Project Title: Research on the impact of school closures in key stage 1
Evaluation Study Plan

Evaluator (institution): NFER
Principal investigator(s): Sue Rose, Rebecca Wheater and Pippa Lord

Education
Endowment
Foundation

PROJECT TITLE	Impact of Key Stage 1 school closures on later attainment and social skills (a longitudinal study)
EVALUATOR (INSTITUTION)	National Foundation for Educational Research
PRINCIPAL INVESTIGATOR(S)	Sue Rose, Rebecca Wheater and Pippa Lord
STUDY PLAN AUTHOR(S)	Lydia Fletcher, Jose Liht and Rebecca Wheater
STUDY DESIGN	Observational longitudinal study
PUPIL AGE RANGE AND KEY STAGE	6 to 9 years, Key Stages 1 and 2
NUMBER OF SCHOOLS	Approximately 87
NUMBER OF PUPILS	Approximately 6,100
PRIMARY OUTCOME MEASURE AND SOURCE	NFER Assessments in reading and maths. Year 2 and 3 in spring 2022, Year 3 and 4 in spring 2023.
SECONDARY OUTCOME MEASURE AND SOURCE	Social skills outcomes. PSMAT (Peer Social Maturity scale) and NFER bespoke questions.

Study Plan version history

VERSION	DATE	REASON FOR REVISION
1.3 [*latest*]	28.02.2023	Revisions following peer review and publication of 2021/22 report
1.2	08.06.2022	Revisions following peer review
1.1	31.03.2022	Revisions following EEF review
1.0 [*original*]	10.03.2022	

Contents

Background and study rationale

In response to rising levels of Covid-19 infections, schools in England were closed to the majority of children from March 2020. From June 2020, selected year groups were allowed to return, if this could be achieved safely by adhering to guidelines aimed at reducing the spread of infections. Schools were closed again to the majority of pupils from January 2021 until March 2021. These events have created a set of unique circumstances for schools where pupils' opportunities for formal learning and social interaction were reduced for a significant part of two school years.

NFER were appointed by EEF to conduct research on the impact of Covid-19 school closures and subsequent support strategies on attainment and socioemotional wellbeing in Key Stage 1 during the 2020/21 academic year. The research found that the disruption to Key Stage 1 pupils' education as a result of the pandemic has resulted in significantly lower achievement in reading and maths compared with pupils before the pandemic. In addition, the partial closures of schools has led to an increase in the disadvantage gap. However, it did also show some evidence of the first steps of recovery in mathematics towards the end of the academic year (Rose *et al.* 2021). With schools continuing to stay open and teachers providing ongoing catch up support throughout 2021/22, we expect pupils to make some progress to recovering learning loss. One of the aims of this study is to understand the extent to which pupils who had learning disrupted in KS1 were able to recover to levels that they should be before the pandemic, and factors that predict this recovery.

The initial one-year project on the impact of Covid-19 school closures on pupil outcomes in Key Stage 1 has provided rich data on the impacts of missed learning, but there are more questions about the extent to which pupils' learning recovers that can be answered by continuing to track the pupils involved in the study. We believe that a focus on pupils affected by school closures during KS1 and the early years of KS2 is of considerable significance, particularly as the youngest pupils had not completed their reception year before the first partial school closures. At this stage, pupils learn school routines and expectations; crucially, for reception children moving into Year 1, Covid-19 disrupted this transition phase which is usually carefully managed by schools (Children's Commissioner, 2020). Children also begin to develop skills, both academic and social, that will be the foundation of future learning (Sylva *et al.*, 2004).

Alongside other work commissioned by EEF looking at the impact of Covid-19 across other key stages, this work will enable a deeper understanding of the long-term impact of school closures on pupil attainment, and providing information to support this cohort as they move through school.

Overview of the study

This longitudinal study will be unique in following the youngest school-age children affected by the pandemic, with a baseline from during the pandemic and comparisons to pre-pandemic standardisation samples, with an aim to understand how quickly children catch up to where they might be expected to be, had the pandemic not happened. This research aims to provide evidence on the impact of school closures on attainment and any differential impact on subgroups (with a particular focus on the disadvantage gap), and on reading and mathematics/specific domains within reading and mathematics. The study will be a combination of quantitative research looking at pupil attainment derived from NFER test data, supplemented with evidence of school practices (from quantitative and qualitative data) and teachers' perspectives of a subsample of pupils' social skills (quantitative data).

