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

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

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 and Pippa Lord
STUDY PLAN AUTHOR(S)	Susan Rose, Jose Liht and Pippa Lord
STUDY DESIGN	Observational longitudinal study
PUPIL AGE RANGE AND KEY STAGE	Age 8 to 11 years, Key Stage 2
NUMBER OF SCHOOLS	Approximately 82
NUMBER OF PUPILS	Approximately 6,200
PRIMARY OUTCOME MEASURE AND SOURCE	NFER Tests in reading and maths. Year 4 and 5 in spring 2024, Year 5 and 6 in spring 2025, Year 6 in spring 2026.
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
1.0	30.10.2023	Updated for grant agreement
1.1	08.12.2023	For peer review
1.2	22.12.2023	Updated for peer review
1.3	19.01.2024	Amended following peer review

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 open again and teachers providing ongoing catch up support pupils continued to make progress to recovering learning loss in the academic years 2021/22 and 2022/23. One of the aims of this study is to understand the extent to which pupils who had learning disrupted in KS1 are able to recover to levels that they should be before the pandemic, and factors that predict this recovery.

Both the initial one-year project on the impact of Covid-19 school closures on pupil outcomes in Key Stage 1 and the subsequent two-year project following the same pupils into Key Stage 2 have 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 continue to move through the primary phase of 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. The study will be a combination of quantitative research looking at pupil attainment derived from NFER assessment 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 2024, spring 2025 and spring 2026 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 2024 (for Year 4 and Year 5 pupils) and spring 2025 (for Year 5 and Year 6 pupils). Compared to pre-pandemic standardisation data (i.e. 2017 for Year 4 and 5 spring tests & 2019 for Year 6 spring tests). 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), spring 2023 (for pupils in Year 3 and Year 4), spring 2024 (for pupils in Year 4 and Year 5), spring 2025 (for pupils in Year 5 and Year 6) and spring 2026 (for pupils in Year 6).
2. RQ2a. What is the attainment gap between disadvantaged pupils and their peers in reading and in maths in spring 2024, in spring 2025 and in spring 2026? RQ2b. To what extent do different groups recover by spring 2024, by spring 2025 and by spring 2026; 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 2024, spring 2025 and spring 2026, 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, spring 2023, spring 2024, spring 2025 and spring 2026.
3. In the 2023/24, in 2024/25 and again in the 2025/26 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
4. 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 2024, spring 2025 and spring 2026 on a (random) sample of 12 pupils from each year group in each school. The spring PSMAT scores for each year of the study can be compared with the previous years and with the spring 2022 and 2023 scores, as well as with the pre-pandemic norms[1]. If cell count allows, PSMAT scores will be analysed by pupil characteristics; gender and free school meals eligibility.

[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.

Design overview

This longitudinal study follows the progress of pupils in Year 4 and Year 5 in the 2023/24 academic year, in Year 5 and Year 6 in 2024/25 and in Year 6 in the 2025/26 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 tests will be used to provide attainment data in spring 2024, spring 2025 and spring 2026. 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 2024 (and again in spring 2025 and in spring 2026) 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 2024 (and again in the spring of 2025 and 2026) tests, 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		82 schools and approximately 6,200 pupils as of October 2023
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[2]

[2] Using the Peer Social Maturity Scale (PSMAT) (Fink et al., 2013) and NFER bespoke questions

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 had 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 the subsequent longitudinal project. In October 2021, schools were invited to participate in a two year longitudinal study.

In year 1 of the first follow up study (2021/22), teachers were asked to mark and upload test data to the NFER progress tool. We anticipated the following factors would incentivise participation:

- provision of free spring tests to schools as a pre-incentive
- a discount for future NFER tests if schools successfully uploaded item-level data (available for schools who uploaded data in year 1, and for schools who provided 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 tests in order to do that. In 2021/22, we identified removal of NFER marking as the biggest cost saving we could 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 saving teachers' time during the uncertainties of partial school closures and re-openings, and to ensure that there were no barriers to participation for schools and to provide schools with diagnostic level data. Schools in 2022/23 felt that the offer of having their tests marked was more beneficial and therefore this was reinstated.

