Education Endowment Foundation

Improving Power Calculations in Educational Trials

2023

Authors

***Durham University**: Akansha Singh, Germaine Uwimpuhwe, Dimitrios Vallis, Nasima Akhter, Tahani Coolen-Maturi, Steve Higgins, Jochen Einbeck,*

***Sheffield Hallam University**: Martin Culliney, Sean Demack*

Durham University

Sheffield Hallam University

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

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

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

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

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

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

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Acknowledgments

This work on power calculation in educational trials was carried out as a collaborative project between Durham University and Sheffield Hallam University. Within Durham University, researchers from the School of Education, Department of Mathematical Sciences, Department of Anthropology, School of Government and International Affairs as well as the Durham Research Methods Centre (DRMC) with track records in evaluation of educational interventions, meta-analysis of evidence in education and advanced quantitative methods were part of this study. Researchers from Sheffield Hallam University have extensive experience of conducting educational trials funded by the EEF. We have partnered with the EEF on improving educational attainment of pupils from disadvantaged backgrounds for more than five years by providing methodological support and translation of evidence to educational stakeholders.

For more information about this report please contact:

Prof Jochen Einbeck (Co-Director, Durham Research Methods Centre)
Durham University
Department of Mathematical Science
Lower Mountjoy
Stockton Road
DURHAM
DH1 3LE

0191 334 3125

jochen.einbeck@durham.ac.uk

www.dur.ac.uk/researchmethodscentre
https://www.durham.ac.uk/departments/academic/mathematical-sciences/

Contents

Figures

Tables

Appendix tables

Executive summary

The aim of this study was to investigate and empirically derive parameters commonly used for statistical power and sample size calculations to better inform future trial design.

Towards achieving this aim, the research project leveraged the richness of the National Pupil Database (NPD) and the Education Endowment Foundation (EEF) Archive to:

I) Estimate unconditional and conditional school-level intra-cluster correlation coefficients (ICCs) for English and maths attainment outcomes at four educational Key Stages – Early Years Foundation Stage (EYFS), Key Stage 1 (KS1), Key Stage 2 (KS2) and Key Stage 4 (KS4).

II) Estimate correlation coefficients between test scores at pupil and school level for English and maths for three subsequent Key Stages – EYFS to KS1, KS1 to KS2, and KS2 to KS4, along with explanatory power of three subsequent Key Stage pre-test scores at school and pupil level.

III) Draw from NPD derived estimates from objectives I and II to provide examples of MDES calculations for each Key Stage.

IV) Build on Allen et al.'s (2018) work which analysed the test properties of the major commercial assessments that have been used by the EEF. The aim was to assess the value of using a commercial pre-test in explaining variation at post-test, and to examine the absolute reduction in the minimum detectable effect size (MDES) achieved by commercial tests relative to NPD data.

The empirical estimation of ICCs and correlations was also conducted for pupils eligible for free school meals (FSM).

Data and Outcomes

Four sets of data were analysed: i) all English schools (using the whole NPD dataset), ii) a random sample of English schools (also derived from the NPD dataset), iii) schools that have participated in an EEF trial (all trials combined, and individual trials available in the EEF Archive) and iv) schools that have participated in an EEF trial with equivalent NPD data.

The test outcomes for the analysis were English/literacy and maths for all Key Stages (EYFS, KS1, KS2, and KS4). The analyses were replicated for eight academic years (2011/12 to 2018/19).

Key findings

Intra-cluster correlations

We provide summarised **unconditional ICC results** for datasets (i) to (iii) in Table 1. Unconditional ICC estimates the proportion of variance in an outcome that is found between clusters (or schools) without any covariates. This table reports median values of ICCs over the years 2012 to 2019 which are robust to outlying results in individual years.

Table 1: Summary of unconditional ICC estimates, median (2012-19)

	Data	Median ICC	
		English	Maths
EYFS	NPD whole	0.06	0.08
	NPD sample*	-	-
	EEF studies	0.09	0.16
KS1	NPD whole	0.05	0.03
	NPD sample	0.04	0.04
	EEF studies	0.08	0.13
KS2	NPD whole	0.12	0.11
	NPD sample	0.12	0.11
	EEF studies	0.12	0.10
KS4	NPD whole	0.13	0.10
	NPD sample	0.10	0.10
	EEF studies	0.10	0.11

*Since yearly data for EYFS was limited, no NPD sample ICC analysis was done for EYFS pupils.

Detailed results of the year-by-year analyses are provided in the Results section of this report, but we provide the key points here, in addition to the information provided in Table 1. For the EYFS, the NPD analyses show that, between 2012 and 2019, the unconditional ICC estimates ranged between 0.05 and 0.06 for English and between 0.07 and 0.08 for maths, except for 2012 where the estimates were greater than 0.10. In KS1, the unconditional ICC estimates for English and maths were broadly comparable. In KS2 and KS4, the NPD analyses further shows that the unconditional ICC estimates were larger than seen in KS1 and were again broadly comparable for English and maths. For EEF studies, unconditional estimates for EYFS suggest slightly higher ICCs for maths compared to literacy. This is echoed for KS1 estimates. As we move to KS2, unconditional ICCs become slightly higher for English than what we would expect to find for educational studies. For KS4, the unconditional ICC estimates in 2014 stand out with relatively higher values for both English and maths; but for other years, they range between 0.05 and 0.13. A comparison of NPD samples with EEF studies shows that the unconditional ICC estimates converged for KS2 and KS4 in recent years. Similar convergence for KS2 and KS4 estimates was observed for FSM-eligible pupils as well. However, the EYFS and KS1 estimates for the NPD and EEF samples were not of similar magnitude.

Conditional ICC analyses show that ICC estimates for conditional models (considering pre-test as a covariate) were able to explain a large amount of the school- and pupil-level variation in outcomes (e.g., explanatory power for pre-test at the school level was more than 0.62, and at the residual level more than 0.39, for KS4 English estimates in the NPD samples). It is important to mention that in comparison to school-level pre-test, pupil-level pre-test explains more of the variation of outcomes, while including pre-test at both the pupil- and school-level resulted in some further reduction in variance. Additional covariates, such as for FSM eligibility, special education needs (SEN) or English as an additional language (EAL), did not explain any additional variation in the school-level or pupil-level residual variance once pre-test at pupil- and school-level were included in the model. Further, conditional ICC estimates for models with pre-test as a covariate and unconditional models were similar for most years for all three Key Stages (KS1, KS2 and KS4) for both NPD samples and EEF studies.

Correlation and explanatory power

Correlations between successive Key Stage outcomes are summarised for datasets (i) to (iii) in Table 2, again using the median of all annual results obtained for the years 2012 to 2019.

Table 2: Summary of pupil-level correlation estimates, median (2012-19)

	Data	Median Correlation	
		English	Maths
EYFS-KS1	NPD whole	0.52	0.38
	NPD sample	0.45	0.35
	EEF studies	0.75	0.64
KS1-KS2	NPD whole	0.66	0.70
	NPD sample	0.66	0.70
	EEF studies	0.57	0.57
KS2-KS4	NPD whole	0.57	0.66
	NPD sample	0.58	0.65
	EEF studies	0.64	0.70

Correlation analysis between Key Stages shows reasonable consistency between estimates from the whole NPD and sampled NPD datasets. Correlation estimates for EYFS and KS1 have increased over time. There is a strong correlation between KS1 and KS2 English outcomes (greater than 0.60). Interestingly, correlations at pupil and school levels are reasonably similar for both English and maths in all Key Stages. For EEF studies, estimates for the correlation between EYFS and KS1 are notably higher but similar for both English (range: 0.54 to 0.88) and maths (range: 0.50 to 0.88). Correlations for both English and maths, for KS1 and KS2 as well as KS2 and KS4, ranged mostly between 0.50 and 0.70 over the years. The correlation estimates for EYFS-KS1 are higher for EEF studies than for NPD data in most years. Correlation estimates for the NPD and EEF samples converged much better for KS1-KS2 and KS2-KS4 than EYFS-KS1. For FSM-eligible pupils, correlation estimates for KS1-KS2 and KS2-KS4 obtained from the NPD and EEF data are very close for most of the years. For EYFS-KS1, there are larger differences, similar to what has been observed for all pupils' data.