Research Design

Research questions

Research question	Data source
1. RQ1a. To what extent does pupils' attainment in reading and maths recover by spring 2022 and spring 2023 compared to pre-pandemic levels? RQ1b. How is pupils' performance changing over time during the course of our study?	RQ1a. NFER Tests reading and maths raw and standardised scores, spring 2022 (for Year 2 and Year 3 pupils) and spring 2023 (for Year 3 and Year 4 pupils). Compared to pre-pandemic standardisation data (i.e. 2019 for Year 2 spring assessments & 2017 for Year 3 and Year 4 spring assessments). Cross-sectional analysis. RQ1b. NFER Tests reading and maths standardised scores. Repeated measures analysis: baseline at spring 2021 (for pupils in Year 1 and Year 2), spring 2022 (for pupils in Year 2 and Year 3), and spring 2023 (for pupils in Year 3 and Year 4).
2. RQ2a. What is the attainment gap between disadvantaged pupils and their peers in reading and in maths in spring 2022 and in spring 2023? RQ2b. To what extent do different groups recover by spring 2022, and by spring 2023; in particular, how is the gap between disadvantaged children and their peers changing over time during the course of our study?	RQ2a. NFER Tests reading and maths raw and standardised score data from spring 2022 and spring 2023, matched to pupil background data from schools. Cross-sectional analysis. Groups to be identified by gender and free school meal eligibility. Individual models for maths and reading by year group. RQ2b. NFER tests reading and maths standardised scores, repeated measures analysis from spring 2021, spring 2022 and spring 2023.
3. Is attainment in some domains in reading/maths changing or recovering at a different rate from others?	NFER Tests reading and maths item-level data.
4. In the 2021/22 and again in the 2022/23 academic years, what practices have been adopted and what learning opportunities have been provided by schools to help pupils catch up; and what challenges have been faced by staff?	Annual school-level surveys
5. Are social skills at or behind expectations, and to what extent do they improve between subsequent academic years? To what extent do the socio-emotional skills of younger pupils, who spent large amounts of time at home during the pandemic, improve between subsequent academic years?	Teachers complete the Peer Social Maturity scale (PSMAT) and additional questions targeting wider socio-emotional skills in spring 2022 and spring 2023 on a (random) sample of 12 pupils from each year group in each school. The 2023 spring PSMAT scores can be compared with the spring 2022 scores, and scores can be compared with the pre-pandemic norms[1]. If cell count allows, PSMAT scores will be analysed by pupil characteristics; gender and free school meals eligibility. In year 1 of the study, the additional questions will be analysed for reliability and, if suitably reliable, comparisons can be made between 2023 and 2022 average scores.

---

[1] Norms sourced from Fink, E., de Rosnay, M., Peterson, C., & Slaughter, V. (2013). Validation of the Peer Social Maturity Scale for assessing children's social skills. *Infant and Child Development*, *22*(5), 539-552.

This longitudinal study follows the progress of pupils in Year 2 and Year 3 in the 2021/22 academic year, and in Year 3 and Year 4 in the 2022/23 academic year, using data collected in the 2020/21 academic year in the original study when the pupils were in Year 1 and Year 2 as a baseline. NFER assessments will be used to provide attainment data in spring 2022 and 2023. Comparisons will be made to the spring data from the previous academic year to ascertain the progress of the pupils. As per the existing study, the main focus will be on the measurement of two attainment gaps:

- The 'Covid-19 gap'; the extent to which pupils' attainment in reading and maths has been impacted by partial school closures. This is measured by the difference between pupil performance in spring 2022 (and spring 2023) compared with the performance of the pre-covid standardisation sample of the equivalent year group.

- The 'disadvantage gap; the extent to which FSM pupils show lower reading and maths performance compared to their non-FSM peers. This is measured by the difference in attainment between pupils who are eligible and not eligible for free school meals in the spring 2022 (and again in the spring 2023) assessments, and how the gap is changing over time during the course of this study.

A teacher-completed pupil-level social skills development survey will also be administered to a sub-sample of pupils within each school. Additional information will be collected through a school-level survey to identify school practices, pupil support and any catch-up activities being undertaken with the pupils as well as challenges for staff. This will provide context to the attainment and social skills findings.

Table 1: Research Design

Design		Longitudinal Observational study
Unit of analysis (school, pupils)		Schools and pupils
Number of Units to be included in analysis		86 schools and approximately 6,100 pupils as of February 2022
Outcome 1	variable	**Maths attainment**
	measure (instrument, scale, source)	NFER spring raw and standardised test scores
Outcome 2	variable(s)	**Reading attainment**
	measure(s) (instrument, scale, source)	NFER spring raw and standardised test scores
Outcome 3	variable(s)	**Social skills and wellbeing?**
	measure(s) (instrument, scale, source)	Teacher-completed social skills questionnaire

Participants

All schools which participated in any wave of the research into the impact of school closures in 2020/2021 were invited to take part in this study. Wave 3 of the Covid baseline study has a sample of 155 schools and 5843 pupils currently in Year 1 and 5916 pupils currently in Year 2. There were 168 schools involved in the first wave of the original project (wave 1), therefore by wave 3, 92% of the schools were retained at the end of Wave 3 indicating that the targeted schools were highly engaged in the original project and we anticipated that many of the schools would sign up to this longitudinal project. In October 2021, schools were invited to participate in this longitudinal study.

