inequalities in attainment in primary schools during COVID-19.

Research questions

We have three research aims, each of which addresses a general question that is addressed specifically for the different phases of the COVID-19 pandemic.

1. Quantify changes in inequalities since the onset of COVID-19[17]

RQ1: How did the attainment gap between disadvantaged pupils and their peers change over each phase of the pandemic? Specifically,

[15] Results were reweighted so that the survey sample reflected the population in terms of observed characteristics.
[16] Students are roughly 15 years old.
[17] The study plan contains two additional questions, not addressed here, focusing on attainment in spring 2021 and summer 2021. If these assessments go ahead, we will provide addenda as per the study plan.

(a) during the initial closure period—as assessed at November/December 2019 and September 2020?
(b) during the reopening period—September 2020 to November/December 2020?
(c) during the second closure period—November/December 2020 to March 2021?
(d) after the end of the second closure—March 2021 to June 2021?

2. Describe how schools responded to COVID-19-related disruptions and their consequences

RQ2: How did schools respond to support pupils' learning during each phase of the pandemic? Specifically,
(a) during the initial closure period—March to May 2020?
(b) during the reopening period—September to December 2020?
(c) during the second closure period—January to March 2021?

3. Explore associations between disadvantage attainment gaps and school responses

RQ3: What was the association between a school's response to supporting learning and the change at school level in the disadvantage gap over that same period? Specifically,
(a) during the initial closure period—November/December 2019 to September 2020?
(b) during the reopening period—September 2020 to November/December 2020?
(c) during the second closure period—November/December 2020 to March 2021?

The analysis we present here was pre-specified in a study plan available on the EEF website.[18] Deviations from (and extensions to) the plan are summarised in the section on statistical analyses below.

Ethics and data protection

FFT was responsible for all communications with schools, data collection, and analysis and the study was conducted in accordance with its research ethics and data protection principles. Further details can be found in Appendix A: (p.39).

Due to the very tight timelines involved in setting up this project—and the limits on data permissions—the data for this study will not be available for archiving.

18 https://educationendowmentfoundation.org.uk/public/files/Projects/EEF_COVID_Partial_Closure_Impact_Response_4.1_clean.pdf

Methods

Study design

This is an observational study. The primary outcome is pupil attainment in maths and reading captured at four different time points during the COVID-19 pandemic between November 2019 and June 2021. We used the PIRA, PUMA, and NTS assessments from Rising Stars for pupils in Year 1 to Year 6 from an opportunity sample of 120 primary schools in England that were already using these assessments. Our focus is on the attainment gap between disadvantaged pupils and their peers, and how that gap changes over the different periods, in both maths and reading. For our purposes, disadvantaged pupils are those who had been in receipt of free school meals (determined by family income) at any point in the six years prior to the study, approximately 30% of the primary school population at that time.

We collected questionnaire responses from a sample of teachers in the same schools about the different ways they had responded to the challenges of providing learning support at three time points: during the periods when the majority of pupils were not in school (March to May 2020), were back in school (September to December 2020), and were out again (January to March 2021).

By matching the teacher questionnaires and pupil assessments at school level for each period, we were able to investigate associations between the different ways schools responded to the pandemic and the changes in their disadvantage attainment gap.

Participants

This section describes our recruitment strategy and the analysis samples we use to address each research question.

School recruitment

Our sample contains state-funded primary schools in England. School recruitment ran from May 2020 to July 2020 and involved two phases. First we used a list provided by Rising Stars (RS) to identify schools that had administered RS assessments in reading and/or maths in Year Groups 1, 2, 3, 4, or 5 in autumn 2019.[19] Second, we contacted these schools and asked if they would be willing to:

1. administer tests in September 2020 and November/December 2020;
2. allow us to collect test data from RS; and
3. allow us to collect pupil context data from their school management information system (MIS).[20]

One hundred and forty-five schools volunteered and were offered a financial incentive to participate in the study[21] along with a tailored report comparing attainment changes in their school with the broader research sample (see Appendix G: Example School Report). Eight schools withdrew before the 2020/2021 academic year[22] and five schools failed to provide data. In total, 132 schools supplied some data for the project. The samples we use to answer specific research questions are described later in this section (see Figures 2 to 5).

Teacher recruitment

We recruited a sample of teachers from within the set of 145 volunteer schools. In July 2020 we contacted volunteer schools and asked them to provide email addresses for up to three teachers in each relevant year group (Years 1 to 5 in 2019/2020). We received 950 email addresses from 135 schools.

[19] Our original study plan stated that we would be examining years 2-5 (in Autumn 2019). We originally did not include year 1 as we thought that the sample size of schools administering assessments in year 1 would be much smaller than other year groups. However, when we received the data from schools it became clear that we had as much data for year 1 as we had for year 2. We therefore decided to include year 1 in an effort to provide a fuller picture of achievement across primary school.
[20] Recruitment materials are presented in Appendix B:.
[21] 250 pounds to be paid in January 2020, and a further 100 pounds, paid at the conclusion of the study.
[22] The reasons for drop out were due to schools not planning to use Rising Stars tests and a misunderstanding from the original communication.

Shortly after assembling the list of teacher email addresses, we disseminated Teacher Survey 1 using a Google Form. The questions focused on the details of the teacher's remote learning provision during March to May 2020 (see Appendix D: Teacher Survey 1). In total, 539 teachers from 137 primary schools responded to the survey.

The same set of teachers were sent Teacher Survey 2 at the end of December, 2020. The questions for Teacher Survey 2 focused on how schools responded to the challenges of returning to face-to-face instruction while COVID-19 continued to cause disruptions (see Appendix E: Teacher Survey 2). In total, 454 teachers from 93 schools responded to the survey.

The same set of teachers were contacted again in March 2021 for Teacher Survey 3. This was not part of the original study plan. The extra teacher survey was added in response to the (partial) closure of schools in January 2021, an event we had not foreseen when setting the study up in mid-2020. The questions for Teacher Survey 3 focused on the details of the teacher's second period of remote learning provision (January to March 2021, see Appendix F: Teacher Survey 3). In total, 343 teachers from 55 primary schools responded to the survey.

Table 2 illustrates the number of responses per school for each survey. In some schools, we were able to get a relatively large number of responses: for example, in 17 schools we had eight or more teachers respond to Teacher Survey 1. Our approach here was to maximise information, that is, to get responses from as many teachers as possible in our sample of recruited schools (while trying to complete recruitment and surveying as quickly as possible).

Table 2: Number of teacher survey responses per school

Number of teacher responses per school	Survey 1 count of schools	Survey 2 count of schools	Survey 3 count of schools
1	39	16	3
2	19	10	6
3	22	11	5
4	9	18	9
5	15	9	9
6	12	7	3
7	4	6	2
8 or more	17	16	18
Total	**137**	**93**	**55**

Overview of analysis samples

Analysis sample for Research Question 1

To address RQ1, the composition of our analysis sample ideally needed to be constant at each of the five assessment points. If the composition of the sample were allowed to change over time, then any changes in estimates of the disadvantage attainment gap could be due to sample compositional changes rather than shifts in the underlying attainment gap between FSM6 pupils and their peers.

Our starting sample was defined by pupils who:

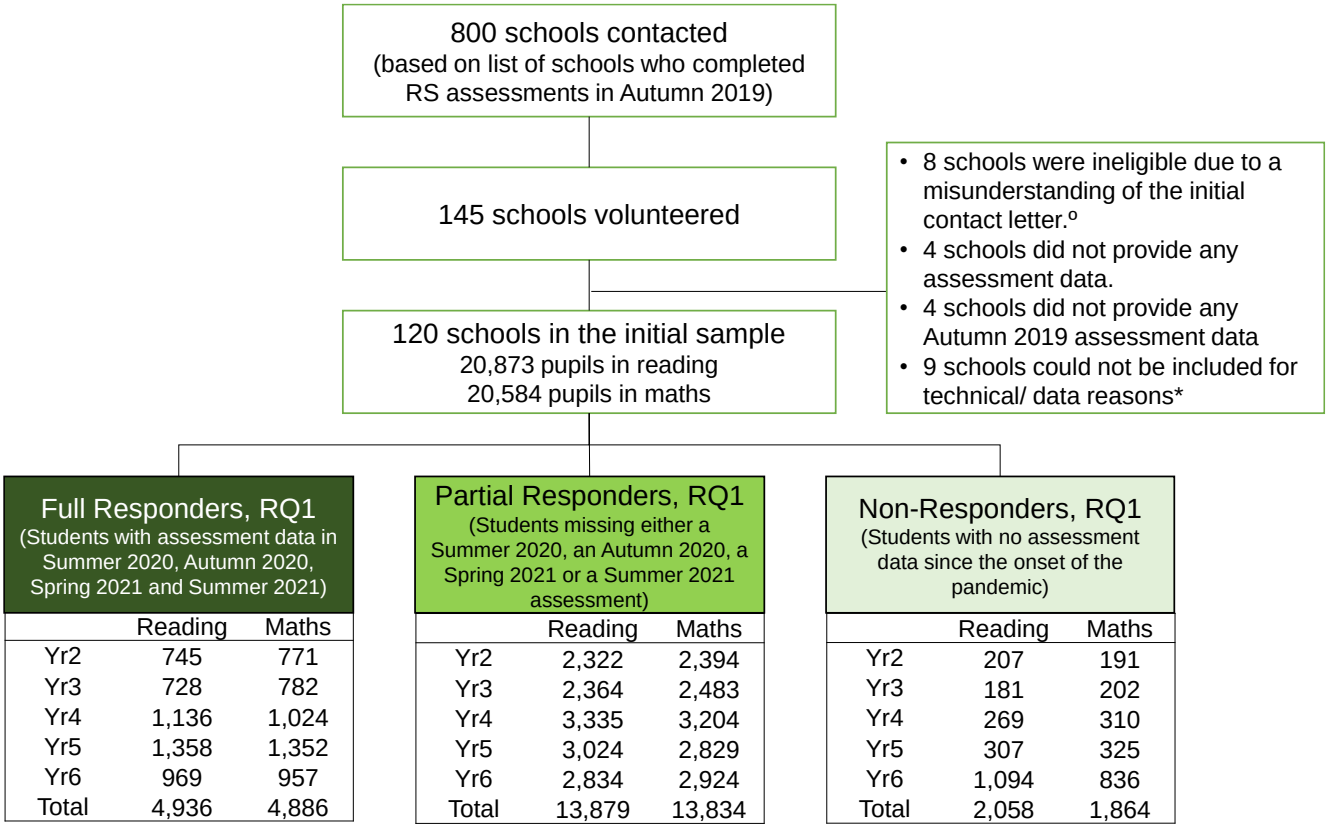
- had a valid assessment in reading or maths in autumn 2019; and
- were on roll during the autumn term 2020 and had a disadvantage flag (FSM6, yes/no) present.

One hundred and twenty of the initial set of 145 schools were included in the starting sample; 20,873 students who were part of these were included for reading, and 20,584 for maths. In autumn 2020, many pupils faced disruptions, including to their reading and maths assessments. This is discussed in the section on missing data. Here we note that, for the purposes of RQ1, we specified three categories of pupil:

- full responders—those for whom a result is observed for all four post-covid assessments: autumn 2019, summer 2020, autumn 2020, spring 2021, and summer 2021;
- partial responders—those for whom at least one assessment is missing (from summer 2020, autumn 2020, spring 2021 and summer 2021); and
- non-responders—those for whom we have baseline data but no test results since the onset of COVID-19.

These samples are summarised in Figure 1.

Figure 1: Overview of sample for RQ1



	Reading	Maths
Yr2	745	771
Yr3	728	782
Yr4	1,136	1,024
Yr5	1,358	1,352
Yr6	969	957
Total	4,936	4,886

	Reading	Maths
Yr2	2,322	2,394
Yr3	2,364	2,483
Yr4	3,335	3,204
Yr5	3,024	2,829
Yr6	2,834	2,924
Total	13,879	13,834

	Reading	Maths
Yr2	207	191
Yr3	181	202
Yr4	269	310
Yr5	307	325
Yr6	1,094	836
Total	2,058	1,864

º Misunderstandings were related to (a) not having PIRA/PUMA/NTS data from autumn 2019 or (b) schools that signed up, but that were intending to switch to PIXL assessments. Technical/data reasons are (a) not providing contextual data about pupils (b) not providing a column of data indicating which pupils are disadvantaged, and (c) not being able to link the test data to the contextual data due to lack of (or mismatching) identifiers.

In this report, our primary analysis for RQ1 focuses on the sample of 'Full Responders + Partial Responders'. We did not analyse data from students for whom we have no assessment data after the onset of COVID-19. To ensure that the sample remains constant across time periods, we impute outcomes for partial responders as described in the Missing Data section.[23]

Analysis sample for Research Question 2

We used all available responses from Teacher Surveys 1, 2, and 3 to address RQ2. Each survey asked different questions, although there was considerable overlap between Teacher Surveys 1 and 3.

Starting from a possible sample of 145 schools, we received data from 137 schools in Teacher Survey 1 ($n_{teacher}$ = 539), 93 schools in Teacher survey 2 ($n_{teacher}$ = 454), and 55 schools for Teacher survey 3 ($n_{teacher}$ = 343). Unfortunately, we were unable to identify teachers and link their responses to the different surveys.

Analysis sample for Research Question 3

Addressing Research Question 3 required overlap between assessment data and teacher survey data. By 'overlap' we mean that we needed teacher survey data from the schools in which we had assessment data. The numbers of schools, pupils, and teachers in our analysis sample for each of the three phases are shown in Table

[23] Note that in the **interim report**, our primary analysis focused on Full Responders only. The reason for the change is that this report considers four post-Covid-19 assessment periods (compared to two in the interim report) and there were widespread disruptions to the spring 2021 and summer 2021 assessments. These disruptions meant that focusing on the Full Responder sample would ignore a large amount of data. To check the robustness of our results to this analytical decision, the section on Missing Data presents the results of the analysis when only the Full Responders are considered.

3. A full participant flow-chart of the starting points and reasons for attrition of the various elements of each sample can be found in Appendix H.

Table 3: Analysis samples for RQ3

	RQ3a: initial closure		RQ3b: reopening		RQ3c: second closure	
Baseline attainment	Autumn 2019		Summer 2020		Autumn 2020	
Outcome attainment	Summer 2020		Autumn 2020		Spring 2021	
Teacher survey	Teacher Survey 1, Jul 2020		Teacher Survey 2, Dec 2020		Teacher Survey 3, Mar 2021	
	Maths	Reading	Maths	Reading	Maths	Reading
Number of schools	90	88	58	58	44	35
Number of pupils	12,564	13,261	7,312	7,743	6,104	5,028
Number of teachers	355	356	290	284	270	216

Comparing analysis samples to broader samples and populations

We were able to compare our achieved samples with national data in two main ways: first, by comparing school-level characteristics for with national data on variables such as region, school size, attainment, percentage FSM6, etc; second, by comparing teacher responses to our survey for RQ2 with a large sample of teachers using TeacherTapp. We were also able to compare the full RQ2 sample with the RQ3 analysis sample, which was limited to teachers in schools where we also had attainment data.

For the first comparison, we compared each of our seven achieved samples (for: RQ1; RQ2a, b, and c; RQ3a, b, and c) with national data for all state-funded primary schools in England, on the following variables:

- % junior schools;
- % London schools;
- % northern schools;
- mean cohort size (pupils);
- mean prior attainment (KS1 average points);
- % disadvantaged pupils;
- % first language other than English;
- % expected standard in read/write/maths; and
- % higher standard in read/write/maths.

Our samples are close to the national data on most measures (for full details see Appendix I). The biggest differences are that we have slightly lower numbers of northern schools (in our samples ranging from 22%–27%, vs 32% nationally), slightly more disadvantaged pupils (33–35% vs 30%), slightly more with first language other than English (26–35% vs 21%), and slightly fewer pupils judged to be at the expected standard (62–63% vs 66%). All these differences seem acceptably small and on all the other variables the differences are even smaller. Hence, we are confident our analysis samples are similar to the population on these dimensions.

For the second comparison, we were able to take advantage of the fact that several of our Teacher Survey 2 questions were also asked to a wider sample of teachers in England via TeacherTapp, a daily survey app. There were six common questions, with 20 response options between them. None had mean differences of greater than 15 percentage points.[24] The question with the biggest difference was the percentage of teachers who reported spending no extra time on English and maths (38% in our sample, 53% in the broader TeacherTapp sample), and there was a corresponding tendency for teachers in our sample to report more extra time being devoted to English and maths than in the wider TeacherTapp sample. However, overall, the similarity between the two samples gives us some confidence that our sample was not unusual in terms of teachers' response to COVID-19-related disruptions.

[24] The full comparison is shown in Appendix I. More details on the Teacher Surveys are provided in the section on Measures and Appendices D, E, and F.

Finally, the comparison between the responses of all teachers who returned our surveys (as analysed in RQ2) and those we were able to match with school-level attainment data (as analysed in RQ3) shows that this restriction of the latter group did not change their observed characteristics very much. Again, the full comparison is shown in Appendix I.

Measures

Attainment data

All schools in the sample used tests in reading and/or maths provided by RS Assessment.[25] Three tests were used: Progress in Reading Assessment (PIRA), Progress in Understanding Mathematics Assessment (PUMA), and NTS Assessments (national test style) reading and maths papers.[26] Schools in the sample therefore had the choice of three types of standardised test in reading and three types of standardised test in maths. In reading the tests were:

- PIRA 2016;
- PIRA 2021; or
- NTS Reading.

In maths the tests were:

- PUMA 2016;
- PUMA 2021; or
- NTS Maths.

The tests take around 45–50 minutes and provide coverage of the revised national curriculum for each year group. The tests can be taken online although, according to the test provider, the majority were taken using pen and paper. Similar to the arrangements for Key Stage 2 tests, all pupils in a year group are typically expected to take the test, with the exception of any pupils who were absent or were unable to access the test for another reason. Scripts were marked by teachers using published mark schemes and the scores for each pupil were entered into MARK, an online reporting and analysis tool provided by RS Assessment, or Aspire Pupil Tracking, a reporting tool provided by FFT.

The technical manuals report reliability coefficients (Cronbach's Alpha) around 0.9 for all the tests we analyse (McCarty & Cooke, 2015; McCarty & Ruttle, 2016).[27]

Standardising attainment data

Although all Rising Stars tests report standardised scores, the tests were standardised at different times, with different norming samples. Our analysis approach required us to be able to treat scores on different versions of the test within each subject as interchangeable. In order to mitigate any possible dependence of our results on the comparability of these standardisations, we considered three different approaches to equating scores and compared them as part of our robustness checking. The three approaches were:

- linear equating—scaling to ensure each test has the same mean and standard deviation; in accordance with the study plan, this is our preferred approach;
- equipercentile equating scaling based on ranks, within each combination of:
    - test (for example, PIRA16);
    - time period (for example, autumn19); and
    - year group (for example, Year 2); and
- scaling based on ranks, within each combination of: test-time-year group *and* school.

[25] **https://www.risingstars-uk.com/rs-assessment**

[26] PIRA, PUMA, and NTS are widely used termly assessments used by over 6,000 primary schools in England. In autumn 2020, RS Assessment introduced a new suite of PIRA and PUMA tests known as PIRA 2021 and PUMA 2021.

[27] During the standardisation process, the PIRA tests were correlated with teacher assessments. Correlations were in the range 0.72 to 0.79. In PUMA, the correlation between the Summer Year 6 test and the national Key Stage 2 maths test was 0.83.

Technical details of how each approach was implemented can be found in Appendix J.

School responses: teacher surveys

Teacher Survey 1

Questions about the initial lockdown were based on the idea that several factors may influence the effectiveness of remote education and that schools may make different choices about how to respond to the pandemic. Some choices were appreciably more expensive or difficult than others, so it seemed valuable to know whether there were associations between choices and pupil outcomes. We asked a total of 16 multiple-choice questions, summarised below. The full questions, along with the response options are in Appendix D.

Overview of Teacher Survey 1 questions

1. How was work set (for example, website download, online learning platform, through live video lesson)?
2. How did pupils send completed work back to you (for example, email, online platform, Facebook)?
3. How far was the school's pre-existing curriculum reduced, paused, or stopped?
4. Which of the following resources were used as part of the home learning you set for students (for example, subscription site, free online resources, Oak National Academy)?
5. To what extent were literacy or English resources adapted or created for parents to use?
6. To what extent could literacy or English resources be completed by children without supervision from their parents?
7. Did you have any form of daily registration for students learning at home?
8. How much of a structured daily timetable did you use (for example, required, suggested, not used)?
9. During the first half of summer term, could pupils: take part in live lessons (with or without talking); take part in live or chat-based social interaction; watch a pre-recorded video of their teacher; none of these?
10. How often did children (that is, not the parents) typically speak to you (their class teacher) on the phone?
11. Did you share examples of student work somewhere for all students to see (for example, in newsletter, on website, email, social media, or not)?
12. What kind of feedback did you give on pieces of work submitted by pupils (for example, general praise/encouragement, specific feedback to support learning on some/all work, or none given)?
13. How frequently did you suggest that parents or pupils send in work?
14. How much do you agree with the following statement: 'During lockdown whilst most of my students were learning at home, it was easy for me to monitor who was, and wasn't, completing work'?
15. How much do you agree with the following statement: 'When setting work for remote learning, I found it difficult to differentiate to the lowest attainers in my class'?
16. Overall, how good do you feel the home learning experience was for your class?

As many of the questions were interrelated, we began our analysis by examining how responses co-varied. This involved two steps. First, we recoded the survey responses from the 16 questions into 29 binary and ordinal variables (as shown in Table 10 in Appendix D). We then used exploratory factor analysis to condense these 29 variables into a smaller number of underlying latent variables that describe primary school practice during the March to May lockdown. This factor analysis was performed at the level of the individual teacher response using a principal factors method (with rotation) to reduce our 29 variables to just eight factors. Table 11 in Appendix D shows the rotated factor loadings, which gives an indication of which variables tend to co-vary.

The factor analysis did not produce a particularly clear factor structure, nor were the factors obviously interpretable. Our goal was to derive measures that captured important and interpretable aspects of school practice—which is partly a matter of judgement. Also relevant to this was the within-school intra-cluster correlation (ICC), an indicator of the extent to which teachers in the same school give similar responses to a question, relative to the variation between different schools.

With these considerations in mind, we used the factor analysis to create a small set of interpretable, ad hoc variables to describe a school's provision for remote learning, focusing on those factors that seem to be the most consistent across a school.[28] Table 4 describes the five variables we use in our analysis.

[28] The goal of providing ad-hoc interpretable measures was specified in our study plan, available **here**.

Table 4: School response variables for Teacher Survey 1

Response variable	What it measures (initial period of school closures, Mar–July 2020)	Scale	Q*	ICC	Agreement°
Phoning students	Did teachers speak to students (not parents) directly on the phone? 1= either one or twice during the half term; 2 = about once a week or more.	0–2	Q10	0.465	58%
Platform use	Did schools use a platform: to set work; to receive work; to share examples of student work; for daily registration; for text chat interactions; for live audio/video chat interactions? One point for each.	0–6	Q1, Q2, Q4, Q7, Q11	0.531	67%
Timetable	What timetabling approach did schools have? 0 = no timetable; 1 = a suggested timetable; 2 = instructed daily timetable.	0–2	Q8	0.337	54%
Videos/live lessons	Did teachers have pre-recorded video lessons (scored 1) or live lessons (scored 2)? If neither, the variable is coded as zero.	0–2	Q9	0.267	56%
Work submission	How often did teachers expect work to be submitted? 0 = no recommendation about work submission; 1 = suggested work submission once a week or fortnight; 2 = several times a week; 3 = every day.	0–3	Q13	0.410	47%

* See Appendix D for full list of questions.

° This measures the percentage of times that pairs of teachers *within the same school* submitted the same set of responses on the questions underlying each of the five variables. The agreement statistic was based on 1,568 pairs of responses from 98 schools. In the case of platform use (which is on a scale of 0 to 6) we coded teachers as submitting the same response if they were within one point of each other.

Teacher Survey 2

Survey 2 asked teachers about the way schools responded in the autumn 2020 term, a period in which COVID-19 cases were still prevalent but most children were back at school. Given the level of disruption among schools and the demands placed on teachers, we decided to create a shorter questionnaire to minimise the burden on teachers. We asked seven questions, two of which were excluded due to low within-school correlation.[29] The final questions are listed in Appendix E. With fewer questions, we judged that it was not necessary to conduct a factor analysis to reduce the number of variables.[30] Instead, we combined questions that focused on the same element of school responses, creating a composite measure. For example, we combined all three prompts that asked about whether schools had increased their use of small group interventions in an effort to compensate for lost learning (see Appendix E for details).

The correlation of items at the school level is presented in Appendix E, Table 13 (p.51). We ultimately created five ad-hoc variables for analysis, summarised in Table 5. The table includes intra-class correlations (ICCs), which were used to check how consistent practices were within schools.[31]

Table 5: School response variables for Teacher Survey 2

Response variable	What it measures (autumn 2020 term)	Scale	ICC	Agreement within schools°
Absence provision for individuals	Did schools provide access to video/streaming of classes when individual children were absent? 0 = no access to any video/streaming of class; 1 = video/streaming class.	0–1	0.336	72%
Absence provision for classes	Did schools provide access to video/streaming of classes when classes needed to isolate? 0 = no access to any video/streaming of class; 1 = video/streaming class; 2 = no class isolation needed within school.	0–2	0.727	78%
Extra time	Did schools extend the school day (+1); run extra lunchtime learning activities in order to compensate for lost learning (+1); remain open during holidays (+1)?	0–3	0.428	80%
Reduced curriculum	Did schools reduce the curriculum, and focus more time on English and maths in the autumn 2020 term?	0–4	0.315	65%

[29] The two questions we asked that were not included due to low ICC were: (a) 'In which of these subjects have you returned to (re)teaching topics and skills that the class had missed as a result of lockdown during summer term 2020? Tick all that apply' and (b) 'Last term, how much time each day have you typically spent supporting remote learning for isolating students, whilst also teaching a class?'

[30] The study plan suggested that we could do a factor analysis. This is a deviation.

[31] We dropped a survey question due to very low ICCs, asking teachers, 'In which of these subjects have you returned to (re)teaching topics and skills that the class had missed as a result of lockdown during Summer term 2020?'

Small group interventions	Did schools make greater (or less) use of small group and individual face-to-face interventions (for example, run by TAs)? 0 = fewer interventions; 1 = same number as normal; 2 to 4 is more than normal, either during lessons, or outside lessons, or both.	0–4	0.332	68%

° This measures the percentage of times that pairs of teachers *within the same school* submitted the same set of responses on the questions underlying each of the five variables. The agreement statistic was based on 1,562 pairs of responses from 77 schools. In the case of 'reduced curriculum' and 'small group intervention' (which have scales ranging from 0–4) we coded teachers as submitting the same response if they were within one point of each other.

Teacher Survey 3

Survey 3 asked teachers about the way schools responded to the second period of widespread school closures. We decided to use a very similar set of survey questions as we used in the first survey (see Appendix F for details on the questions). This left us with five variables that were extremely similar to those that were analysed in Teacher Survey 1. The variables for Teacher Survey 3 are summarised in Table 6. Once again, the table includes ICCs, which were used to check whether practices were reasonably consistent within schools.

Table 6: School response variables for Teacher Survey 3

Response variable	What it measures (second period of closures, Jan–Mar 2021)	Scale	ICC	Agreement within schools°
Phoning students	Did teachers speak to students (not parents) directly on the phone? 1 = once, twice or several times; 2 = about once a week or more.	0–2	0.191	42%
Platform use	Did schools use a platform: to set work; to receive work; to share examples of student work; for daily registration; for text chat interactions; for live audio/video chat interactions? One point for each.	0–6	0.213	64%
Timetable	What timetabling approach did schools have? 0 = no timetable; 1 = a suggested timetable; 2 = instructed daily timetable.	0–2	0.321	66%
Live lessons	Did teachers provide live lessons? 0 = no; 1 = one, two, or several live lessons; 2 = most days; 3 = more than one most days.	0–3	0.667	62%
Work submission	How often did teachers expect work to be submitted? 0 = no recommendation about work submission; 1 = suggested work submission once a week or fortnight; 2 = several times a week; 3 = every day.	0–3	0.310	72%

° This measures the percentage of times that pairs of teachers *within the same school* submitted the same set of responses on the questions underlying each of the five variables. This was based on 1,348 pairs of responses from 52 schools. In the case of platform use (which is on a scale of 0 to 6) we coded teachers as submitting the same response if they were within one point of each other.

Aggregating teacher responses into school-level variables

For each school, we calculate the mean of the teacher responses. Note that within schools there were disagreements between teachers as to what 'school practice' looked like. Disagreements within a school could be due to measurement error or reflect genuine differences across teachers and classes. In both cases, these disagreements act as a barrier to our analysis of how these practices are associated with learning and changes in attainment. We summarise these disagreements in two ways, using the intra-cluster correlation (ICC) and a measure of 'agreement' (see, for example, Table 4).

Statistical analyses

This section provides an overview of our statistical methods. Full details of the models can be found in Appendix K.

RQ1: How did the attainment gap between disadvantaged pupils and their peers change over each phase of the pandemic?

We address this question separately for maths and reading. Initially, we focus on summarising how gaps have changed since the onset of the pandemic, for all year groups combined. In addition, following our analysis plan, we address this question separately for each of the four-time intervals, in each subject (maths and reading), and for each year group (Year 2 to Year 6).

For each of these combinations, and for each pupil who has attainment data at both relevant time points, we calculate the difference between their two scores then fit a linear mixed model with a random school intercept to account for clustering, and terms to control for the different time gaps between the two assessments. From this

we estimate the mean change overall and include an interaction term to capture the additional change in attainment for disadvantaged pupils—this last coefficient being our estimate of the change in the gap between the attainment of disadvantaged pupils and others.

As there are five assessment points, and some missing data from each, we need to mitigate the threat of non-response bias. We use multiple imputation using chained equations (MICE) with 20 imputed datasets for each model combination and calculate the average effect estimate. A robustness check about the impact of using MICE is described in Appendix Q.

RQ2: How did schools respond to support pupils' learning during each phase of the pandemic?

We address RQ2 using simple descriptive statistics. The school response variables are described in the Measures section and Appendices D, E, and F.

RQ3: What was the association between a school's response to supporting learning and the change at school level in the disadvantage gap over that same period?

For RQ3, the models are similar to those in RQ1, with separate estimation for each time interval, subject, and year group. The left hand side of the model is again the pupil-level difference in before and after attainment scores. However, there are three key differences in the models for RQ3.