School-level explanatory power estimates obtained from conditional models that included pre-tests at both pupil and school levels were consistently lower than those obtained from the correlation estimates. Residual (within-school, between-pupils) explanatory power estimates were slightly greater compared with those obtained from the pupil-level correlation estimates. The reasons for this seem clear. The conditional models were multivariate, and so school-level explanatory power estimates drew on pre-test variance to account for school-level variance in an outcome. By way of contrast, the estimates obtained from the (squared) school-level correlations were bivariate and did not take account of any covariance between pupil- and school-level pre-tests. The (squared) pupil-level correlation estimates closely reflected estimates for total explanatory power across all NPD analyses. However, trial sensitivity draws on residual rather than total explanatory power: total explanatory power is an estimate of the proportion of explained variance at both pupil and school levels, whilst residual explanatory power is an estimate of the proportion of explained variance that is within schools, between pupils (the variance that remains after between-school variance is accounted for).

MDES and sample size

Overall, the aim of this study was to obtain estimates of the key study design parameters such as unconditional and conditional ICCs, pre/post-test correlations and/or explanatory power estimates from the NPD and EEF Archive data. These estimates can be used to inform the design of 2-level cluster randomised trials along with pupil-randomised, multisite trials and quasi-experimental designs to identify the MDES for a fixed sample size of schools and pupils. Using these estimates to provide an applied example of MDES estimates across all Key Stages was the third objective of this study.

The resulting analysis shows that when including an NPD pre-test for KS1 maths, KS2 English and KS4 English/maths, a sample of 80+ schools with 20+ pupils per school or that of 100+ schools with 10+ pupils per school is typically needed to achieve a MDES of 0.20. Sample size requirements are larger for EYFS English/maths when including an NPD pre-test, where a sample of 100+ schools with 20+

pupils per school or that of 150+ schools with 10+ pupils per school is required for the same MDES. KS1 English is the outcome for which including an NPD outcome lowers the sample size requirements the most, with only 50+ schools with 20+ pupils per school needed for a MDES of 0.20.

Furthermore, detecting a MDES of 0.10, which is a more commonly observed effect size in EEF trials, requires a sample of 180+ schools with 30+ pupils per school for KS4 English, 200+ schools with at least 30+ pupils per school for KS1 English and maths, 230+ schools with 30+ pupils per school for KS4 maths, 270+ schools with 30+ pupils per school for KS2 English, and more than 300+ schools with 30+ pupils per school in EYFS and KS2 maths.

For FSM-eligible pupils, detecting a MDES of 0.20 requires 80+ schools with 10+ FSM-eligible pupils per school in KS2 English and KS4 English/maths, while 80+ schools with 30+ pupils per school are needed for KS2 maths. However, it is important to mention here that it is practically infeasible to power trials for FSM-eligible pupils with a MDES of 0.10.

MDES estimates for the given sample size in the report are estimated using unconditional ICCs and explanatory power estimates for the most recent three years of NPD data. However, estimates for correlation, ICC and variance from 2012 to 2019 for NPD and EEF Archive data are available and provided as Excel spreadsheets which can be utilised by evaluators and researchers conducting educational trials for the estimation of MDES relevant for their study.

Commercial pre-test

To address the fourth objective, a **comparison of the MDES for commercial and NPD pre-test models** was performed. These results suggested that MDES estimates obtained using NPD test data as a baseline indicator were marginally higher/lower than the MDES obtained using a commercial pre-test, except for a few trials where large differences were observed due to small sample sizes. There is a strong positive relationship between the MDES obtained using commercial and NPD pre-tests for both English and maths outcomes. This finding suggests that replacing a commercial pre-test with NPD pre-test data may not make a large difference in the MDES required for studies evaluated using commercial tests.

Study rationale and background

Statistical power analysis refers to the equivalent questions of identifying the sample size needed to detect a given effect size with a certain probability, or to identify the minimum effect size that can be detected with a given sample size, which is commonly referred to as the minimum detectable effect size (MDES). It helps researchers and funders to balance between recruiting too few or too many participants (in education research, often schools or pupils), ensuring an adequate sample size for an effect size that is deemed appropriate for (the cost of) the intervention. Trials that fail due to not recruiting enough participants or large trials with negligible effect size are not good value for resources.

This collaborative project between the University of Durham and Sheffield Hallam University focused on providing useful and up to date parameter estimates which can be used for the design of randomised control trials (RCTs) in educational contexts. This includes cluster randomised trials (CRTs), where randomisation is done at a school level with pupils clustered into schools, but also multisite trials or quasi-experimental trials designs (QEDs), etc. In essence, the aim was to investigate and empirically derive parameters such as intra-cluster correlation (ICC) and correlation commonly used for statistical power and sample size calculations to better inform future trial design. Estimation of ICC is important to appropriately account for clustering in educational outcomes, as children from one school are likely to be more similar in their educational outcomes when compared to children from other schools. Further, accurate estimates of correlation coefficients between test scores and explanatory power can be used to reduce unwarranted variation and consequently improve the power of trials.

Along with this, this study also assessed implications of using commercial tests for the estimation of MDES for educational trials relative to the national-level standardised Key Stage scores for different Key Stages. Building on a previous study conducted by Allen et al. (2018), this specific analysis examined all RCT trials funded by the EEF that have used commercial assessments as outcome measures. Allen et al. (2018) mainly investigated the predictive validity of the commercial test scores for Key Stage test scores and estimation of ICC across selected trials both in terms of commercial tests, KS2 examinations and the magnitude of achievement gaps between demographic groups.

Objectives

Towards achieving this aim, the research project leveraged the richness of the National Pupil Database (NPD) and the Education Endowment Foundation (EEF) Archive to:

I) Estimate unconditional and conditional school-level intra-cluster correlation coefficients (ICCs) for English and maths attainment outcomes at four educational Key Stages – Early Years Foundation Stage (EYFS), Key Stage 1 (KS1), Key Stage 2 (KS2) and Key Stage 4 (KS4).

II) Estimate correlation coefficients between test scores at pupil and school level for English and maths for three subsequent Key Stages – EYFS to KS1, KS1 to KS2, and KS2 to KS4, along with explanatory power of three subsequent Key Stage pre-test scores at school and pupil level.

III) Draw from NPD derived estimates from objectives I and II to provide examples of MDES calculations for each Key Stage.

IV) Build on Allen et al.'s (2018) work which analysed the test properties of the major commercial assessments that have been used by the EEF. The aim was to assess the value of using a commercial pre-test in explaining variation at post-test, and to examine the absolute reduction in the minimum detectable effect size (MDES) achieved by commercial tests relative to NPD data.

Empirical estimation of ICCs and correlations was also conducted for FSM-eligible pupils. It is important to understand the variation in these key study design parameters for FSM pupils to be able to improve analyses of this subgroup in trials.