In year 1 of this current study (2021/22), teachers were asked to mark and upload test data to the NFER progress tool. We scheduled reminder strategies where test data are not forthcoming, however, we anticipate the following factors will incentivise participation:

- provision of free spring assessments to schools as a pre-incentive
- a discount for future NFER tests if schools successfully upload item-level data (available for schools who upload data in year 1, and for schools who provide their data for marking in year 2 of the study)
- summary results based on schools' item-level domain analysis disseminated through a feedback leaflet

This is a different incentive package from the baseline study that was conducted in 2020/21, which provided diagnostic information and required NFER to mark the assessments in order to do that. In year of this current study (2021/22), we identified removal of NFER marking as the biggest cost saving we can make for this study, whilst still collecting valuable data. It was important that we offered marking to schools in the baseline study in order to ensure we were not taking away teachers' during the uncertainties of partial school closures and reopenings, and to ensure that there were no barriers to participation for schools with, e.g. high levels of FSM pupils, and to provide schools with diagnostic level data. Instead, we have offered schools a discount on future tests in each year of their participation (not offered in the baseline study).

To ensure good participation in year 2 (2022/23) of the current study, NFER will mark the assessments. As in year 1 (2021/22) schools will be provided with: free spring assessments, a discount on a future order (as per above), and a summary feedback document.

Sample size calculations

Standardised mean difference effect sizes seen so far for the Covid-19 gap are -0.17, -0.14, -0.24 and -0.19 (Year 2 pupils reading and maths autumn 2020, and reading and maths spring 2021, respectively).

We have carried out simulations to estimate the MDES for the longitudinal sample, these are based on the size of effects obtained from the original Covid baseline study, power of 0.80, alpha of 0.05, ICC = 0.12 and FSM proportion of 0.16 for a longitudinal design of 3 waves (baseline, plus year 1 and year 2 of the current project).

The following table shows the estimated minimum detectable effect sizes for achieved samples.

Number of schools analysed (i.e. achieved sample)	MDES for change in disadvantage gap	MDES for Covid gap (effect size between project sample and pre-Covid standardisation sample)	MDES for Covid gap (effect size between project sample at two time points, e.g. spring 2021 and spring 2022)
80	0.22	0.19	0.08
100	0.20	0.18	0.08
120	0.18	0.17	0.07
150	0.16	0.17	0.06

The simulations indicate that even with 80 schools, we have a viable project, as they suggest we will be able to detect educationally relevant changes in the Covid gap – the MDES for the Covid gap is of similar magnitude to those seen in the previous study. But we are unlikely to detect the **changes** we expect to see in the disadvantage gap, not because they were comparably smaller, but due to the imbalance between FSM and none FSM pupils and thus the loss in effective sample size. To note, the baseline study saw changes over a four to six month period, whilst this study will look at changes over 12 months.

To estimate the power of the study to detect standardised mean difference effect sizes, a power analysis by simulation was undertaken (Arnold et. al., 2011). Power was calculated separately for the Covid gap and the disadvantage gap, for all combinations of 80 to 150 schools (after attrition) and effect sizes between 0.01 and 0.3 in intervals of 0.01 (i.e. 0.01, 0.02, 0.03… 0.28, 0.29, 0.3). For each of the N/effect size combinations, 1000 data sets were simulated in the proposed longitudinal design of three time points (baseline, plus year 1 and year 2 of the current project) and 38 pupils per school. The ICC was taken to be 0.12. Setting the sum of the school level and residual variances to be 1 meant that the school level and residual variances used to simulate the variability in the data were the ICC and 1-ICC respectively. This also ensured that the coefficients of models were on the effect size scale.

For the Covid gap, a difference of the desired size was induced at one of the post baseline time points. A linear mixed effects model was fit to each of the simulated data sets with school as the random effect and time point as the fixed effect. All between time point contrasts were tested for significance at a 5% significance threshold.

For the disadvantage gap, 16% of simulated pupils were labelled as FSM and the desired effect size was induced in the FSM pupils at one of the time points. A linear mixed effects model was fit to each of the simulated data sets with school as the random effect and time point, FSM and their interaction as the fixed effects. The difference between FSM and Not FSM pupils was compared between all pairwise combinations of time points and tested for significance at a 5% significance threshold.

For both the Covid gap and the disadvantage gap, the power for a given combination of number of schools and effect size was calculated as the proportion of the 1000 simulated data sets where all comparisons involving the time point where the effect was induced were declared as significant. The MDES for a particular number of schools was the smallest effect size where the power was greater than 80%.

These calculations highlight the importance of each participating school remaining within the project for the three waves for the robustness of the proposed analysis.

Following analysis after year 1 of the study, we will determine whether the remaining sample size will provide adequate statistical power and useful results in order for the study to proceed.

Outcome measures and other data

OUTCOME MEASURES

The outcome measures will be attainment data from NFER tests in reading and maths for individual pupils[2]. Test data will be collected in the spring term for Year 2 and Year 3 pupils in 2022, and for Year 3 and Year 4 in 2023. Tests will be administered by the schools following the usual NFER guidance in how to administer the assessments. In year 1 of the study (2021/22), the tests will be marked by

---

[2] Information on NFER assessments can be found in here for KS1 assessments (Year 2) https://www.nfer.ac.uk/for-schools/products-services/nfer-tests/key-stage-1-assessments/ and here for KS2 assessments (Year 3) https://www.nfer.ac.uk/for-schools/products-services/nfer-tests/key-stage-2-assessments/

teachers. As noted earlier, this was a change in procedure from the original Covid baseline study (2020/21), where NFER marked and coded responses. The assessments are designed to be marked by teachers. In year 1 of the current study, we are providing marking support in case teachers have any questions about how to apply the mark scheme: by running a webinar in advance of marking and providing a helpdesk for any queries during the marking period. In year 2, to mitigate attrition and where possible to encourage schools from the baseline year to re-join the study, NFER will mark the assessments. Schools appreciated this in the baseline year of the study, and it was felt by the research team that re-employing this strategy would help with retaining the longitudinal sample. The following table identifies the time required to complete these assessments.