To ensure good participation in this current study (2023/24, 2024/25 and 2025/26), NFER will mark the tests. As in previous years schools will continue to be provided with: free spring tests, a discount on a future order and a summary feedback document.

Sample size calculations

Power calculation assumptions

Power calculations were carried out separately for the longitudinal analyses (using simulations) and the comparison between project and pre-Covid standardisation samples (using calculations). The simulated power calculations took into account the following overall assumptions: a power of 0.80, alpha of 0.05, school ICC = 0.09 and FSM proportion of 0.16, for a longitudinal design of 4 waves (baseline, plus three more waves). Additionally, we assumed that there are, on average, 38 pupils per school based on the number of pupils and schools in earlier waves. These results should hold conservatively for further waves. For the calculation method assumptions were the same as above but an ICC = 0.12, with a different number of pupils and schools (discussed below). The simulations and calculations are outlined further below.

Simulations for changes over time, for a range of school sample sizes

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 changes in the Covid gap and changes in the disadvantage gap over time, for all combinations of 45, 60 and 80 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 four time points (baseline, plus year 1, year 2 and year 3 of the current project) and 38 pupils per school. The school level ICC was taken to be 0.09 and the pupil level ICC

was take to be 0.65. These came from a model run in the previous study in 2022/2023. Setting the sum of the school and pupil level and residual variances to be 1 meant that the school and pupil level and residual variances used to simulate the variability in the data were ICC_{school}, ICC_{pupil} - ICC_{school} *and 1* - ICC_{pupil} respectively. This ensured that the coefficients of models were on the effect size scale.

For changes in the Covid gap over time, 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 changes in the disadvantage gap over time, 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 non FSM pupils was compared between all pairwise combinations of time points and tested for significance at a 5% significance threshold.

For both changes in the Covid gap and changes in 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% (i.e., a statistically significant effect was inferred in more than 800 of the simulated datasets).

MDES calculation for the Covid-gap (project sample and pre-Covid sample)

The calculations of MDES for the Covid gap (effect size between project sample and pre-Covid standardisation sample) have not used the simulation method, as the number of pupils was different for the two samples (the simulations assumed balance between the two groups). Instead, they were calculated using NFER's standard power calculation spreadsheets for an unbalanced design. This calculation used an ICC of 0.12, and an average class size of 38 in the project sample and an average class size of 26 in 76 schools for our standardisation sample.

MDES summary

Table 2 shows the calculated and simulated minimum detectable effect sizes for achieved samples.

Table 2: Calculated and simulated MDES

Number of schools analysed (i.e. achieved sample)	MDES for Covid gap (effect size between project sample and pre-Covid standardisation sample) (calculated)	MDES for change in Covid gap (effect size between project sample at two time points, e.g. spring 2021 and spring 2022) (simulated)	MDES for change in disadvantage gap (simulated)
45	0.22	0.08	0.20
60	0.21	0.07	0.17
80	0.19	0.06	0.15

The results from the calculations and simulations for the Covid-gap (cross-sectional and over time) indicate that even with 60 schools, the project would be viable, as they suggest we would 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 Wave 1 (Rose *et al*., 2021).

We note that the disadvantage gap between FSM and non-FSM pupils in any given year of this study (i.e. at single time points) has been found to have an effect size of around 0.4 (e.g. Rose *et al*., 2021), which is of a magnitude well above a minimum detectable for this gap. However, the results of the simulated MDES for changes in the disadvantage gap over time undertaken for Table 2 above, indicate that it is unlikely that changes in the disadvantage gap would be detected with these sample sizes, not because they are comparably smaller, but due to the imbalance between FSM and non-FSM pupils and thus the loss in effective sample size.