Ethics and registration

Ethical approval for this study is provided by the Department of Anthropology, Durham University. The data used in the quantitative analyses were extracted from the EEF Archive generated by the Fischer Family Trust and provided to Durham and Sheffield Hallam University as part of the EEF Archive and Database project through the Office for National Statistics (ONS) Secure Research Service (SRS).

Data protection

The EEF commissioned this project and is the data controller for this project and the EEF Data Archive. Durham University and Sheffield Hallam University are both the data processors. The legal basis for processing this data by EEF is 'Legitimate Interest' while for Durham and Sheffield Hallam University is 'Public Task' as defined in Article 6(1e) of the General Data Protection Regulations (GDPR).

The research team at Durham and Sheffield processed the pseudonymised extracts from the EEF Archive and matched them with additional data from the NPD including Pupil Matching Reference and Schools IDs (Unique Reference Number), which are made available by the Department for Education (DfE) in the ONS SRS.

NPD data variables of sensitivity level C (FSM-eligibility), D and E (exam results) were used in this study as mentioned in the NPD data tables. All the analyses for this project were performed in a secure environment as per ONS and DfE guidelines. This work was produced using statistical data from the ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce ONS aggregates.

Methods

Data and outcomes

Data: The analyses were conducted for i) all non-selective, mainstream English schools (using the whole NPD dataset), ii) a random sample of all non-selective, mainstream English schools (also derived from the NPD dataset), iii) schools that have participated in an EEF trial (all trials combined, and individual trials available in the EEF Archive), and iv) schools that have participated in an EEF trial with equivalent NPD data. All these data were made available through the secure research environment of the ONS.

- **NPD whole**: Analysis for all non-selective mainstream English schools[1] provides an appropriate understanding of the statistical parameters for all pupils in English schools at the national level. These findings are generalised for all pupils in England, and they are robust and useful for designing any future study.

- **NPD sample**: A random sample of non-selective mainstream English schools was selected to generate estimates for each indicator using an appropriate probabilistic sampling approach. The required number of NPD sample schools was equal to the total number of schools available in EEF Archive data at the time of analysis. This number was broken down by the Key Stages (KS1, KS2 and KS4) using the EEF Archive data. A similar number of representative schools from each Key Stage were then selected from the NPD data using a two-stage approach. First, the total number of schools was stratified based on nine geographical regions in England (i.e., East Midlands, East of England, London, North East, North West, South East, South West, West Midlands, Yorkshire and the Humber). Then, the total number of schools to be selected from each specific region was calculated by multiplying the total number of schools by the percentage of schools in that region. In the second stage, the required number of schools from each region was selected using a probability proportionate to size (PPS) sampling method. This sampling approach ensured that a representative sample of schools is obtained whilst maintaining the required heterogeneity.

- **EEF studies**: Further, a separate analysis including all schools in the EEF Archive was carried out. Analysis for EEF schools was done by considering schools and pupils that have taken part in an EEF study and complemented with additional inputs through the NPD. By using key identifiers, such as Pupil Matching Reference (PMR), the Durham and Sheffield Hallam University teams merged NPD data with the EEF data. Data from EEF trials (including cluster and multisite trials) which were conducted during 2011-2019 were used in this study. All the observations available from each of these trials were utilised. More details of the EEF studies data are available in Appendix A.

- **EEF NPD:** The EEF studies data file was adapted so that for trials with a commercial test outcome, this was replaced by a suitable NPD test outcome.

It is known that convenience samples are common across EEF trials, which may limit the generalisability of their results to the broader population of English schools (EEF, 2018). Therefore, this study also aims to assess the external validity of EEF studies by comparing them to the general NPD population. Given that the NPD dataset represents the complete data for the entire population of schools in England, it is possible that significant differences in the number of schools included in the EEF studies and the NPD dataset could arise. To address this issue and provide a more precise comparison, an equivalent random sample of schools was drawn from the NPD for each Key Stage, mirroring the EEF sample.

[1] All mainstream, non-selective primary and secondary schools in England included in the NPD data files.

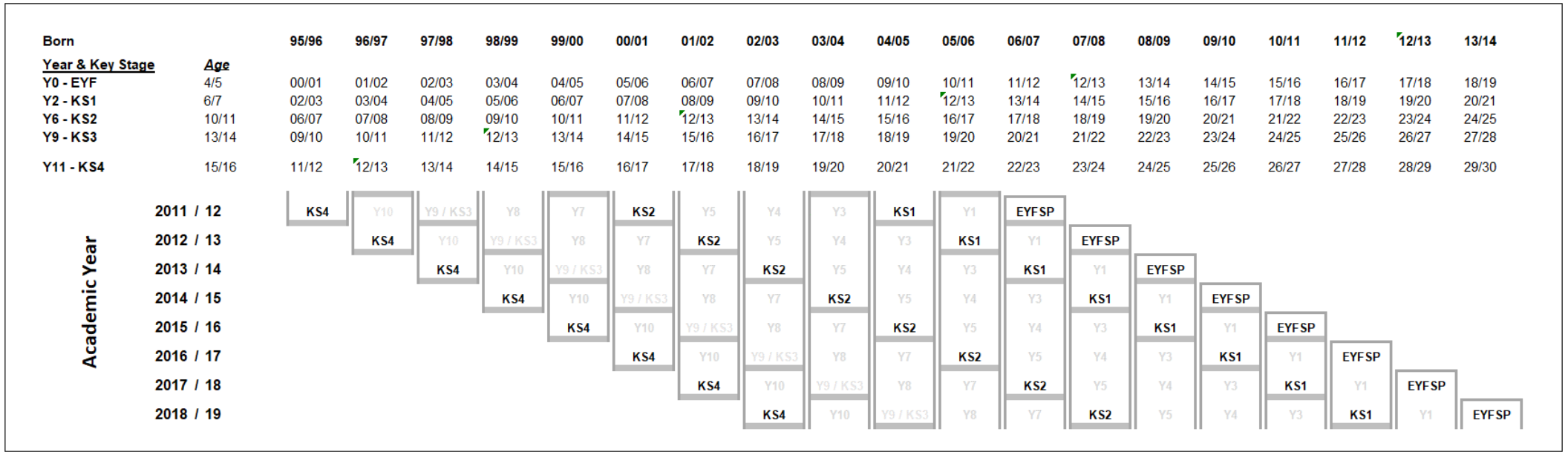
This refined comparison aimed to determine whether the parameter estimates obtained from the EEF data align closely with those of the general population sample. It is important to mention here that the size of the NPD random sample is equal to the sample size of EEF studies. The main aim of performing this comparison was to examine whether the statistical parameters generated from the NPD random sample of English schools (using probabilistic sampling methods) would converge or diverge from the estimates generated from all EEF schools. By assessing the convergence or divergence of the estimates, it can be determined whether the EEF study population is representative of the broader population. If the estimates from the EEF and NPD data converge and demonstrate similarity, it suggests that the EEF estimates closely reflect the population. Conversely, if the estimates diverge significantly, it indicates that the EEF estimates may not be generalisable to the broader population.

All statistical analysis for the project was done using R and STATA software. Analyses for the NPD and EEF Archive were undertaken within the SRS and publication clearance for outputs was obtained.

Outcomes: The outcomes for the analysis were English/literacy and maths test scores for all Key Stages (EYFS, KS1, KS2, and KS4). The analyses were replicated over the span of eight academic years (2011/12 to 2018/19) and drew on data from 19 pupil cohorts (Figures 1 and 2 summarise both the Key Stages and timeframe of the analysis).