Assessment	Duration of paper 1 (mins)	Duration of paper 2 (mins)	Duration of paper 3 (mins)	Total (mins)
Maths Year 2 spring	20 (arithmetic)	35 (reasoning)	N/A	55
Reading Year 2 spring	40	50	N/A	90
Maths Year 3 spring	25 (arithmetic)	30 (Test 1)	30 (Test 2)	85
Reading Year 3 spring	75	N/A	N/A	75

These tests have a strong alignment to the English national curriculum in reading and mathematics and have robust technical properties; outcomes include standardised scores and age standardised scores (i.e. scores based on a large, nationally representative samples). Schools use the tests to monitor termly and yearly progress of their pupils and to identify gaps in learning and misconceptions.

At the time of writing the study plan, all schools are open and maintaining normal teaching practices. We will review the sample of schools and the completion of tests if any local restrictions affect school closure or how tests might be administered.

NON-ATTAINMENT OUTCOMES

Survey of social skills development

Alongside attainment outcomes, this longitudinal study will capture pupils' social skills and wellbeing surrounding their school experiences and learning recovery. This is important as school staff have reported challenges with pupil wellbeing over the course of the 2020/21 academic year (Rose et al, 2021; Lucas et al, 2020; Nelson et al, 2021). This measurement will track the social skills and wellbeing recovery of the pupils as they move into Year 2 and Year 3.

To explore this non-attainment outcome, we will collect data on pupil social skills via a pupil-level survey completed by teachers in spring 2022 and spring 2023 using a validated instrument and additional NFER bespoke questions. This will be completed for a random sub-sample of 12 pupils per year group selected from pupil lists provided by schools, or, if schools do not provide updated pupil lists, the 12 pupils will be those that were selected for the CSBQ in the Covid baseline study. The instrument that will be used to measure social skills is the Peer Social Maturity Scale (PSMAT; Peterson et al, 2007; Fink et al, 2013). This is a 7-item scale of children's maturity when interacting with their peers, compared to an average child of the same age. The response options lie on a 7-point scale:

- 1: Very much less mature than the average child this age
- 2: Less mature than the average child this age
- 3: A little less mature than the average child this age
- 4: About average for children this age
- 5: A little more mature than the average child this age
- 6: More mature than the average child this age

- 7: Very much more mature than the average child this age

Results from a validation of the PSMAT are available for a sample of Australian children aged from 5 to 8 years 5 months, and the measure has subsequently been used for children up to age 13.

The PSMAT measures social skills with regard to a child's interaction with their peers. However, the research team felt it was important to capture other aspects of social skills and wellbeing, such as interactions with adults and emotional regulation. An additional seven bespoke questions were created by mapping the elements contained in the CSBQ (the measure used for the Key Stage 1 study) to PSMAT, identifying the domains not covered by PSMAT, and creating relevant items for these domains. These questions will be asked in addition to the PSMAT in an exploratory way, to yield additional data on the pupils. The additional domains were:

- Attention / focus
- Interactions with relevant adults
- Independence
- Persistence
- Emotional regulation.

ADDITIONAL DATA COLLECTIONS

Pupil background data

Schools are asked to provide basic pupil background data; name, DOB, UPN, gender, year group, class name, school name and FSM.

Data on eligibility for FSM will thus be ideally collected concurrent to each assessment point (e.g. current FSM status in 2021, 2022 and 2023) and will also be available for the 2020 pre-pandemic baseline. Consequently, particularly for students joining a school in 2022 or afterwards, FSM status at baseline will not be available.

Since we don't and will not have any available concurrent (Autumn 2020) data for students missing pre-pandemic FSM status (students joining the study in 2022 and 2023), we do not think imputing pre-pandemic FSM can be achieved successfully. We therefore will run the repeated measures analysis with two different samples: (a) pupils with test data at Spring 2021 and/or Spring 2022 and a valid response for FSM eligibility at March 2020, excluding students for whom we don't have pre-pandemic FSM as a primary analysis; (b) use current FSM as the pre-pandemic FSM and include all pupils with test data at Spring 2021 and/or Spring 2022 even without a valid response for FSM eligibility at March 2020 as a sensitivity analysis. We will report how both of these samples differ from the original 2021 sample. Excluding students for whom we don't have pre-pandemic FSM will allow us to weight in regards to the baseline characteristics of the population. Weighting will aim at having the same distribution of KS2 school achievement in both the pre-pandemic standardisation sample and the study's sample.