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

Outcome measures and other data

OUTCOME MEASURES

The outcome measures will be attainment data from NFER tests in reading and maths for individual pupils[3]. Assessment data will be collected in the spring term for Year 4 and Year 5 pupils in 2024, Year 5 and Year 6 pupils in 2025 and for Year 6 pupils in 2026. Tests will be administered by the schools following the usual NFER guidance in how to administer the tests. In all three years of this study, the tests will be marked by NFER to mitigate attrition and where possible to encourage schools from the baseline year to re-join the study. 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. Table 3 identifies the time required to complete these tests.

Table 3: Assessments

Assessment	Duration of paper 1 (mins)	Duration of paper 2 (mins)	Duration of paper 3 (mins)	Total (mins)
Maths Year 4 spring	25 (arithmetic)	35 (Test 1)	35 (Test 2)	95
Reading Year 4 spring	45	30	N/A	75
Maths Year 5 spring	30 (arithmetic)	40 (Test 1)	40 (Test 2)	110
Reading Year 5 spring	60	N/A	N/A	60
Maths Year 6 spring	30 (arithmetic)	40 (Test 1)	40 (Test 2)	110
Reading Year 5 spring	60	N/A	N/A	60

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.

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

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 4 and Year 5 in 2023/24, Year 5 and Year 6 in 2024/25 and Year 6 in 2025/26.

To explore this non-attainment outcome, we will collect data on pupil social skills via a pupil-level survey completed by teachers in spring 2024, spring 2025 and spring 2026 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 (Fink et al., 2013).

The PSMAT measures social skills with regard to a child's interaction with their peers. However, the research team developed additional seven bespoke questions 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 created relevant items for these domains. These questions were asked in addition to the PSMAT in Wave 2 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.

The bespoke items were assessed for reliability using Cronbach's alpha, using the absolute cut-off score of at least 0.7 (deemed acceptable; Bland and Altman, 1997). The mean of the sumscores for these items were also reported in 2021/22. The means of the sumscore were sufficiently reliable that they formed a baseline, which was subsequently compared to pupils in Year 3 and Year 4 in Spring Term 2023, to track progress in their social skills and wellbeing recovery/development. The PSMAT and bespoke items will be used for this current study with the same sub-sample of 12 pupils from each year in each school with this results being compare to those from the 2021/22 and 2022/23 study.

Additional Data Collection

Pupil background data

Schools are asked to provide basic pupil background data; name, DOB, UPN, gender, year group, class name, school name and FSM eligibility[4].

Data on eligibility for FSM will thus be ideally collected concurrent to each assessment point (e.g. current FSM status in each year from 2021 to 2026) Since we don't and will not have any available concurrent (Autumn 2020) data for students who did not participate in Wave 1 (students joining the study from 2022 onwards), 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 2024, spring 2025 and/or Spring 2026 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 each spring point 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.

For cross-sectional analyses, we are 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 allows us to assess what are the COVID and FSM gaps at each time point by comparing each sub-group with the standardised norm. Given the use of alternative FSM indicators to answer two different research questions we will ensure reporting does not produce conflicting interpretations. If there is a conflict we will ensure there are potential explanations for this.

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 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 2023, school type, geographical location, proportion of pupils with SEN, proportion of pupils with English as an additional language and academy status. The distribution of these characteristics within participating schools will be compared against all primary schools[5]. If there are statistically significant differences between the two groups based on school-level performance at Key Stage 2 in 2023 (using the variable KS2rwmExp_23 the proportion of pupils meeting the expected standard in reading, writing and maths available from DFE website), we will weight our analysis accordingly using scaling weights, particularly

[4] It was decided to use FSM eligibility, rather than FSM Ever, as it was felt this would provide more reliable data from school collections.

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

for the analysis that will estimate the Covid-19 gap. 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 tests (i.e. Year 4 and Year 5 spring tests) and any limitations of this method will be discussed in the final report.