Figure 1: **A single pupil cohort; Y0 (age 4/5) to Y11 (age 15/16)**

Figure 2: **Pupil cohorts included in analyses**



The ICC and correlation estimates were computed for each specific year, allowing for a better understanding of the impact of EYFS, KS1 and GCSE assessment and scoring changes between years on the parameters of interest. New KS1 scores from 2016-19 were recoded as a continuous variable, GCSE point scores in 2017-19 were treated as a continuous variable, and GCSE grades in 2012-16 were recoded as continuous variables to estimate the required indicators. Full details are presented in Table 8.

For EEF studies, English/literacy and maths outcomes for the trials were either outcomes measured with a commercial test or the NPD test scores for different Key Stages. To analyse the different outcomes for various trials together over time, the outcomes were standardised trial-wise. This standardisation process allowed a statistical conversion of individual pupil test scores into Z-scores. This means that individual pupil test scores for any of the outcomes were converted into distance from the population mean, which was then divided by the standard deviation (SD) of the population mean score to derive the Z-score.

Objective I: Intra-cluster correlations

In a clustered design with two levels, ICC estimates the proportion of variance in an outcome that is found between clusters (or schools). The ICC can also be defined as the strength of clustering of variance in an outcome at the school level. The ICC value is an important estimate in the design of clustered trials as the statistical sensitivity of a CRT decreases with an increasing ICC. For this analysis, a multilevel random intercept model with the schools as random effects was used to obtain accurate information on the ICCs. Both unconditional and conditional ICCs were estimated using the following equation:

$$\mathrm{ICC} = \frac{\mathrm{Variance}_{\mathrm{school}}}{\mathrm{Variance}_{\mathrm{school}} + \mathrm{Variance}_{\mathrm{pupil}}} \quad (1)$$

Unconditional ICC

Unconditional (using an empty or null multilevel model) ICCs were estimated across the four datasets (NPD whole, NPD sample, EEF studies and EEF NPD) mentioned above. An empty multilevel model (i.e., a multilevel model without any covariate) was fitted for each of the EYFS, KS1, KS2, KS4 outcomes as dependent variables. Estimates were derived using Equation 1, based on the school- and pupil-level variances from empty multilevel models. These unconditional variance estimates were also drawn on to estimate the explanatory power from including a pre-test (measure of prior attainment) at pupil and school levels to supplement the bivariate correlation estimates in objective II.

Conditional ICC

According to the multilevel modelling literature (Bloom et al., 2007; Hedges and Hedberg, 2013), including covariates can reduce the variance to be explained (which would reduce the MDES), but this could also affect the ICCs. However, the direction of the change of the ICCs is unknown and their subsequent effect on MDES can be ambiguous. Therefore, it is important to explore the effect of including covariates on the estimation of ICCs to better understand their impact in power calculations.

Conditional ICCs were estimated using both the NPD data and EEF studies data. A series of multilevel random intercept models were fitted, and covariates were added in a stepwise approach, with ICCs being estimated for each model. The most important covariate at the pupil level was previous attainment, as it is a strong predictor of current academic attainment (Hemmings et al., 2011; Mujis and Dunne, 2010). Therefore, it is also the EEF's preferred analytical model (EEF, 2022). Several other education-based studies also suggest that previous attainment can significantly reduce the MDES and the number of randomised schools required for a certain level of precision (Bloom et al., 2007; Hedges and Hedberg, 2007; Hedges and Hedberg, 2013). The stronger the correlation between pre- and post-test, the greater the gains in sensitivity.

The explanatory power provided by a measure of prior attainment (pre-test) can be at cluster (school) and individual (pupil) levels. At both levels, statistical sensitivity increases with increasing explanatory power from a pre-test. However, for a CRT design, sensitivity is influenced more by school-level explanatory power than pupil-level explanatory power. This is clearly shown in the MDES equation below (Equation 6). A pre-test might be included in a model in different ways:

- First, the pre-test was included as a raw score at the pupil level.
- Second, the pre-test was included as a raw score at the pupil level and as an aggregated (mean) raw score at the school level.
- Third, the pre-test was included as centred score at both pupil- and school-levels. In this final centred model, at the pupil level the raw score is centred around the school mean (the school mean is subtracted from the raw score for a particular pupil). At the school-level, the school mean is centred around the grand mean of aggregated pre-test scores (the grand mean of aggregated pre-test scores is subtracted from the mean score for a particular school).

Estimates for conditional ICCs and explanatory power for these three approaches were compared with their respective unconditional ICC and bivariate correlation estimates. Comparing the first two approaches illustrates whether including a school-level measure of attainment leads to gains in explanatory power (and hence statistical sensitivity) and helps to clarify the impact on ICCs. The third approach uses centred versions of attainment at pupil- and school-levels. Centering removes any correlation between the pupil- and school-level measures and so avoids potential problems of multicollinearity (see Demack 2018; Hedges and Hedberg, 2013). There are methodological advantages for centering the variables if school-level means are strongly correlated with their pupil-level scores. Since schools included in the NPD and EEF analyses were non-selective mainstream primary or secondary schools, it seems unlikely that the correlation between school- and pupil-level attainment will be very high[2]. Comparing the third (centred) with the second (raw) approach illustrates if/how the approaches resulted in different estimates for explanatory power and conditional ICCs.

The analysis was extended to include other standard covariates like FSM-eligibility, special education needs (SEN), and English as an additional language (EAL), which are commonly included in educational RCTs conducted in England. Several studies highlight significant variation in education outcomes for these standard covariates (Strand et al., 2015; Gorard, 2018). Since CRT design is commonly used in EEF trials, all these covariates are also considered at the school level along with the pupil level to examine the explanatory power of these covariates at that level. These additional analyses can be found in the accompanying Excel files.

Table 3 summarises all the variables used for this analysis and Table 4 shows the 13 models used for conditional ICC estimation by the research team. The use of pre-test covariates to help maximise the statistical power/sensitivity of a 2-level clustered RCT design is widespread across EEF trials whilst the use of other pupil- and school-level covariates to maximise statistical power is less common. Therefore, we foreground the findings from conditional models that only included pre-test covariates here. Findings from conditional models that included variables other than a pre-test (such as FSM, EAL, SEN) can be found in the accompanying Excel files. We do this because our findings from conditional models that only included a pre-test are likely to be of use for most designers of 2-level clustered RCTs in educational settings, whilst conditional models that include variables other than a pre-test are less widely used.

[2] If the schools were highly selective based on attainment, pupils with higher attainment would be clustered into schools with higher mean attainment (i.e., positively correlated). The more 'comprehensive' the system is, the lower this correlation will be.

Table 3: Covariates used in conditional ICC models

Variable	Description
Pre-test	Pupil-level raw pre-test score
Pre-test (School)	School-level mean pre-test
Pre-test (Centred)	Pupil-level pre-test centred around the school mean
Pre-test (School Centred)	School-level pre-test centred around the grand mean of aggregated pre-test scores
SEN	Pupil-level binary variable for Special Education Needs
EAL	Pupil-level binary variable for English as an Additional Language
FSM	Pupil-level binary variable for Free School Meals
%SEN/EAL/FSM	School-level percentage of SEN/EAL/FSM

Table 4: The set of covariates included in each fitted multilevel conditional model

Model	Pre-test	SEN	EAL	FSM	%SEN	%EAL	%FSM	Pre-test (School)*	Pre-test (Centred)**	Pre-test (School Centred)***
1	x									
2		x	x	x						
3	x	x	x	x						
4								x		
5					x	x	x			
6					x	x	x	x		
7	x							x		
8	x	x	x	x	x	x	x	x		
9									x	
10	x									x
11									x	x
12	x	x	x	x	x	x	x			x
13		x	x	x	x	x	x		x	x

*Pre-test (School): School-level mean pre-test
**Pre-test (Centred): Pupil-level pre-test centred around the school mean
***Pre-test (School Centred): School-level pre-test centred around the grand mean of aggregated pre-test scores

Note: Models 1, 4, 7 and 9-11 are highlighted in grey, as these models all contain the pre-test measure in different ways: at the pupil level as a raw score or centred around the school mean and at the school level as a raw aggregated score or centred around the overall school-level grand mean.