In longitudinal analyses, multilevel models can successfully deal with missing outcome data but not with missing explanatory variables. Therefore, including cases with missing pre-pandemic FSM is not advisable. Although it would be possible to include students for which we don't have pre-pandemic FSM if we were using FSM current at each data point, as a time varying covariate, this analysis would not be addressing the core question on 'how are FSM students changing in comparison to their non-FSM peers after the pandemic's disruption', but rather 'how are changes in FSM status across wave are associated to changes in achievement'. For cross-sectional analyses, we recommend using FSM (current) and including all students for whom we have data at each time point under consideration regardless of whether we have pre-pandemic FSM status for them. Using FSM current for cross-sectional analyses will allow us to assess what are the COVID and FSM gaps at each time point by comparing each sub-group with the standardised norm.

School background data

School background characteristics such as the proportion of children eligible for FSM, the proportion of pupils meeting the expected standard in reading, writing, and maths, the proportion of pupils with special educational needs (SEND), the proportion of pupils with English as an additional language (EAL), the academy status of the school, whether the school is in an urban or rural area, and the geographical region in which a school is located will be obtained from the DFE website where data is freely available to be downloaded.

Representativeness and weighting

The current longitudinal project draws schools from a sample of 168 schools which have taken part during at least one of the waves of the Covid baseline study. In the 155 schools which took part in all 3 waves of the baseline project, approximately 16 per cent of pupils were eligible for free school meals in 2020/21 academic year. In each round of analysis, we will carry out a full representativeness analysis on the final sample of participating schools. As well as FSM, this analysis will include looking at school characteristics that identify KS2 performance in reading and maths from 2019, school type, geographical location, proportion of pupils with SEN, proportion of pupils with English as an additional language and academy status.

We will carry out a full representativeness analysis on the final sample of participating schools. We will weight our analysis accordingly, particularly for the analysis that will estimate the Covid-19 gap. We will ensure the sample of participating schools is representative based on school level performance at Key Stage 2 in 2019. We will use the variable KS2rwmExp_19, the proportion of pupils meeting the expected standard in reading, writing and maths available from DFE website to determine representativeness of the sample to the population of primary schools[3]. To address the issue of analysis being undertaken at pupil level but information on the sample is at school level the analysis to determine representativeness will be weighted by the number of pupils in the school. Therefore, for Year 2 the population will be weighted by the number of pupils on roll in Year 2 in the spring census of 2019 and the schools in our sample will be weighted by the number of pupils who took the test within each school. Whilst not producing analysis ensuring the sample of pupils is representative of pupil population characteristics, it will ensure the sample is not introducing bias because we have too many pupils from schools with particular characteristics, for example, too many pupils from high performing schools. This procedure will be replicated for all other assessments (i.e. Year 3 and Year 4 spring assessments) and any limitations of this method will be discussed in the final report.

SURVEY DATA – CONTEXTUAL FACTORS

In addition to attainment outcomes, we will collect data around support strategies used and challenges faced at school level. The school-level survey will take place in March 2022 and March 2023, and will ask schools about recovery approaches, support and challenges. The IPE section provides further details about the survey. Some contextual school level factors, such as geographical location and the proportion of pupils on free school meals, will be explored in the analysis.

Main analysis

Numbers of pupils and schools included in each stage of the analysis will be reported and we will carry out a full representativeness analysis on the final sample of participating schools.

All cross-sectional analyses will report both standardised scores and raw scores. Standardised scores are reported because their original means of 100 and standard deviations of 15 points make them more interpretable and comparable across year and subject and because they are more familiar for educators. More importantly, standardised scores allow for the reporting of the number of pupils unable to access the assessment (those receiving a score of 69). The proportion of pupils unable to access the assessment is an important indicator of differences between samples. Nevertheless, since

---

[3] The KS2 variable has been put into quintiles of school performance with a further category that identifies schools with missing data.

standardised scores restrict the score range from a minimum of 69 to a maximum of 141 points for the lowest and highest achievers, there is a risk that this restriction can distort group mean comparisons, particularly when the proportion of students below or above the thresholds of 69 and 141 differ between the groups being compared.

To address the potential effect of censoring, all statistical significance tests for the Covid-19 and disadvantage gaps are generated using raw test scores. Raw test scores are simple summations of the number of questions responded correctly. Consequently, when assessing, for example, whether the 2022 Year 2 maths pupil sample differed significantly from the 2019 benchmarking sample used to standardise the test (the Covid-19 gap), the statistical significance is based on the comparison of the mean raw scores for these two samples. Moreover, the significance of the t-tests for the raw scores incorporates the effect of school clustering. It should be noted that the method used in the standardisation means that comparisons are with estimated raw scores for each pupil in the standardisation rather than their actual score for the assessment. Mean group comparisons that do not incorporate the clustering effect that result from sampling schools versus sampling pupils directly overestimate the p-values of comparisons when intra-cluster correlations are high. The significance and confidence intervals of raw scores is obtained using complex survey analysis methodology, which uses inverse-probability weighting and design-based standard errors (Lumley, 2004).