Consideration of National Pupil database data

The aim of this study is to explore the Covid-gap and the disadvantage gap in primary reading and mathematics including how these have changed over time by tracking two cohorts of pupils throughout primary school, using NFER's standardised tests. For the Covid gap, we compare test results to the relevant pre-Covid standardisation samples. For each subject in each year, we also establish the disadvantage gap and explore how these scores and gaps have changed over time through longitudinal analysis.

We considered whether it would be beneficial to also include pupils' Key Stage 2 test data, once they are in Year 6. Analysing KS2 data at the national level to determine Covid and disadvantage gaps is very valuable, however, incorporating this into the longitudinal nature of this study would be problematic. NFER tests are different measures resulting in standardised rather than scaled scores, and they are administered under low rather than high stakes conditions. We note that incorporating scaled scores into the longitudinal nature of the study proved to be a difficulty in the summer analysis in the baseline year of the project where we used KS1 assessment data. KS2 scores are also not available until the autumn of each year when our yearly reports are published. Therefore, considering the Research Questions of our study, we will not include KS2 assessment data in this longitudinal study.

However, NPD data may add some value in allowing formal representativeness checks at the pupil level. To date, we have only been able to check the representativeness of our sample of NFER test users at the school level. Matching it to NPD data, including KS2 results, would allow a formal representativeness check at the pupil level, potentially including a breakdown for FSM pupils. This will be achieved by obtaining KS2 results for all pupils in state-maintained schools and running a multi-level model, with pupils clustered within schools, to determine if there is a statistical difference in mean scaled scores when comparing study pupils and non-study pupils.

Due to the timing of the data release, this will be run for the 2024/25 Year 6 cohort, with data available in autumn 2025, with the results being incorporated into the final report. We would not anticipate running the check on the subsequent Year 6 cohort (2025/26).

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 2024, March 2025 and March 2026, 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 (as per the Representative and Weighting section above) 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

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 2024 Year 4 maths pupil sample differed significantly from the 2017 benchmarking sample used to standardise the test (the Covid-19 gap), the statistical significance test we will report will be based on the comparison of the mean raw scores for these two samples. Moreover, the significance of the t-tests for the raw scores will incorporate 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 tests 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[6] 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 estimating 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 test whether the sample mean is different from the population mean. This will be undertaken on the spring tests for Year 4 and Year 5 pupils in 2023/24, Year 5 and Year 6 pupils in 2024/25 and Year 6 in 2025/26. 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 2024, 2025 and 2026.

The disadvantage gap (RQ2a)

We will be able to estimate 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[7]. 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 2024 compares the attainment of FSM and non-FSM pupils in 2024, and so on for each subsequent year).

[6] Links to standardisations:
https://www.nfer.ac.uk/for-schools/products-services/nfer-tests/nfer-tests-development/

[7] 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 of each year from 2020 to 2026 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[8], 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 each year from spring 2022 to spring 2026 at endpoint (indicated with values 0, 1, 2, 3, 4 and 5 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 contextual data (RQ3)

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 (RQ4)

Social skills development will be assessed using the Peer Social Maturity Scale (PSMAT). As previously discussed, this data collection will be from a sub-sample of approximately 12 pupils for each additional cohort for subsequent waves in each school. The PSMAT will be analysed by comparing the present sample to the results from the original measure validation. 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, to track social skills and wellbeing recovery/development in subsequent waves.

[8] Autumn, spring or summer.

Implementation and process evaluation (IPE) [9] (RQ3)

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 2023/24 academic year, the 2024/25 academic year and again in the 2025/26 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 2024 and again in both March 2025 and March 2026. 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 2024, March 2025 and subsequently in March 2026. The survey for each year will be informed by responses to the survey in the previous year 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):
    - enforced practices and their impact
    - 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 allow 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 will 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

[9] Principles are detailed in the Implementation and Process Evaluation Guidance (2019).

understanding of the attainment results. Themes arising from the coded responses to the 2023 school-level survey will feed into the development of the 2024 survey and this will be repeated each year with the previous year's responses informing the development of subsequent years.