Objective II: Pre/post-test correlations and explanatory power

The main approach for improving the precision of randomised experiments is to utilise information from other key covariates that explain variation in outcomes of interest. Using the predictive power of past information about sample members can help reduce unexplained variation in their future outcomes. In turn, this also reduces the standard error of the impact estimator and its corresponding minimum detectable effect (Bloom et al., 2007).

The previous section on ICCs, shows the importance of using a pre-test or the previous Key Stage score as a covariate for reducing the variance of scores (English/maths) at the school and pupil levels.

Therefore, the aim of this objective was to provide correlation estimates of pre-test scores with their current Key Stage or post-test scores to examine the strength of this association. Additionally, this section will also provide estimates of the explanatory power of pre-test scores at both school and pupil levels, which is a key parameter for the estimation of the MDES in RCT studies in the field of education.

Correlation

The bivariate Pearson correlation for the population of English schools, the random sample of English schools and for the EEF sample were analysed and compared. The correlation was estimated for the population in all the EEF trials combined, and for each EEF trial separately. All these analyses were undertaken at both the pupil and school level and were performed for each year.

The proportion of correlation at the school level ($\mathrm{Corr}_{\mathrm{ratio}}$) was also obtained as the ratio of school-level correlation ($\mathrm{Corr}_{\mathrm{ratio}}$) divided by the total correlation ($\mathrm{corr}_{\mathrm{school}} + \mathrm{corr}_{\mathrm{pupil}}$), as shown in equation 2.

$$\mathrm{Corr}_{\mathrm{ratio}} = \frac{|\mathrm{corr}_{\mathrm{school}}|}{|\mathrm{corr}_{\mathrm{school}}| + |\mathrm{corr}_{\mathrm{pupil}}|} \quad (2)$$

This measure captures the percentage of total correlation at the school-level and is estimated to understand the contribution of correlation estimates at both school and pupil levels. These ratio estimates are not discussed in the report but are provided in the supplementary Excel files.

Explanatory power

The correlation estimates were supplemented by estimates of explanatory power extracted from multilevel analyses for the conditional models 1, 7 and 11 specified in Table 4. Table 5 summarises the set of covariates included in these three conditional models.

Whilst Pearson correlation estimates are useful to draw on, they are bivariate. The explanatory power estimates from models 7 and 11 draw on multivariate analyses by including the pre-test at both pupil- and school-levels in raw (model 7) or centred (model 11) forms.

Table 5: The three conditional models that included prior attainment covariates only

Model	Model 1	Model 7	Model 11
Pupil-level raw pre-test scores	X	X	-
Pupil-level pre-test centred around the school mean	-	-	X
School-level mean pre-test	-	X	-
School-level pre-test centred around the grand mean of aggregated pre-test scores	-	-	X

Similar to Bloom et al. (2007), the proportion of random variance in a trial outcome between and within-clusters (or schools) that is reduced or "explained" by covariates (e.g., a pre-test) can be estimated using the conditional and unconditional (null) models. Specifically, three estimates of covariate explanatory power can be obtained:

$$School\ (Explained\ variance\ between\ schools): \qquad R_C^2 = \frac{\tau^2 - \tau_*^2}{\tau^2} \quad (3)$$

$$Residual\ (Explained\ variance\ within\ schools): \qquad R_R^2 = \frac{\sigma^2 - \sigma_*^2}{\sigma^2} \quad (4)$$

$$Total\ Explained\ Variance: \quad R_T^2 = \frac{(\tau^2 + \sigma^2) - (\tau_*^2 + \sigma_*^2)}{\tau^2 + \sigma^2} \tag{5}$$

where τ^2 and σ^2 are the between- and within-school unconditional variances from a multilevel model without covariates, and τ_*^2 and σ_*^2 are the corresponding conditional variances from a specific multilevel model with covariates.

Objective III: Using ICC and pre-test explanatory power estimates for the design of 2-level CRTs in educational settings (applied example)

The statistical sensitivity of a 2-level CRT design can be estimated using the MDES[3]. The MDES is the smallest effect size that a specified design can detect as being statistically significant (usually set as p<0.05, two-tailed) with a statistical power of 0.80 or higher. The aim of providing this applied example is to provide statistical guidance to educational researchers for calculating MDES using ICC and explanatory power estimates. The unconditional ICC estimates along with pre-test explanatory power at pupil- and school-levels can be used to estimate the MDES for 2-level clustered RCTs for English and maths outcomes in EYFS, KS1, KS2 and KS4.

Equation 6 can be used to estimate the MDES of a 2-level trial and was taken from Bloom et al. (2007).

$$MDES \sim M_{(J-m-2)} \sqrt{\frac{1}{P(1-P)}} \sqrt{\frac{ICC_2(1-R_C^2)}{J} + \frac{(1-ICC_2)(1-R_R^2)}{Jn}} \tag{6}$$

Where:

- P is the proportion of participants allocated to the intervention group (=0.50 when half are randomly allocated);
- ICC_2 is the unconditional school-level ICC coefficient (proportion of variance in an outcome that is found between schools);
- R_C^2 is the explanatory power at the school level; and R_R^2 is the explanatory power at the residual (within-school, between pupils) level;
- J is the total number of schools in the evaluation;
- n is the number of pupils per school;
- m is the number of school-level covariates included in the impact analyses; and
- M is the t-distribution multiplier, which has $(J - m - 2)$ degrees of freedom.

An alternative/addition to using Equation 6 is the PowerUp! software (Dong et al., 2015, Sheet 3.1).

[3] The 'MDES' abbreviation has been used here for brevity but included in quotations to highlight that technically, these are not MDES estimates. This is because a MDES is the effect size that a particular design would be able to detect (i.e., it is prospective) whilst the 'MDES' estimates here are estimated using outcome data (i.e., they are retrospective). The 'MDES' estimates provide an indication of how sensitive a prospective trial design would be, assuming the same sample size, ICC and explanatory power estimates observed from retrospective trial data.

Objective IV: Value of commercial pre-tests

Pre-test covariates are useful to increase power/sensitivity. If a pre-test is available in the NPD, the benefit of a commercial pre-test is mainly to increase covariate explanatory power (and hence power/sensitivity). Allen et al. (2018) illustrated that correlations for commercial pre-tests were not significantly higher than for NPD pre-tests. However, due to recent changes in national assessments and the way data is reported in the NPD, the work from Allen et al. (2018) needs updating.

The commercial pre-test analysis in this report comprised two components:

- For EEF trials where both NPD and commercial pre-tests were available, comparative analysis of the correlation of NPD-based pre-test with commercial post-test versus commercial pre-test with commercial post-test was done. Similarly, conditional ICC and variance values were obtained for these trials by including either a commercial or NPD pre-test, or both, in the multilevel model. These conditional ICC and variance estimates were used to compare the effect of commercial/NPD pre-test on ICC and variance parameters. Furthermore, including both commercial and NPD pre-tests in the same model made it possible to understand whether commercial pre-tests explained any extra variation in the post-test outcome, once we accounted for the NPD baseline data.