Whereas cross-sectional analyses report both standardised and raw scores, longitudinal analyses report only standardised scores. Since the psychometric properties of the assessments are different, raw scores cannot be used. Consequently, there is no way of avoiding the potential effects of standardised score censoring when comparing the performance of pupils across time. Nevertheless, all repeated measures analyses will be produced using multilevel modelling regressions: this takes into account the effect of school clustering and thus the significance of regression coefficients is robust against the effect of sampling schools instead of sampling students directly.

Measuring the Covid-19 and disadvantage gaps

The Covid-19 gap (RQ1a)

We will estimate the Covid-19 gap (RQ1a) counterfactual using the pre-pandemic standardisation sample for each relevant test. Each test was standardised on a representative sample[4] of pupils following the introduction of the new National Curriculum and at the same time of the academic year as the present tests are scheduled. This analysis will address RQ1a and, by taking the weighted mean standardised score and mean raw score for our sample along with its standard error (see section on Representativeness and weighting above), we will be able to determine if the sample mean is different from the population mean. This will be undertaken on the spring tests for Year 2 and Year 3 pupils in 2021/22 and Year 3 and Year 4 pupils in 2022/23. To note, for the Covid-19 gap, the reference group is the pre-pandemic standardisation sample of 2017/2019, which is then compared to the samples observed in 2022 and 2023.

The disadvantage gap (RQ2a)

We will be able to calculate the disadvantage gap by comparing the mean standardised scale and raw score for disadvantaged pupils with advantaged pupils, resulting in the unit effect size for the gap which can be converted into months of progress using EEF's conversion table[5]. For the disadvantage gap, the reference group is the group of non-FSM pupils within the observed sample in any given year (i.e. the disadvantage gap in 2022 compares the attainment of FSM and non-FSM pupils in 2022, and so on).

---

[4] Links to standardisations:
https://www.nfer.ac.uk/for-schools/products-services/nfer-tests/nfer-tests-development/
[5] https://educationendowmentfoundation.org.uk/education-evidence/using-the-toolkits

The Covid-19 and disadvantage gap over time (RQ1b and RQ2b)

Analysis on identifying the Covid-19 gap will be reliant on a representative distribution of nationally standardised scores for Years 1 and 2 in the relevant terms. This has a number of advantages:

- by comparing the mean score to 100 (standardised mean for national standardisation) we can estimate Covid-19 gap in standard deviation units

- by comparing the mean score between disadvantaged and non-disadvantaged pupils, we can estimate disadvantage gap in standard deviation units, and through using DfE method of calculation

- by comparing the raw mean score to the raw mean score for the national standardisation sample we can estimate Covid-19 gap in regards to the full distribution of scores (minimising the likelihood of floor and ceiling effects that can affect standardised scores)

- by using repeated-measures multilevel models, we can track the closing of both the Covid-19 and disadvantage gaps over the course of this study, and how gaps change for different subgroups

- the Covid-19 and disadvantage gaps will be calculated at each time point for each cohort in reading and in maths.

As we are proposing annual data collection sweeps, both the Covid-19 and disadvantage gaps will be tracked (RQ1 & RQ2). For both reading and maths, gaps will be measured across spring 2021, 2022 and 2023 using standardised scores. In order to monitor change over these time periods we propose using a multi-level structure to the models and a repeated-measures design. The models will have three levels; test occurrence[6], pupil and schools and will be run separately for each year group and subject, resulting in four individual models. These will identify how pupils' performance identified at the first time point is changing over the subsequent academic years (RQ1b). The dependent variable will be the reading or maths outcome score. The independent variables entered into the model will be time to identify if there is a significant difference in the change in outcome score between the spring 2021 baseline and spring 2022 at midpoint, and then spring 2023 at endpoint (indicated with values 0, 1 and 2 respectively) and FSM status 2020. An interaction between time and FSM status will indicate if the disadvantaged students are changing at a different rate than their non-disadvantaged peers (RQ2b).

Analysis of test domains (RQ3)

For year 1 of this study, as schools will be marking assessments and uploading them to the NFER progress tool, this will provide access to individual item level data. In year 2 of the study, the marking of the assessments will similarly provide item level data which will allow items to be grouped within particular domains. Each domain contains a number of individual items that can provide greater information on a particular area of learning than trying to assess performance across many individual items.

Analysis will look to identify differences between domains as well as within domains. Analysis will look to see if pupil factors (i.e. background characteristics) are associated with variation in domain scores. For example, do girls perform better than boys across particular domains? Do pupils eligible for free school meals perform differently across the domains?

As assessments are occurring at three time points, analysis on domains can look at how domain scores change over time, compared with the pre-pandemic standardisation sample in 2019. The main analysis is looking at changes in overall test scores by using standardised and raw scores. Analysis will look at how domain scores have changed between the baseline Spring 2021 and each time point (Spring 2022 at midpoint and Spring 2023 at endpoint) to determine if particular domains have seen a bigger change than other domains.

---

[6] Autumn, spring or summer.

Analysis of contextual data (RQ4)

Analysis of contextual data from the head teacher survey will be descriptive. This will give an indication of what schools have been focussing on in terms of catch-up in each academic year, and will provide an opportunity for schools to report anything important about the academic year that they have not already had the opportunity to share. Frequencies will be compared by school factors such as geographical location, school size and FSM quintile (dependant on response rate), to determine whether these factors are influencing school catch-up strategies, experiences or challenges.