The first is that we incorporate a matrix of control factors, at both pupil and school level:

- at the pupil level: gender and EAL;
- at the school level:
    - % pupils achieving expected standard in reading, writing, and maths in 2019;
    - % pupils achieving higher standard in reading, writing, and maths in 2019;
    - inspection ratings (as factor); and
    - % pupils who were FSM6 in 2019.

This allows us to rule out the possibility that any raw associations between teacher survey responses and pupil gains are a result of these pupil- or school-level differences.

The second difference is that we fit a separate model for each of our five response constructs from the teacher survey that describe that school's response to the challenges of remote and disrupted teaching.[32] Each of these models was run separately for each subject (maths and reading) and for each pandemic phase (initial closures, reopening, and second closures), but was run just with all year groups combined, not for separate year groups.

The third difference is that we now allow the school random effect to be estimated as a function of the school's mean survey response for that construct. The coefficient of that response construct score is therefore an estimate of the association between our school-level measure of the school's response to supporting learning and the average gains made by its pupils. As before, we fit an interaction term to capture the additional effect for disadvantaged pupils; in RQ3 this interaction is estimated as a linear function of the construct score for each school. The fixed-effect coefficient of this school-level response construct score is therefore an estimate of the association between the attainment score gains made by disadvantaged pupils (relative to others) in each school and that school's score on our response construct from the teacher survey.

Full technical details of the statistical models used can be found in Appendix K.

Multiple comparisons

Our analyses of RQ1 and RQ3 involve multiple comparisons. For RQ1, we are simultaneously examining the FSM gap for two outcome variables: maths and reading. In each part of RQ3, we simultaneously examine the effect of five variables. In order to control for the Family-Wise Error Rate, in each case we use a Bonferroni

[32] See Measures section above.

adjustment to guard against the possibility of 'false positives'. Note that we only explicitly discuss the adjustment in cases where unadjusted p-values would result in the rejection of null hypotheses.[33]

Study plan deviations and extensions

This section documents changes from our original study plan.

1. The study plan specifies that we will fit model 1a allowing for the parameter estimates to vary by year group. To simplify estimation and interpretation of this 'by year group' analysis, we fit model 1a separately for each year group (g) and subject (k)—for a total of ten sets of estimates.[34] This applies throughout our analysis of RQ1.
2. In addition to estimating changes in disadvantage gaps between assessments (for example, autumn 2019 to summer 2020) we also estimate a 'overall' change covering the entire study period: autumn 2019 to summer 2021.
3. We present summary results averaging across Years 2 to 6. As noted above, this involved estimating $\widehat{\Delta}^{\mathrm{k}}_{T_X}$—an average change in disadvantage gaps across Years 2 to 6 for outcome k, in time period T_X—in addition to the year-level specific estimates $\widehat{\Delta}^{\mathrm{k}}_{T_X g}$. See Appendix O for details of Likelihood Ratio tests. We present $\widehat{\Delta}^{\mathrm{k}}_{T_X}$ estimates in Table 8.
4. The study plan states that we would adjust for the timing of tests using the date of test administration (in RQ1 and RQ3). Unfortunately, one of our sources of date information only reported information on the month in which pupils took tests. To make our data consistent across sources we therefore used 'months' throughout. In the robustness section we provide evidence that the timing of the tests is not an important factor in the evolution of disadvantage gaps—Figure 14 (p.82) presents the results for reading and maths averaged across Years 2 to 6; including time controls barely changes our estimates. As such, we remove these controls for our analysis of RQ3. The associations we estimate are very insensitive to this change. A likely explanation for this lack of sensitivity is that our analysis focuses on a *difference* (the disadvantage *gap*) rather than the *level* of attainment, which is more likely to be systematically influenced by the date of test administration.
5. In RQ3, we did not include 'ethnic background' as a covariate. The reason was that the number of predictors introduced by this categorical variables made models unstable.
6. In RQ3, we present the 'main effects' of each school construct (ϕ_0^c) in addition to the main focus of our analysis, namely the association between school responses and disadvantage gaps. We believe this is of substantive interest.
7. In the analysis plan for RQ3, models 3a, 3b, and 3c specified that the FSM6 indicator would be zero centred. Upon reflection, this made parameter interpretation more difficult so we left the FSM6 indicator as binary.
8. The original study plan did not include plans for a third teacher survey. When schools closed in January 2021, we added a third survey in an effort to describe how teachers were responding to a second period of (largely) remote teaching. This addition was document in an **amended study plan**.
9. The original study plan did not make clear how we would deal with multiple comparisons in RQ1 and RQ3. In this report, we use a Bonferroni adjustment to control for the Family Wise Error Rate.
10. There were several deviations to the 'robustness' section of our study plan:
    a. The study plan said that we would examine whether our results were robust to including individual-level controls for absences. Unfortunately, technical problems prevented us from collecting absence data from the majority of schools within the timeframe of the project. Data was available for less than a third of the pupils included in Research Question 1 and was not available at all prior to autumn term 2020. For this reason we do not control for absences in our main analysis. However, we did repeat the RQ1 analysis on the reduced sample for whom data was available with controls for absences. This analysis indicated that results were robust to these controls; results from the analysis can be seen in Appendix Q (part 4).
    b. The plan said that we would examine the effect of how school responses Z and $\tilde{Z}$ were scaled. However, we ultimately decided not to conduct these sensitivity checks. The survey data we gathered ended up being less detailed than we anticipated when we wrote the study plan. In

[33] We thank the reviewers for clarifying this point.
[34] This deviation applies to all research questions in RQ1.

particular, we had thought there may be a large number of questions for teachers. But, in an effort to minimise the burden on teachers we ultimately had quite short surveys (especially for survey two). This being the case, we did not think it was justified to use IRT on constructs that rely on one or two questions.

c. Finally, we did not run the 'repeated measures models' from the study plan. We wanted to limit the already-large number of analyses we were presenting and believe that the models we present are simpler and preferable to more complex models.

Research findings

Results

Research Question 1

Estimating disadvantage gaps at the start of the study: autumn 2019

To contextualise changes in the magnitude of disadvantage gaps, Table 7 presents the level of disadvantage gaps before COVID-19 disruptions, using equation (1). The scale is Effect Size units. In our sample, the average disadvantage gap is 0.42 ES in reading and 0.45 ES in maths. Using the EEF conversion table, this equates to five months' progress for maths and six months' progress for reading.[35] Estimates from the National Pupil Database suggest that, across primary schools, the average disadvantage gap is around 0.50 ES units (Thomson, 2021). The gaps in the sample we analyse are similar to, but slightly narrower than, the disadvantage gaps in the broader population.

Table 7: Estimated disadvantage gap in Autumn 2019 ($\hat{G}_{Aut19}$) of RQ1 responder sample

	Reading (s.e.)	Maths (s.e.)
Year 2	0.43 (0.04)	0.41 (0.04)
Year 3	0.40 (0.04)	0.39 (0.04)
Year 4	0.48 (0.03)	0.52 (0.03)
Year 5	0.38 (0.03)	0.42 (0.03)
Year 6	0.41 (0.03)	0.52 (0.03)
Overall average (Y2–Y6)	**0.42**	**0.45**

Note: throughout this study, year levels are defined by the 2020/2021 academic year. 'Year 3' in this table refers to the cohort of children who were in Year 2 in 2019/2020. Standard errors are in parentheses. The definition of 'disadvantage gap' is given in equation (1). The scale is Effect Size units. The samples used here are the same as the analysis sample for RQ1 (Full responders + Partial responders, described in Figure 1).

Main analysis for RQ1

Next, we present results from statistical models that directly estimate changes in disadvantage gaps. We start with the most aggregated level: changes in the disadvantage gap for reading and maths, averaged across Years 2 to 6. The results are presented in Table 8, which has five columns. The first column presents estimates of gap changes during the period from autumn 2019 to September 2020. During this 'initial closure period', point estimates for the change in the disadvantage gap are negative (implying a widening) for maths. We estimate the change in the maths gap to be -0.038 ES, with a 95% confidence interval of -0.063 to -0.014. For reading, the point estimate of -0.002 suggests no change in the disadvantage gap with a confidence interval of -0.031 to 0.026.

The second column presents estimates from the 'reopening period' during which schools across the country generally returned to face-to-face instruction, albeit with widespread disruption due to COVID-19. Here we see no signs of changes in disadvantage gaps, with point estimates very close to zero for both reading and maths. In short, during the first term in which students went back to school, we see no clear evidence of gaps closing, or widening further.

The third column presents estimates from spring term of 2020/2021, which was also heavily disrupted by COVID-19 including widespread closures. Again, we see very limited evidence of changes in attainment gaps with point estimates close to zero for both reading and maths. This pattern is repeated for the summer term of 2020/2021: once again we see no clear evidence of changes in disadvantage gaps.

[35] The conversion table is available from **https://educationendowmentfoundation.org.uk/public/files/Toolkit/Toolkit_Manual_2018.pdf**

The fifth column is a summary of columns one to four. From November/December 2019 to June/July 2021 we find evidence that the disadvantage gap for primary maths widened. Our point estimate is -0.048 ES units (-0.077, -0.019), which represents an 11% increase in the pre-COVID-19 gap.[36] A plausible range is a widening of between 4% and 17%.[37]

Our estimate for the change in the disadvantage gap in reading is -0.003 (95% CI: -0.035, 0.029). While there is uncertainty, this suggests that within our sample the gap in reading is similar to pre-COVID-19 levels.

Table 8: Estimated average change in the disadvantage gap

	Autumn 19–Sep 20 (initial closure period)	Sep 20–autumn 20 (reopening period)	Autumn 20–spring 21 (second closure period)	Spring 21–summer 21 (summer term)	**Overall: autumn 2019–summer 2021**
Maths (95% CI)	-0.038 (-0.063, -0.014)	-0.007 (-0.031, 0.016)	-0.006 (-0.026, 0.015)	0.006 (-0.018, 0.03)	**-0.048 (-0.077, -0.019)**
Reading (95% CI)	-0.002 (-0.031, 0.026)	-0.004 (-0.031, 0.024)	-0.010 (-0.037, 0.016)	0.017 (-0.014, 0.047)	**-0.003 (-0.035, 0.029)**

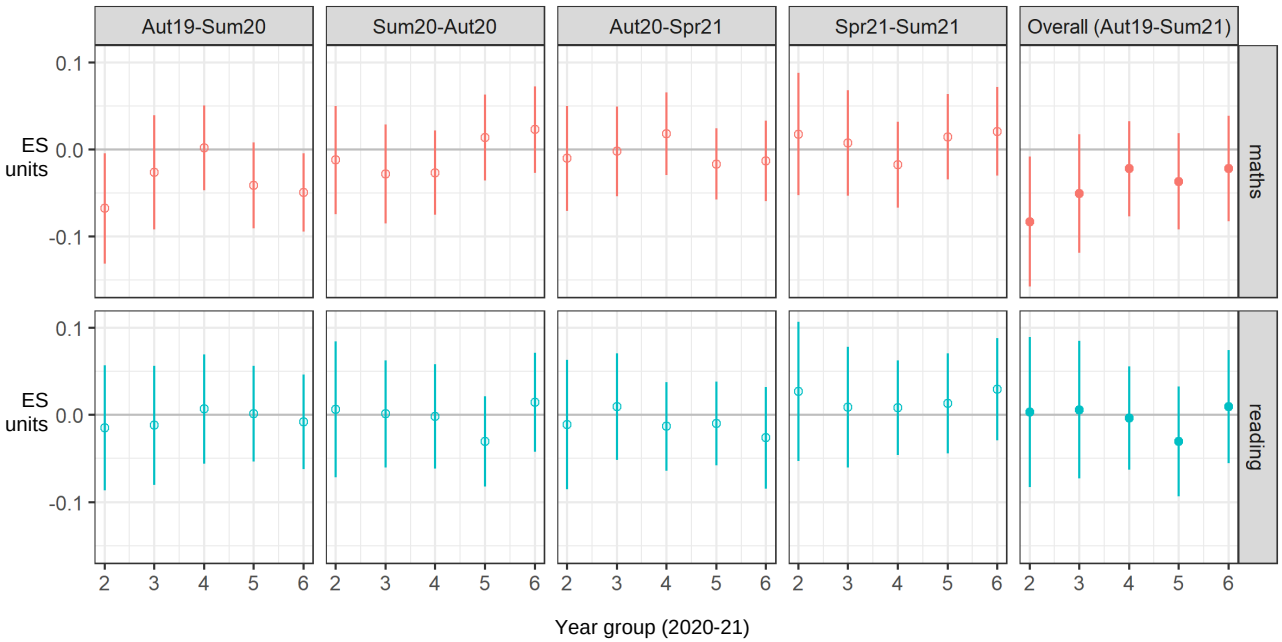
Notes: full regression results underlying these estimates are presented in Appendix F. All analyses rely on the RQ1 analysis sample (Full Responders + Partial Responders—see Figure 1 for details).

Results by year level

Figure 2 presents results disaggregated by year group. The top row of plots shows results for maths, the bottom row for reading. In each case, the first four panes present changes in the attainment gap for four different periods and the panel on the right presents a summary.

The largest point estimates of gap widening are among maths outcomes for students in Year 2, from autumn 2019 to summer 2020. However, as discussed in Appendix O, we fail to reject the hypothesis that there are no effects of year level on changes in attainment gaps for reading and maths. So, while there do appear to be subject differences in our data (between reading and maths), we do not find strong evidence that children of different ages were differentially affected by COVID-19.

Figure 2: Results from RQ1a and RQ1b by year level



Notes: negative estimates indicate a widening of gaps. All estimates use the RQ1 analysis sample (Full Responders + Partial Responders, see Figure 1). For full regression results see Appendix F. The righthand pane is a summary of the overall study period.

[36] The average disadvantage gap for maths in our sample at the autumn 2019 baseline was 0.45 ES units as per Table 7. To calculate the percentage change, we divided 0.048 by 0.45.

[37] Even after making a conservative Bonferroni adjustment—to account for the fact that we are testing both reading and maths outcomes—we find evidence against the null hypothesis that the maths gap has remained constant over the entire study period.

Research Question 2

Teacher Survey 1: school responses during the 'initial closure period' from March to May 2020

We used five variables to describe school responses during the first period of lockdown (variables are described in Table 4). At the teacher level there was a very even spread of responses in all categories, as is clear from tables of teacher counts (reported in Appendix P). There were two minor exceptions: relatively few teachers reported having a daily instructed timetable (80 of 539) or live lessons (66 of 539). Overall, the teacher counts show that there was considerable variability in the response to the initial lockdown.

There was also variability at the school level, demonstrated by the histograms in Appendix P, which plot school-level means. That said, two of the variables (timetabling and phoning students) had a clear, modal response—indicated by the respective spikes in the histograms for those variables—which resulted in less variability at the school level. This limits our statistical power in RQ3.

Teacher Survey 2

This survey focuses on the period from September 2020 to December 2020, during which most students returned to schools. As described in the Measures section, we summarise the teacher survey from this period using five variables (see Table 5 for variable description).

Again, we note that responses at the teacher level are fairly evenly spread. The 'extra time' variable is an exception: most teachers reported that schools had not pursued this strategy in the autumn term (75% of the 454 respondents) and those that did report some extra time; almost all reported learning during lunchtime (rather than longer days or being open in the holidays). At the school level, the five variables all have some variation. In particular, schools showed a varied response to the questions about responding to lost learning by reducing the curriculum or using small-group interventions. There was less variation in providing 'extra time' (this was a relatively rare response). Similarly, only a minority of schools provided video/streamed classes to *pupils* who were absent (as opposed to whole classes that were isolating). For detailed counts and histograms of each variable, see Appendix P.

Teacher Survey 3

This survey focuses on the period from January 2021 to March 2021 and again captures five variables (described in Table 6). Once again, there is variation for each of the questions, at both the teacher and school levels (see Appendix P for histograms and tables). This suggests that the second period of lockdown did not result in a clear standardisation of remote learning practices. That said, there are some trends of teachers converging on common practices. For example, the 'work submission' variable shows that 76% of teachers in our sample were asking their students to submit work once a day (or more). Similarly, 76% of teachers reported completing at least one 'live lesson'.

Compared to the initial period of school closures (captured by Teacher Survey 1), a clear difference in Teacher Survey 3 is that teachers were typically doing *more* of the activities we asked about. This is illustrated in Table 9. The two starkest examples are those mentioned above: live lessons and the frequency of work submission. In the second period of closures, 76% of teachers reported completing at least one live lesson, compared to 12% in the initial lockdown. Similarly, in the second period of closures, 76% of teachers suggested work be submitted daily (or more often) compared to 31% in the first lockdown. Teacher Survey 3 also saw greater use of online platforms and more timetabling. The only variable where teachers reported less activity in the second period of closures was 'phoning students', but these differences were small.

Table 9: Comparison of Teacher Surveys 1 and 3

Question	Response categories	Teacher Survey 1 (Jul 2020)	Teacher Survey 3 (Mar 2021)
Live lessons Did you try teaching a 'live' lesson to your class during the first half of the summer term (with students either speaking or on mute)?	No	52%	5%
	No, but I did pre-record a video of myself for them to watch	35%	19%
	Yes (any number of live lessons)	12%	76%
Phone students Did children (i.e. not the parents) typically speak to you (i.e. their class teacher) on the phone?	No, children typically did not speak to their class teacher	37%	48%
	Yes, I spoke to them each once or twice during the half term	40%	30%
	Yes, I spoke them to each about once a week (or more frequently)	23%	23%
Platform use Did schools use a platform: to set work; to receive work; to share examples of student work; for daily registration; for text chat interactions; for live audio/video chat interactions? (One point for each.)	Count = 0	17%	1%
	Count = 1	15%	4%
	Count = 2	18%	4%
	Count = 3	24%	13%
	Count = 4	17%	22%
	Count = 5	8%	24%
	Count = 6	2%	32%
Timetable Did you provide students with a timetable to follow each day (this could be hourly, lesson-by-lesson, or a daily list)?	No	33%	6%
	Yes, but it was only a suggested timetable and pupils were not required to follow it	52%	31%
	Yes, we asked pupils to follow a daily timetable	15%	63%
Work submission How frequently did you suggest that parents or pupils send in work?	I gave no suggestion about how frequently work should be sent in	34%	5%
	Once a week (or less)	19%	2%
	Several times a week	15%	17%
	Every day (or after each lesson)	31%	76%

Note: the sample for Teacher Survey 1 was n = 539 teachers. For Teacher Survey 3, the sample was n = 343.

Research Question 3

Research Question 3 synthesises the analysis from RQ1 and RQ2. We explore whether there are any associations between the school responses captured in our teacher surveys and changes in attainment gaps.

Analysis of Teacher Survey 1 and outcomes from before and after the 'initial closure period'

We begin by estimating the association between relative **changes in attainment** over the initial closure period (autumn 2019 to summer 2020) and the five school response variables from Teacher Survey 1 ('phoning students', 'use of technology platforms', 'timetabling', 'video/live lessons', and 'frequency of work submission').[38] The left panel of Figure 3 presents these associations, conditional on a set of pupil and other school covariates.[39] We fail to find evidence of conditional association between relative change in attainment and any of the five response variables. None of the point estimates are greater than 0.06 ES in magnitude and all 95% confidence intervals contain zero. Similarly, we find no evidence that the raw associations have large, non-zero associations.[40]

[38] Plots illustrating the distribution of change in attainment are presented in Appendix I.

[39] These are the coefficients from model 3a. At the pupil level, we condition on gender and EAL status. At the school level, we condition on % pupils achieving expected standard in reading, writing, and maths in 2019; % pupils achieving higher standard in reading/writing/maths in 2019; Ofsted results; and % pupils who were FSM6 in 2019.

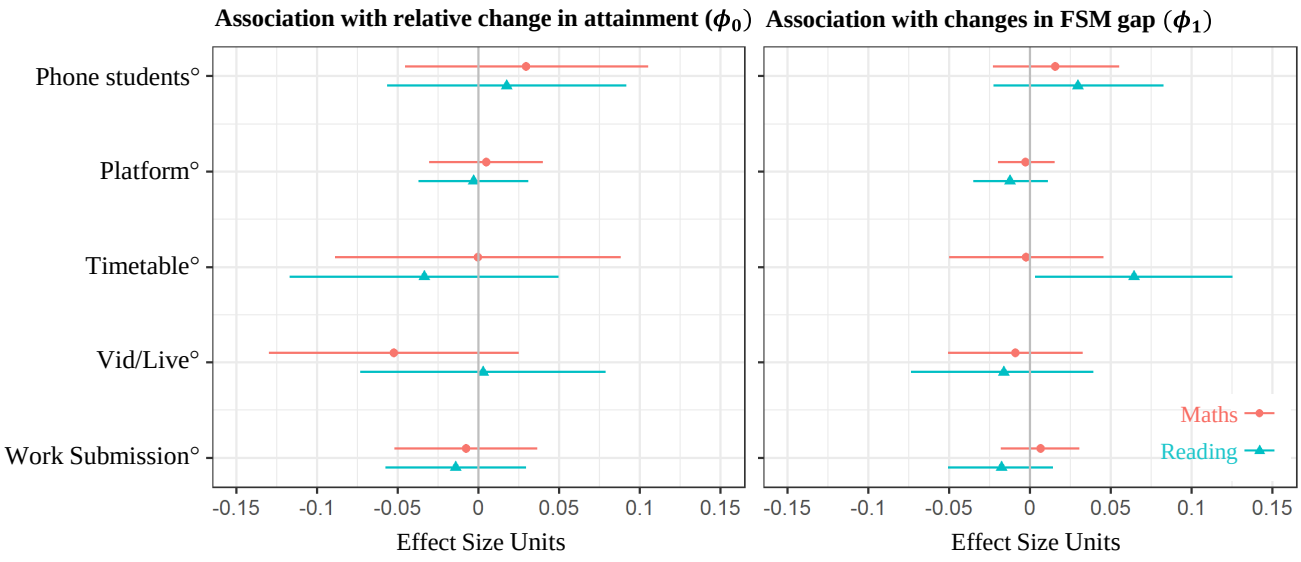
[40] None of the raw correlations between school response and the change in relative achievement ($Y_{ijT_1} - Y_{ijT_0}$) are greater than 0.06 in magnitude. There is no evidence to suggest any of these associations are different from zero. The lack of association does not appear to be a function of conditioning, or modelling.

Next, we examine the conditional association between the school responses from Teacher Survey 1 and **changes in disadvantage gaps** over the initial closure period. These conditional associations are presented in the right panel of Figure 3.[41] Once again we fail to find clear evidence of non-zero associations. All point estimates are smaller than 0.03 ES units with the exception of the association between timetabling and reading, with an estimate of 0.063 ES units (95% CI: 0.001, 0.125).[42] In the absence of a clear a priori hypothesis about why this particular practice should be associated with reading and not mathematics, and the exploratory nature of the study with multiple comparisons, we are wary of interpreting this result as something we would expect to be reproducible. Indeed, in Teacher Survey 3 we asked the exact same question about timetabling during the second period of widespread school closures. As described below, in RQ3c we find no evidence of association between the timetabling measure and changes in disadvantage gaps.

Analysis of Teacher Survey 2 and changes in outcomes across the autumn 2020 term

Our analysis here follows the same pattern as the analysis of Teacher Survey 1. We start by examining the associations between relative **changes in attainment** during the 'reopening period' (autumn term of 2020) and the five school response variables from Teacher Survey 2 ('absence provision for *individuals'*; 'absence provision for *classes*'; 'extra time'; 'reduced curriculum', and 'small group interventions'). The left panel of Figure 4 presents these associations, conditional on a set of pupil and other school covariates.[43] We fail to find evidence of conditional association between relative change in attainment and any of the five response variables. All 95% confidence intervals contain zero. Similarly, we find no evidence that the raw associations are different from zero.[44]

*Figure 3: Associations between **school responses during the initial closure period** and 'change in attainment' (left panel) and 'change in disadvantage gap' (right panel)*



Notes: Both panels are based on model 3a. The dependent variable is $Y_{ijT_1} - Y_{ijT_0}$, the estimated change in learning from autumn 2019 to summer 2020. ° School response variables are defined in the Measure section (see Table 4 for an overview) and have been centred to have a mean of zero. We have retained the original scaling to make the results as interpretable as possible. The sample here is the analysis sample for RQ3a, described in Figure 6 and consisting of 90 schools for reading and 88 schools for maths. Conditional associations between school response variables and changes in learning are presented in the left panel and represent $\hat{\phi}_0$ parameters from model 3a. Condition associations

[41] These are ϕ_1 coefficients from model 3a.

[42] We thank an anonymous reviewer who pointed out the following post-hoc hypothesis to explain the reading/maths difference in timetabling. Large numbers of schools use Accelerated Reader and use a timetabled approach (such as 20 minutes per day). Such interventions are more prevalent for literacy than maths. While debate exists over the wider efficacy of such programmes, this is a potential explanation for effect on reading and a higher level of timetabling, namely that the structure was associated with the involvement of a programme that revealed itself to be helpful during lockdown.

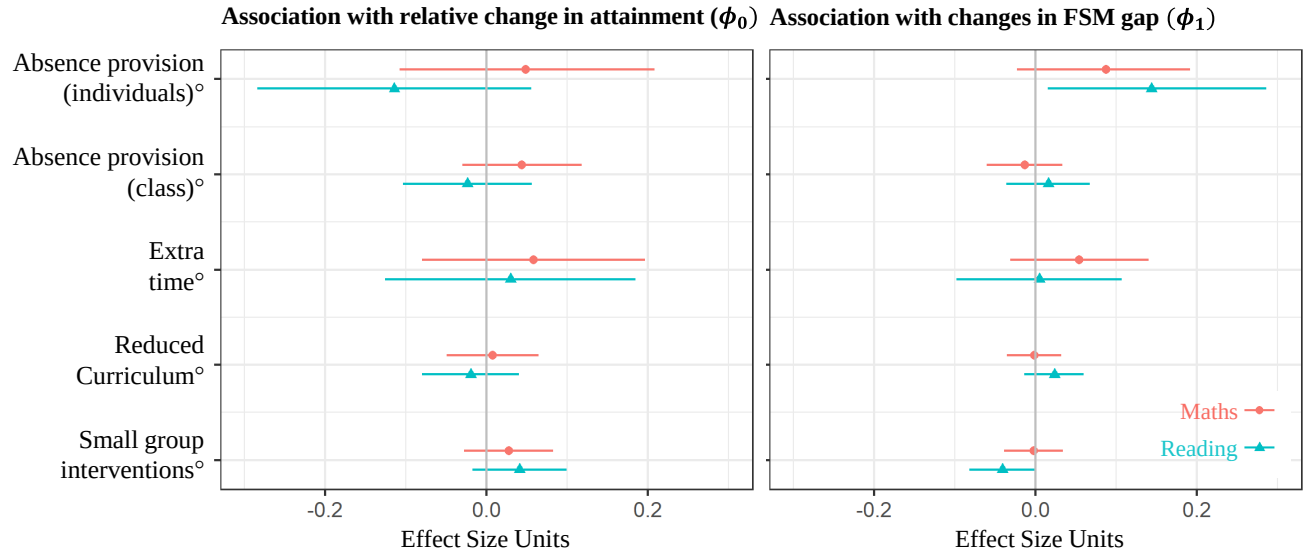
[43] These are the coefficients from model 3b. At the pupil level, we condition on gender and EAL status. At the school level, we condition on % pupils achieving expected standard in reading, writing, and maths in 2019; % pupils achieving higher standard in reading, writing, and maths in 2019; Ofsted results; and % pupils who were FSM6 in 2019.

[44] None of the raw correlations between school response and the change in relative achievement ($Y_{ijT_2} - Y_{ijT_1}$) are greater than 0.08 in magnitude. There is no evidence to suggest any of these associations are different from zero. In short, the lack of association in the left panel of Figure 4 does not appear to be a function of conditioning, or modelling.

between school response variables and changes in disadvantage gaps are presented in the right panel and represent $\hat{\phi}_1$ from model 3a. Full regression results are in Appendix L:

The right-hand panel of Figure 4 presents the association between **changes in disadvantage gaps** over the reopening period (autumn 2020 term) and the five school response variables we captured in Teacher Survey 2. Once again, we fail to find clear evidence of non-zero associations. There may be weak evidence that providing videos or live lessons to *absent pupils* is associated with a narrowing of disadvantage gaps. In schools where absent pupils had access to video or live recordings of lessons, we estimate that gaps in the autumn term narrowed in maths by 0.087 (-0.023, 0.191) and in reading by 0.143 (0.015, 0.286). These positive estimates are robust to different sets of covariates in model 3a. However, we again note that these are highly uncertain findings and, in a context of multiple comparisons, could easily have arisen due to chance. Broadly, the main finding from the analysis of RQ3b is that we fail to find clear evidence of associations.

*Figure 4: Associations between **school responses during the reopening period (autumn 2020 term)** and 'change in attainment' (left panel) and 'change in disadvantage gap' (right panel)*



Notes: Both panels are based on model 3b. The dependent variable is $Y_{ijT_2} - Y_{ijT_1}$, the estimated change in learning from summer 2020 to autumn 2020. ° School response variables are defined in the Measure section (see Table 5 for an overview) and have been centred, to have a mean of zero. We have retained the original scaling to make the results as interpretable as possible. The sample here is 'analysis sample for RQ3b' initially discussed in Figure 7 and consisting of 58 schools. Conditional associations between school response variables and changes in learning are presented in the left panel and represent $\hat{\phi}_0$ parameters from model 3b. Conditional associations between school response variables and changes in disadvantage gaps are presented in the right panel and represent $\hat{\phi}_1$ from model 3b. Full regression results are in Appendix M:.