- The benefits of a commercial pre-test were assessed further based on the change in MDES. As a marginal change in MDES is harder to achieve when the MDES is lower, the importance of commercial pre-tests was evaluated for the trials with different numbers of schools (say, N1, N2, N3, N4, N5 and N6). The change in MDES achieved by commercial tests relative to NPD data was assessed using the ICC and variance estimates obtained from the multilevel models in the first component, for EEF trials with different numbers of schools and where the information for both commercial pre-test and NPD equivalent scores were available.

It is important to note that this analysis was applicable only to EEF trials that have used a commercial pre and post-test scores.

Results

This section presents key results and reflections on the ICC and correlation analyses done using NPD and EEF Archive data. Please note that the detailed results, including all relevant parameters such as school and residual variance, assessment correlations and ICCs (unconditional and conditional), obtained from the analysis are also summarised in the accompanying Excel spreadsheets for researchers who are keen to use these results for the purposes of their work.

Whilst the scope of the project is outlined in the Methods section above, coverage is summarised in Table 6 and Table 7 below, for the ICC and correlation analyses respectively.

Table 6: **Coverage of ICC estimates across Key Stages**

	NPD whole	NPD sample	EEF studies	EEF NPD
EYFS	2012; 2017 to 2019 (unconditional ICC only)	None	2014 (English only); 2017	None
KS1	2012 to 2019	2012 to 2019	2013 to 2015; 2017; 2018 (maths only); 2019 (English only)	2013 to 2015; 2017 to 2019
KS2	2012 to 2019	2012 to 2019	2013 to 2018; 2019 (English only)	2013 to 2019
KS4	2012 to 2019	2012 to 2019	2013 to 2014; 2016 to 2017	2013 to 2014; 2016 to 2017

Note: The ICC estimates for the whole NPD dataset and NPD sample are complete, except for the EYFS and missing years for the EEF analyses (EEF studies and EEF NPD).

Table 7: Coverage of correlation estimates across Key Stages

	NPD whole	NPD sample	EEF studies	EEF NPD
EYFS-KS1	2012 to 2019	2012 to 2019	2013 to 2014; 2015 (English only); 2017; 2018 (maths only) 2019 (English only)	2013 to 2015 2017 to 2019
KS1-KS2	2012 to 2019	2012 to 2019	2013 to 2018; 2019 (English only)	2013 to 2019
KS2-KS4	2012 to 2019	2012 to 2019	2013 to 2014; 2016 to 2017	2013 to 2014; 2016 to 2017

Note: The whole NPD and sample NPD analyses are complete whilst the EEF analyses (EEF studies and EEF NPD) have some missing years.

Key Stage Outcomes

Table 8 summarises the NPD English and maths outcomes used in the analyses. It is important to highlight once more that between 2012 and 2019, assessment changes were introduced at all four Key Stages:

- EYFS: New Early Years Foundation Stage Profile introduced in 2013[4].
- KS1: New curriculum introduced in 2014, KS1 SATs changed from 2016[5].
- KS2: New curriculum introduced in 2014, KS2 SATs change from 2016[6].
- KS4: New 0 to 9 scale used in GCSE English and GCSE maths from 2017[7].

Table 8: **English and maths outcomes used in ICC and correlation analyses, NPD data**

Class year, Key Stage	Subject	Date	Variable	Scale
Y0, EYFS	English	2012	Score in 'Communication, Language & Literacy (CLL)'	0-36
		2013-19	Mean score for 'EYFS Reading & Writing'	0-6
	Maths	2012	Score in 'Mathematical Development'	0-9
		2013-19	Mean score for EYFS Number & 'Shape, space, measures'	0-6
Y2, KS1	English	2012-15	KS1 Reading	1-6
		2016-19	KS1 Reading*	BLW-GDS (1-6)
	Maths	2012-15	KS1 Maths	1-6
		2016-19	KS1 Maths*	BLW-GDS (1-6)
Y6, KS2	English	2012-15	KS2 Reading	0 - 50
		2016-19	KS2 Reading	0 - 50
	Maths	2012-15	KS2 Maths	0-110
		2016-19	KS2 Maths	0-110
Y11, KS4	English	2012-16	GCSE English*	A*-U
		2017-19	GCSE English	0-9
	Maths	2012-16	GCSE Maths*	A*-U
		2017-19	GCSE Maths	0-9

*More details for each categorical variable are available in Appendix B.

[4] Early Years Foundation Stage Profile - https://www.gov.uk/government/statistics/early-years-foundation-stage-profile-results-2012-to-2013
[5] KS1 Assessments - https://www.gov.uk/government/statistics/phonics-screening-check-and-key-stage-1-assessments-england-2016
[6] KS2 Assessments - https://www.gov.uk/government/statistics/national-curriculum-assessments-key-stage-2-2016-revised
[7] Revised GCSE - https://www.gov.uk/government/statistics/revised-gcse-and-equivalent-results-in-england-2016-to-2017

Objective I: Intra-cluster correlations

Unconditional ICC estimates

Table 9 to Table 12 provide unconditional ICC estimates for English and maths outcomes in all Key Stages using NPD and EEF Archive data.

Table 9: Unconditional ICC estimates for Early Years Foundation Stage (EYFS)

	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	Maths	English	Maths	English	Maths	English	Maths
2012	0.15	0.17	-	-	-	-	-	-
2013	-	-	-	-	-	-	-	-
2014	-	-	-	-	0.07	-	-	-
2015	-	-	-	-	-	-	-	-
2016	-	-	-	-	-	-	-	-
2017	0.06	0.08	-	-	0.10	0.16	-	-
2018	0.05	0.07	-	-	-	-	-	-
2019	0.05	0.07	-	-	-	-	-	-
Max	0.15	0.17	-	-	0.10	-	-	-
Min	0.05	0.08	-	-	0.07	-	-	-
Median	0.06	0.08	-	-	0.09	0.16	-	-

Notes: A new EYFS Profile was introduced in 2013. Cells with a '-' sign indicate that no data is available for those years.

Table 10: **Unconditional ICC estimates for Key Stage 1 (KS1)**

	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	Maths	English	Maths	English	Maths	English	Maths
2012	0.05	0.02	0.05	0.04	-	-	-	-
2013	0.05	0.02	0.04	0.04	0.19	0.19	0.04	0.05
2014	0.04	0.02	0.04	0.03	0.08	0.13	0.05	0.05
2015	0.04	0.02	0.04	0.03	0.06	0.05	0.62	0.67
2016	0.06	0.07	0.07	0.07	-	-	-	-
2017	0.05	0.05	0.04	0.04	0.08	0.12	0.06	0.06
2018	0.04	0.04	0.03	0.03	-	0.20	0.08	0.10
2019	0.04	0.04	0.03	0.02	0.17	-	0.09	0.05
Max	0.06	0.07	0.07	0.07	0.19	0.20	0.62	0.67
Min	0.04	0.02	0.03	0.02	0.06	0.05	0.04	0.05
Median	0.05	0.03	0.04	0.04	0.08	0.13	0.07	0.06

Notes: A new KS1 curriculum was introduced in 2014; KS1 SATs changed from 2016. Cells with a '-' sign indicate that no data is available for those years.