Analysis of social skills development (RQ5)

Social skills development will be assessed using the Peer Social Maturity Scale (PSMAT) and additional bespoke items. As previously discussed, this data collection will be from a sub sample of approximately 12 pupils from Year 2 and 12 pupils from Year 3 in each school. The PSMAT will be analysed by comparing the present sample to the results from the original measure validation. The bespoke items will be assessed for reliability using Cronbach's alpha, using a cut-off of at least 0.7 (acceptable; Bland & Altman (1997)). If the items are suitably reliable, the mean of the sumscores for these items will be reported. Descriptives for individual items will also be included in the report. If reliable, the mean of the sumscore will form a baseline which can be referred back to when the pupils progress to Years 3 and 4, to track social skills and wellbeing recovery/development.

Implementation and process evaluation (IPE) [7] (RQ4)

IPE research questions

The IPE will provide understanding around what has happened at a school level during each academic year as well as any support strategies in place to aid effective learning. This data will be reported as described below and it will be used to provide context to the assessment results. The research question for the IPE is RQ4:

> RQ 4   In the 2021/22 academic year, and again in the 2022/23 academic years, what practices have been adopted and what learning opportunities have been provided by schools to help pupils catch up; and what challenges have been faced by staff?

IPE research methods

In the exploration of school-level practices occurring during the two academic years, we believe headteacher-rated measures will give the most accurate contextual data for this age group. The survey will be sent to the headteacher in all the schools taking part in the study, in March 2022 and again in March 2023. We will collect data on the strategies implemented by schools to aid learning/recovery, the challenges faced and anything else about the academic year that schools wish to report via a school-level survey in March 2022 and subsequently in March 2023. The survey for 2023 will be informed by responses to the survey in 2022 and any relevant context that may have changed over the course of the academic year.

The survey will collect information about topics such as:

- remote learning, including how schools are supporting vulnerable children not in school or those missing large periods of school-based learning

- new practices post partial school closures (divided into):

  o   enforced practices and their impact

  o   practices schools have chosen to retain because they have found they are a better way of working

---

- challenges for staff, for instance coping with staff absences and any additional CPD requirements as a result of the pandemic

- social and emotional support for pupils

- how schools are approaching tutoring

- catch-up strategies/recovery actions for reading and maths

- parental engagement and whether it has been sustained (both in terms of capability and willingness).

One open question will allows schools to tell us about anything additional happening in their school which they think is relevant. Whilst this would not necessarily form part of the analysis, it would be useful to inform the school survey for the following year.

We will use our online survey software Questback (QB) for developing and hosting the school-level survey.

IPE analysis

As outlined in the main analysis section earlier, survey responses will be analysed using descriptive statistics and tables included in the report. These responses may also provide some contextual understanding of the attainment results. Themes arising from the coded responses to the 2022 school-level survey will feed into the development of the 2023 survey.

Data protection

The legal basis for processing personal data is covered by GDPR Article 6 (1) (f):

Legitimate interests: the processing is necessary for your (or a third party's) legitimate interests unless there is a good reason to protect the individual's personal data which overrides those legitimate interests.

We have carried out a legitimate interest assessment, which demonstrates that the research fulfils one of our core business purposes (undertaking research, evaluation and information activities). The research project has broader societal benefits and will contribute to improving the lives of learners by identifying if any pupil level factors are associated with the degree of impact of the Covid-19 school closures on pupils' attainment and their recovery over the academic year. Personal data is required for the research and its processing will not cause damage or distress to the data subjects.

NFER will provide a memorandum of understanding to schools, explaining the nature of the data being requested of schools and children, how it will be collected, and how it will be passed to and shared with NFER. For the purpose of research, UPN and test outcome data for all pupils taking NFER assessments will be linked with information about pupils, including EAL status, free school meal eligibility and, for Year 2 pupils, their Key Stage 1 outcomes in reading and maths assessments (obtained from their school). This data will be shared with NFER, EEF and EEF's data archive contractor FFT Education, and potentially, in an anonymised form to the UK Data Archive. Pupil data will be treated with the strictest confidence. Neither we, nor any of the named parties, will use pupil names or the name of any school in any report arising from the research.

On conclusion of our project, the Fischer Family Trust (see http://www.fft.org.uk/) will collate and de-identify the data for upload to the EEF data archive. The archived data will be available in a de-identified form with restricted access for research purposes only. NFER handles personal data in accordance with the rights given to individuals under data protection legislation. Individual rights are respected.