Analysis of Teacher Survey 3 and outcomes from before and after the 'secondary closure period'

Our analysis of Teacher Survey 3 follows the same pattern. To start, we examine the associations between relative **changes in attainment** during the 'second closure period' (January to March 2021) and the five school response variables from Teacher Survey 3 ('live lessons', 'phoning students', 'use of technology platforms', 'timetabling', and 'frequency of work submission'). The left panel of Figure 5 presents these associations, conditional on a set of pupil and other school covariates.[45] We again fail to find evidence of conditional association between relative change in attainment and any of the five response variables. All 95% confidence intervals contain zero and we find no evidence that the raw associations are different from zero.[46]

The right-hand panel of Figure 5 presents the association between **changes in disadvantage gaps** over the second closure period and the five school response variables we captured in Teacher Survey 3. Two of the results—for phoning students and platform use—suggest that these variables were associated with *widening* gaps

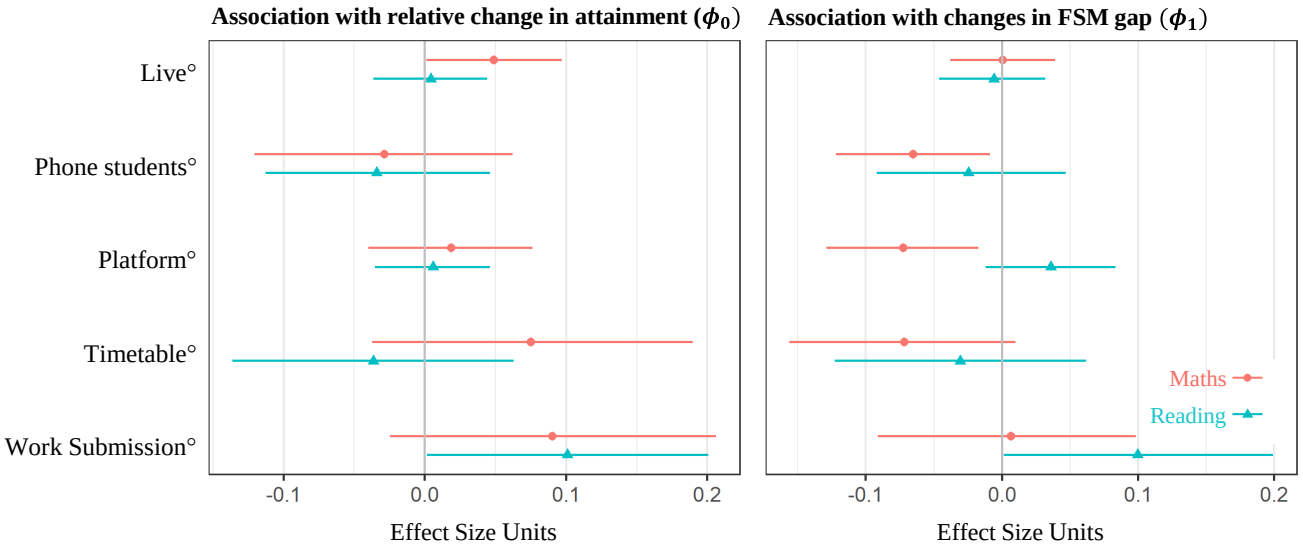
[45] These are the coefficients from model 3c. At the pupil level we condition on gender and EAL status. At the school level we condition on % pupils achieving expected standard in reading, writing, and maths in 2019; % pupils achieving higher standard in reading, writing, and maths in 2019; Ofsted results; % pupils who were FSM6 in 2019.

[46] None of the raw correlations between school response and the change in relative achievement ($Y_{ijT_3} - Y_{ijT_2}$) are greater than 0.07 in magnitude. There is no evidence to suggest any of these associations are different from zero. In short, the lack of association in the left panel of Figure 5 does not appear to be a function of conditioning, or modelling.

in maths (but not for reading). In other words, the results indicate that increases in platform use and phoning students were associated with widening gaps. However, these are marginal results and, given the number of comparisons we are testing, we suggest that they are treated with caution. Using the Bonferroni adjustment we fail to reject the composite null at $\alpha = 0.05$ (VanderWeele & Mathur, 2019). The smallest p-value we observe in the right-hand panel of Figure 5 is 0.01, which is larger than $\frac{\alpha}{10} = 0.005$. Moreover, the magnitude of the point estimates are quite small, with all estimates having an absolute value of less than 0.1 Effect Size units. In short, we interpret the findings for RQ3c as broadly in line RQ3a and RQ3b: an absence of evidence for clear associations between school responses and changes in disadvantage gaps, conditional on a set of student and school controls.

Finally, it is worth noting that none of the school responses were randomised so the estimated associations we present rely entirely on the conditional independence assumption for their causal credibility. While there is growing evidence from within-study comparisons that this assumption is often plausible in education settings—and that typical levels of selection bias in education may be smaller than previously thought (e.g. Cook et al., 2008; Weidmann & Miratrix, 2020; Wong et al., 2017)—given that we are not conditioning on a student-level measure of attainment, we place limited credence in any causal interpretation of our results.

*Figure 5: Associations between **school responses during the second closure period** and 'change in attainment' (left panel) and 'change in disadvantage gap' (right panel)*



Notes: Both panels are based on model 3c. The dependent variable is $Y_{ijT_3} - Y_{ijT_2}$, the estimated change in learning from autumn 2020 to spring 2021. ° School response variables are defined in the Measure section (see Table 6 for an overview) and have been centred to have a mean of zero. We have retained the original scaling to make the results as interpretable as possible. The sample here is 'analysis sample for RQ3c' initially discussed in Figure 8 and consisting of 35 schools for maths and 44 schools for reading. Conditional associations between school response variables and changes in learning are presented in the left panel and represent $\hat{\phi}_0$ parameters from model 3c. Conditional associations between school response variables and changes in disadvantage gaps are presented in the right panel and represent $\hat{\phi}_1$ from model 3c. Full regression results are in Appendix M:.

Robustness and further analysis

We used a number of robustness checks to test the sensitivity of our results to different assumptions or methods. Specifically, we compared different approaches to scaling the attainment scores, different analyses to control for the variation in dates of test administration, and different approaches to dealing with missing data. Full details and results from all these robustness checks can be found in Appendix Q. Here we present a summary of the results of the three robustness checks.

Standardisation and scaling of outcomes

In this analysis we check whether the results for RQ1 are sensitive to the method of scaling attainment outcomes by comparing our core findings using three different scaling approaches:

- standardising scores (this method is represented in the main results);
- ranks, defined by assessment, time period, and year group; and
- ranks, defined by assessment, time period, year group, and school.[47]

A full comparison of the effect size estimates from each standardisation approach over each phase of the pandemic can be found in Appendix Q. While there are some minor differences across scaling methods—for example, the point estimates for the change in maths gaps are closer to zero when we use with 'within-school rank'—the main results are not very sensitive to our choice of scaling. The main conclusion is similar across all three approaches: the disadvantage gap in maths appears to have widened in the initial lockdown and then stabilised; for reading, the disadvantage gap seems to be relatively stable throughout.

Methods to control for the date of test administration

To assess whether our results are robust to different methods of controlling for test administration date, we compare the results of model 1a with a model that has no time controls.[48] Here, the results are very clear: removing the controls for timing of assessments makes pretty much no difference to any of our estimates.

Missing data

RQ1 missingness analysis

As noted in the Participants section, our starting sample for RQ1 was defined by pupils who:

- had a valid assessment in reading or maths in Autumn 2019; and
- were on roll during the autumn term 2020 and had a disadvantage flag (yes/ no) present.

We asked schools to test pupils in each of reading or maths using the summer 2020, autumn 2020, spring 2021 and summer 2021 test suites. Based on the response patterns we classified pupils into three groups:[49]

- full responders—those for whom results are observed in all four test suites;
- partial responders—those for whom between one and three further test results are observed; and
- non-responders—those for whom no further test results are observed.

Comparing non-responders with others, we find that, overall, their baseline (autumn 2019) scores are somewhat lower. This difference becomes more extreme for non-responders who are in schools in which the overall response rate was good. By contrast, the majority of non-responders who are in schools where a majority of the other pupils in the school did not respond (in other words, where the non-response is largely a school-level phenomenon), are hardly different from the partial and full responder samples. Other characteristics of non-

[47] This method was not mentioned in the study plan but is included to provide a more detailed examination of robustness to scaling.
[48] See Appendix Q for details, and Appendix K for main description of models.
[49] This was done separately for reading and maths.

responders are that they are more likely to be disadvantaged, to have special educational needs, or come from London.

In order to address this differential non-response, our main analyses used multiple imputation using chained equations (MICE). As a robustness check, we compared these estimates with estimates from the full responder sample, with no imputation. While we observe some differences, the overall conclusions for the two sets of results are very similar. For maths, both sets of results show evidence of gaps widening over the study period (autumn 2019 to summer 2021). In both cases this is driven by the change in the first period—from autumn 2019 to summer 2020. In subsequent periods, there is very little evidence of change in disadvantage gaps. For reading, the main pattern of results is consistent across the 'main' and 'no imputation' analysis: we find no clear evidence of disadvantage gaps widening or shrinking throughout the study period. However, there is a tendency for the 'no imputation' estimates of gap widening in maths to be slightly bigger and in reading to be slightly smaller than estimates from the full sample using MICE. However, these differences are not large and the confidence intervals overlap substantially.

RQ3 missingness analysis

We have already presented comparisons between the RQ3 and RQ2 samples (see Appendix I). To recap: it does not appear that the schools we analyse in RQ3 (for which we have overlapping data from teacher surveys and attainment) were unusual in terms of their response to COVID-19 disruptions.

We examine this more formally using logistic regression (see Appendix Q for details). This analysis suggests that, conditional on other observed characteristics, FSM6 students were more likely to be missing from the RQ3 sample (that is, not have post-COVID-19 attainment data, despite attending schools represented in the teacher surveys) than their peers. There were no other clear patterns in the data.

We also compare our estimates for RQ3 to an alternative approach using multiple imputation (MICE). Once again, we find that the estimates of the associations between school responses and changes in attainment gaps from both approaches (with and without imputation) are very similar.

Discussion

Why has the gap grown in maths but not in reading?

Before conducting this study, we had no particular reason to expect that COVID-19 would affect attainment in maths and reading differently. A rapid evidence assessment of existing studies, published at about the time we were designing this study but not including any data from schools affected by COVID-19, found no evidence of differential effects of school closures on disadvantage gaps by subject (EEF, 2020). Other studies that have emerged since then have reported mixed results. Of the two best U.K. studies, DfE (2021) did find a larger effect in maths than reading, but Rising Stars (2020) found the gap change in both subjects was about the same. In the U.S., both Sass and Goldring (2021) and Pier et al. (2020) found a larger widening of disadvantage gaps in reading than in maths.

It is also not clear that there is any convincing, a priori reason why school closures and remote learning would lead to widening of disadvantage gaps in maths but not in reading. In Cooper et al.'s (1996) meta-analysis of the effects of school summer closures, a larger overall drop in maths than in reading was explained by suggesting that mathematical performance is both more dependent on school instruction than home life (relative to reading) and that the worst affected subject areas were those that were most dependent on practice and fluency, rather than conceptual understanding. Contrastingly, they also reported no differential effects by family income in maths, but did find these interactions in reading: 'Middle-class students showed a nonsignificant gain in grade-level equivalent reading scores, while lower-class students showed a significant loss' (p.261). Our result may therefore be considered unexpected, and perhaps surprising.

Would we expect gaps to grow within a cohort, regardless of COVID-19?

Our analysis tracks how disadvantage gaps have changed over a year for a constant group of students (spanning Years 2 to 6). It is worth considering how these direct estimates of gaps typically evolve over time in the absence of COVID-19. Such an analysis was recently published using NPD data (Thomson, 2021). The analysis examines two separate cohorts in the National Pupil Database: those who finished Key Stage 2 in 2019 and those who finished Key Stage 2 in 2018. The results differ slightly by cohort, but in both cases the magnitude of the average annual change in the disadvantage gap was smaller than 0.005 ES units. In other words, in the absence of COVID-19, gaps within a cohort change extremely gradually from year to year. This suggests that the widening of the maths gap we observe in Table 8 is substantial in comparison to expected changes and is likely a result of COVID-19 disruptions.

Relative sizes of gaps for different year groups

The largest point estimates of gap widening are among maths outcomes for students in Year 2, from autumn 2019 to summer 2020. This echoes other work suggesting that younger children from disadvantaged backgrounds may have been particularly affected by COVID-19 disruptions (Curriculum Associates, 2020; Juniper Education, 2020; Rising Stars, 2021; Tomasik et al., 2020). That said, we do not believe our analysis warrants strong conclusions about whether gaps have grown more for younger children. As discussed in Appendix O:, we fail to reject the hypothesis that there are no effects of year level on changes in attainment gaps for reading and maths. So, while there do appear to be subject differences in our data (between reading and maths), we do not find strong evidence that children of different ages were differentially affected by COVID-19.

RQ3 findings: understanding and interpreting the null results

Looking across all the findings for RQ3, the dominance of null results is arguably surprising. We examined what we believed were strong candidates for school-level variables that could predict changes in attainment and disadvantage gaps. This included resource-intensive efforts, such as providing live-steam videos during the initial lockdown.[50] However, it is not clear that any of these variables had strong associations with changes in attainment, or its social gradient.

[50] We acknowledge that education evaluations often provide null results so perhaps we should not be surprised. However, most evaluations study relatively limited interventions affecting a small minority of instruction time. In this case, we examined larger differences (for instance 'receiving online lessons' vs 'no lessons at all'). In other words, we were examining what we believed to be a sharper contrast than is the case in a typical evaluation.

Our starting belief was that differences in the practices adopted by schools would predict the attainment gains of their pupils during a period of school closures. This belief seemed to be shared by policymakers in England: specifically, that during the first period of closures many schools were not providing online lessons and that they should be required to do so.[51] A key question is whether our failure to find such relationships should be interpreted as a challenge to the view that these practices make a difference, or whether the null result can more readily be explained by measurement and methodological shortcomings of our study.

A notable feature of our data is the similarity in the questions between Teacher Surveys 1 and 3. These surveys ask very similar questions about the initial lockdown (March to May 2020) and the second period of widespread school closures (January to March 2021). The samples for each analysis share substantial overlap, although the Teacher Survey 3 sample is quite a bit smaller. The fact that both analyses broadly report null results increases our belief that the measures we collected are not strongly associated with changes in attainment or changes in disadvantage gaps.

Among possible explanations for these null findings, we note four in particular. First, our measures of 'school' practices are noisy. As noted in the Measures section, teachers within a school sometimes provided different responses. Ideally, our analysis would have been conducted at the class level so that we could directly link teacher reports to pupil attainment. Unfortunately, we were unable to link teacher surveys to specific students, which meant that we had to perform our analysis at the school level.[52] This introduced significant measurement error into our analysis.

Second, for some of the variables there was a lack of variation in school-level responses. In some cases, relatively few schools pursued a particular approach. For example, very few schools reported extending the school day during the autumn 2020 term (4 of 93). This limits our ability to test whether these practices were associated with changes in learning inequalities.

Third, our examination of practices during the initial lockdown involves tests taken in November and December 2019 and September 2020. The school responses we examined were only operating during a fraction of this period. Note that this reasoning is less applicable to our analysis of Teacher Survey 2 and Teacher Survey 3 as assessments were taken immediately before and after the period of schooling our survey describes.

Finally, throughout all three periods we analyse, school operations were substantially disrupted by COVID-19. Schools' ability to implement any of these responses we measure would have been limited by extremely difficult circumstances. We were not able to capture any qualitative information about the various school practices we analyse but we imagine there was substantial variation across schools in how they were implemented. This variation may also have contributed to our null results.

These methodological challenges should not be underestimated. That said, we interpret our results as providing evidence against the idea that there were strong associations between the practices captured in our teacher surveys and changes in attainment gaps. Of course, this does not rule out the possibility that there were weak associations, nor that some of these practices—if implemented well—had a strong positive causal impact on learning during the pandemic.

Lessons learned and limitations

The study was conceived in the initial lockdown (April 2020) and was set up as quickly as possible. Our goal was to recruit a longitudinal sample in a short period of time and then administer teacher surveys and tests with a minimal burden on schools. The research was also conducted in a period of considerable disruption and high rates of student absences (discussed in Appendix Q). It is important to note that the study has several important limitations.

[51] See, for example, **https://www.bbc.co.uk/news/education-54363069** , **https://www.independent.co.uk/news/uk/politics/covid-live-teaching-gavin-williamson-b1789106.html**, **https://www.thetimes.co.uk/article/education-secretary-gavin-williamson-tells-parents-to-report-schools-to-ofsted-if-online-learning-is-not-good-enough-66mv59q07**

[52] Our data provided no way of knowing which teachers taught which children, and collecting this would have been an additional burden on schools. Moreover, our teacher survey was anonymous.

First, we only have access to a non-random sample of schools and pupils. This does not affect our ability to quantify how inequalities in academic attainment have changed for this set of schools (internal validity)—but there is a risk that a similarly constructed study examining the full population may find different results (generalisability). We take some comfort from the fact that our analysis samples are similar to broader populations in terms of observable characteristics such as prior attainment (on Key Stage 1) and FSM6%. However, our geographic coverage is not as representative. Moreover, there may be unobserved differences between our sample and the population.

Second, as noted above, in our analysis of the association between school responses and changes in attainment (RQ3), some of the practices we examine were reported by a small number of schools. Moreover, the way in which practices are implemented surely impacts their effect. In a study setting where we lack information about how things were done, and where we do not have large samples, it is important not to over-interpret null results.

Third, as previously noted, there was considerable measurement error in the variables capturing school response. Unfortunately, we were unable to link teachers to specific students (thereby linking teacher practice directly with student outcomes). The best we could do was to create school-level variables. Disagreements in survey responses among teachers within schools limited our ability to draw strong conclusions from the analysis in RQ3. In future studies, we strongly recommend that researchers find a way to link teacher responses to pupil outcomes.

Fourth, we used a somewhat novel approach to collecting data for this project. We aimed to minimise administrative burdens on schools by using data from normal, routine tests and using existing tools to collect the data. There are disadvantages to this approach. Schools had discretion over when to administer such tests, and with which year groups, classes, or pupils. This limited the numbers of pupils tested at all five periods. In this final report, to maintain a large sample we have relied on imputing missing values. While our robustness checks suggest that missingness did not substantially alter our results, if we were repeating a similar exercise in future, we would recommend collecting from schools at the outset of a project a list of which tests they plan to administer (while acknowledging that in difficult circumstances, such as those presented by COVID-19, plans would likely change).

Last, we would suggest that studies of future periods of remote learning (or periods of ‘catch-up’) use the survey questions we report here *as a starting point*. Future studies will undoubtedly be interested in aspects of remote or face-to-face learning that differ from the ones we focused on.

Conclusion

Summary of findings

Research question	Finding
How did 'disadvantage gaps'—that is, attainment gaps between FSM6 pupils and their peers—change between autumn 2019 and summer 2021?	We find evidence that disadvantage gaps for primary maths have widened since the onset of the COVID-19 pandemic. On average, for Years 2 to 6, the maths gap between disadvantaged pupils and their peers widened by an estimated 0.05 ES units (with a plausible range of 0.02 to 0.08). Before the pandemic, the maths gap in our sample was 0.45 ES units, so the widening represents an increase of between 4% to 17% of the pre-COVID-19 disadvantage gap. There was no discernible change in the disadvantage gap for reading.
Were there associations between how schools responded to 'remote learning' and changes in disadvantage gaps?	We examined five variables describing primary schools' responses to remote learning: 'phoning students', 'timetabling', 'frequency of work submission', 'use of technology platforms', and providing live lessons. We examined these variables in two periods: the initial lockdown (March to May 2020) and the second period of school closures (January 2021 to March 2021). We did not find clear associations in either period between these variables and changes in disadvantage gaps.
Were there associations between school responses during the reopening period (September to December 2020) and changes in the disadvantage gap in December 2020?	We examined five variables describing primary school practice in the autumn 2020 term: 'providing videos/live streams for *absent pupils*', 'providing videos/live streams for absent *classes*', 'extra learning time', 'reducing the curriculum', and 'more small group interventions'. Overall, we did not find clear associations between these variables and changes in disadvantage gaps.

References

Allen, R., Jerrim, J. and Sims, S. (2020) 'How Did the Early Stages of the COVID-19 Pandemic Affect Teacher Wellbeing?': https://repec-cepeo.ucl.ac.uk/cepeow/cepeowp20-15.pdf

Andrew, A., Cattan, S., Costa-Dias, M., Farquharson, C., Kraftman, L., Krutikova, S., . . . Sevilla, A. (2020) 'Learning During the Lockdown: Real-Time Data on Children's Experiences During Home Learning'. https://ifs.org.uk/publications/14848

Bacher-Hicks, A., Goodman, J. and Mulhern, C. (2020) 'Inequality in Household Adaptation to Schooling Shocks: Covid-Induced Online Learning Engagement in Real Time', *Journal of Public Economics*, 193, https://www.sciencedirect.com/science/article/pii/S0047272720302097

Chetty, R., Friedman, J., Hendren, N. and Stepner, M. (2020) 'The Economic Impacts of COVID-19: Evidence From a New Public Database Built from Private Sector Data', Opportunity Insights, Cambridge, MA: National Bureau of Economic Research.

Clark, A. E., Nong, H., Zhu, H. and Zhu, R. (2021) 'Compensating for Academic Loss: Online Learning and Student Performance During the COVID-19 Pandemic', *China Economic Review*, 68, 101629.

Cook, T. D., Shadish, W. R. and Wong, V. C. (2008) 'Three Conditions Under Which Experiments and Observational Studies Produce Comparable Causal Estimates: New Findings from Within-Study Comparisons', *Journal of Policy Analysis and Management*, 27 (4), pp. 724–750.

Cullinane, C. and Montacute, R. (2020) 'COVID-19 and Social Mobility Impact Brief #1: School Closures': https://www.suttontrust.com/wp-content/uploads/2021/01/School-Shutdown-Covid-19.pdf

Curriculum Associates (2020) 'Understanding Student Needs': https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-diagnostic-results-understanding-student-needs-paper-2020.pdf

DfE (2021) 'Understanding Progress in the 2020/21 Academic Year': https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/994364/Understanding_Progress_in_the_2020_21_Academic_Year_Initial_Report_3_.pdf

Domingue, B., Hough, H., Lang, D. and Yeatman, D. (2021) 'Changing Patterns of Growth in Oral Reading Fluency During the COVID-19 Pandemic': https://files.eric.ed.gov/fulltext/ED612595.pdf

EEF (2020) 'Impact Of School Closures on the Attainment Gap: Rapid Evidence Assessment', London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/EEF_(2020)_-_Impact_of_School_Closures_on_the_Attainment_Gap.pdf

Engzell, P., Frey, A. and Verhagen, M. D. (2021) 'Learning Loss Due to School Closures During the COVID-19 Pandemic', *Proceedings of the National Academy of Sciences*, 118 (17).

Gore, J., Fray, L., Miller, A., Harris, J. and Taggart, W. (2021) 'The Impact of COVID-19 on Student Learning in New South Wales Primary Schools: An Empirical Study', *Australian Educational Researcher*, pp. 1–33.

Hammerstein, S., König, C., Dreisörner, T. and Frey, A. (2021) 'Effects of COVID-19-Related School Closures on Student Achievement—A Systematic Review': https://doi.org/10.3389/fpsyg.2021.746289

Hupkau, C. and Petrongolo, B. (2020) 'Work, Care and Gender During the Covid-19 crisis', *Fiscal Studies*, 41 (3), pp. 623–651.

Juniper Education (2020) 'National Dataset Report': https://junipereducation.org/resource/downloads/national-dataset-report/

Kogan, V. and Lavertu, S. (2021) 'The COVID-19 Pandemic and Student Achievement on Ohio's Third-Grade English Language Arts Assessment': https://glenn.osu.edu/sites/default/files/2021-09/ODE_ThirdGradeELA_KL_1-27-2021.pdf

Lucas, M., Nelson, J. and Sims, D. (2020a) 'Pupil Engagement in Remote Learning': https://www.nfer.ac.uk/media/4073/schools_responses_to_covid_19_pupil_engagement_in_remote_learning.pdf

Maldonado, J. and De Witte, K., 2020. The effect of school closures on standardised student test. *British Educational Research Journal*.

McCarty, C. and Cooke, C. (2015) *Progress in Understanding Mathematics Assessment*, Hodder Education.

McCarty, C. and Ruttle, K. (2016) *PIRA Stage Two (Tests 3–6) Manual - 2ED (Progress in Reading Assessment)*, Hodder Education.

NFER (2021) ‘Impact of School Closures and Subsequent Support Strategies on Attainment and Socio-Emotional Wellbeing in Key Stage 1’: Interim Paper 1: https://educationendowmentfoundation.org.uk/public/files/Publications/Covid-19_Resources/Impact_of_school_closures_KS1_interim_findings_paper_-_Jan_2021.pdf.

Pier, L., Hough, H., Christian, M., Bookman, N., Wilkenfeld, B. and Miller, R. (2021) ‘Evidence on Learning Loss From the CORE Data Collaborative’: https://edpolicyinca.org/newsroom/covid-19-and-educational-equity-crisis

Renaissance (2020) ‘How Kids Are Performing’: https://www.renaissance.com/how-kids-are-performing/

Rising Stars (2020) ‘The Impact of School Closures on Autumn 2020 Attainment’: https://www.risingstars-uk.com/media/Rising-Stars/Assessment/RS_Assessment_white_paper_2021_impact_of_school_closures_on_autumn_2020_attainment.pdf

Rising Stars (2021) ‘The Impact of School Closures on Spring 2021 Attainment’: https://www.risingstars-uk.com/getmedia/8181effc-58ef-48f7-9f78-94186578efa5/The_Impact_Of_School_Closures_May_2021

Sass, T. and Goldring, T. (2021) ‘Student Achievement Growth During the COVID-19 Pandemic’: https://scholarworks.gsu.edu/gpl_reports/9/

Thomson, D. (2021) ‘Do Primary Schools Close the Attainment Gap for Disadvantaged Pupils?’: https://ffteducationdatalab.org.uk/2021/03/do-primary-schools-close-the-attainment-gap-for-disadvantaged-pupils/

Tomasik, M. J., Helbling, L. A. and Moser, U. (2020) ‘Educational Gains of In-Person vs. Distance Learning in Primary and Secondary Schools: A Natural Experiment During the COVID-19 Pandemic School Closures in Switzerland’, International Journal of Psychology: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7753520/

van der Weele, T. J. and Mathur, M. B. (2019) ‘Some Desirable Properties of the Bonferroni Correction: Is the Bonferroni Correction Really So Bad?’, *American Journal of Epidemiology*, 188 (3), pp. 617–618.

Weidmann, B. and Miratrix, L. (2020) ‘Lurking Inferential Monsters? Quantifying Selection Bias in Evaluations of School Programs’, *Journal of Policy Analysis and Management*. 40(3), pp.964-986

Wong, V. C., Valentine, J. C. and Miller-Bains, K. (2017) ‘Empirical Performance of Covariates in Education Observational Studies’, *Journal of Research on Educational Effectiveness*, 10 (1), pp. 207–236.

Appendices

Appendix A: Ethics and data protection procedures

Ethics

FFT was responsible for school recruitment and data collection. Schools opted in to the project via an online form that was completed by the headteacher. The online form was sent in an invitation email to 800 FFT Aspire schools that we believed would also have Rising Stars assessment data.[53] The email outlined the project aims and requirements and invited schools to complete an online form with the contact information required for the project. Additionally, it required headteachers signing up to tick a box confirming their schools participation in the project as well as agreeing the data already shared with FFT can be used for this research.

Due to the speed required for project setup, participating schools were required to be an FFT Aspire customer as the project relies on the data schools already share with FFT and the terms and conditions schools sign up to for FFT Aspire to enable the research and analysis to take place. These terms and conditions enable FFT to use schools' data to undertake research. The welcome email along with the online form headteachers completed to be a part of the research can be seen in Appendix B:.

Data protection

The legal basis for processing the personal data used in this research project is legitimate interest Article 6(1). FFT Aspire customers agree to Terms and Conditions and Terms of Use that outline how data shared with FFT can be used to undertake research into how education systems function. This is:

- Necessary for our customers' legitimate interests (analysis of pupil performance and requirements, school performance, staff performance and ensuring equality of opportunity and treatment of pupils)
- Necessary for our legitimate interests (to be able to undertake research for public benefit)
  Link to privacy notice can be found here: **https://fft.org.uk/privacy/**

Participating schools have volunteered to take part in this research and understand that the data shared with FFT Aspire will also be used for this research.

The personal data for this research will only be accessible by FFT colleagues through FFT's IT infrastructure. FFT is certified and accredited to ISO 27001 (the international standard for Information Security) and is registered as a data controller with the Information Commissioner's office. FFT is also certified to the government's Cyber Essentials Plus standard and completes an annual IT Health Check of our complete internal and external-facing IT systems, websites and infrastructure which is reviewed and approved by the Department as part of our accreditation to receive DFE ASP data.

Data being used for this project is defined as Customer Data (as noted in FFT's privacy policy). This means that the customer is the Data Controller and FFT is the Data Processor. The data is only accessible by FFT colleagues through FFT's IT infrastructure and all pupil data is encrypted both in transit and at rest. No pupil data was shared with collaborators at the EEF or TeacherTapp.

Customer data is provided by schools for use within the Aspire platform, subject to the terms and conditions agreed between the customer and FFT. Data retention periods for Customer data are detailed in the FFT privacy notice. Customer Data will be deleted no later than 48 months after termination of a customer's subscription term.

[53] FFT provided Rising Star with a list of schools that are FFT Aspire customers. Rising Stars then used this list to indicate schools that had purchased Rising Stars assessments in previous years and in 2020.

Appendix B: Recruitment documents

Link to email sent to schools to invite them into the project: **https://mailchi.mp/fft/action-required-ffteef-covid-19-research-project?e=%5bUNIQID%5d**

Dear colleague,

I am emailing to request your school's participation in a new national research project investigating **the impact of Covid-19 and school closures on pupil attainment.** The project is being run jointly by FFT and the Education Endowment Foundation (EEF) and your school is of particular importance for the research as you already use FFT Aspire and the standardised tests from RS Assessment (e.g. PIRA, PUMA or NTS tests).

Participating in the research project will be straightforward. Over the next 6 months, we will just need your school to:

1. Have some **baseline data** from RS Assessment tests (PIRA/PUMA/NTS) from December 2019 and test these same pupils using **standardised tests** (PIRA/PUMA/NTS) from RS Assessment in September 2020 and December 2020.
2. Ask your teachers and headteacher to complete a **short 10-minute survey** about practices when your school was closed.
3. Give permission for FFT to use your **school's pupil data in FFT Aspire** including the data which we can access from your school MIS and the RS Assessment MARK platform.

Your school will benefit from collaborating with FFT and EEF on this research:

- Each school will **receive a £250 payment** as a thank you for your participation
- Each school will receive a **'special report' in FFT Aspire** which analyses the impact of COVID-19 on the attainment of your pupils compared to other schools

Your school can choose the number of year groups that participate in this research. You can have one year group participating or multiple year groups.

To register your school for the programme, please click the button below to complete the short online registration form by **15th July**.

Register my school >

If you have any questions or would like any further information about the research project, please do get in touch with Laura James at FFT by emailing: **Covid-1919research@fft.org.uk**.

I hope that you will be able to collaborate with us on this important piece of research. I know that the research will provide a unique and valuable insight into the impact of COVID-19 on pupil outcomes for primary schools.