Table 11: **Unconditional ICC estimates for Key Stage 2 (KS2)**

	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	**Maths**	**English**	**Maths**	**English**	**Maths**	**English**	**Maths**
2012	0.12	0.10	0.12	0.10	-	-	-	-
2013	0.11	0.11	0.11	0.11	0.25	0.21	0.12	0.11
2014	0.12	0.10	0.12	0.10	0.12	0.08	0.22	0.22
2015	0.11	0.10	0.11	0.10	0.11	0.11	0.12	0.10
2016	0.13	0.14	0.13	0.14	0.12	0.07	0.13	0.13
2017	0.12	0.13	0.12	0.13	0.10	0.12	0.12	0.14
2018	0.10	0.12	0.10	0.12	0.17	0.09	0.11	0.12
2019	0.09	0.11	0.09	0.11	0.03	-	0.08	0.09
Max	0.13	0.14	0.13	0.14	0.25	0.21	0.22	0.22
Min	0.09	0.10	0.09	0.10	0.03	0.07	0.08	0.09
Median	0.12	0.11	0.12	0.11	0.12	0.10	0.12	0.12

Notes: A new KS2 curriculum was introduced in 2014; KS2 SATs change from 2016. Cells with a '-' sign indicate that no data is available for those years.

Table 12: **Unconditional ICC estimates for Key Stage 4 (KS4)**

	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	**Maths**	**English**	**Maths**	**English**	**Maths**	**English**	**Maths**
2012	0.11	0.11	0.10	0.11	-	-	-	-
2013	0.18	0.11	0.10	0.10	0.10	0.13	0.12	0.10
2014	0.14	0.11	0.10	0.10	0.39	0.33	0.11	0.10
2015	0.18	0.10	0.11	0.09	-	-	-	-
2016	0.19	0.10	0.11	0.09	0.06	0.05	0.01	0.01
2017	0.11	0.10	0.11	0.10	0.09	0.09	0.02	0.02
2018	0.11	0.10	0.10	0.10	-	-	-	-
2019	0.11	0.10	0.10	0.10	-	-	-	-
Max	0.19	0.11	0.11	0.11	0.39	0.33	0.12	0.10
Min	0.11	0.10	0.10	0.09	0.06	0.05	0.01	0.01
Median	0.13	0.10	0.10	0.10	0.10	0.11	0.07	0.06

Notes: New 0 to 9 scale used in KS4 (GCSE English and maths) from 2017. Cells with a '-' sign indicate that no data is available for those years.

Reflections on unconditional ICC estimates

NPD (whole and sample)

For the EYFS, the NPD analyses were restricted to four academic years and carried out only for the whole NPD dataset. One of the years was 2012, which was prior to the introduction of the EYFS. The unconditional ICC estimates reflect how measures in 2012 are different from those in 2017. In 2012, ICC estimates were notably higher in both English (0.15) and maths (0.17), compared with estimates from 2017 onwards which ranged between 0.05 and 0.06 for English and between 0.07 and 0.08 for maths. All EYFS analyses found slightly higher ICC estimates for maths compared with English (see Table 9).

For KS1 (Table 10), the ICC estimates included data for all eight years, were relatively small (0.05 or lower since 2017) and were broadly comparable between the English and maths estimates (between 0.04 and 0.05 since 2017).

For KS2 (Table 11), the ICC estimates also included data for all eight years and were larger than seen in KS1 but still relatively small (0.13 or lower since 2017). Again, the English and maths estimates were broadly comparable (between 0.09 and 0.13 since 2017).

For KS4 (Table 12), the ICC estimates also included data for all eight years and were slightly smaller than seen in KS2 and remarkably consistent for English (between 0.10 and 0.11) and maths (0.10) since 2017.

EEF studies

Unconditional estimates for the EYFS (Table 9) suggest slightly higher ICCs for maths (0.16 in 2017) compared to English (0.10 in 2017). This is echoed for KS1 estimates (Table 10), which ranged between 0.06 and 0.19 for English and between 0.05 and 0.20 for maths.

For KS2, ICCs become slightly higher for both English and maths, which is what we would expect to find for educational studies, ranging from 0.07 to 0.25 (with the exception of 2019), as shown in Table 11.

For KS4 (Table 12), 2014 stands out with a relatively higher ICC for both English and maths (0.39 and 0.33, respectively). Otherwise, ICC estimates for the other available years were in the broad range of 0.05 to 0.13.

Some EEF trials collect commercial test scores for English/maths, while others collect NPD scores only. As mentioned in the Methods section above, we conducted a sensitivity analysis for EEF trial data by replacing all commercial test scores with the relevant NPD test scores to understand the implications of having mixed scores on ICC and correlation measures. The EEF NPD column presented in the tables of this report refers to the results obtained from this analysis. Results for KS2 show no significant difference between ICC estimates obtained with the original test scores used in EEF studies and those obtained by replacing EEF test scores with NPD data: the ICC estimates were consistent and broadly comparable over the years. For KS1, the results obtained using equivalent NPD data show slightly lower ICC estimates compared to the actual EEF trial data. With some exceptions, ICC estimates for KS4 English and maths over the years do not vary much between the EEF studies or EEF NPD datasets.

Note that it was not possible to conduct this sensitivity analysis for the EYFS, as there is an insufficient number of Early Years trials in the EEF Archive.

Comparison between NPD sample and EEF studies

In KS1, unconditional ICC estimates for English or maths were generally larger for EEF studies compared with the NPD sample. Since 2017, ICC estimates for KS1 English ranged between 0.03 and 0.05 for the NPD whole and NPD sample analyses, between 0.08 and 0.17 for the EEF studies analysis and between 0.06 and 0.09 for the EEF NPD analysis. For KS1 maths, ICC estimates ranged between 0.02 and 0.05 for the NPD whole and NPD sample analyses, between 0.12 and 0.20 for the EEF studies analysis and between 0.05 and 0.10 for the EEF NPD analysis.

For KS2, results showed that the NPD and EEF ICC estimates are comparable for most years. In the most recent three years (2017-2019), ICC estimates for KS2 English ranged between 0.09 and 0.12 for the NPD whole and NPD sample analyses, between 0.03 and 0.17 for the EEF studies analysis and between 0.08 and 0.12 for the EEF NPD analysis. For KS2 maths, ICC estimates ranged between 0.11 and 0.13 for the NPD whole and NPD sample analyses, between 0.09 and 0.12 for the EEF studies analysis and between 0.09 and 0.14 for the EEF NPD analysis.

Apart from 2014, this pattern was the same for KS4 outcomes as well. In the most recent three years, ICC estimates for KS4 English ranged between 0.10 and 0.11 for the NPD whole and NPD sample

analyses and was 0.09 and 0.02 (2017 only) for the EEF studies and EEF NPD analyses, respectively. For KS4 maths, ICC estimates were 0.10 for the NPD and NPD sample analyses, 0.09 for the EEF studies analysis (2017 only) and 0.02 for the EEF NPD analysis (2017 only).

Overall, unconditional ICC estimates based on NPD and EEF data converged for KS2 and KS4. However, KS1 estimates for NPD and EEF samples were not similar in magnitude. It is important to mention that the KS1 sample size for EEF studies is low.

Note that it was not possible to compare ICC estimates between NPD and EEF samples for the EYFS, as there is an insufficient number of Early Years trials in the EEF Archive.