For further information, please see the privacy notices:
Parents: https://www.nfer.ac.uk/media/4652/llon_parent_privacy_notice.pdf
Schools: https://www.nfer.ac.uk/media/4653/llon_schools_privacy_notice.pdf

Personnel

Name	Institute	Roles and responsibilities
Rebecca Wheater (RW)	NFER	Project Director 2021/22 and planning for 2022/23, responsible for directing the NFER team and quality of delivery.
Pippa Lord (PL)	NFER	Project Director (2022/23), responsible for directing the NFER team and the quality of delivery.
Susan Rose (SR)	NFER	Project manager, responsible for overseeing the day to day running of the project.
Ben Styles (BS)	NFER	Statistical oversight and QA
Liz Twist (LT)	NFER	Assessment oversight and QA
Lydia Fletcher (LF)	NFER	Process evaluation lead (2021), responsible for managing the process evaluation activities and analysis and social skills measurement

Rob Ager (RA)	NFER	Process evaluation lead (2022 onwards), responsible for managing the process evaluation activities and analysis and social skills measurement
Alison Hale (AH)	NFER	Test and Schools administration lead (2021/22), responsible for overseeing recruitment, school contact and testing
Sarah Millar (SM)	NFER	Test and Schools administration lead (2022/23), responsible for overseeing recruitment, school contact and testing
Jose Liht (JL)	NFER	Senior Statistician, responsible for statistical analysis

Timeline

Date	Activity
Jul-21	Warm-up communication to LENT schools about longitudinal study
Sep-21 to Jan-21	Set up Data sharing agreement Draft study plan Preparation of school communications School communications / engagement including survey about testing plans School-level questionnaire design Selection of social skills survey and instrument design for any gaps
Feb-22	Draft study plan Send spring tests to schools Y2 and Y3 spring tests in schools School questionnaires in schools Teacher-completed PSMAT and NFER social skills survey with subsample in schools
Mar-22	Finalise study plan Y2 and Y3 spring tests in schools School questionnaires in schools Social skills survey with subsample in schools
Apr-22	Spring data collection from analysis tool NFER mark some schools' tests Coding of survey open response items
May-22	Feedback to schools on tests Analysis
Jun-22	Analysis Reporting
Jul-22	Reporting

Aug-22	**Publish Y2/3 report (NB published autumn 2022)** **Decision point for continuation to 22/23**
Sep-22 to Jan-23	Preparation of school communications School communications / engagement including survey about testing plans School-level questionnaire design
Feb-23	Send spring tests to schools Y3 and Y4 spring tests in schools School questionnaires in schools Teacher-completed PSMAT and NFER social skills survey with subsample in schools
Mar-23	Y3 and Y4 spring tests in schools School questionnaires in schools Social skills survey with subsample in schools
Apr-23	NFER mark schools' tests Coding of survey open response items
May-23	Feedback to schools on tests Analysis
Jun-23	Analysis
Jul-23	Reporting
Aug-23	Reporting
Sep-23	**Publish Y3/4 report**

References

Arnold, B.F., Hogan, D.R., Colford, J.M. and Hubbard A.E. Simulation methods to estimate design power: an overview for applied research. *BMC Med Res Methodol* **11,** 94 (2011). https://doi.org/10.1186/1471-2288-11-94.

Bland, J. M., and Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ*, *314*, 572.

Children's Commissioner Briefing, 2020. *Putting children first in future lockdowns.* [pdf]. Available at <https://www.childrenscommissioner.gov.uk/wp-content/uploads/2020/08/cco-putting-children-first-in-future-lockdowns.pdf> [Accessed 16 November 2020].

Fink, E., de Rosnay, M., Peterson, C. and Slaughter, V. (2013). Validation of the Peer Social Maturity Scale for assessing children's social skills. *Infant and Child Development*, 22 (5), 539-552.

Lucas, M., Nelson, J. and Sims, D. (2020). *Schools' responses to Covid-19: Pupil engagement in remote learning* [online]. Available: https://www.nfer.ac.uk/media/4073/schools_responses_to_covid_19_pupil_engagement_in_remote_learning.pdf

Nelson, J., Andrade, J. and Donkin, A. (2021). *The impact of Covid-19 on schools in England: experiences of the third period of partial school closures and plans for learning recovery: Graphs and commentary on questions posed to the NFER Teacher Voice Omnibus Survey panel, March 2021* [online]. Available: https://www.nfer.ac.uk/media/4435/the_impact_of_covid_19_on_schools_in_england.pdf

Peterson, C. C., Slaughter, V. P., and Paynter, J. (2007). Social maturity and theory of mind in typically developing children and those on the autism spectrum. *Journal of Child Psychology and Psychiatry*, *48*(12), 1243-1250. Available: https://acamh.onlinelibrary.wiley.com/doi/epdf/10.1111/j.1469-7610.2007.01810.x

Rose, S., Badr, K., Fletcher, L., Paxman, T., Lord, P., Rutt, S., Styles, B. and Twist, L. (2021). *Impact of School Closures and subsequent support strategies on attainment and socio-emotional wellbeing in Key Stage 1: Research report* [online]. Available: https://www.nfer.ac.uk/impact-of-school-closures-and-subsequent-support-strategies-on-attainment-and-socio-emotional-wellbeing/

Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. and Taggart, B., 2004. *The effective provision of pre-school education (EPPE) project: final report: a longitudinal study funded by the DfES 1997-2004.* [pdf] Available at <https://discovery.ucl.ac.uk/id/eprint/10005309/1/sylva2004EPPEfinal.pdf> [Accessed 16 November 2020].