Many thanks,

Paul Charman
Managing Director, FFT

Link to online sign up form: **https://fftedu.wufoo.com/forms/ffteef-Covid-1919-research-project-copy/**

Text from the school sign up form can be seen below

FFT/EEF COVID-19 Research Project Copy

Please read and complete this short form to register your school for the FFT/EEF COVID-19 research project.

Please confirm that you will be able to share the following data and information with FFT for the research project: *

☐ ☐ Baseline data – you must have November/December 2019 RS Assessment (PIRA, PUMA, & GAPS or NTS Assessments)

☐ ☐ Outcome data – we will need you to complete two further RS Assessments for these same pupils (September 2020 and November/December 2020)

☐ ☐ Teacher survey – Short survey with headteacher and class teachers about the learning provision for pupils while schools have been closed

☐ ☐ Data collection – Data Exchange connection between FFT Aspire and your School MIS and enabling test data sharing in MARK to FFT Aspire Pupil Tracking

School name *

School DFE number (7 digits) *

Enter a value between **1000000** and **9999999**.

Local Authority *

Barking and Dagenham | Barking and Dagenham

Lead contact within your school for this research project: * First Last

Email address for the lead contact: *

Lead contact: job role *

- Headteacher / senior leader
- Middle leader
- Teacher
- School administrator

Year groups to be included in the research

For each year group participating in the research we will need:

(1) PIRA/PUMA/GAPS/NTS Assessments test results in MARK ;

(2) your teachers to complete a short survey

Please select the year groups which will be able to participate in the research: *

- Current Year 1 (Year 2 in Sept)
- Current Year 2 (Year 3 in Sept)
- Current Year 3 (Year 4 in Sept)
- Current Year 4 (Year 5 in Sept)
- Current Year 5 (Year 6 in Sept)

Headteacher and Teacher survey: email addresses

We will need your headteacher and teachers to complete a short 10 minute online survey about teacher practices during the partial school closures.

Headteacher name *First Last

Headteacher email address *

Year 1 teachers

Please provide the name and email address for each of your Year 1 teachers

Name First Last

Email address

Name First Last

Email address

Name First Last

Email address

Year 2 teachers

Please provide the name and email address for each of your Year 2 teachers

Name First Last

Email address

Name First Last

Email address

Name First Last

Email address

Year 3 teachers

Please provide the name and email address for each of your Year 3 teachers

Name First Last

Email address

Name First Last

Email address

Name First Last

Email address

Year 4 teachers

Please provide the name and email address for each of your Year 4 teachers

Name First Last

Email address

Name First Last

Email address

Name First Last

Email address

Year 5 teachers

Please provide the name and email address for each of your Year 5 teachers

Name First Last

Email address

Name First Last

Email address

Name First Last

Email address

IMPORTANT: permission to participate in the research project

Name of person completing this form for the school *First

Last

Job role of person completing this form *

- Headteacher / senior leader
- Middle leader
- Teacher
- School administrator

By ticking this box, you confirm that you are authorised on behalf of your school to participate in the FFT/EEF COVID-19 research project. You confirm that the pupil and teacher information shared with FFT can be used for this research project under the GDPR-compliant data sharing agreement in place between FFT and the school for FFT Aspire. *

- I agree that my school can participate in this research project

Appendix C: Collecting data

Attainment data
For the purposes of reporting and analysis, schools in the sample used either (or both) MARK (RS Assessment's online analysis tool) or FFT Aspire Pupil Tracking (APT). We collected data from MARK via an interface (API) once schools authorised us to do so. Data was directly extracted from APT, again from schools authorising us to do so. In exceptional circumstances, schools supplied us with some test data in Excel format.

The key variables extracted from the MARK API for each pupil were:

- Test suite (PIRA/ PUMA/NTS Reading/ NTS maths)
- National curriculum year (NCY) of the test
- Term of the test (Autumn, Spring, Summer)
- Standardised score
- Test date

Similar variables were also extracted from APT.

Pupil characteristics
Data on pupil characteristics, such as disadvantage, gender and ethnicity, were collected directly from schools' management information systems (MIS) using Aspire Data Exchange (ADX). Weekly extracts were taken from September 2020 onwards. However, some schools only authorised collection later in the Autumn 2020 term. We did not have the facility to retrospectively collect data.

Published School Level Data
Further relevant school characteristics to control for in models, such as inspection ratings, region, historic attainment and percentage disadvantage were sourced from published government datasets.

Cleaning and Transformation

A pseudonymised dataset was constructed from the source data in order to conduct analysis. This involved a number of stages to select, clean and transform the data to make it suitable for analysis. Stages of data cleaning included:

- Identifying relevant test data
- Adding linked pupil characteristic data
    - Determine which FSM6 flag to use
- Equating test scores

Identifying relevant test data
We first identified relevant test data in both MARK and APT. This meant identifying:

- The type of test (e.g. PIRA, PUMA) and the subject
- The national curriculum year of the test
- The term of the test
- The date it was taken

In some cases, we found that data on the term of test clashed with the date it was taken (e.g. summer tests taken in November). This was particularly the case in data sourced from APT. We adjudicated based on the user-provided test name field.

Where there were multiple results for a pupil for a test, we gave precedence to results that were in the date window we expect, have a UPN, have a non-null score and have the closest NCY to that of the pupil at the time they took the test.

The process we followed was as follows:

- Identify relevant tests taken by pupils in Years 2 to 5 in Autumn 2019
- Identify relevant tests taken by pupils in Years 3 to 6 in 2020/21 academic year
- Criteria for determining the year/ subject/ term of test
    - Subject – test type parsing (e.g. PUMA/NTSM tests are classified as maths)
    - Term – MARK and Excel data use the term as provided from source (the term field); APT looks at whether the person providing the data commented on the term they were submitting for in the test_name field or uses the date window otherwise
    - Test NCY – MARK uses the ncyear_id field; APT is a parse of the test_name or falls back on the ncyear_id field (from test source) – these are then compared with the pupil's NCY in school roll data to determine which is more likely.

Linking pupil characteristic data

Test data was linked to data on pupil characteristics using UPN and school identifier (DfE number).

From the weekly snapshots of data collected since the start of Autumn 2020 we:

- Identified those on roll at each participating school. This includes those who were on roll at the start of the 2019/20 academic year, and anyone else who appears in test data
- Classified any pupils flagged as FSM6 at any point in the Autumn 2020 term as disadvantaged
- Classified pupils according to their SEN status in the first snapshot from the Autumn 2020 term
- Identified other characteristics (ethnicity, EAL, gender, month of birth) based on the modal values across all snapshots.

Appendix D: Teacher survey 1

Table 10 – Survey items relating to learning during the first half of Summer term 2020

1.	How did you set work for your class to complete? Tick the response that best describes how MOST families would usually find out what work is set. • By post or collection from the office • Via a page on the school website • Via an email • Via an online learning or communication platform (e.g. Google classroom, Microsoft Teams, Class Dojo, Seesaw, Firefly etc…) • Via a social media site (e.g. a school Facebook page) • Via a video or phone call/lesson • Some other way • Not relevant – No work was set for my class	`comms_techplatform` Coded 1 if set or received work via an online learning or communication platform. Coded 0 otherwise, including if no work is set.
2.	How did pupils send completed work back to you? Tick the response that best describes how families would usually send you work. • By post or collection from the office • Via an email • Via an online learning or communication • platform (e.g. Google classroom, , Microsoft Teams, Class Dojo, Seesaw, Firefly etc…) • Via a social media site (e.g. a school Facebook page) • Via a video or phone call/lesson • Some other way • Not relevant – No work was sent in from families or I had no class	

3.	How much did you try to stick to your school's pre-existing curriculum? Tick the response that most closely applies.
	• We tried to stick to the content and pace of our school's curriculum as much as possible
	• We SOMEWHAT reduced the content and/or the pace of the school's curriculum during lockdown
	• We LARGELY PAUSED or ENTIRELY STOPPED the school's curriculum during lockdown and switched mostly to revisiting old topics and retrieval activities
	• We LARGELY PAUSED or ENTIRELY STOPPED the school's curriculum during lockdown and aligned with another curriculum (e.g. Oak National or BBC Bitesize topics)
	• None of the above statements align with our practice during lockdown
	• Not relevant / cannot answer
		`curric_div_slowed`
		Coded 1 if teacher some reduced content or slowed pace (0 otherwise).
		
		`Curric_div_paused`
		Coded 1 if teacher largely paused or stopped curriculum to align with other curriculum resources (0 otherwise).
		
		`Curric_div_retrieval`
		Coded 1 if teacher largely switched to revisiting topics or retrieval practice (0 otherwise).

4.	Which of the following resources have you used as part of the home learning you set for students? Tick all that apply
	• An app or online SUBSCRIPTION education site (e.g. Numbots)
	• An app or online education site WITHOUT subscription needed
	• Videos or streamed content of you teaching
	• Worksheets and tasks that you or your colleagues have created
	• Worksheets and tasks created by someone outside your school
	• BBC Bitesize shows or resources
	• Oak National Academy resources
	• None of the above
	• Not relevant / cannot answer
		`res_site_subs`
		`res_site_free`
		`res_bbc`
		`res_video`
		`res_sheet_ext`
		`res_sheet_int`
		`res_oak`
		Seven binary indicators of whether they used the resource.

5.	Thinking back to the literacy or English resources you provided for home learning, were they originally developed for classroom teaching or for parents to use? Tick the response that most closely applies.
	• Majority of resources and activities were adapted from classroom teaching activities
		`resource_adapt_fully`
		Coded 1 if majority of resources and activities were created with parents in mind (0 otherwise).
		
		`Resource_adapt`

	• Majority of resources and activities were created with parents in mind • Even balance – some were created for parents and some for teachers • Not relevant / cannot answer	Coded 1 if either even balance of adaption or fully adapted for parents (0 otherwise).
6.	Thinking back to the literacy or English resources you provided for home learning, could children complete the activities you set without supervision from their parents? Tick the response that most closely applies. • Children could complete all the activities without parental supervision • Children could complete many of the activities without parental supervision • Children could complete some of the activities without parental supervision • Children could not generally complete the activities unless they had parental supervision • Not relevant / cannot answer	childindependent Scale of 0-3 where 0 indicates parental supervision needed for everything and 3 indicates all activities designed for working independently.
7.	Did you have any form of daily registration for students learning at home? Tick the response that most closely applies. • Yes, we asked pupils to log onto an online platform or send an email (or similar) every school day • No, but we suggested pupils to log onto an online platform or send an email (or similar) every school day • No daily registration • Not relevant / cannot answer	interaction_register Binary indicator coded 1 if a required or suggested daily check-in system was in place.
8.	Did you provide students with a timetable to follow each day (this could be hourly, lesson-by-lesson, or a daily list)? • Yes, we asked pupils to follow a daily timetable • Yes, but it was only a suggested timetable and pupils were not required to follow it • No • Not relevant / cannot answer	interaction_timetable Binary indicator coded 1 if there was a suggested or required timetable to follow. Interaction_timetable_high Binary indicator coded 1 if it was a require timetable to follow.
9.	Were pupils able to do any of the following with their class teacher during the first half of Summer term? (Tick any that apply) • Take part in a 'live' online lesson where they could talk • Take part in a 'live' online lesson where they could not talk • Take part in a 'live' social video chat or check-in time • Take part in a text-based online chat where they type conversation • Watch a pre-recorded video of their teacher talking • None of the above were possible • Not relevant / cannot answer	interaction_livelesson interaction_livechat interaction_textchat interaction_video Four binary indicators of how students could interact with their teacher during lockdown.
10.	Did children (i.e. not the parents) typically speak to you (i.e. their class teacher) on the phone? • No, children typically did not speak to their class teacher • Yes, one or twice during the half term • Yes, about once a week • Yes, more than once a week • Not relevant / cannot answer	interaction_phone Binary indicator for whether student (not parent) spoke to teacher on the phone. Interaction_phone_regular Binary indicator for whether student spoke weekly to the teacher on the phone.
11.	Did you share examples of student work somewhere for all students to see? Tick any that apply. • Yes – on school website • Yes – in school newsletter sent by email • Yes – in an email to the class (or similar) • Yes – within our online learning platform • Yes – via social media (e.g. a facebook page or twitter) • Yes – somewhere else not listed above • No • Not relevant / cannot answer	interaction_worksharing Binary indicator for whether the school shared examples of student work by any means.
12.	Which best describes the type of feedback you felt able to give on pieces of work submitted by pupils? • I didn't give feedback on individual pieces of work • Feedback was essentially all general praise and encouragement	feedback_praise Binary indicator for whether feedback given was mostly praise and encouragement. Feedback_specific

	• I gave specific feedback intended to support learning on SOME pieces of work • I gave specific feedback intended to support learning on MANY pieces of work • I gave specific feedback intended to support learning on all/ALMOST ALL pieces of work • Not relevant / cannot answer	Binary indicator for whether teacher gave any specific feedback on work intended to support learning.

13.	How frequently did you suggest that parents or pupils send in work?
	• Every day (or after each lesson)
	• Several times a week
	• Once a week
	• Once a fortnight
	• Less than once a fortnight
	• I gave no suggestion about how frequently work should be sent in
	• Not relevant / cannot answer
		`feedback_given`
		Binary indicator for whether or not the teacher encouraged work to regularly be submitted.
		
		`Feedback_given_daily`
		Binary indicator for whether or not the teacher encouraged the daily submission of work.

14.	How much do you agree with the following statement: "During lockdown whilst most of my students were learning at home, it was easy for me to monitor who was, and wasn't, completing work." • Strongly agree • Somewhat agree • Slightly agree • Slightly disagree • Somewhat disagree • Strongly disagree • Cannot answer / not relevant	`monitoring_ease` Scale from 1-6 where 6 indicates high ease in monitoring students.
15.	How much do you agree with the following statement: "When setting work for remote learning, I found it difficult to differentiate to the lowest attainers in my class." • Strongly agree • Somewhat agree • Slightly agree • Slightly disagree • Somewhat disagree • Strongly disagree • Cannot answer / not relevant	`differentiate_ease` Scale from 1-6 where 6 indicates that the teacher found differentiation a challenge during lockdown.
16.	Overall, how good do you feel the home learning experience was for your class? Tick the response that most closely aligns with your feelings. • Very successful – all or almost all pupils were consistently completing school work during the first half of Summer term • Successful – the majority of pupils seemed to be completing school work • Mixed – whilst many pupils did seem to be completing school work, a significant portion of the class were not doing so much • Not so good overall – most pupils were clearly completing far less work than we had set for them • Other…	`Outcome` Coded from 0 to 3 but not used in analysis.

Notes: (1) Cannot answer / not relevant responses generally coded as missing
(2) Preceding these questions, information on the teacher's school, class and job role were collected.

Table 11: Rotated factor loadings (pattern matrix) – Variables created from Survey 1

Variable	Factor1 λ=2.95	Factor2 λ=1.29	Factor3 λ=1.16	Factor4 λ=1.12	Factor5 λ=0.97	Factor6 λ=0.81	Factor7 λ=0.61	Factor8 λ=0.45
comms_techplatform	-0.03	**0.27**	-0.01	0.02	-0.08	0.02	-0.02	-0.16
curric_div_retrieval	0.00	0.05	**0.37**	0.02	0.07	-0.09	0.00	0.01
res_site_subs	0.00	-0.01	0.05	0.02	-0.03	0.08	0.04	**0.16**
res_site_free	0.02	-0.02	-0.04	0.03	-0.03	**0.19**	0.04	0.03
res_bbc	0.00	0.01	0.01	-0.02	0.00	**0.36**	-0.02	0.03

res_video	-0.01	-0.03	0.00	-0.04	0.06	0.03	**0.35**	0.04
res_sheet_ext	-0.01	0.01	0.00	0.01	-0.04	**0.18**	0.13	0.00
res_sheet_int	-0.02	0.03	-0.02	0.00	-0.01	0.02	0.10	0.00
res_oak	0.02	0.00	0.02	0.01	0.07	**0.22**	-0.10	-0.13
resource_adapt	0.00	0.02	0.01	**0.44**	0.01	0.03	0.03	0.00
resource_adapt_fully	0.01	0.03	-0.03	**0.44**	0.01	-0.01	-0.03	-0.03
childindependent	0.00	0.09	0.02	-0.06	0.02	0.02	-0.07	-0.07
interaction_register	0.02	**0.11**	-0.01	0.04	0.10	0.01	-0.05	0.04
interaction_timetable	0.00	0.02	-0.03	0.03	0.03	0.03	0.03	**0.24**
interaction_timetable_high	0.00	0.07	-0.06	0.00	0.07	-0.02	-0.07	**0.22**
interaction_livelesson	-0.02	0.05	-0.02	-0.03	**0.16**	0.03	-0.06	0.03
interaction_livechat	-0.02	**0.07**	0.02	-0.02	0.04	0.10	0.04	0.04
interaction_textchat	0.00	**0.11**	0.04	-0.02	-0.03	0.00	0.00	-0.07
interaction_video	0.01	0.01	0.00	0.03	-0.03	-0.06	**0.34**	-0.05
interaction_phone	0.00	-0.03	0.00	0.00	**0.35**	0.07	0.00	-0.09
interaction_phone_regular	0.01	-0.04	0.03	0.02	**0.33**	-0.08	0.03	0.07
interaction_worksharing	-0.02	0.01	0.00	-0.03	**0.11**	0.02	0.06	0.01
feedback_specific	**0.52**	0.09	0.02	-0.01	-0.01	-0.14	0.14	-0.20
feedback_praise	**-0.46**	0.27	-0.02	-0.04	0.05	-0.09	0.15	-0.14
feedback_given	0.01	**0.19**	-0.01	0.00	-0.03	-0.06	-0.05	0.20
feedback_given_daily	-0.02	**0.23**	0.03	0.01	-0.02	0.04	-0.02	0.12
monitoring_ease	-0.02	**0.25**	0.01	0.01	0.02	0.02	0.06	-0.11
differentiate_ease	0.01	-0.02	0.04	0.00	-0.10	0.02	0.01	**0.14**
curric_slowdown	0.02	0.00	**0.50**	-0.03	-0.02	0.04	0.01	0.04

Appendix E: Teacher survey 2

Table 12 – Mapping of Teacher Survey 2 questions to variables used for analysis

Survey question	Responses (and coding)	Variable
1. Which of the following statements are true about your class and school absence so far since September?	2=no class isolation needed within school; 1=class isolation with live lessons; 0=class isolation without live lessons	Absence provision Covid-19
2. Which of the following types of provision were available for an individual child who needed to isolate whilst their peers were in class?	0=no access to any video/streaming of class; 1=video/streaming of class	Class provision for Covid-19
3. Has your school done any of the following to compensate for lost learning time last year?		
3a. Removing items from the curriculum	0=no; 1=yes	Reduced curriculum
3b. Running extra lunchtime learning activities	0=no; 1=yes	Extra time
3c. Extending the school day	0=no; 1=yes	Extra time
3d. Being open during holidays	0=no; 1=yes	Extra time
3e. Small group interventions DURING lessons	0=no; 1=yes	Small group interventions
3f. Small group interventions OUTSIDE lesson time	0=no; 1=yes	Small group interventions
4. How many extra minutes a day (over and above a normal year) were you spending on English and maths last term to cope with missed learning as a result of shutdown and absences? Tick the response that most closely applies.	0=no extra time; 1=15 mins a day; 2=30 mins a day; 3=45 mins a day or more	Extra time
5. Was your school making greater or less use of small group and individual face-to-face interventions (e.g. run by TAs) last term?	0=fewer interventions; 1=same number as normal; 2=more interventions	Small group interventions

Table 13 – Correlation of items from Teacher Survey 2

Variable description		q1*	q2	q3a	q3b	q3c	q3d	q3e	q3f	q4
class_absence provision	q2	0.06								
curriculum (remove items)	q3a	-0.06	-0.12							
extra time (lunch)	q3b	0.02	-0.12	0.1						
extra time (school day)	q3c	0.01	0.09	0.02	0.01					
extra time (holidays)	q3d	0.13	0.01	0.08	0.1	0.16				
small group (during lessons)	q3e	-0.15	0.08	0.15	0.07	0.08	0.14			
small group (outside lesson time)	q3f	0.2	-0.16	0	0.17	-0.06	0.04	0.27		
extra 51nglish/maths	q4	0.01	0.12	0.44	-0.01	0.18	0.15	0.06	-0.09	
extra small group interventions	q5	0.16	0.21	-0.11	0.18	-0.13	0.1	0.45	0.44	-0.11

Notes: q1* is “pupil_absence provision

Appendix F: Teacher survey 3

Question	Responses (and coding)	Variable
1. Did you use a technology platform (e.g. Google classroom, Microsoft Teams, Class Dojo, Seesaw, Firefly etc…) to communicate with families in any of the following ways? You may tick more than one response	• To set pupils' work • To receive pupils' work submissions • To share examples of student work with everyone • We held a daily registration or check-in system • Parents or pupils typed messages for me reply to via the platform • I spoke to parents or pupils using video/audio chat via the platform Any of the above responses received a +1 in the final variable. There was also a final option: • No, I did not use a technology platform for communication with families	Platform Use
2. Did you provide students with a timetable to follow each day (this could be hourly, lesson-by-lesson, or a daily list)?	• Yes, we asked pupils to follow a daily timetable (=2) • Yes, but it was only a suggested timetable and pupils were not required to follow it (=1) • No (=0)	Timetable
3. How frequently did you suggest that parents or pupils send in work?	• Every day, or after each lesson (=3) • Several times a week (=2) • Once a week or less (=1) • I gave no suggestion about how frequently work should be sent in (=0)	Work Submission
4. Did children (i.e. not the parents) typically speak to you (i.e. their class teacher) on the phone?	• No, children typically did not speak to their class teacher (=0) • Yes, I spoke to them each once, twice or several times during the half term (=1) • Yes, I spoke them to each about once a week (or more frequently) (=2)	Phone students
5. Did you try teaching a 'live' lesson to your class in January, February or March (with students either speaking or on mute)?	• No OR No, but I did pre-record a video of myself for them to watch (=0) • Yes, I held one, two or several live lessons during this time (=1) • Yes, I held a live lesson about once a week* (=1) • Yes, I held one live lesson most days (=2) • Yes, I held more than one live lesson most days (=3)	Live Lessons

Notes: *This category had a small number of respondents, so we elided it with the next closest category in our coding. **In addition to the 5 core questions which we derived from Teacher Survey 1 and listed in this table, Teacher Survey 3 included 5 additional questions. In this report, we follow our analysis plan and focus on the five constructs that are analogous to those analysed in Teacher Survey 1.

Appendix G: Example school report

education datalab

COVID-19 Disruption

Primary School (9999000)

Autumn 2019

120 primary schools participated in a project to examine the impact of COVID-19 disruption on the attainment gap between disadvantaged pupils and their peers in 2020/21. All these schools used tests provided by Rising Stars Assessments (PUMA, PIRA and NTS).

This report provides some summary data about the test scores of pupils who took the tests at your school in each year group in Autumn 2019 and Summer 2020. We only include pupils who have taken the expected test for their national curriculum year. Comparisons are made to a) a wider sample of schools that administered Rising Stars tests and b) the set of other schools which participated in the study.

Standardised Scores: Autumn 2019 - Summer 2020

The tables below shows the average standardized score of pupils who took tests at your school in each year group. You can compare results at your school to a wider national sample of schools that administer Rising Stars tests.

Reading

Year group	Subject	Term	Test	Number of pupils	Standardised score	National score
Year 2	Reading	Autumn 19	PiRA 2016	15	103.0	105.2
		Summer 20	PiRA 2016	15	111.6	95.6
Year 3	Reading	Autumn 19	PiRA 2016	16	103.1	104.5
		Summer 20	PiRA 2016	17	108.2	96.9
Year 4	Reading	Autumn 19	PiRA 2016	15	102.7	102.7
		Summer 20	PiRA 2016	16	99.1	97.3
Year 5	Reading	Autumn 19	PiRA 2016	19	91.2	102.3
		Summer 20	PiRA 2016	18	95.8	96.7
Year 6	Reading	Autumn 19	PiRA 2016	14	90.8	103.0
		Summer 20	PiRA 2016	15	98.0	99.0

* results suppresed due cohort size below 5

Maths

Year group	Subject	Term	Test	Number of pupils	Standardised score	National score
Year 2	Maths	Autumn 19	PUMA 2016	15	117.0	103.4
		Summer 20	PUMA 2016	15	107.9	98.1
Year 3	Maths	Autumn 19	PUMA 2016	16	106.9	101.9
		Summer 20	PUMA 2016	17	98.5	94.3
Year 4	Maths	Autumn 19	PUMA 2016	15	103.0	104.5
		Summer 20	PUMA 2016	16	102.8	98.1
Year 5	Maths	Autumn 19	PUMA 2016	19	98.3	103.3
		Summer 20	PUMA 2016	18	91.2	99.1
Year 6	Maths	Autumn 19	PUMA 2016	14	98.4	103.4
		Summer 20	PUMA 2016	15	95.5	100.8

* results suppresed due cohort size below 5

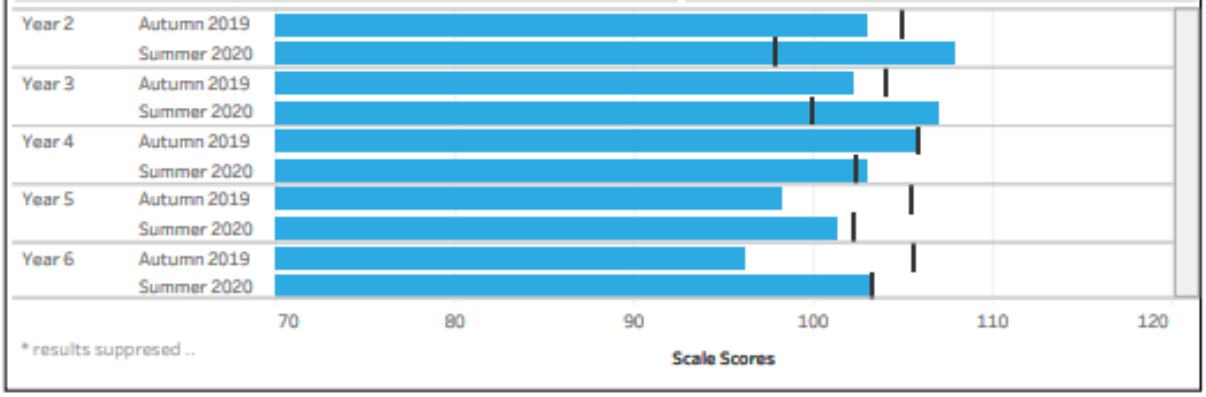
COVID-19 Disruption
Primary School (9999000)

Autumn 2019
Summer 2020

In Aspire Pupil Tracking, we convert standardized scores in different tests into a common currency to enable comparability. For Year 1 and Year 2 tests we use Key Stage 1 scaled scores and for Years 3 to 6 we use Key Stage 2 scaled scores. Note that KS1 and KS2 scaled scores are not directly comparable. Below we show the scaled scores for matched pupils. These are pupils who took tests in both Autumn 2019 and Summer 2020. We have also included averages for pupils in other schools that participated in the study where pupils have entered both tests. This information is presentedin the tables and graphically in the charts below. The black bar..

Reading Scaled Scores

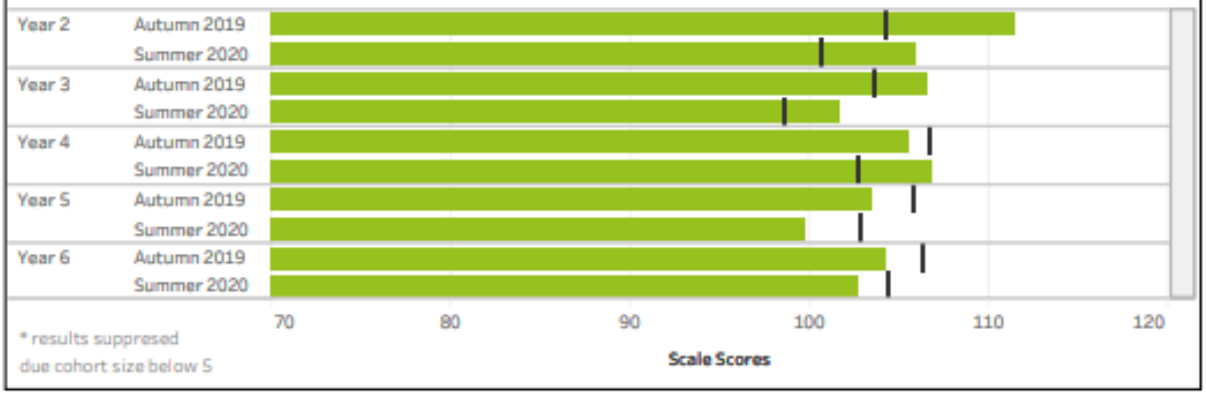
Year Group	Matched pupils	Autumn 2019	Summer 2020	Difference	Other participating schools (reading) Autumn 2019	Summer 2020	Difference
Year 2	14	102.9	107.9	+5.0	104.8	97.7	-7.1
Year 3	16	102.2	106.9	+4.7	104.0	99.8	-4.1
Year 4	15	105.7	103.0	-2.7	105.8	102.3	-3.5
Year 5	18	98.2	101.3	+3.1	105.4	102.2	-3.2
Year 6	14	96.1	103.2	+7.1	105.5	103.2	-2.3



* results suppresed ..