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 tests will be linked with information about pupils, including EAL status, free school meal eligibility. This data will be shared with NFER, EEF and EEF's data archive contractor FFT Education, and potentially, in an pseudonymised 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/1ldpu0bi/llun_parent_privacy_notice.pdf
Schools: https://www.nfer.ac.uk/media/a0hatv45/llun_school_privacy_notice.pdf

Personnel

Name	Institute	Roles and responsibilities
Pippa Lord (PL)	NFER	Project Director 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
Rob Ager (RA)	NFER	Process evaluation lead responsible for managing the process evaluation activities and analysis and social skills measurement
Jo Stringer (JS)	NFER	Test and Schools administration lead responsible for overseeing recruitment, school contact and testing
Jose Liht (JL)	NFER	Senior Statistician, responsible for statistical analysis

Timeline

Date	Activity
Oct-23	Warm-up communication to sample schools about longitudinal study
Nov-23 to Dec-23	Set up Data sharing agreement Draft study plan Preparation of school communications School communications / engagement School-level questionnaire design Preparation of social skills survey instrument
Dec-23	Finalise study plan
Feb-24	Send Y4 and Y5 spring tests to schools School questionnaires in schools Teacher-completed PSMAT and NFER social skills survey with subsample in schools
Mar-24	Y4 and Y5 spring tests in schools School questionnaires in schools Social skills survey with subsample in schools
Apr-24 to May-24	NFER mark tests and return results to schools Coding of survey open response items
May-24 to Jun-24	Data cleaning Analysis
Jun-24 to Sep-24	Analysis of assessment and survey data Draft report Draft school feedback paper
Oct 24	Publish 2023/24 report of Year 4 and Year 5 data Publish school feedback paper **Decision point for continuation to 2024/25**

Oct-24	Warm-up communication to sample schools about longitudinal study
Nov-24 to Dec-24	Set up Data sharing agreement Preparation of school communications School communications / engagement School-level questionnaire design Preparation of social skills survey instrument
Feb-25	Send Y5 and Y6 spring tests to schools School questionnaires in schools Teacher-completed PSMAT and NFER social skills survey with subsample in schools
Mar-25 to Apr 25	Y5 and Y6 spring tests in schools School questionnaires in schools Social skills survey with subsample in schools
Apr-25 to May-25	NFER mark tests and return results to schools Coding of survey open response items
May-25 to Jun-25	Data cleaning Analysis
Jun-25 to Sep-25	Analysis of assessment and survey data Draft report Draft school feedback paper
Oct 25	Publish 2024/25 report of Year 5 and Year 6 data Publish school feedback paper **Decision point for continuation to 2025/26**
Oct-25	Warm-up communication to sample schools about longitudinal study
Nov-25 to Dec-25	Set up Data sharing agreement Preparation of school communications School communications / engagement School-level questionnaire design Preparation of social skills survey instrument

Feb-26	Send Y6 spring tests to schools School questionnaires in schools Teacher-completed PSMAT and NFER social skills survey with subsample in schools
Mar-26 to Apr 26	Y6 spring tests in schools School questionnaires in schools Social skills survey with subsample in schools
Apr-26 to May-26	NFER mark tests and return results to schools Coding of survey open response items
May-26 to Jun-26	Data cleaning Analysis Carry out representativeness check of the 2024/25 sample using NPD data
Jun-26 to Sep-26	Analysis of assessment and survey data Draft report Draft school feedback paper
Oct 26	Publish 2025/26 report of Year 6 data including analysis of representativeness check of the 2024/25 sample using NPD data Publish school feedback paper
Jan 27	Archive data

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

Lumley, T., 2004. Analysis of complex survey samples. *Journal of statistical software*, *9*, pp.1-19. Available: https://www.jstatsoft.org/article/view/v009i08

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].