Maths Scaled Scores

Year Group	Matched pu..	Autumn 2019	Summer 2020	Difference	Other participating schools (maths) Autumn 2019	Summer 2020	Difference
Year 2	14	111.4	105.9	-5.5	104.2	100.6	-3.6
Year 3	16	106.5	101.6	-4.9	103.5	98.5	-5.0
Year 4	15	105.5	106.8	+1.3	106.6	102.6	-3.9
Year 5	18	103.4	99.7	-3.7	105.8	102.8	-3.0
Year 6	14	104.2	102.7	-1.5	106.3	104.3	-2.0



* results suppresed
due cohort size below 5

education datalab

COVID-19 Disruption
Primary School (9999000)

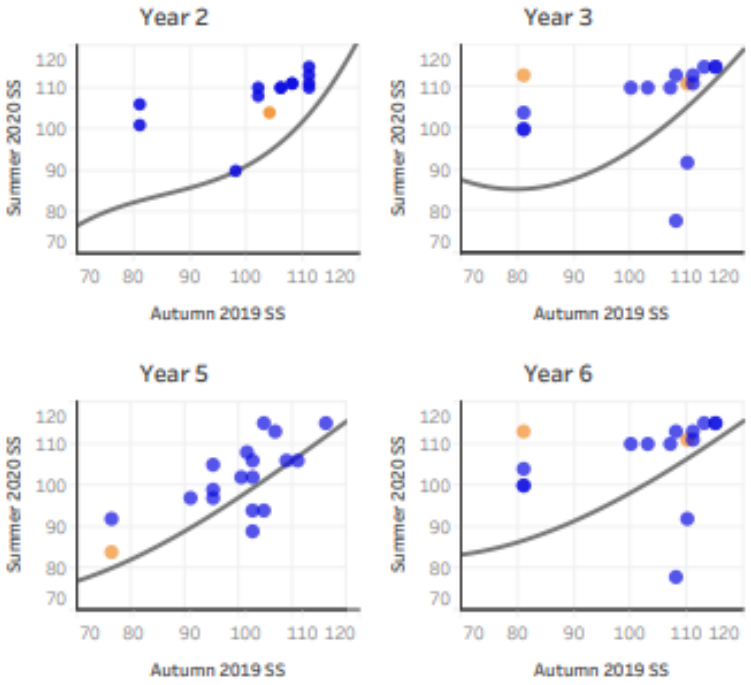
Autumn 2019

The scatterplots show the Autumn 2019 and Summer 2020 scaled scores for pupils at your school. We have colour-coded them according to whether they were flagged as disadvantaged (FSM6/Pupil Premium) in the MIS data we collected from your school. Approximate progression lines are also shown. They show an approximate average scaled score for Summer 2020 based on Autumn 2019 scaled score for pupils in schools which participated in the study.

Reading

Year Group	Matched pupils
Year 2	14
Year 3	16
Year 4	15
Year 5	18
Year 6	14

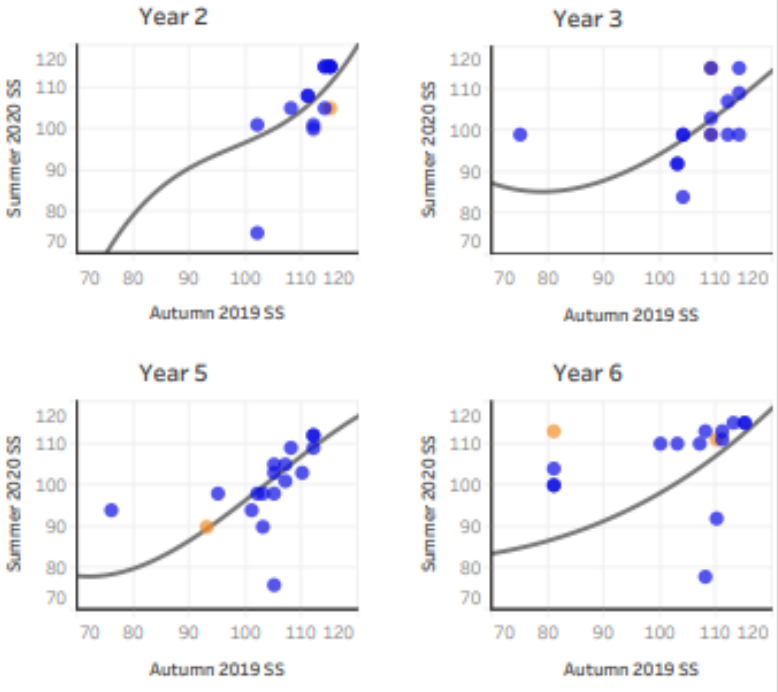
Not disadvantaged
Disadvantaged



Maths

Year Group	Matched pupils
Year 2	14
Year 3	16
Year 4	15
Year 5	18
Year 6	14

Not disadvantaged
Disadvantaged



Appendix H: Achieved samples and participant flow diagrams for RQ3

Figure 6 – Overview of sample for RQ3a

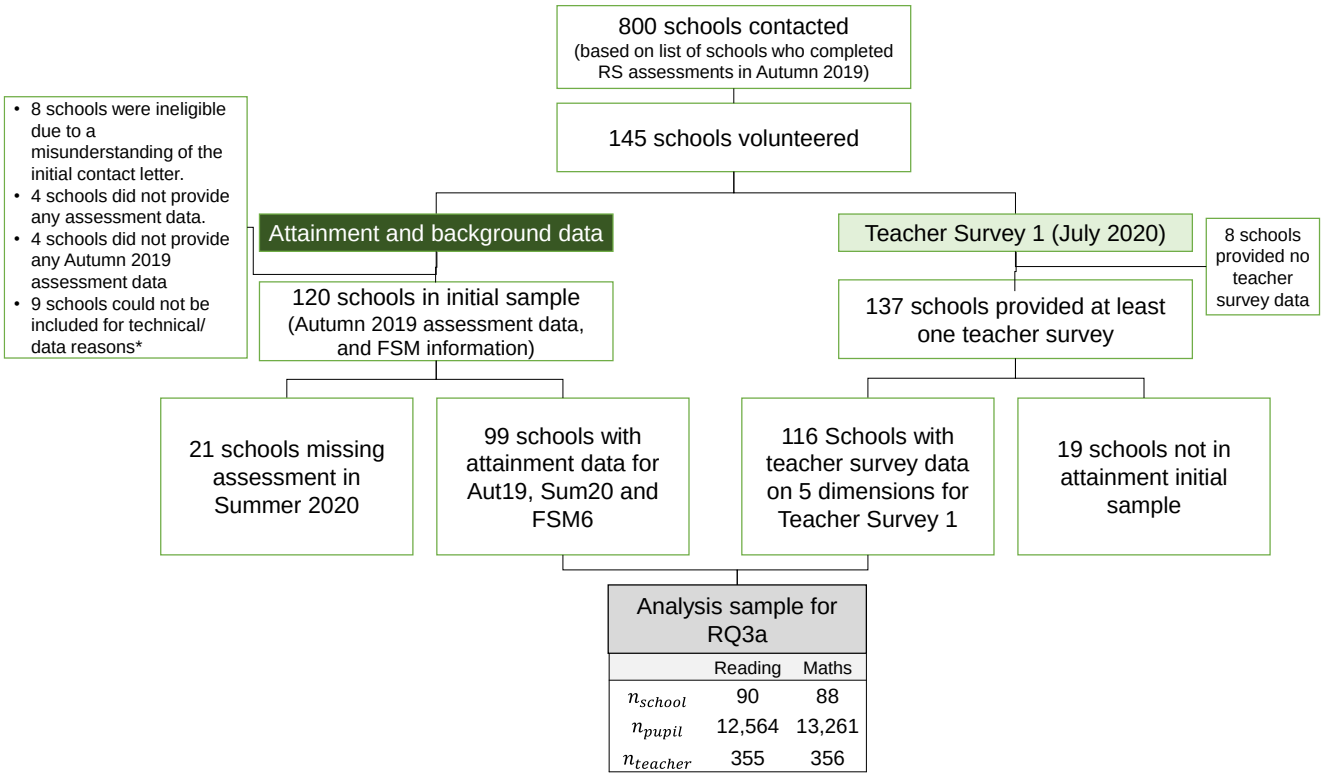


Figure 7 – Overview of sample for RQ3b

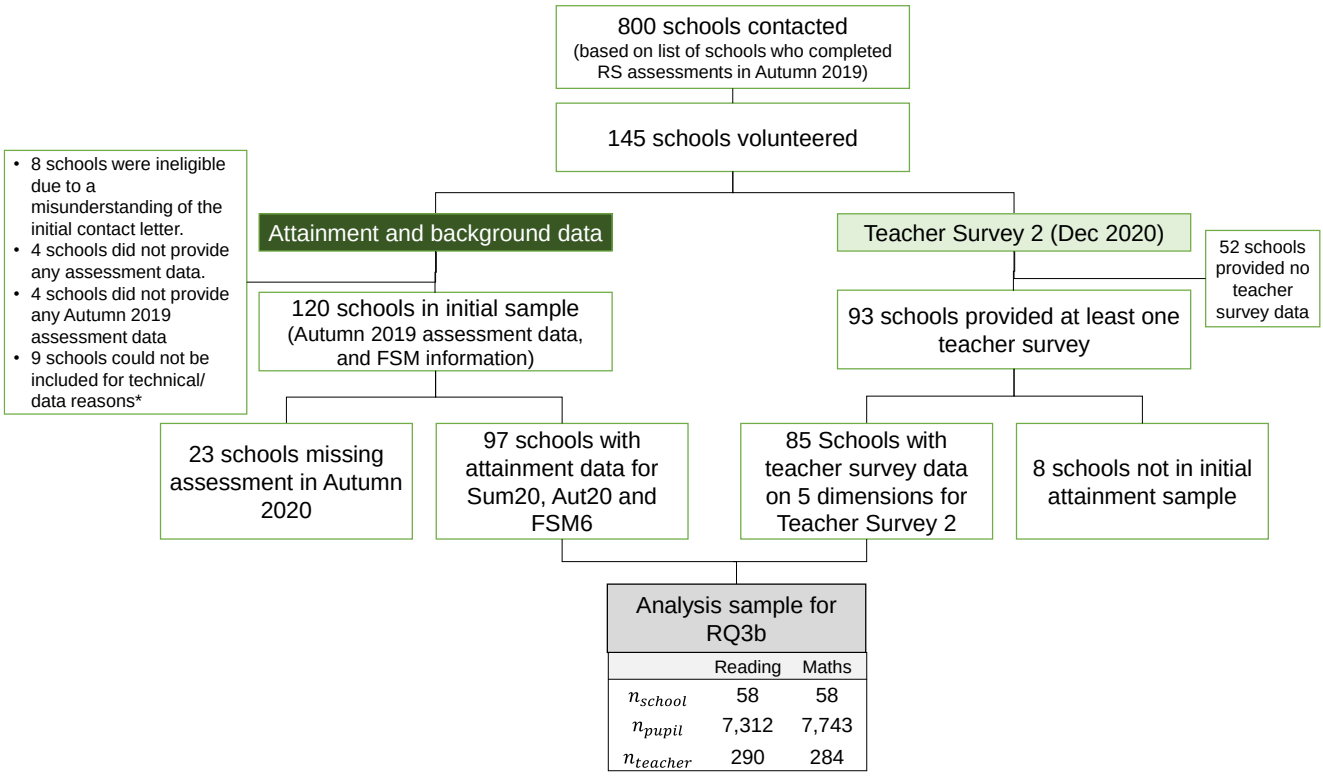
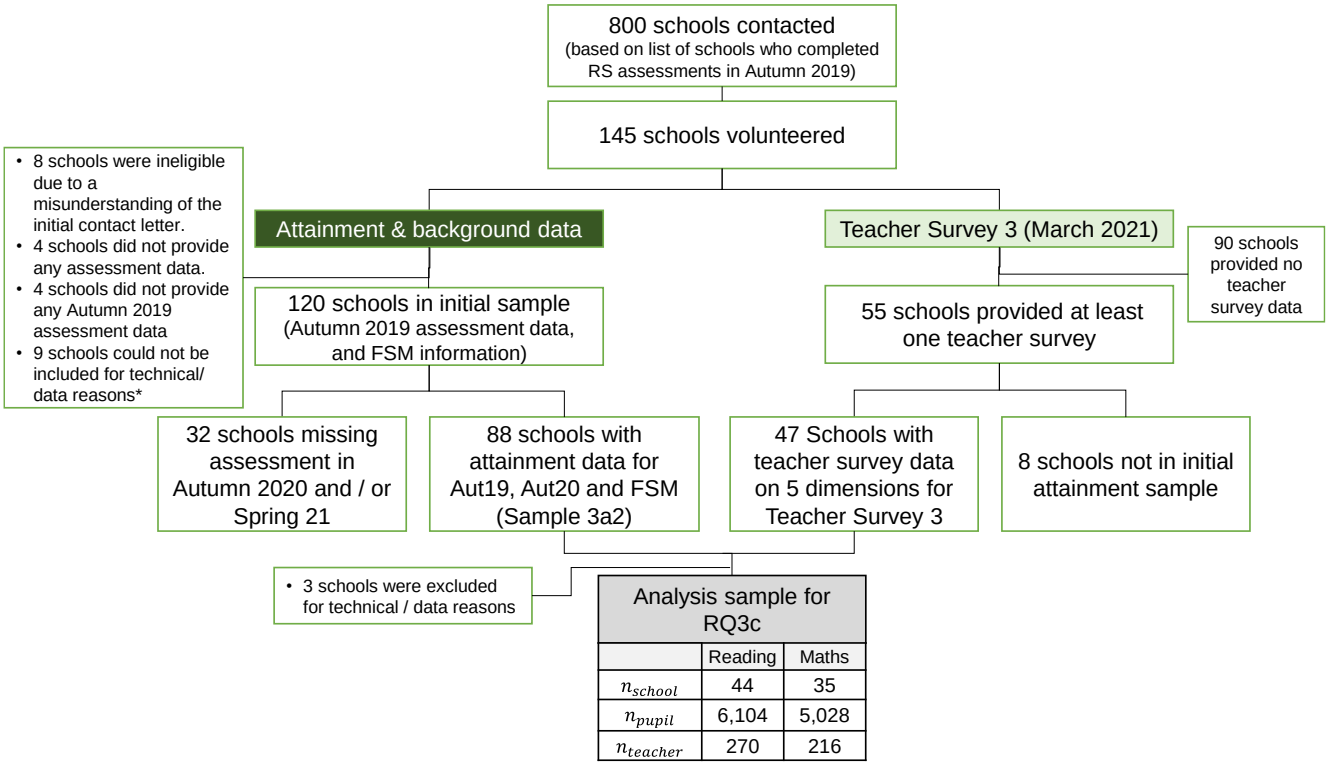


Figure 8 – Overview of sample for RQ3c



	Reading	Maths
n_{school}	44	35
n_{pupil}	6,104	5,028
$n_{teacher}$	270	216

Appendix I: Comparison between achieved samples and national data

Table 14 – Characteristics of schools in RQ1, RQ2 and RQ3 compared to the population of state-funded primary schools

	RQ1	RQ2a	RQ2b	RQ2c	RQ3a	RQ3b	RQ3c	**Population**
% junior schools	8%	10%	6%	9%	8%	8%	9%	**7%**
% London schools	8%	12%	10%	11%	9%	6%	13%	**11%**
% northern schools	22%	24%	27%	27%	25%	23%	24%	**32%**
Mean cohort size (pupils)	46.6	45.9	45.5	43.2	47.0	47.4	43.2	**41.7**
Mean prior attainment (KS1 ave. points)	16.0	16.2	16.0	16.1	16.1	15.9	16.1	**16.3**
% disadvantaged pupils	34%	33%	35%	34%	34%	35%	35%	**30%**
% first language other than English	29%	26%	31%	32%	28%	32%	35%	**21%**
% expected standard in read/write/maths	62%	62%	62%	62%	63%	62%	63%	**66%**
% higher standard in read/write/maths	9%	10%	10%	10%	9%	9%	9%	**11%**
Number of schools	120	139	93	55	95	64	45	**15,580**

Notes: for a description of recruitment and participant flow, see Figure 1, Figure 6, Figure 7 and Figure 8. The "population" represents all state-funded primary schools and data comes from the UK government **website**.

Table 15 – School practices in study schools compared to a broader sample of teachers in England.

Question	**Response categories**	**RQ2 sample**	**Teacher Tapp**
Did you provide students with a timetable to follow each day?	No	33%	34%
	Yes, but it was only a suggested timetable and pupils were not required to follow it	52%	54%
	Yes, we asked pupils to follow a daily timetable	15%	12%
	Number of responses	539	1449
How frequently did you suggest that parents or pupils send in work?	I gave no suggestion about how frequently work should be sent in	34%	39%
	Once a week (or less)	19%	11%
	Several times a week	15%	18%
	Every day (or after each lesson)	31%	33%
	Number of responses	539	1436
Did children (i.e. not the parents) typically speak to you (i.e. their class teacher) on the phone?	No, children typically did not speak to their class teacher	37%	46%
	Yes, I spoke to them each once or twice during the half term	40%	30%
	Yes, I spoke them to each about once a week (or more frequently)	23%	23%
	Number of responses	539	1431
Did you try teaching a 'live' lesson to your class during the first half of the Summer term (with students either speaking or on mute)?	No	52%	61%
	No, but I did pre-record a video of myself for them to watch	35%	29%
	Yes (any number of live lessons)	12%	10%
	Number of responses	539	1442
Is your school making greater or less use of small group and individual face-to-face interventions (e.g. run by TAs) so far this year?	There are more intervention groups running than in a normal year so far	25%	20%
	Similar number of intervention groups running compared to a normal year so far	33%	38%
	There are fewer intervention groups running compared to a normal year so far	41%	42%
	Number of responses	423	1611
How many extra minutes a day (over and above a normal year) are you spending on English and maths right now to cope with missed learning as a result of shutdown and absences?	Zero minutes – we have not reallocated lesson time at all	38%	53%
	Extra 15 minutes	23%	15%
	Extra 30 minutes on English and maths	25%	21%
	Extra 45 minutes (or more) on English and maths	14%	11%
	Number of responses	403	1432

Note: this table does not include results from Teacher Survey 3, as we do not have access to a comparison sample from Teacher Tapp in March 2021.

Appendix J: Standardisation approaches for equating the different tests

We considered three approaches to equating scores from different tests (PIRA, PUMA and NTS) conducted at different periods (Autumn 19, Summer 20, Autumn 20, Spring 21, Summer 21):[54]

1. Linear equating: scaling to ensure each test has the same mean and standard deviation. In accordance with the study plan, this is our preferred approach.
2. Equipercentile equating: scaling based on ranks, within each combination of:
    a. test (e.g. PIRA16)
    b. time period (e.g. Autumn19)
    c. year group (e.g. year 2)
3. Scaling based on ranks, within each combination of: test-time-year group *and* school

Note that the impact of equating is discussed in the Robustness section. We find that our results are largely insensitive to our method of standardising scores.

Linear Equating: Standardised scores (main approach)

For all tests (PIRA, PUMA and NTS) we received scores from Rising Stars (RS) that had been normed relative to a representative sample of students in England.[55]

Let Y_{ijgta}^{NORM} be the scores we received from RS for pupil i, in school j, for year group g (defined in terms of their year group in 2020-21), at time t (Autumn19, Summer20, Autumn20, Spring21 and Summer21), measured by assessment a (PIRA16, PIRA21, PUMA16, PUMA21, NTS Reading, NTS Maths). Because the norming was done in different years, we take the extra step of adjusting scores to have the same means and standard deviations of the large samples of students taking these tests in the three assessment points we analyse. Specifically, we calculate:

$$Y_{ijgta} = \frac{Y_{ijgta}^{Norm} - \hat{\mu}_{tga}}{\hat{\sigma}_{tga}}$$

where $\hat{\mu}_{tga}$ is the mean score on assessment a at time t in year group g in the full sample of children who sit RS tests.[56] This leaves us with our main measure of attainment Y_{ijgt}^{k}, in which the "assessment" subscript a is replaced by a "subject" superscript k which indicates reading or maths.

Robustness checks: ranks

To check whether our results are robust to different methods of scaling, we converted attainment scores into ranks. We did this in two different ways. First, we looked at ranks within each combination of assessment-time-year_group. Let n_{ijgta}^{k} be the ordinal rank of pupil i in year group g, at time period t, in assessment a (for subject k) in terms of Y_{ijgta}^{k}. We calculate:

$$R_{ijgt}^{k} = 1 + floor\left(100 \times \frac{n_{ijgta}^{k} - 0.5}{N_{gta}^{k}}\right) \quad (A)$$

[54] See **Error! Reference source not found.** for detailed information on dates.
[55] Note that this norming process took place in different years. For example, PIRA16 norming took place in 2013-14; PUMA16 norming took place in 2014-15. PIRA21 and PUMA21 were both normed in 2020.
[56] We were unable to source $\hat{\mu}_{tga}$ and $\hat{\sigma}_{tga}$ for $t = Summer2020$ and $a = (NTS\ Reading, NTS\ Maths)$. In both these cases, we used estimates of the mean and standard deviation of Y^{Norm} from our own sample.

where N_{gta}^{k} is the number of students in our analysis sample at time t for year group g in assessment a. In order to make the results easily comparable across scaling approaches we converted the rank measure R_{ijgt}^{k} into standard deviation units, using the normal CDF (Φ): [57]

$$Y_{ijgt}^{k,RANK} = \Phi\left(\frac{R_{ijgt}^{k}}{100}\right) \quad (B)$$

As a second robustness check, we calculate "within-school ranks". We follow the same procedure outlined in equations (A) and (B), but calculate ranks within school:

$$R'^{k}_{ijgt} = 1 + floor\left(100 \times \frac{n_{ijgta}^{k} - 0.5}{N_{jgta}^{k}}\right)$$

where N_{jgta}^{k} is the number of students in school j who provided data for subject k at time t, and n_{ijgta}^{k} is the rank of pupil i within their year-group g, in school j.

[57] As noted in the Robustness section, we also ran our analyses using ranks R. This made no difference to the pattern of results.

Appendix K: Statistical models

All analyses are run in R.

Research Question 1

Let Y_{ijgt}^{k} be a standardised attainment score for pupil i, in school j, for year group g (defined in terms of their year group in 2020-21), at time t on subject k (reading or maths). The standardisation of attainment scores is discussed in the Measures section. Let F_i be a binary indicator equal to 1 if student i was classified as FSM6 in October 2020.[58]

We define the following five time points (see **Error! Reference source not found.** for an overview of the t imeline):

- T_0 = December 2019 (modal date of Autumn 2019/20 assessment)
- T_1 = September 2020 (modal date of Summer 2020 assessment)
- T_2 = December 2020 (modal date of Autumn 2020/21 assessment)
- T_3 = March 2021 (modal date of Spring 2020/21 assessment)
- T_4 = June 2021 (modal date of Summer 2020/21 assessment)

Our focus is the attainment gap between FSM6 pupils and their peers. We define this as G_{tg}^{k} – the difference in Y_{ijgt}^{k} between the mean attainment of FSM6 pupils and their peers for year group g:

$$\mathrm{G}_{tg}^{\mathrm{k}} = \frac{\Sigma_i Y_{ijgt}^{k} \cdot (1 - F_i)}{\Sigma_i (1 - F_i)} - \frac{\Sigma_i Y_{ijgt}^{k} \cdot F_i}{\Sigma_i F_i} \quad (1)$$

Our first research aim is to estimate G_t^k (the average attainment gap in primary schools for subject k) at different points in time, and then to estimate this quantity for different year levels, i.e. to estimate G_{tg}^{k}.

Focusing on the year-level specific estimates, let the change in G_{tg}^{k} from December 2019 to September 2020 be given by:[59]

$$\Delta_{T_1 g}^{\mathrm{k}} = G_{T_1 g}^{k} - G_{T_0 g}^{k} \quad (2)$$

We estimate these quantities using statistical models. $\Delta_{T_1 g}^{\mathrm{k}}$ can be estimated using model 1a. We fit this model once for each year level (and subject k) using the lme4 package. In other words, we fit the following model 10 times:

$$Y_{ijT_1}^{k} - Y_{ijT_0}^{k} = \alpha_j + \delta_{T_1 g}^{k} + \Delta_{T_1 g}^{\mathrm{k}} F_i + month_{iT_0} + \beta month\_difference_i + e_{ijg}^{k} \quad (model\ 1a)$$

$$\alpha_j \sim N(0, \sigma_\alpha^2)$$

$$e_{ijg}^{k} \sim N(0, \sigma^2)$$

In each case:

[58] The proportion of children who are designated FSM6 may have increased since Autumn 2019, due to Covid-19. This change in the composition of the two groups (FSM6 and non-FSM6) could influence our estimate of how the disadvantage gap has changed over time. To control for this, our headline results define FSM6 at a single point in time – October 2020. This ensures that the percentage of FSM6 children is constant at different time points and allows us to focus on how attainment in these two groups has changed. It may have been preferable to focus on FSM6 status in Autumn 2019, but unfortunately, as noted in the study plan, we did not have access to this data.

[59] In the study plan, the start date was listed as November 2019. However, the modal month of Autumn assessments in 2019-20 turned out to be December.

- $\delta^{k}_{T_1 g}$ is the average change in Y (from to T_0 to T_1) for non-FSM6 pupils in year level g in subject k
- $\Delta^{\mathrm{k}}_{T_1 g}$ is the average additional change in Y (from to T_0 to T_1) for FSM6 pupils in year level g in subject k
- $month_{iT_0}$ is a fixed effect for the month in which the Autumn 2019 assessment was taken by pupil *i* in subject k.[60]
- $month\_difference_i$ is the length of time, in months, between the administration of the Autumn 2019 and Summer 2020 assessment
- α_j is a random effect for school j, to account for clustering of pupils within schools
- e^{k}_{ijg} are normally-distributed disturbances at the pupil level

To present readers with a high-level summary of our results, we also estimate $\widehat{\Delta}^{\mathrm{k}}_{T_1}$ – an average change in disadvantage gaps across year levels 2 to 6 for outcome k.[61]

$$Y^{k}_{ijT_1} - Y^{k}_{ijT_0} = \alpha_j + \delta^{k}_{T_1 g} + \Delta^{\mathrm{k}}_{T_1} F_i + month_{iT_0} + \beta month\_difference_i + e^{k}_{ijg} \quad (model\ 1a')$$
$$\alpha_j \sim N(0, \sigma^2_\alpha)$$
$$e^{k}_{ijg} \sim N(0, \sigma^2)$$

Our analysis for research questions 1b, 1c and 1d is analogous, but compares different time points, e.g. for research question RQ1b we fit the following model:

$$Y^{k}_{ijgT_2} - Y^{k}_{ijgT_1} = \alpha_j + \delta^{k}_{T_2 g} + \Delta^{k}_{T_2 g} F_i + month_{iT_1} + \beta month\_difference_i + e^{k}_{ijg} \quad (model\ 1b)$$
$$\alpha_j \sim N(0, \sigma^2_\alpha)$$
$$e^{k}_{ijg} \sim N(0, \sigma^2)$$

Similarly, when comparing gaps across the entire period ($T_0 \rightarrow T_4$) we use the same model structures (which we label model 1c).

Research Question 2

We address RQ2 using simple descriptive statistics. The school response variables are described in the Measures section and in Appendix D:, Appendix E: and Appendix F:.

Research Question 3

We answer RQ3a, RQ3b and RQ3c using multilevel models. Consider model 3a, in which cohort and subject indices have been suppressed for simplicity. For each construct Z^c (for c = 1,…,5) we fit the following model:

$$Y_{ijT_1} - Y_{ijT_0} = \alpha_j + \Delta_j F_i + \boldsymbol{\beta X_{ij}} + e_{ij} \quad (model\ 3a)$$
$$\alpha_j = \delta_a + \phi^c_0 Z^c_j + \gamma_j$$
$$\Delta_j = \Delta_\alpha + \phi^c_1 Z^c_j + \eta_j$$

Where:

- F_i is a binary indicator of FSM6 status for pupil i
- δ_a is the mean change in Y for non-FSM6 pupils in a typical school

[60] In cases where a very small number of pupils fall in a particular month, we aggregate these fixed effects into longer time periods.
[61] To check whether it was sensible to summarise the data in this way, we conducted Likelihood Ratio tests comparing a constrained model 1a' to an unconstrained model (in which $\Delta^{\mathrm{k}}_{T_1 g}$ can deviate by year level, i.e. model 1a). These tests are reported in Appendix O:. We do not find evidence against the hypothesis that changes in disadvantage gaps are similar across year levels.

- $\Delta_a + \delta_a$ is the mean gain score for FSM pupils in a typical school
- $\boldsymbol{X_{ij}}$ is a matrix of student and school controls (at the pupil level: gender and EAL; at the school level: % pupils achieving expected standard in reading, writing and maths in 2019, % pupils achieving higher standard in Reading/Writing/Maths in 2019, inspection ratings (as factor), % pupils who were FSM6 in 2019)
- Z_j^c is a zero-centered school construct describing school responses from March to May 2020 (see Measures section for more detail)
- ϕ_0^c measures the association between school-level construct Z^c and the mean change in Y for non-FSM students
- ϕ_1^c measures the association between school-level construct Z^c and the mean change in Y for non-FSM students (i.e. the association between Z^c and the disadvantage gap)
- $e_{ij} \sim N(0, \sigma^2)$
- γ_j and η_j are school-level random effects, where:

$$\begin{pmatrix} \gamma_j \\ \eta_j \end{pmatrix} \sim N\left(\mathbf{0}, \begin{matrix} \sigma_\alpha^2 & \sigma_{a\eta} \\ \sigma_{a\eta} & \sigma_{\Delta_j}^2 \end{matrix}\right)$$

- γ_j is the school-effect for school j for non-FSM pupils [i.e. the mean change in Y at school j for non-FSM pupils]
- $\gamma_j + \eta_j$ is the school-effect for school j for FSM pupils [i.e. the change in Y at school j for FSM pupils]
- η_j is the estimand for the differential impact school j has on the change in Y of FSM (compared to non-FSM pupils) after controlling for the variables in $\boldsymbol{X}$

Our analysis focuses on point estimates of ϕ_1 along with 95% confidence intervals, using profile likelihood.

The analysis for RQ3b uses the same setup, but makes two changes:

- The survey variables come from Teacher Survey 2 (summarised as $\tilde{Z}^c$).
- The LHS variable is the change in achievement from September/Oct 2020 (T_1) to December 2020 (T_2)

$$Y_{ijT_2} - Y_{ijT_1} = \alpha_j + \Delta_j F_i' + \boldsymbol{\beta X_{ij}} + e_{ij} \quad (model\ 3b)$$
$$\alpha_j = \delta_a + \phi_0^c \tilde{Z}_j^c + \gamma_j$$
$$\Delta_j = \Delta_\alpha + \phi_1^c \tilde{Z}_j^c + \eta_j$$

Similarly, the analysis for RQ3c uses the same approach as RQ3b, but:

- The survey variables come from Teacher Survey 3 (summarised as $\boldsymbol{Z}^*$)
- The LHS variable is the change in achievement from Dec 2020 (T_2) to March 2021 (T_3)

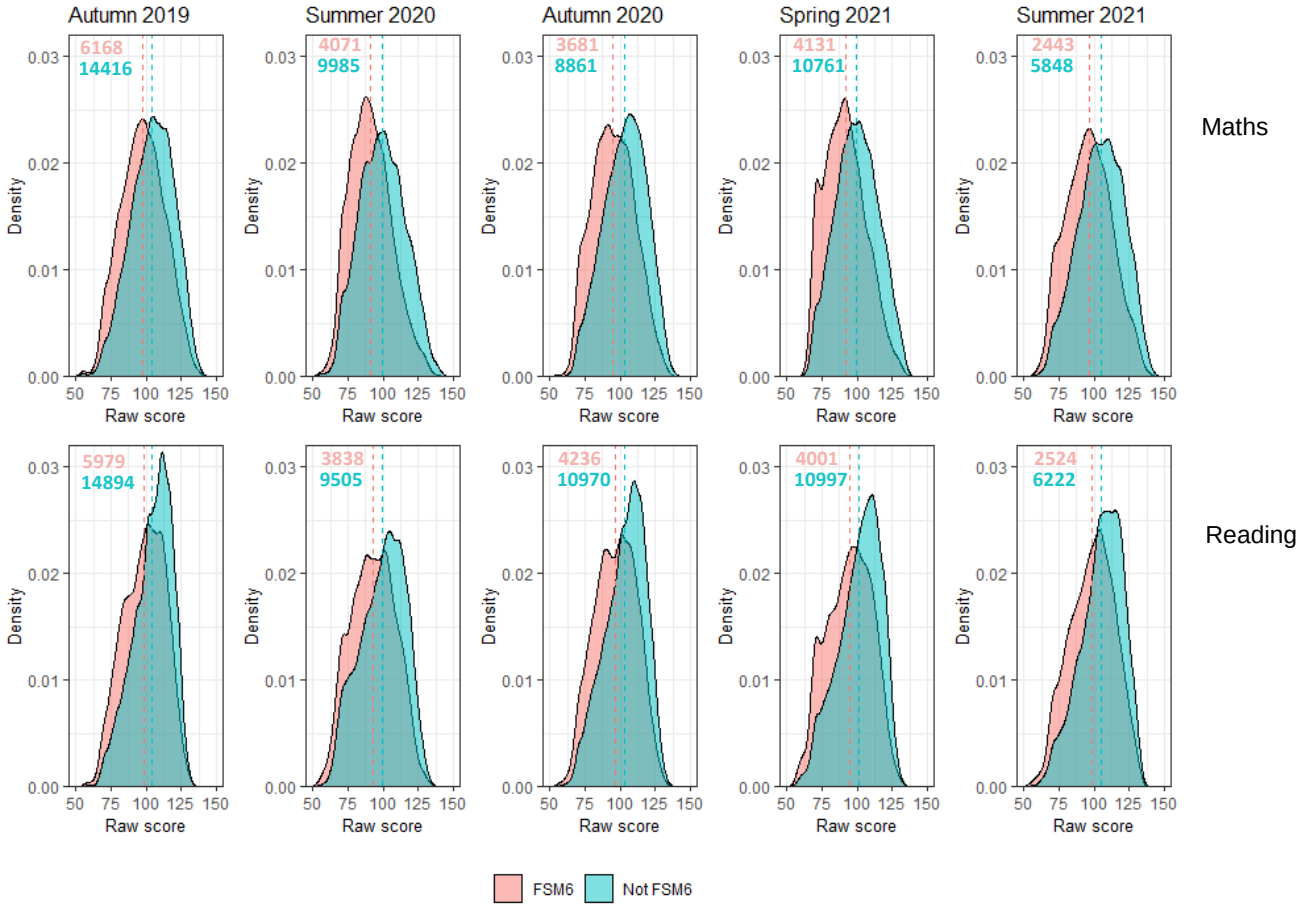
$$Y_{ijT_3} - Y_{ijT_2} = \alpha_j + \Delta_j F_i' + \boldsymbol{\beta X_{ij}} + e_{ij} \quad (model\ 3c)$$
$$\alpha_j = \delta_a + \phi_0 Z_j^* + \gamma_j$$
$$\Delta_j = \Delta_\alpha + \phi_1 Z_j^* + \eta_j$$

Appendix L: Detailed results for Research Question 1

Distributions of attainment data

We begin by plotting the distributions of RS assessment data for the five assessment periods (Figure 9). These distributions cover years 2-6 and do not appear to suffer from ceiling or floor effects. In all five periods we observe a gap between FSM6 pupils and their peers (indicated by the difference between the dotted lines on each plot, which represent means).

Figure 9 - Distribution of attainment scores (Y^{Norm})



Notes: "Raw score" is Y^{Norm}, described in the Measures section. This plot presents all raw scores for pupils who provided data. The number of pupils represented in each period, by FSM6 status, is represented in the top left of each plot (n_{FSM} on top, n_{notFSM} underneath). The dotted lines on each plot represent the mean of each distribution.

Mean attainment, by FSM6, during 5 assessment windows

Next we present information about mean attainment, for FSM and non-FSM students, in maths and reading (**Error! Reference source not found.**).[62] These simple means represent averages across year levels. They p rovide an initial sense of how disadvantage gaps have changed over time, without using models to control for things like 'the timing of assessments'. We use standardised scores, as described in the measures section. Note that these scores differ from the raw attainment data we received from RS Assessment, presented in Figure 9.

Table 16 - Mean standardised scores in reading ($Y_{ij}^{Reading}$) and maths (Y_{ij}^{Maths}) by FSM status, in RQ1 analysis sample

	Reading	Maths

[62] See the Measures section for definitions of the standardised attainment outcomes.

	Not FSM6 (s.e.)	FSM6 (s.e.)	Gap	Not FSM6 (s.e.)	FSM6 (s.e.)	Gap
Autumn 2019	0.09 (0.01)	-0.33 (0.01)	0.42	0.08 (0.01)	-0.37 (0.01)	0.45
Summer 2020	0.15 (0.01)	-0.30 (0.01)	0.45	0.16 (0.01)	-0.37 (0.01)	0.53
Autumn 2020	0.15 (0.01)	-0.31 (0.01)	0.45	0.14 (0.01)	-0.39 (0.01)	0.53
Spring 2021	0.08 (0.01)	-0.39 (0.01)	0.47	0.07 (0.01)	-0.46 (0.01)	0.53
Summer 2021	0.07 (0.01)	-0.37 (0.01)	0.44	0.04 (0.01)	-0.47 (0.01)	0.51
n_{pupils}	13518	5297		13237	5483	

Note: standard errors for mean estimates are in parentheses. All numbers are in ES units. The sample reported here is the analysis sample for RQ1 (Full Responders + Partial Responders, described in Figure 1).

Table 17 – Regression results for main analysis of Research Question 1 (using statistical models)

	Reading					Maths				
	Aut19-Sum20	Sum20 - Aut20	Aut20-Spr21	Spr21-Sum21	Aut19-Sum21	Aut19-Sum20	Sum20 - Aut20	Aut20-Spr21	Spr21-Sum21	Aut19-Sum21
Constant	-0.254	-0.383	-0.196	-0.305	-1.017	-0.326	-0.509	-0.463	-0.146	-0.222
	(0.174)	(0.048)	(0.133)	(0.056)	(0.321)	(0.165)	(0.050)	(0.100)	(0.07)	(0.360)
FSM6	-0.002	-0.004	-0.010	0.017	-0.003	-0.038	-0.007	-0.006	0.006	-0.048
	(0.014)	(0.014)	(0.014)	(0.015)	(0.016)	(0.012)	(0.012)	(0.010)	(0.012)	(0.015)
Year 3	0.054	-0.073	0.002	-0.004	0.000	0.011	-0.034	0.049	-0.069	-0.025
	(0.020)	(0.021)	(0.021)	(0.020)	(0.022)	(0.018)	(0.017)	(0.017)	(0.019)	(0.020)
Year 4	0.046	0.009	-0.018	-0.007	0.028	-0.008	0.029	0.013	-0.047	-0.008
	(0.020)	(0.018)	(0.018)	(0.019)	(0.022)	(0.017)	(0.017)	(0.017)	(0.018)	(0.020)
Year 5	0.043	-0.024	-0.003	-0.009	0.008	-0.011	0.033	0.053	-0.052	0.028
	(0.020)	(0.020)	(0.019)	(0.020)	(0.023)	(0.017)	(0.017)	(0.017)	(0.017)	(0.020)
Year 6	0.010	-0.007	0.001	-0.017	-0.007	-0.035	0.027	0.057	-0.064	-0.005
	(0.020)	(0.019)	(0.019)	(0.021)	(0.024)	(0.018)	(0.018)	(0.017)	(0.021)	(0.024)
Aut 19 test date: Nov19	0.041				-0.016	-0.066				-0.048
	(0.034)				(0.033)	(0.030)				(0.035)
Aut 19 test date: Oct19 or earlier	0.232				0.207	0.152				0.170
	(0.068)				(0.072)	(0.076)				(0.082)
Difference in test dates	0.026	0.043	0.033	0.098	0.052	0.045	0.082	0.097	0.062	0.010
	(0.020)	(0.015)	(0.032)	(0.017)	(0.018)	(0.018)	(0.015)	(0.024)	(0.020)	(0.019)
Sum 20 test: Sept20		0.312					0.283			
		(0.045)					(0.044)			
Sum 20 test: Aug20 or earlier		0.185					0.181			
		(0.053)					(0.051)			
Aut 20 test: Dec20 or later			0.035					0.034		
			(0.035)					(0.028)		
Aut 20 test: Oct 20 or earlier			0.145					0.119		
			(0.073)					(0.075)		
Spr 21 test: Feb21 or earlier				-0.464					-0.225	
				(0.689)					(0.404)	
Spr 21 test: Apr21 or later				0.005					-0.127	
				(0.034)					(0.044)	
Observations	18815	18815	18815	18815	18815	18720	18720	18720	18720	18720

Table 18 - MATHS results by year level for Research Question 1

	Change in score, Aut 19 – Sum 20					Change in score, Sum 20 – Aut 20					Change in score, Aut 19 – Aut 20				
	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)
Aut 19 test date: Nov 19	0.111 (0.079)	-0.056 (0.085)	-0.085 (0.06)	-0.015 (0.067)	0.073 (0.071)										
Aut 19 test date: Oct19 or earlier	-0.062 (0.231)	0.264 (0.242)	0.195 (0.115)	0.25 (0.127)	-0.147 (0.139)										
Sum 20 test date: Sept 20						0.32 (0.094)	0.252 (0.089)	0.382 (0.09)	0.295 (0.086)	0.334 (0.07)					
Sum 20 test date: Aug20 or earlier						0.232 (0.111)	0.16 (0.116)	0.327 (0.106)	0.149 (0.112)	0.177 (0.087)					
Aut 20 test date: Dec 20 or later											0.003 (0.067)	0.008 (0.075)	0.092 (0.06)	0.045 (0.057)	-0.017 (0.061)
Aut 20 test date: Oct 20 or earlier											0.181 (0.144)	0.128 (0.166)	0.076 (0.138)	0.125 (0.134)	0.079 (0.106)
Difference in test dates	0.114 (0.046)	0.093 (0.051)	0.077 (0.036)	0.05 (0.043)	0.038 (0.039)						0.059 (0.053)	0.065 (0.063)	0.144 (0.051)	0.099 (0.049)	0.047 (0.055)
FSM6	-0.067 (0.032)	-0.026 (0.033)	0.002 (0.025)	-0.041 (0.025)	-0.049 (0.023)						-0.01 (0.031)	-0.002 (0.026)	0.018 (0.024)	-0.017 (0.021)	-0.013 (0.023)
Constant	-1.065 (0.41)	-0.775 (0.462)	-0.636 (0.319)	-0.42 (0.38)	-0.33 (0.34)	-0.625 (0.112)	-0.511 (0.101)	-0.516 (0.103)	-0.472 (0.119)	-0.536 (0.082)	-0.323 (0.224)	-0.286 (0.265)	-0.661 (0.212)	-0.426 (0.201)	-0.199 (0.225)
Observations	3165	3265	4228	4181	3881	3165	3265	4228	4181	3881	3165	3265	4228	4181	3881

[Table continued over page]

	Change in score, Spr21 – Sum 21					Change in score, Aut 19– Sum 21 (OVERALL)				
	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)
Spr 21 test date: Apr 21 or later	-0.093	-0.094	-0.142	-0.155	-0.099					
	(0.064)	(0.075)	(0.07)	(0.06)	(0.086)					
Spr 21 test date: Feb 21 or earlier	-0.16	-0.584	-0.179	-0.126	-0.248					
	(0.755)	(0.775)	(0.873)	(0.472)	(0.86)					
Aut 19 test date: Nov 19						0.065	0.084	-0.044	-0.072	-0.006
						(0.093)	(0.075)	(0.063)	(0.063)	(0.063)
Aut 19 test date: Oct19 or earlier						-0.058	0.532	0.308	0.131	-0.04
						(0.261)	(0.256)	(0.137)	(0.138)	(0.175)
Difference in test dates	0.068	0.084	0.071	0.034	0.063	0.016	-0.016	-0.022	0.023	0.009
	(0.038)	(0.036)	(0.029)	(0.028)	(0.035)	(0.043)	(0.046)	(0.031)	(0.032)	(0.037)
FSM6	0.018	0.007	-0.018	0.014	0.021	-0.083	-0.051	-0.022	-0.037	-0.022
	(0.036)	(0.031)	(0.025)	(0.025)	(0.026)	(0.038)	(0.035)	(0.028)	(0.028)	(0.031)
Constant	-0.156	-0.281	-0.211	-0.11	-0.214	-0.39	0.202	0.349	-0.438	-0.206
	(0.128)	(0.117)	(0.099)	(0.093)	(0.116)	(0.783)	(0.854)	(0.569)	(0.591)	(0.675)
Observations	3165	3265	4228	4181	3881	3165	3265	4228	4181	3881

Table 19 - READING results by year level for Research Question 1

	Change in score, Aut 19 – Sum 20					Change in score, Sum 20 – Aut 20					Change in score, Aut 19 – Aut 20				
	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)
Aut 19 test date: Nov 19	-0.063 (0.086)	-0.006 (0.088)	-0.006 (0.061)	0.063 (0.08)	0.108 (0.073)										
Aut 19 test date: Oct19 or earlier	0.122 (0.306)	-0.004 (0.202)	0.28 (0.118)	0.489 (0.161)	0.142 (0.149)										
Sum 20 test date: Sept 20						0.326 (0.089)	0.33 (0.102)	0.385 (0.092)	0.396 (0.092)	0.336 (0.127)					
Sum 20 test date: Aug20 or earlier						0.157 (0.114)	0.194 (0.118)	0.262 (0.118)	0.312 (0.111)	0.118 (0.15)					
Aut 20 test date: Dec 20 or later											0.053 (0.1)	0.103 (0.086)	0.062 (0.068)	0.013 (0.067)	-0.053 (0.075)
Aut 20 test date: Oct 20 or earlier											-0.019 (0.16)	0.266 (0.179)	0.074 (0.414)	0.079 (0.59)	0.138 (0.175)
Difference in test dates	0.041 (0.046)	0.088 (0.056)	0.066 (0.034)	0.061 (0.042)	0.006 (0.045)	0.046 (0.038)	0.032 (0.037)	0.012 (0.039)	0.011 (0.038)	0.111 (0.041)	0.059 (0.091)	0.064 (0.079)	0.075 (0.058)	0.003 (0.057)	-0.018 (0.065)
FSM6	-0.015 (0.037)	-0.012 (0.035)	0.007 (0.032)	0.001 (0.028)	-0.008 (0.028)	0.006 (0.04)	0.001 (0.031)	-0.002 (0.031)	-0.031 (0.026)	0.015 (0.029)	-0.011 (0.038)	0.009 (0.031)	-0.013 (0.026)	-0.01 (0.025)	-0.026 (0.03)
Constant	-0.321 (0.415)	-0.768 (0.509)	-0.557 (0.304)	-0.561 (0.371)	-0.074 (0.394)	-0.401 (0.107)	-0.441 (0.108)	-0.366 (0.103)	-0.392 (0.111)	-0.567 (0.127)	-0.301 (0.377)	-0.354 (0.323)	-0.366 (0.238)	-0.078 (0.236)	0.037 (0.267)
Observations	3067	3092	4471	4382	3803	3067	3092	4471	4382	3803	3067	3092	4471	4382	3803

[Table continued over page]

	Change in score, Spr21 – Sum 21					Change in score, Aut 19– Sum 21 (OVERALL)				
	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)	(Year 2)	(Year 3)	(Year 4)	(Year 5)	(Year 6)
Spr 21 test date: Apr 21 or later	0.03 (0.082)	-0.073 (0.09)	-0.008 (0.056)	-0.022 (0.069)	0.09 (0.077)					
Spr 21 test date: Feb 21 or earlier	-0.555 (0.782)	-0.491 (1.207)	-0.27 (0.87)	-0.095 (1.463)	-0.255 (0.71)					
Aut 19 test date: Nov 19						-0.112 (0.09)	0.052 (0.085)	-0.05 (0.066)	0.062 (0.074)	-0.05 (0.077)
Aut 19 test date: Oct19 or earlier						-0.066 (0.287)	0.062 (0.204)	0.311 (0.136)	0.65 (0.167)	0.164 (0.161)
Difference in test dates	0.135 (0.041)	0.089 (0.042)	0.081 (0.034)	0.074 (0.038)	0.088 (0.034)	0.048 (0.044)	-0.002 (0.047)	0.046 (0.042)	0.041 (0.042)	0.035 (0.038)
FSM6	0.027 (0.041)	0.009 (0.035)	0.008 (0.028)	0.013 (0.029)	0.029 (0.03)	0.003 (0.044)	0.006 (0.04)	-0.004 (0.03)	-0.03 (0.032)	0.01 (0.033)
Constant	-0.425 (0.136)	-0.271 (0.137)	-0.249 (0.107)	-0.223 (0.12)	-0.299 (0.113)	-0.883 (0.811)	-0.035 (0.862)	-0.847 (0.765)	-0.834 (0.767)	-0.684 (0.687)
Observations	3067	3092	4471	4382	3803	3067	3092	4471	4382	3803

Appendix M: Regression results for Research Question 3

Table 20 - Regression results for Research Question 3a

	Change in maths score, Aut 19 – Sum 20					Change in reading score, Aut 19 – Sum 20				
	Platform Use	Time-tabling	Work Sub-missions	Phone calls	Vid/Live lessons	Platform Use	Time-tabling	Work Sub-missions	Phone calls	Vid/Live lessons
FSM6	-0.046**	-0.045**	-0.046**	-0.046**	-0.045**	-0.012	-0.009	-0.010	-0.011	-0.011
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.018)	(0.018)	(0.019)	(0.019)	(0.019)
% pupils achieving expected standard in R/W/M in 2019	0.004	0.003	0.011	0.010	0.020	0.067	0.053	0.078	0.075	0.063
	(0.228)	(0.228)	(0.229)	(0.227)	(0.225)	(0.224)	(0.223)	(0.223)	(0.224)	(0.224)
% pupils achieving higher standard R/W/M in 2019	-0.059	-0.078	-0.099	-0.024	-0.064	-0.193	-0.184	-0.241	-0.157	-0.173
	(0.478)	(0.473)	(0.476)	(0.473)	(0.466)	(0.483)	(0.476)	(0.486)	(0.477)	(0.476)
Ofsted 2	-0.041	-0.042	-0.041	-0.043	-0.036	0.084	0.084	0.081	0.091	0.086
	(0.104)	(0.104)	(0.104)	(0.103)	(0.102)	(0.104)	(0.104)	(0.104)	(0.104)	(0.104)
Ofsted 3	-0.044	-0.040	-0.036	-0.044	-0.035	0.050	0.054	0.050	0.046	0.046
	(0.121)	(0.120)	(0.120)	(0.118)	(0.117)	(0.121)	(0.120)	(0.120)	(0.120)	(0.120)
Ofsted 4	0.043	0.044	0.045	0.028	0.038	0.200	0.194	0.187	0.197	0.202
	(0.172)	(0.172)	(0.172)	(0.172)	(0.170)	(0.171)	(0.171)	(0.171)	(0.171)	(0.171)
% pupils who were FSM6 in 2019	-0.320	-0.334	-0.333	-0.334	-0.314	-0.369*	-0.366*	-0.360*	-0.357*	-0.356*
	(0.185)	(0.182)	(0.178)	(0.177)	(0.176)	(0.187)	(0.182)	(0.179)	(0.179)	(0.179)
Female	-0.001	-0.001	-0.001	-0.001	-0.001	0.027*	0.027*	0.027*	0.027*	0.027*
	(0.011)	(0.011)	(0.011)	(0.011)	(0.011)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
EAL	0.004	0.003	0.011	0.010	0.020	0.067	0.053	0.078	0.075	0.063
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Main effect of School Response Variable (Z)	0.005	-0.001	-0.008	0.030	-0.053	-0.003	-0.034	-0.014	0.017	0.003
	(0.018)	(0.045)	(0.022)	(0.038)	(0.039)	(0.017)	(0.042)	(0.022)	(0.037)	(0.038)
Interaction of School Response and FSM (Z:FSM6)	-0.003	-0.002	0.007	0.015	-0.009	-0.013	0.064*	-0.018	0.029	-0.016
	(0.009)	(0.024)	(0.012)	(0.019)	(0.021)	(0.012)	(0.031)	(0.016)	(0.026)	(0.028)
Constant	0.244	0.251	0.246	0.243	0.229	0.057	0.064	0.054	0.040	0.052
	(0.202)	(0.203)	(0.201)	(0.201)	(0.199)	(0.204)	(0.203)	(0.201)	(0.202)	(0.202)
Observations	13,261	13,261	13,261	13,261	13,261	12,564	12,564	12,564	12,564	12,564

Key: *p<0.05, **p<0.01

Table 21 - Regression results for Research Question 3b

	Change in maths score, Sum 20 – Aut20	Change in reading score, Sum 20 – Aut 20

	Pupil absence	School closures	Reduce curriculum	Extra time	Small group	Pupil absence	School closures	Reduce curriculum	Extra time	Small group
FSM6	-0.031*	-0.031	-0.030	-0.029	-0.030	-0.006	-0.001	-0.004	-0.003	-0.005
	(0.016)	(0.017)	(0.017)	(0.016)	(0.017)	(0.019)	(0.019)	(0.018)	(0.019)	(0.019)
% pupils achieving expected standard in R/W/M in 2019	-0.329	-0.357	-0.363	-0.349	-0.345	0.033	0.078	0.080	0.081	0.111
	(0.210)	(0.206)	(0.208)	(0.207)	(0.207)	(0.223)	(0.220)	(0.221)	(0.221)	(0.219)
% pupils achieving higher standard R/W/M in 2019	1.129**	1.050*	1.109*	1.018*	1.115**	0.778	0.860	0.797	0.822	0.853
	(0.430)	(0.427)	(0.436)	(0.430)	(0.425)	(0.457)	(0.464)	(0.470)	(0.463)	(0.456)
Ofsted 2	-0.252*	-0.297*	-0.274*	-0.245*	-0.241*	-0.225	-0.175	-0.175	-0.181	-0.174
	(0.119)	(0.117)	(0.117)	(0.118)	(0.119)	(0.146)	(0.151)	(0.149)	(0.150)	(0.148)
Ofsted 3	-0.226	-0.274*	-0.255	-0.218	-0.229	-0.222	-0.170	-0.173	-0.171	-0.172
	(0.136)	(0.133)	(0.133)	(0.136)	(0.134)	(0.163)	(0.166)	(0.165)	(0.168)	(0.163)
Ofsted 4	-0.299	-0.341	-0.330	-0.320	-0.304	-0.229	-0.175	-0.179	-0.179	-0.172
	(0.183)	(0.177)	(0.180)	(0.178)	(0.180)	(0.207)	(0.209)	(0.210)	(0.209)	(0.206)
% pupils who were FSM6 in 2019	0.375*	0.443*	0.367*	0.389*	0.416*	0.219	0.175	0.221	0.217	0.282
	(0.170)	(0.181)	(0.171)	(0.170)	(0.176)	(0.180)	(0.196)	(0.182)	(0.182)	(0.185)
Female	-0.017	-0.017	-0.017	-0.017	-0.017	-0.026	-0.026	-0.027	-0.026	-0.026
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
EAL	0.028	0.029	0.028	0.027	0.029	-0.005	-0.005	-0.004	-0.004	-0.001
	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)
Main effect of School Response Variable ($\tilde{Z}$)	0.049	0.043	0.008	0.058	0.027	-0.115	-0.024	-0.020	0.029	0.041
	(0.079)	(0.037)	(0.028)	(0.069)	(0.028)	(0.085)	(0.040)	(0.030)	(0.078)	(0.029)
Interaction of School Response and FSM ($\tilde{Z}$:FSM6)	0.087	-0.013	-0.002	0.054	-0.003	0.143*	0.016	0.023	0.005	-0.041*
	(0.052)	(0.023)	(0.017)	(0.043)	(0.018)	(0.065)	(0.026)	(0.018)	(0.052)	(0.021)
Constant	0.208	0.250	0.255	0.225	0.199	-0.027	-0.096	-0.106	-0.105	-0.152
	(0.206)	(0.196)	(0.198)	(0.198)	(0.204)	(0.234)	(0.232)	(0.233)	(0.234)	(0.236)
Observations	7,743	7,743	7,743	7,743	7,743	7,312	7,312	7,312	7,312	7,312

Key: *p<0.05, **p<0.01

Table 22 - Regression results for Research Question 3c

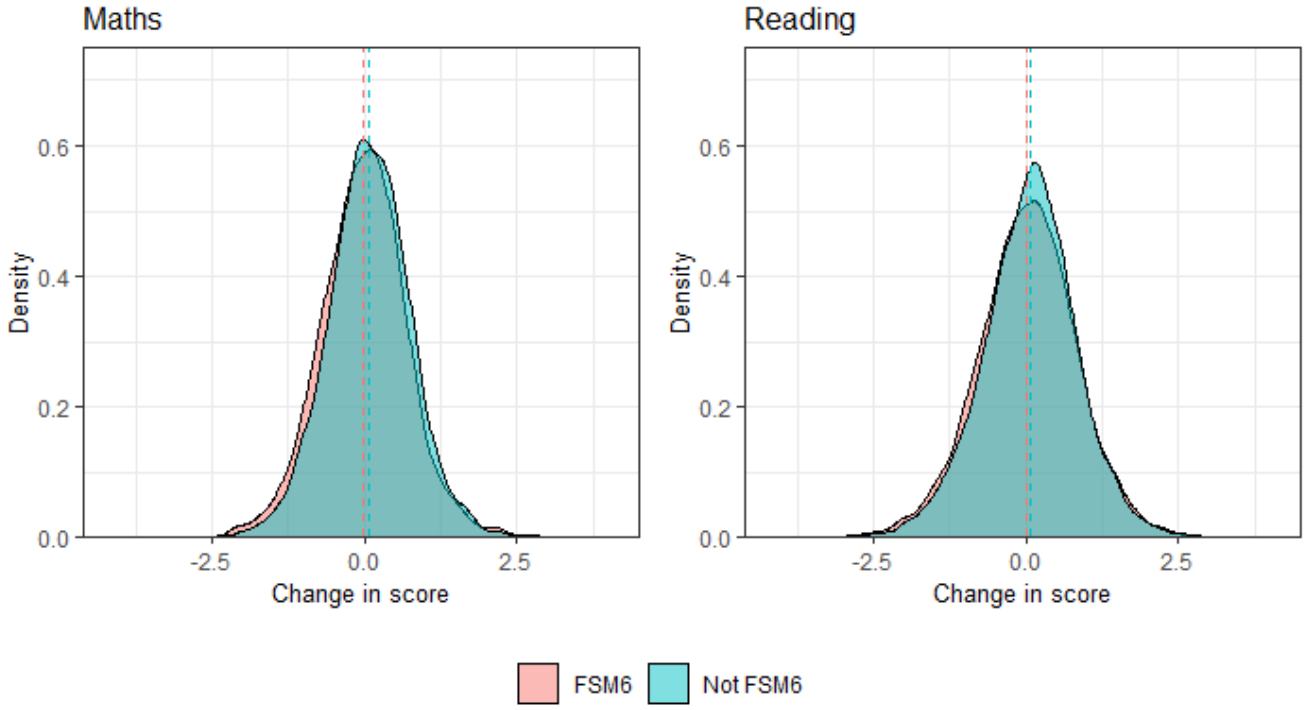
	Change in maths score, Aut 20 – Spr 21					Change in reading score, Aut 20 – Spr 21				
	Platform	Timetable	Work submission	Phoning	Live Lessons	Platform	Timetable	Work submission	Phoning	Live Lessons
FSM6	0.030	0.024	0.025	0.027	0.025	-0.004	-0.005	-0.009	-0.004	-0.004
	(0.019)	(0.019)	(0.021)	(0.019)	(0.021)	(0.021)	(0.022)	(0.022)	(0.022)	(0.022)
% pupils achieving expected standard in R/W/M in 2019	0.230	0.259	0.228	0.214	0.277	0.049	0.044	0.040	0.028	0.059
	(0.185)	(0.181)	(0.177)	(0.183)	(0.175)	(0.167)	(0.168)	(0.161)	(0.169)	(0.169)
% pupils achieving a higher standard in R/W/M in 2019	-0.439	-0.483	-0.362	-0.532	-0.312	-0.659*	-0.638	-0.579	-0.728*	-0.665*
	(0.377)	(0.372)	(0.366)	(0.389)	(0.359)	(0.328)	(0.332)	(0.318)	(0.337)	(0.331)
Ofsted 2	-0.113	-0.145	-0.259	-0.106	-0.046	-0.235*	-0.208	-0.400**	-0.225*	-0.224*
	(0.114)	(0.114)	(0.144)	(0.111)	(0.109)	(0.104)	(0.107)	(0.131)	(0.104)	(0.107)
Ofsted 3	-0.186	-0.201	-0.330*	-0.173	-0.089	-0.241*	-0.221	-0.417**	-0.240*	-0.232
	(0.133)	(0.131)	(0.162)	(0.130)	(0.130)	(0.122)	(0.123)	(0.147)	(0.121)	(0.126)
Ofsted 4	-0.061	-0.130	-0.256	-0.079	0.051	-0.315*	-0.283	-0.506**	-0.305*	-0.296
	(0.162)	(0.164)	(0.193)	(0.162)	(0.165)	(0.148)	(0.152)	(0.173)	(0.148)	(0.157)
% pupils who were FSM6 in 2019	0.054	0.089	0.026	0.018	0.108	-0.009	-0.038	-0.041	-0.039	-0.016
	(0.159)	(0.160)	(0.154)	(0.159)	(0.153)	(0.144)	(0.146)	(0.140)	(0.145)	(0.146)
Female	-0.0001	0.0004	0.0002	0.0001	-0.0001	0.033*	0.034*	0.033	0.034*	0.034*
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
EAL	0.015	0.015	0.014	0.013	0.014	0.037	0.038	0.036	0.038	0.038
	(0.023)	(0.023)	(0.023)	(0.023)	(0.023)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Main effect of School Response Variable (Z^*)	0.019	0.075	0.090	-0.029	0.049*	0.006	-0.037	0.101*	-0.034	0.004
	(0.029)	(0.056)	(0.057)	(0.045)	(0.024)	(0.020)	(0.050)	(0.049)	(0.039)	(0.020)
Interaction of School Response and FSM (Z^*:FSM6)	-0.073**	-0.072	0.006	-0.065*	0.001	0.036	-0.031	0.099*	-0.024	-0.006
	(0.028)	(0.040)	(0.046)	(0.029)	(0.018)	(0.024)	(0.047)	(0.050)	(0.035)	(0.019)
Constant	-0.211	-0.204	-0.068	-0.192	-0.346*	0.148	0.130	0.319*	0.164	0.132
	(0.152)	(0.145)	(0.172)	(0.148)	(0.152)	(0.142)	(0.141)	(0.162)	(0.143)	(0.149)
Observations	5,028	5,028	5,028	5,028	5,028	6,104	6,104	6,104	6,104	6,104

Key: *p<0.05, **p<0.01

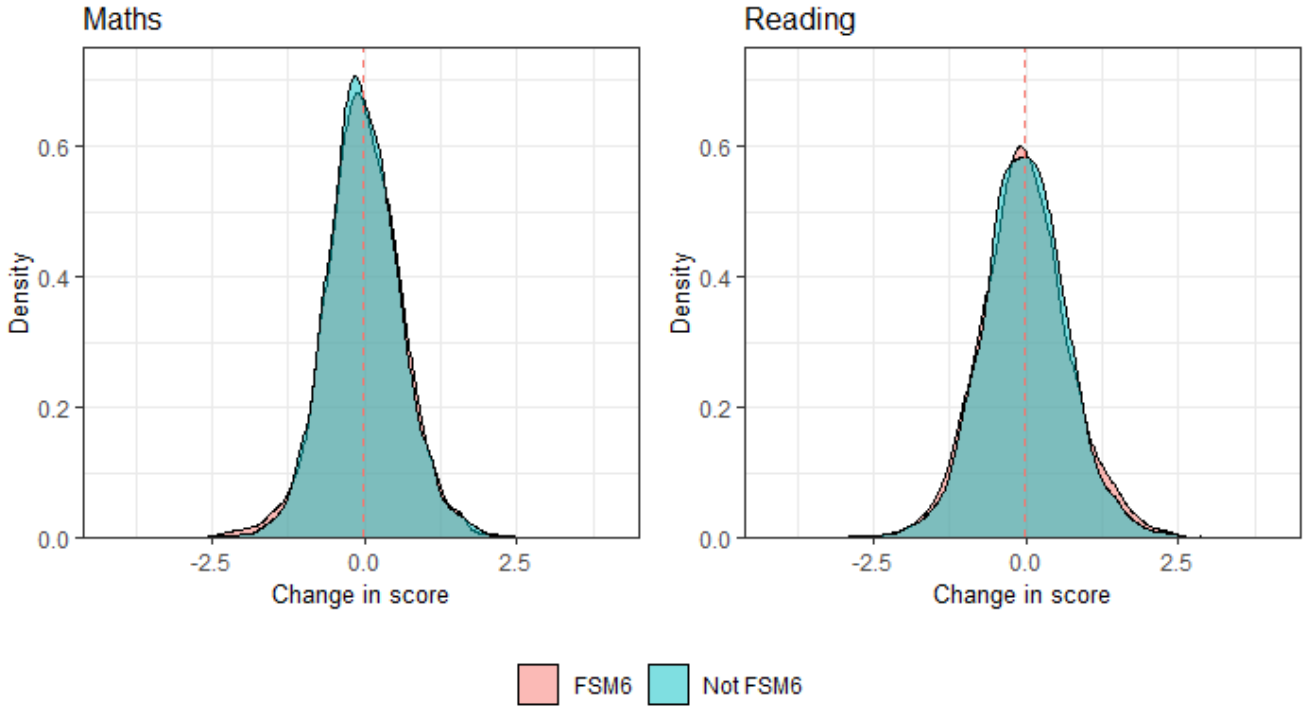
Appendix N: Distribution of change in attainment

This appendix shows the distribution of change in attainment for two periods (relevant to Research Questions 3a and 3b).

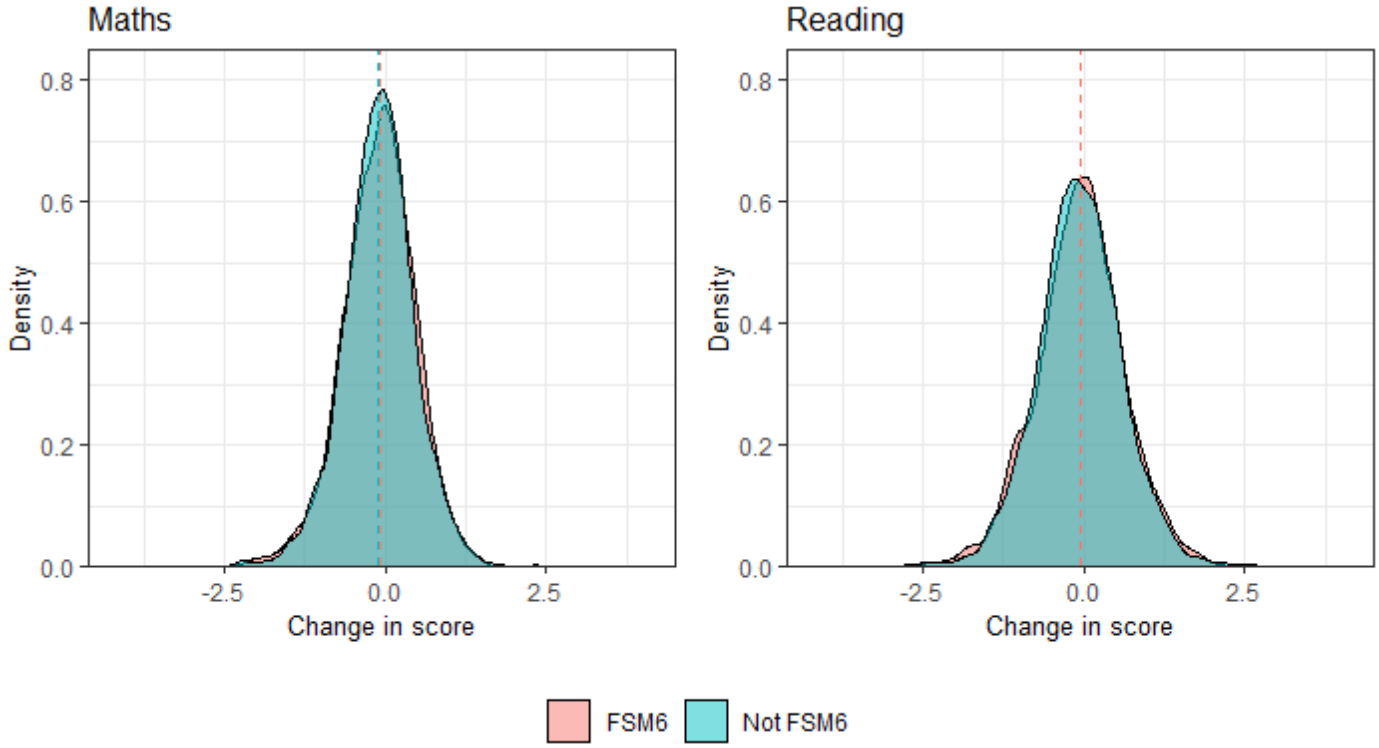
RQ3a: distribution of change in attainment from Autumn 2019 to Summer 2020



RQ3b: distribution of change in attainment from Summer 2020 to Autumn 2020



RQ3c: distribution of change in attainment from Spring 2021 to Summer 2021



Appendix O: Likelihood ratio tests for Research Question 1

In this appendix, we compare two sets of models:

1. Models where we estimate Δ_{g}^{k} separately for each year group, e.g. having the following structure:[63]

$$Y_{ijT_1}^{k} - Y_{ijT_0}^{k} = \alpha_j + \delta_{T_1 g}^{k} + \Delta_{T_1 g}^{\mathrm{k}} F_i + month_{iT_0} + \beta month\_difference_i + e_{ijg}^{k}$$

$$\alpha_j \sim N(0, \sigma_\alpha^2)$$

$$e_{ijg}^{k} \sim N(0, \sigma^2)$$

2. Models where we constrain Δ^k parameter to be constant across year levels (to provide an "average" change in disadvantage gaps), e.g. having the following structure:[64]

$$Y_{ijT_1}^{k} - Y_{ijT_0}^{k} = \alpha_j + \delta_{T_1 g}^{k} + \Delta_{T_1}^{\mathrm{k}} F_i + month_{iT_0} + \beta month\_difference_i + e_{ijg}^{k}$$

$$\alpha_j \sim N(0, \sigma_\alpha^2)$$

$$e_{ijg}^{k} \sim N(0, \sigma^2)$$

The output comparing these models using Likelihood Ratio tests is below.

Table 23 - LR tests for reading

	Autumn 19 – Summer 20	Summer 20 – Autumn 20	Autumn 20 – Spring 21	Spring 21 – Summer 21	Autumn 19 – Summer 21
Fixed parameters in model 1	13	13	13	13	13
AIC	44,262.920	39,630.680	39,236.230	38,744.910	46,478.220
BIC	44,364.870	39,732.630	39,338.180	38,846.860	46,580.170
logLik	-22,118.460	-19,802.340	-19,605.120	-19,359.460	-23,226.110
deviance	44,236.920	39,604.680	39,210.230	38,718.910	46,452.220
Chisq	2.332	0.721	2.101	2.092	2.235
Df	4.000	4.000	4.000	4.000	4.000
Pr(> Chisq)	**0.675**	**0.949**	**0.717**	**0.719**	**0.693**

Table 24 - LR tests for maths

	Autumn 19 – Summer 20	Summer 20 – Autumn 20	Autumn 20 – Spring 21	Spring 21 – Summer 21	Autumn 19 – Summer 21

[63] This example is for the period Aut19-Sum20.
[64] This example is for the period Aut19-Sum20.

Fixed parameters in model 1	13	13	13	13	13
AIC	38,996.800	34,254.630	31,923.120	33,078.760	41,297.390
BIC	39,098.690	34,356.510	32,025.010	33,180.640	41,399.270
logLik	-19,485.400	-17,114.310	-15,948.560	-16,526.380	-20,635.690
deviance	38,970.800	34,228.630	31,897.120	33,052.760	41,271.390
Chisq	4.834	7.078	0.616	0.872	5.442
Df	4.000	4.000	4.000	4.000	4.000
Pr(> Chisq)	**0.305**	**0.132**	**0.961**	**0.929**	**0.245**

Appendix P: School-level means of coded responses to Teacher Survey 1

*Figure 10 – Description of responses to Teacher Survey 1: tables of **teacher counts** and **histograms of school means***

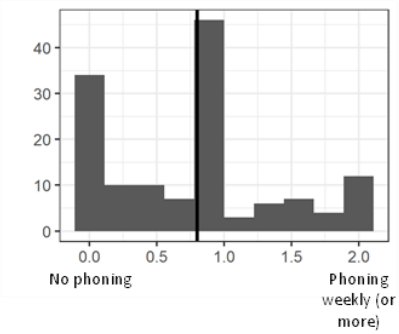
Phoning students

Did teachers speak to students (not parents) directly on the phone? 0 = not at all during the half-term; 1= either one or twice during the half term; 2 = about once a week or more.

Teacher responses (counts)

0 (no phoning)	1	2 (weekly or more)
200	216	123

Histogram of school means

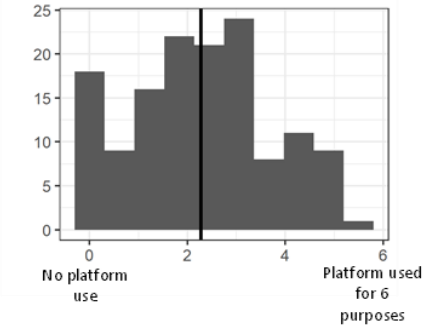


Number of purposes for online Platforms *(Platform)*

Did schools use a platform: to set work (+1); to receive work (+1); to share examples of student work (+1); for daily registration (+1); for text chat interactions (+1); for live audio/video chat interactions(+1)? One point for each, for a maximum of 6 total points.

Teacher responses (counts)

0 uses	1	2	3	4	5	6 uses
91	80	94	130	92	41	11

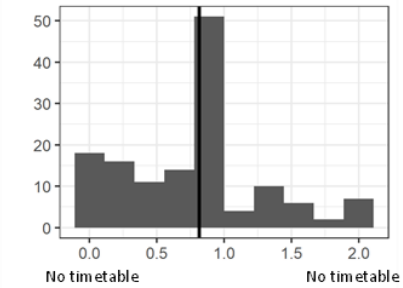


Timetabling approach *(Timetable)*

What timetabling approach did schools have?
0 = no timetable; 1 = a suggested timetable; 2 = instructed daily timetable.

Teacher responses (counts)

0 (no timetable)	1	2 (instructed daily timetable)
179	280	80

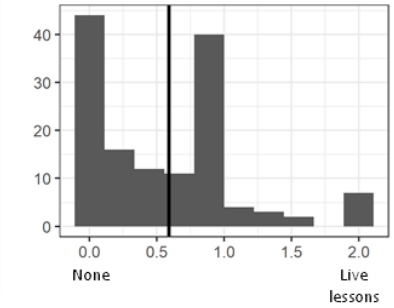


Videos/Live Lessons *(Vid/live)*

Did teachers have pre-recorded video lessons (scored 1) or live lessons (scored 2)? If neither, the variable is coded as zero.

Teacher responses (counts)

0 (none)	1 (video)	2 (live)
282	191	66



Frequency of work submission *(Work submission)*

How often did teachers expect work to be submitted? 0 = no recommendation about work submission; 1 = suggested work submission once a week or fortnight; 2 = several times a week; 3 = every day.

Teacher responses (counts)

0 (no requirement)	1	2	3 (daily)
185	104	84	166

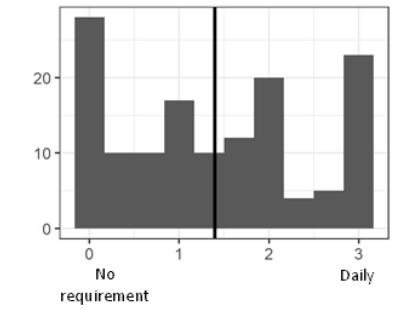


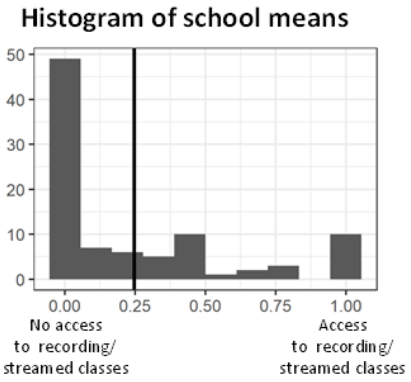
*Figure 11 – Description of responses to Teacher Survey 2: tables of **teacher counts** and **histograms of school means***

Absence provision for *individuals*

Did schools provide access to video/streaming of classes when individual children were absent? 0 = no access to any video/streaming of class; 1=video/streaming class

Teacher responses (counts)

0 (no access to video/stream)	1 (video/streamed lessons)
346	108

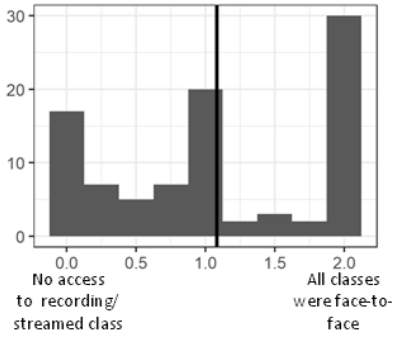


Absence provision for *classes*

Did schools provide access to video/streaming of classes when who classes needed to isolate? 0=no access to any video/streaming; 1=video/streaming; 2=no class isolation needed

Teacher responses (counts)

0 (no video lessons)	1 (recording/stream)	2 (face-to-face)
169	131	154

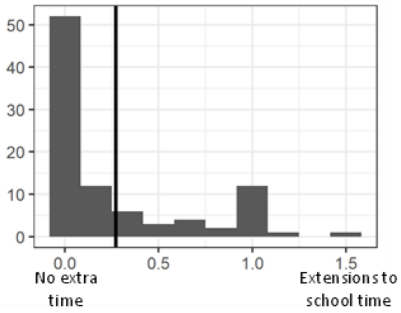


Extra time

Did schools extend the school day (+1); run extra lunchtime learning activities in order to compensate for lost learning (+1); remain open during holidays (+1)?

Teacher responses (counts)

0 (no extra time)	1	2	3 (multiple extensions to school time)
341	104	6	3

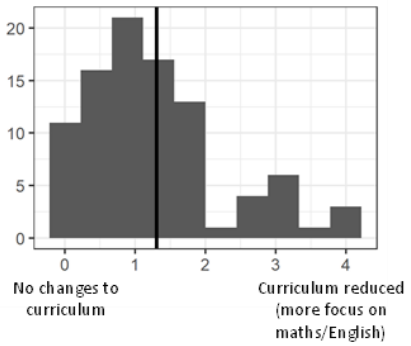


Reduced curriculum

Did schools reduce the curriculum, and focus more time on English and maths in the autumn 2020 term? Composite measure of "removing items from the curriculum" (+1) and "how many extra minutes per day were spent on English and maths (15 mins=1, 30 mins=2, 45+ mins=3, zero mins=0)

Teacher responses (counts)

0 (no change)	1	2	3	4 (reduced curriculum)
171	102	93	51	37



Small group interventions

Did schools make more/less use of small group and individual interventions (e.g. run by TAs)? 0=fewer interventions; 1=same number as normal; 2 to 4 is more than normal, either during lessons, or outside lessons, or both

Teacher responses (counts)

0 (less small group)	1	2	3	4 (more small group)
73	102	90	117	72

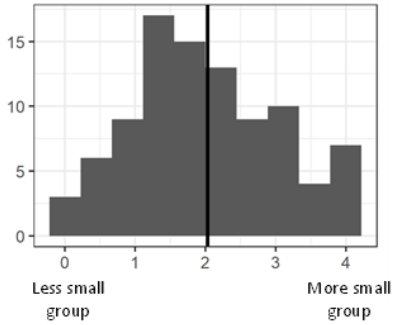


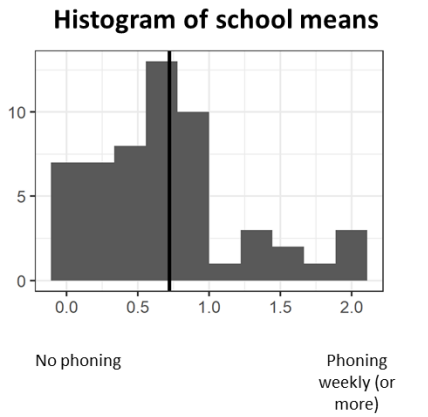
*Figure 12 – Description of responses to Teacher Survey 3: tables of **teacher counts** and **histograms of school means***

Phoning students

Did teachers speak to students (not parents) directly on the phone? 0 = not at all during the half-term; 1= either one or twice during the half term; 2 = about once a week or more.

Teacher responses (counts)

0 (no phoning)	1	2 (weekly or more)
163	102	78

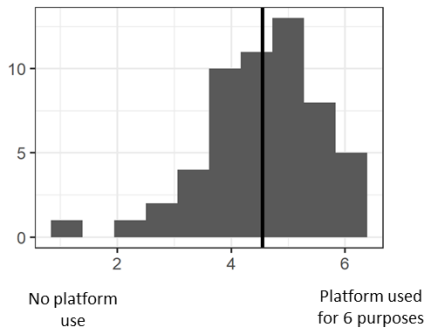


Number of purposes for online Platforms *(Platform)*

Did schools use a platform: to set work (+1); to receive work (+1); to share examples of student work (+1); for daily registration (+1); for text chat interactions (+1); for live audio/video chat interactions(+1)? One point for each, for a maximum of 6 total points.

Teacher responses (counts)

0 uses	1	2	3	4	5	6 uses
3	13	15	43	76	83	110

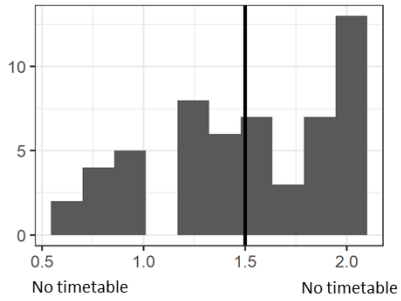


Timetabling approach *(Timetable)*

What timetabling approach did schools have?
0 = no timetable; 1 = a suggested timetable; 2 = instructed daily timetable.

Teacher responses (counts)

0 (no timetable)	1	2 (instructed daily timetable)
21	105	217

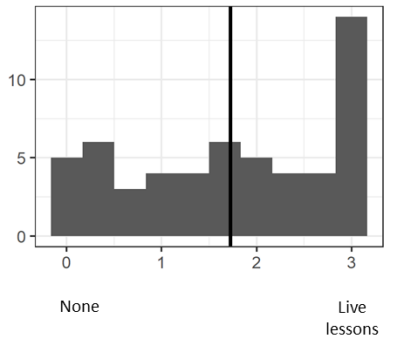


Live Lessons

Did you try teaching a 'live' lesson to your class in January, February or March? 0 = no or 'pre-recorded only'; 1 = "up to once a week, but no more"; 2 = most days; 3 = more than one most days

Teacher responses (counts)

0 (none)	1	2	3 (more than 1 a day)
85	57	68	133

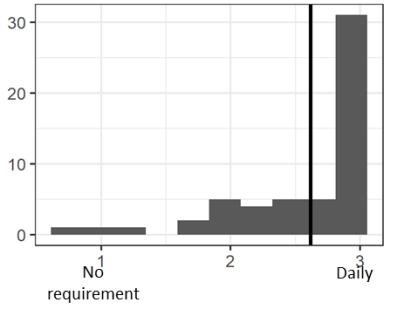


Frequency of work submission *(Work submission)*

How often did teachers expect work to be submitted? 0 = no recommendation about work submission; 1 = suggested work submission once a week or fortnight; 2 = several times a week; 3 = every day.

Teacher responses (counts)

0 (no requirement)	1	2	3 (daily)
16	7	58	262



Appendix Q: Robustness checks

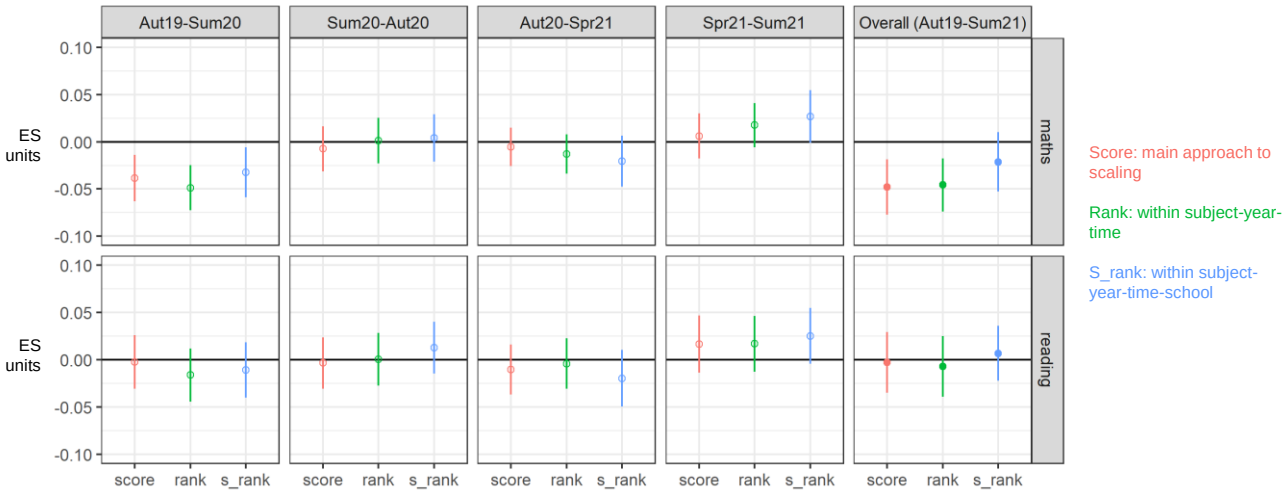
1. Standardisation/scaling of outcomes

As an initial robustness check we examine whether the results for RQ1 are sensitive to the method of scaling attainment outcomes by comparing our core findings using three different scaling approaches:

- Standardising scores (this method is represented in the main results)
- Ranks, defined by assessment/time period/year group
- Ranks, defined by assessment/time period/year group/school[65]

Figure 13 presents the results for these three different methods, illustrating the changes in disadvantage gaps for reading and maths in three periods (mirroring the structure of Table 8). In all cases we present average impacts across year groups 2 to 6. The top row of charts presents results for maths, with reading on the bottom row. The right-hand pane presents an average across the entire study period.

Figure 13 – Estimates of changes in disadvantage gaps, using 3 different approaches to scaling



Notes: all estimates average across year levels 2 to 6, as described in the methods section. Point estimates have 95% confidence intervals. All estimates are from the Q1 analysis sample (combining Full and Partial responders, see Figure 1).

2. Methods to control for the date of test administration

To assess whether our results are robust to different methods of controlling for test administration date, we compare the results of model 1a with a model that has no time controls (model 1a''):[66]

$$Y^k_{ijgT_1} - Y^k_{ijgT_0} = \alpha_j + \delta^k_{T_1 g} + \Delta^{\mathrm{k}}_{T_1 g} F_i + e^k_{ijg} \qquad (model\ 1a'')$$

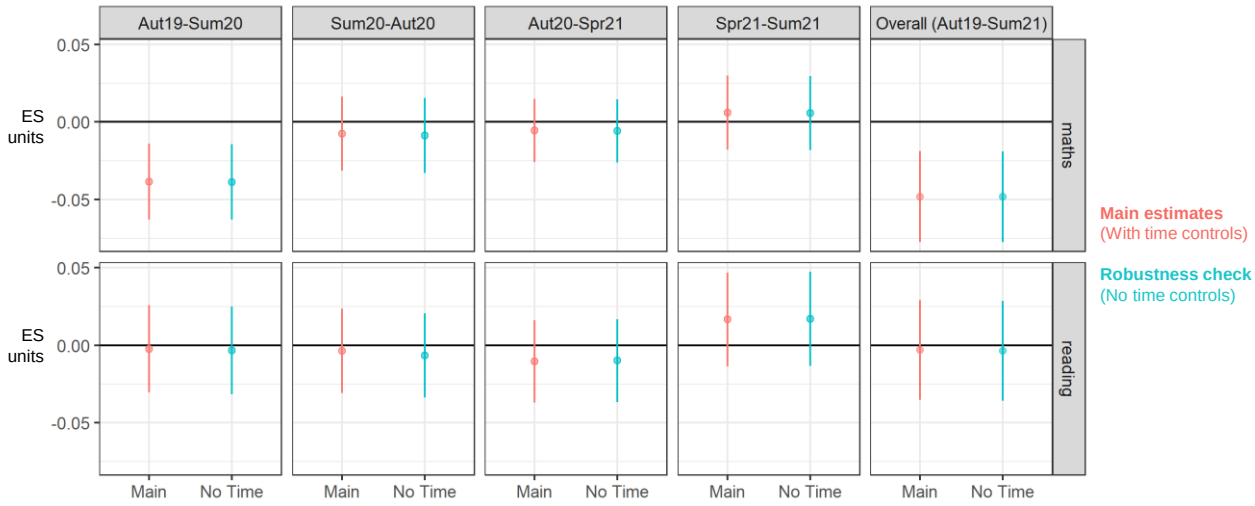
$$\alpha_j \sim N(0, \sigma^2_\alpha)$$

$$e^k_{ijg} \sim N(0, \sigma^2)$$

Figure 14 presents the results for reading and maths averaged across year groups 2 to 6. Including time controls barely changes our estimates.

[65] This method was not mentioned in the study plan, but is included to provide a more detailed examination of robustness to scaling.
[66] We fit analogous models for model 1b and model 1c

Figure 14 – Exploring the effect of time controls on $\hat{\Delta}_{T1}$



Notes: main estimates come from the models described in the section on Statistical Analysis. Estimates with no time controls come from equivalent models with the time terms ($month, month\_difference$) removed, see for example model 1a''.

3. Missing data

RQ1 Missingness analysis

As noted in the Participants section, our starting sample for RQ1 was defined by pupils who:

- had a valid assessment in reading or maths in Autumn 2019, and
- were on roll during the Autumn term 2020 and had a disadvantage flag (yes/ no) present

We asked schools to test pupils in each of reading or maths using the Summer 2020, Autumn 2020, Spring 2021 and Summer 2021 test suites. Based on the response patterns we classified pupils into three groups:[67]

- Full responders: those for whom results are observed in all four test suites
- Partial responders: those for whom between one and three further test results are observed
- Non responders: those for whom no further test results are observed

Figure 1 illustrates this process. Table 25 provides addition information, broken down into year groups.

Table 25 – Response by year group

	Response	Y2	Y3	Y4	Y5	Y6	Total
Reading	Full	745	728	1136	1358	969	4936
	Partial	2322	2364	3335	3024	2834	13879
	Non-response	207	181	269	307	1094	2058
	Total	3274	3273	4740	4689	4897	20873
Maths	Full	771	782	1024	1352	957	4886
	Partial	2394	2483	3204	2829	2924	13834
	Non-response	191	202	310	325	836	1864
	Total	3356	3467	4538	4506	4717	20584

[67] This was done separately for reading and maths.

The principal reason for pupil non-response is school non-response. These are cases where at least two-thirds of pupils in the same national curriculum year were classified as non-responders (mostly attrition at school/ year group level). Of the 3,922 non-responders in both subjects, 2,555 (65%) fall into this category.

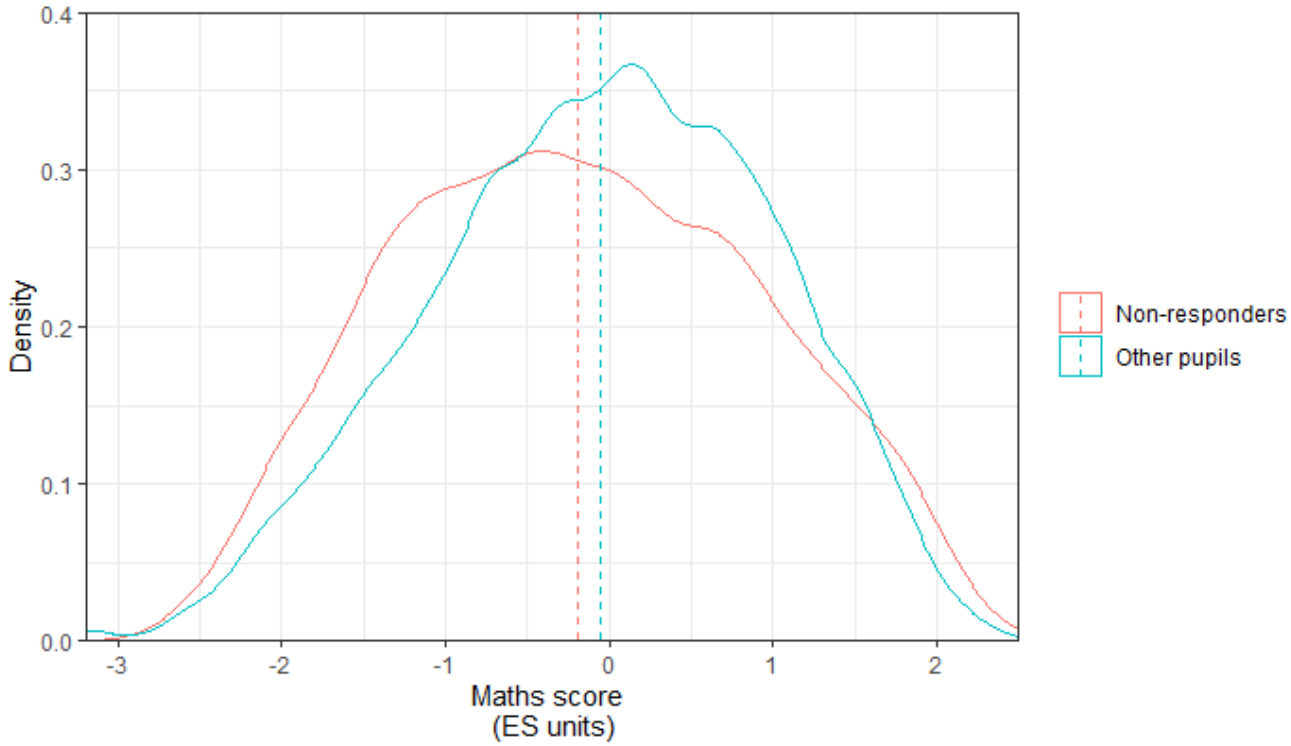
Table 26 shows that the Autumn 2019 attainment of this group in maths was fractionally below average (-0.03 ES units) but was higher than non-responders in schools with <66% non-responders. The table presents the mean Autumn 2019 maths score alongside a version which absorbs the school mean. This suggests that non-responders in schools with a non-response rate of <40% tend not only to be low-attaining but also low-attaining given the school they attend. By contrast, non-responders in schools with a 40% to 66% non-response rate tend to be lower attaining but not especially different to other pupils in the school. This could arise from one class in a year group participating in the project and another one not.

Table 26 – Non-responders in maths by school non-response rate

School % non-responders	Autumn 2019 score		
	Actual	Absorbing school mean	Number
<10%	-0.67	-0.43	282
<40%	-0.47	-0.19	87
<66%	-0.27	-0.01	318
>=66%	-0.03	0.01	1177
All non-responders	-0.19	-0.07	1864

The distribution of Autumn 2019 maths scores for non-responders is compared to other pupils in Figure 15. The non-responder distribution is shifted to the left, suggesting lower baseline performance.

Figure 15 – distribution of prior-attainment for non-responders and others for RQ1



Note: dotted mean lines represent means.

Several other forms of non-response can be identified. In Table 27 we show the number of missing test results in each subject in each term (non-responders plus the partial responders who did not respond in the

given term). By far the largest category is school non-response. These are cases where at least two-thirds of pupils in the same national curriculum year at a given school were classified as non-responders.

Table 27 – Other sources of non-response for RQ1

		Summer 2020	Autumn 2020	Spring 2021	Summer 2021
Reading	Leavers	67	80	101	283
	School non-response	6179	4508	4647	11032
	School response pupil non-response	1284	1079	1127	812
	Total missing	7530	5667	5875	12127
Maths	Leavers	60	87	93	288
	School non-response	5214	7003	4570	11359
	School response pupil non-response	1254	952	1029	646
	Total missing	6528	8042	5692	12293

Reasons for pupil non-response in responding schools include:

- Working below the level of the test
- Invalid test results
- Absence
- Technical issues linking data

However, we cannot reliably distinguish between the four.

Next, we turn to the question of whether our RQ1 analysis sample differs from the broader sample of schools who took part in the study. Table 28 examines whether observable characteristics differ between our analysis sample ("Full responders and Partial responders) and non-responders.

Table 28 – Observed characteristics of responder and non-responder samples

		Full responders for RQ1	Partial responders for RQ1	Non-responders for RQ1	All pupils
Reading	Autumn 2019 score	0.02	-0.04	-0.07	-0.03
	% disadvantaged	28%	28%	33%	29%
	% EAL	24%	25%	29%	25%
	% SEN	15%	15%	18%	16%
	% female	49%	50%	48%	49%
	% London schools	6%	19%	38%	18%
	% northern schools	23%	23%	17%	22%
	Number of pupils	4936	13879	2058	20873
Maths	Autumn 2019 score	0.04	-0.08	-0.19	-0.06
	% disadvantaged	28%	30%	37%	30%
	% EAL	22%	24%	30%	24%
	% SEN	17%	16%	19%	17%
	% female	48%	49%	48%	49%
	% London schools	5%	16%	27%	14%
	% northern schools	20%	25%	21%	23%
	Number of pupils	4886	13834	1864	20584

Non-responders are more likely to be disadvantaged, and to have been lower attaining in Autumn 2019. Pupils attending London schools are disproportionately more likely to be non-responders. We illustrate these patterns more formally using logistic regression. We examine how the characteristics in Table 28 are associated with 'being excluded from the analysis sample'. Let $P_{ij} = 0$ if student i (from school j) is in the

"Non-Responder" samples for RQ1 (and $P_{ij} = 1$ if they're either in the Full Responder or Partial Responder samples):

$$\Pr(P_{ij} = 1) = logit^{-1}(\boldsymbol{\beta X_{ij}})$$

where $\boldsymbol{X_{ij}}$ are the variables listed in Table 28.The results of these regressions are presented in Table 29. Children who were higher attaining were more likely to remain in our sample and those in London were less likely to be represented. Conditional on other predictors, FSM6 status was negatively associated with being in our analysis sample.

Table 29: Logistic regression to predict probability of missingness

	Dependent Variable: Full OR Partial Responder (Binary variable)	
	Maths	Reading
Autumn 19 score	0.090**	-0.021
	(0.025)	(0.025)
FSM6	-0.306**	-0.246**
	(0.052)	(0.052)
EAL	-0.122*	0.089
	(0.057)	(0.056)
SEN	-0.113	-0.246**
	(0.068)	(0.066)
Female	0.029	0.031
	(0.050)	(0.048)
London school	-0.900**	-1.239**
	(0.062)	(0.056)
Northern school	-0.026	-0.043
	(0.063)	(0.065)
Constant	2.635**	2.618**
	(0.049)	(0.048)
Observations	20,584	20,873
Note:		*p < 0.05, **p < 0.01

Sensitivity of main RQ1 results to missingness imputation

In this section we see if the results presented for RQ1 are robust to the way in which we have used multiple imputation using chained equations (MICE). Specifically, we compare estimates from Table 8 to equivalent results when we only focus on the Full Responder sample.

In our main estimates, we impute 20 datasets for each combination of year-level/subject/period.[68] For each dataset, we fit the relevant model (1a, 1b or 1c) and calculate the relevant $\hat{\Delta}$ parameter. We then take the average estimate across these 20 imputed datasets.

[68] For example, for model 1a we impute 20 datasets for the cohort of children entering year 2 in 2020-21, who have missing maths data in the Summer 2020 period. Similarly, we impute 20 maths datasets for the cohort entering year 3, year 4 and so on. Note that this number of dataset imputations is a deviation from our analysis plan, which suggested we would follow the rule-of-thumb recommended by White et al. (2011), and have the same number of imputed datasets as the average percentage of missingness (at

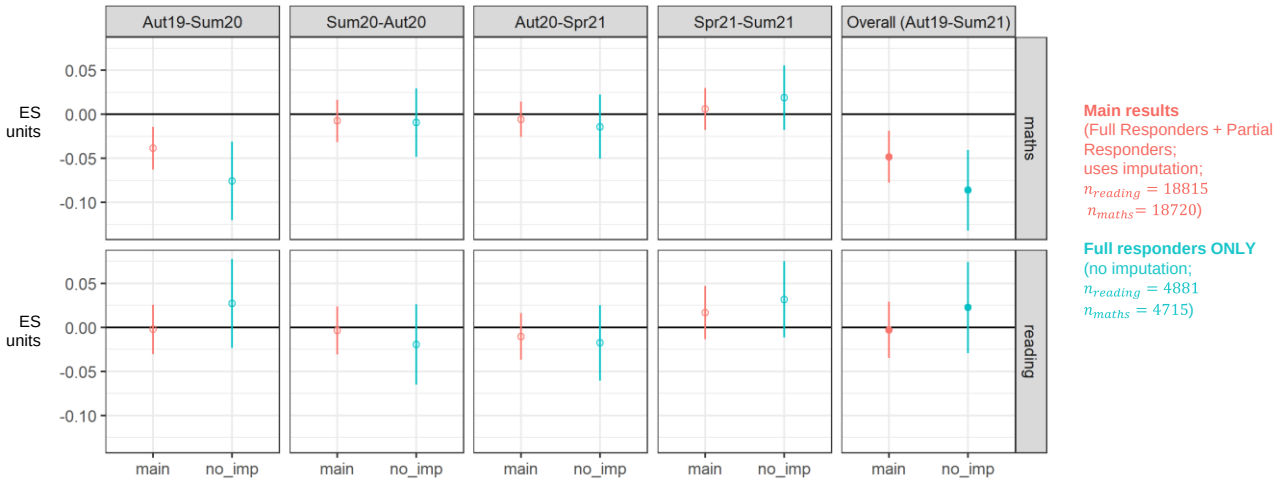
To see how this impacts the findings, in this section we present a set of results that have no imputation (i.e. they only use the "Full Responder" sample, for whom there is no missingness). The results of this comparison are presented in Table 8. The "main" estimates are the original estimates from Table 8, and the "no_imp" figures are the results when we only focus on Full Responders.

While we observe some differences, the overall conclusions for the two sets of results are very similar. For maths, both sets of results show evidence of gaps widening over the study period (Autumn 2019 to Summer 2021). In both cases this is driven by the change in the first period – from Autumn 2019 to Summer 2020. In subsequent periods, there is very little evidence of change in disadvantage gaps. For reading, the main pattern of results is consistent across the 'main' and 'no imputation' analysis: we find no clear evidence of disadvantage gaps widening or shrinking throughout the study period.

As noted above, despite the robustness of the main conclusions, there are minor differences in the point estimates of gap changes between the main analysis and 'no imputation' results (illustrated in Figure 16 – Comparing main results (using imputation) with results from the Full Responder sample (no imputation)). These differences don't follow a clear pattern. For maths, the 'no imputation' estimates find that gaps have widened by more than our main analysis suggests. For reading, the effect goes in the opposite direction: the 'no imputation' estimates suggests that disadvantage gaps became narrower than the main analysis suggests. Overall, the imputed analysis results in a more moderate difference between maths and reading gap changes. Finally, we note that the magnitude of these differences between the imputed and 'no imputation' estimates are not large, and that all confidence intervals of main and non-imputed results are overlapping.

Finally we note that the main results have narrower confidence intervals than the 'no imputation' results. This is as expected, given that the main analyses use data from more students.

Figure 16 – Comparing main results (using imputation) with results from the Full Responder sample (no imputation)



Notes: negative estimates indicate a widening of gaps. Main results (in red) use estimates come from the RQ1 analysis sample (Full Responders + Partial Responders, see Figure 1 for details). The righthand pane is a summary of the overall study period.

RQ3 Missingness analysis

This section examines the robustness of RQ3 to missingness and mirrors the analysis examining RQ1 missingness. We use logistic regression, where $P'_{ij} = 0$ if student i is in school j that has data for RQ2 (teacher sample) but lacks outcome information for analysis in RQ3 (and $P'_{ij} = 1$ for students who are included in RQ3 analysis). We fit the following model:

student level). This proved to be very time consuming, so instead we present results that are averaged across 20 imputed data samples. The results are not sensitive to this choice.

$$\Pr\left(P'_{ij} = 1\right) = logit^{-1}(\boldsymbol{\beta X_{ij}})$$

where $\boldsymbol{X_{ij}}$ covariates.[69]

The results of these regressions are presented in Table 30. They suggest that, conditional on other observed characteristics, FSM6 students were more likely to attrit (i.e. not have post-Covid-19 attainment data, despite attending schools represented in the Teacher Surveys) than their peers. There were no other clear patterns in the data.

Table 30: Results of logistic regression to predict missingness for RQ3

	Dependent Variable: Responders for RQ3					
	Maths RQ3a	Reading RQ3a	Maths RQ3b	Reading RQ3b	Maths RQ3c	Reading RQ3c
Autumn 19 score	1.369**	0.765**				
	(0.138)	(0.158)				
Summer 20 score			-0.077	0.245**		
			(0.290)	(0.038)		
Autumn 20 score					0.207**	0.209**
					(0.029)	(0.027)
FSM6	-1.400**	-1.337**	-1.259*	0.046	-0.234**	-0.144*
	(0.250)	(0.322)	(0.559)	(0.076)	(0.060)	(0.056)
EAL	-0.097	-0.268	0.357	0.534**	-0.078	0.071
	(0.251)	(0.323)	(0.653)	(0.087)	(0.059)	(0.056)
SEN	-0.046	0.135	1.198	-0.149	0.087	-0.009
	(0.247)	(0.363)	(1.076)	(0.092)	(0.080)	(0.075)
Female	0.496*	0.442	0.310	0.150*	0.129*	0.084
	(0.232)	(0.315)	(0.548)	(0.070)	(0.056)	(0.052)
Constant	6.685**	6.511**	6.515**	1.795**	1.074**	1.124**
	(0.299)	(0.329)	(0.518)	(0.060)	(0.050)	(0.045)
Observations	13,481	12,741	7,893	7,893	6,978	8.279

**p < 0.05, **p < 0.01

Note: we did not include London and Northern indicators, as these made models unstable.

Sensitivity of main RQ3 results to missingness (assuming Missing-At-Random)

Last, we examine the impact of imputing missing values, conditional on observed characteristics, using MICE. Specifically, we compare estimates presented in Figure 3 and Figure 4 to equivalent results when we use MICE to impute outcomes for students who were represented in the teacher survey, but lacked post-Covid-19 outcome data.[70] For each analysis (RQ3a, RQ3b and RQ3c) we impute 20 datasets. For each dataset, we fit the relevant models (3a, 3b or 3c) and calculate $\hat{\phi}_0$ and $\hat{\phi}_1$ for the school response variables ($\boldsymbol{Z}$, $\widetilde{\boldsymbol{Z}}$ or $\boldsymbol{Z}^*$). We then take the average estimate across these 20 imputed datasets and compare the results to those in the results section (Figure 3 or Figure 4).

For RQ3a, the results of this procedure are presented in Figure 17. The top row (circles) shows the new estimates, using imputed data. The bottom row (triangles) shows the main estimates. The imputation

[69] We removed London and Northern indicators, as these were very rare in our missingness model and led to unstable estimates.
[70] For RQ3a we were concerned with missing outcome data for September 2020 ('Summer 2020'); for RQ3b, the relevant missingness was for November/December 2020 ('Autumn 2020').

procedure produces results that are extremely similar. The same is true for the RQ3b (Figure 18) and RQ3c (Figure 19).

Figure 17 – Comparison of imputed and main results for RQ3a

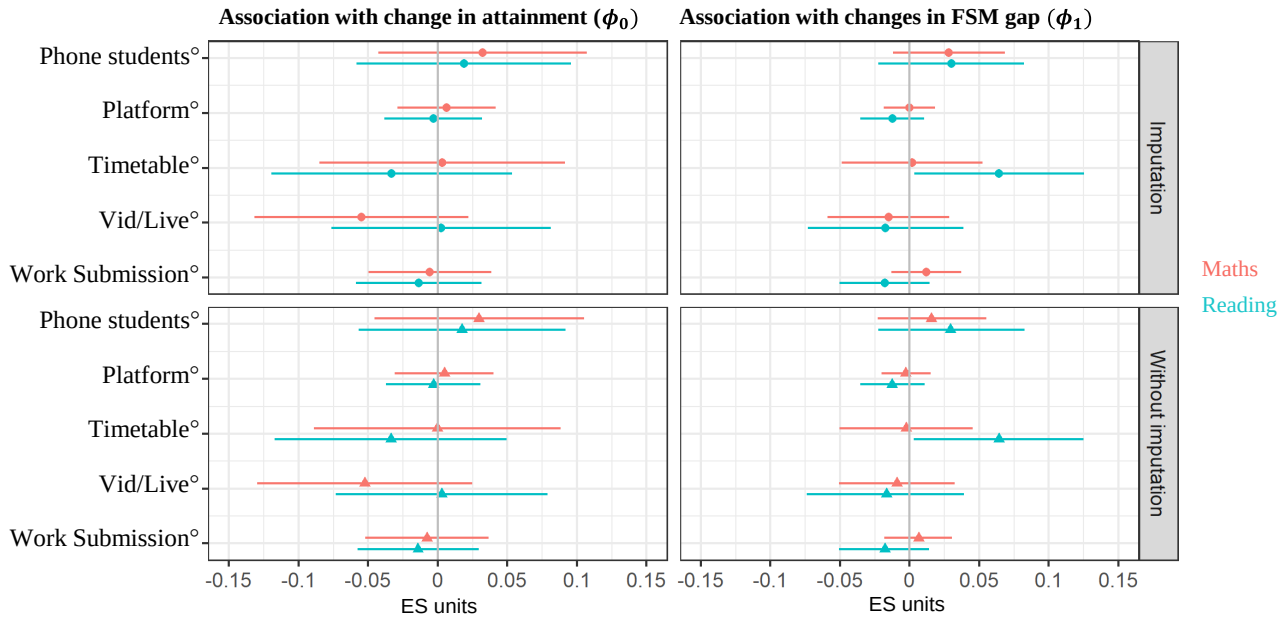


Figure 18 – Comparison of imputed and main results for RQ3b

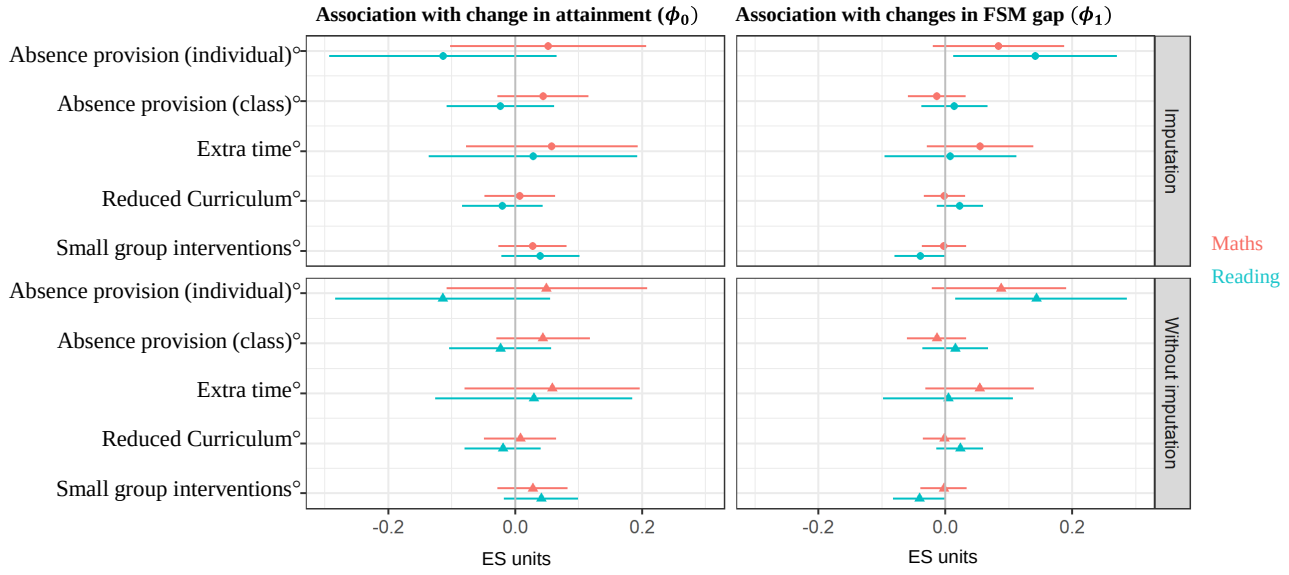
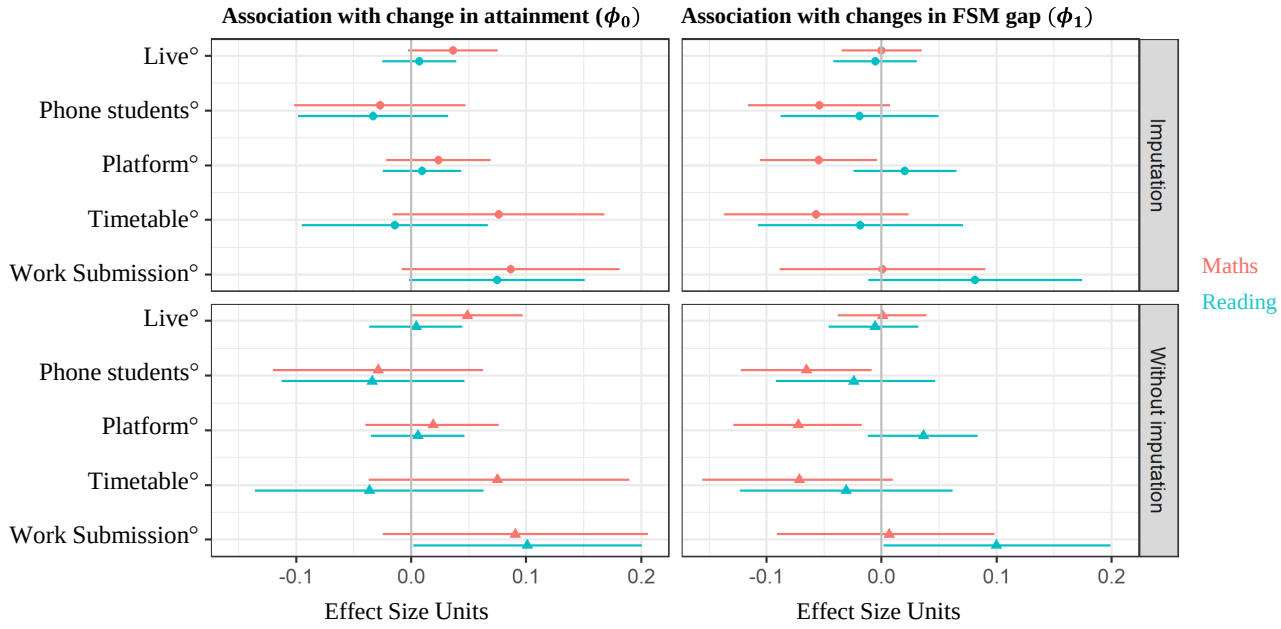


Figure 19 – Comparison of imputed and main results for RQ3c



4. Absences

Data on absences

We were able to obtain information on absences for a subset of pupils in the 2020-21 year. Unfortunately, absence data was unavailable for some schools across the entire study period. Similarly, we were unable to collect any absence data for 2019-20. Table 31 summarises the sample for whom we have absence data in 2020-21.

Table 31 - Summary of absence data collected for 2020-21

		Maths	Reading
RQ1 analysis sample	Pupils	18,720	18,815
	Schools	114	117
Sample with absence data	Pupils	5,622	5,954
	Schools	48	51

Our measure of absence is at the individual level, and is defined as "proportion of school sessions attended during the 2020/21 academic year". The distribution of absences for FSM pupils and their peers is summarised in Figure 20. As is clear from the plots, a vast majority of students attended at least 80% of sessions. Absence rates were slightly higher among FSM pupils.

Robustness check

In this section, we focus on the sub-sample of students for whom we have absence data and ask: "are the results of RQ1 robust to the inclusion of individual-level controls for absence?". We compare the main RQ1 analyses, to analyses from models that have a single extra control "$absences_i$" (the proportion of school sessions attended in 2020/21). The results are presented in Figure 21. For the sub-sample of schools where we have absence data, including this data in our analysis makes virtually no difference to pattern of results. This provides some reassurance that the results of RQ1 are not affected by differences in attendance patterns for FSM and non-FSM pupils.

Figure 20 - Absence rates in 2020-21

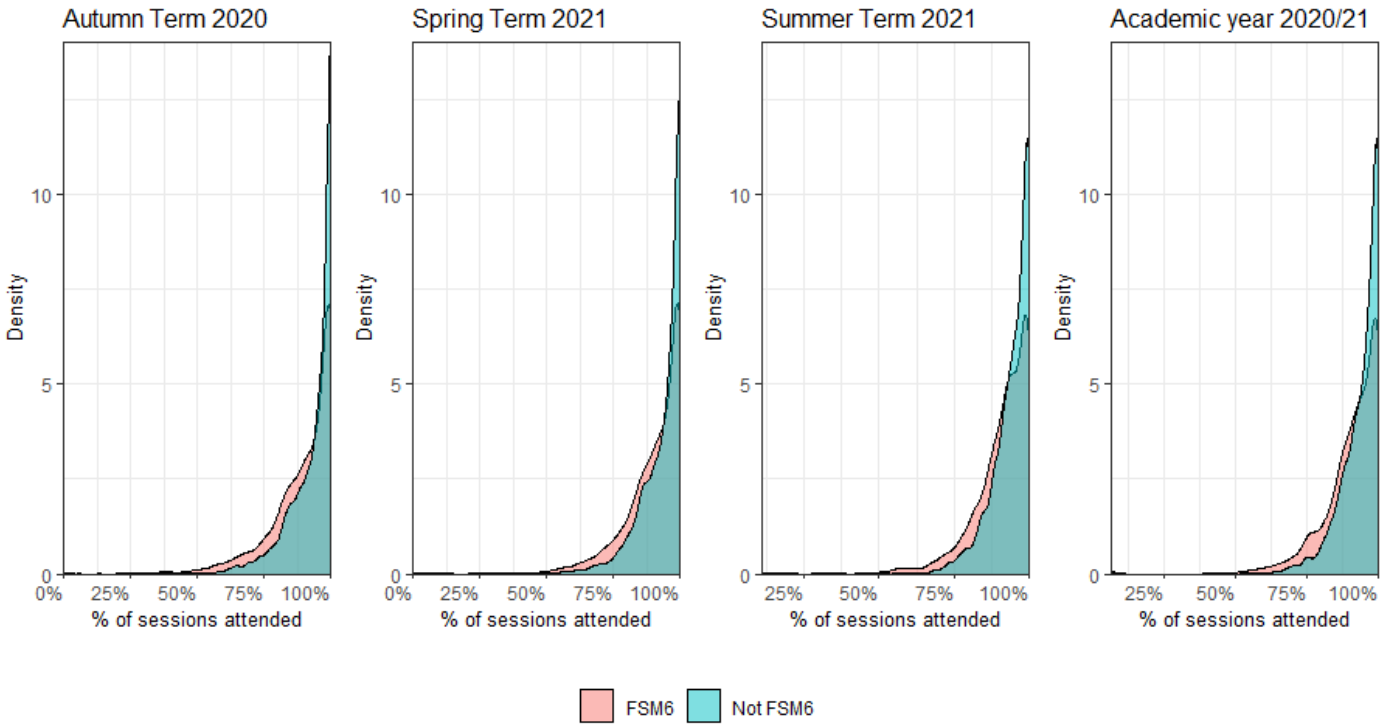
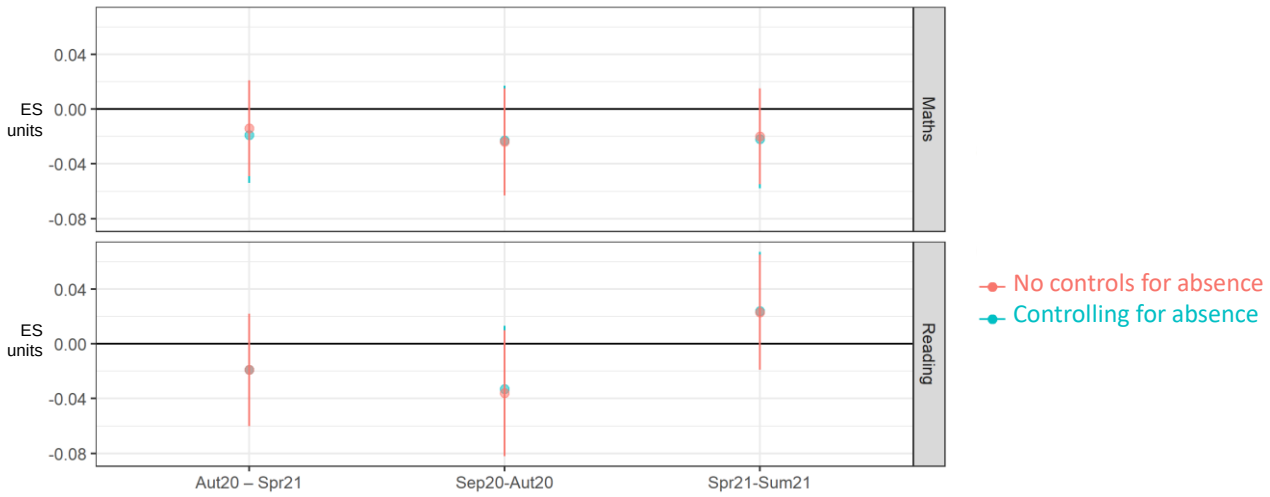


Figure 21 - Comparing RQ1 results with/without controls for absence



Note: analysis completed on subset of students for whom we have absence data (see Table 31). Only 2020-21 considered, as no absence data was available for 2019-20.

Appendix R: Timeline

Dates	Activity
July 2020	School Recruitment and first teacher survey administered
September/October 2020	Schools administer Spring 2020 Rising Stars assessments
November/December 2020	Schools administer Autumn 2020 Rising Stars assessments
December 2020	Second teacher survey administered
January 2021	Collection of assessment data from schools
March 2021	Draft report delivered
March/April 2021	Third teacher survey administered
May 2021	Schools received reports from FFT
June 2021	Schools administer Summer 2021 assessments
September 2021	Final Report submitted to EEF

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk
@EducEndowFoundn
Facebook.com/EducEndowFoundn