FSM subgroup analysis

Table 13 provides unconditional ICC estimates for KS1 to KS4 English and maths outcomes for FSM-eligible pupils using NPD and EEF Archive data. Since there was no significant difference between the ICC estimates obtained from the NPD whole and NPD samples for the overall analysis, we reported NPD sample estimates only for the FSM subgroup analysis. It is evident from the KS1-KS4 results that the ICC estimates for NPD and EEF samples are consistent to what has been observed for data on all pupils. In general, unconditional ICC estimates for KS2 and KS4 were higher than for KS1 in both samples. For EEF studies, ICC estimates for KS1 were mostly greater than 0.05 for all years and in the broad range of 0.05-0.19. For the NPD sample, ICC estimates for KS1 English were as low as 0.02 in 2013 and as high as 0.08 in 2016. This shows that NPD estimates for KS1 were lower than estimates from EEF studies. These differences were significantly reduced for KS2 and KS4; for example, in 2017, the ICC estimates from the NPD sample was 0.10 for KS2 English and 0.14 for KS2 maths, whereas the estimates for EEF studies were 0.09 and 0.12, respectively.

Note that there was not enough data from EEF Early Years studies to perform the EYFS analysis on the FSM subgroup over time.

Table 13: **Unconditional ICC estimates for FSM pupils by Key Stage and outcome**

	Year	KS1 English	KS1 Maths	KS2 English	KS2 Maths	KS4 English	KS4 Maths
NPD sample	2012	0.04	0.03	0.11	0.09	0.08	0.08
	2013	0.02	0.03	0.10	0.11	0.08	0.07
	2014	0.03	0.03	0.10	0.10	0.09	0.07
	2015	0.03	0.03	0.10	0.11	0.07	0.06
	2016	0.08	0.06	0.11	0.15	0.08	0.07
	2017	0.03	0.03	0.10	0.14	0.09	0.07
	2018	0.03	0.04	0.09	0.12	0.08	0.07
	2019	0.04	0.02	0.09	0.11	0.06	0.06
	Max	0.08	0.06	0.11	0.15	0.09	0.08
	Min	0.02	0.02	0.09	0.09	0.06	0.06
	Median	0.03	0.03	0.10	0.11	0.08	0.07
EEF studies	2013	-	-	-	-	-	-
	2014	0.05	0.05	0.05	0.00	0.11	0.06
	2015	0.05	0.06	0.11	0.11	-	-
	2016	-	-	0.11	0.08	0.05	0.04
	2017	0.08	0.12	0.09	0.12	0.06	0.05
	2018	-	0.19	-	0.07	-	-
	2019	0.10	-	0.54	-	-	-

	Max	0.10	0.19	0.05	0.12	0.11	0.06
	Min	0.05	0.05	0.54	0.00	0.05	0.04
	Median	0.07	0.09	0.11	0.08	0.06	0.05

Note: Cells with a '-' sign indicate that no data is available for those years.

Conditional ICC estimates

ICC estimates were obtained for the conditional models specified in Table 5 and full details are available in the accompanying Excel files. Table 14 to Table 16 below compare the unconditional ICC estimates with those obtained from the three key conditional models that included a pre-test (models 1, 7 and 11).

Table 14: **Unconditional and conditional ICC estimates for Key Stage 1 English and maths**

	Year	KS1 English				KS1 Maths			
		Uncond-itional	Conditional			Uncond-itional	Conditional		
		NULL	M1	M7	M11	NULL	M1	M7	M11
NPD whole	2012	0.05	0.05	0.05	0.05	0.02	0.02	0.02	0.02
	2013	0.05	0.05	0.04	0.04	0.02	0.02	0.02	0.02
	2014	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02
	2015	0.04	0.04	0.04	0.04	0.02	0.02	0.02	0.02
	2016	0.06	0.10	0.09	0.09	0.07	0.11	0.10	0.10
	2017	0.05	0.07	0.07	0.07	0.05	0.08	0.07	0.07
	2018	0.04	0.06	0.05	0.05	0.04	0.06	0.05	0.05
	2019	0.04	0.05	0.05	0.05	0.04	0.06	0.05	0.05
NPD sample	2012	0.05	0.05	0.05	0.05	0.04	0.03	0.03	0.03
	2013	0.04	0.04	0.04	0.04	0.04	0.03	0.03	0.03
	2014	0.04	0.05	0.04	0.04	0.03	0.03	0.03	0.03
	2015	0.04	0.04	0.04	0.04	0.03	0.03	0.03	0.03
	2016	0.07	0.11	0.11	0.11	0.07	0.10	0.09	0.09
	2017	0.04	0.07	0.07	0.07	0.04	0.06	0.05	0.05
	2018	0.03	0.05	0.05	0.05	0.03	0.05	0.04	0.04
	2019	0.03	0.05	0.05	0.05	0.02	0.04	0.04	0.04
EEF studies	2012	-	-	-	-	-	-	-	-
	2013	0.19	0.04	0.04	0.04	0.19	0.16	0.14	0.14
	2014	0.08	0.08	0.08	0.08	0.13	0.16	0.16	0.16
	2015	0.06	0.12	0.11	0.11	0.05	-	-	-
	2016	-	-	-	-	-	-	-	-
	2017	0.08	0.10	0.10	0.10	0.12	0.17	0.17	0.17
	2018	-	-	-	-	0.20	0.13	0.12	0.12
	2019	0.17	0.06	0.06	0.06	-	-	-	-

Notes: NULL refers to the model with no covariates. M1 refers to the model with pupil-level raw pre-test score as a covariate. M7 refers to the model with pupil-level raw and school-level mean pre-test scores as covariates. M11 refers to the model with pupil-level pre-test centred around the school mean and school-level pre-test centred around the grand mean of aggregated pre-test scores as covariates. Cells with a '-' sign indicate that no data is available for those years.

Table 15: **Unconditional and conditional ICC estimates for Key Stage 2 English and maths**

	Year	KS1 English				KS1 Maths			
		Uncond-itional	Conditional			Uncond-itional	Conditional		
		NULL	M1	M7	M11	NULL	M1	M7	M11
NPD whole	2012	0.12	0.14	0.14	0.14	0.10	0.17	0.17	0.17
	2013	0.11	0.12	0.12	0.12	0.11	0.19	0.18	0.18
	2014	0.12	0.13	0.13	0.13	0.10	0.17	0.17	0.17
	2015	0.11	0.12	0.12	0.12	0.10	0.17	0.17	0.17
	2016	0.13	0.14	0.14	0.14	0.14	0.23	0.22	0.22
	2017	0.12	0.13	0.13	0.13	0.13	0.22	0.21	0.21
	2018	0.10	0.12	0.12	0.12	0.12	0.20	0.20	0.20
	2019	0.09	0.11	0.11	0.11	0.11	0.18	0.18	0.18
NPD sample	2012	0.12	0.14	0.14	0.14	0.10	0.17	0.17	0.17
	2013	0.11	0.12	0.12	0.12	0.11	0.19	0.18	0.18
	2014	0.12	0.13	0.13	0.13	0.10	0.17	0.17	0.17
	2015	0.11	0.13	0.13	0.13	0.10	0.17	0.17	0.17
	2016	0.13	0.14	0.14	0.14	0.14	0.23	0.22	0.22
	2017	0.12	0.13	0.13	0.13	0.13	0.22	0.21	0.21
	2018	0.10	0.12	0.12	0.12	0.12	0.20	0.19	0.19
	2019	0.09	0.11	0.11	0.11	0.11	0.18	0.18	0.18
EEF studies	2012	-	-	-	-	-	-	-	-
	2013	0.25	0.24	0.24	0.24	0.21	0.15	0.15	0.15
	2014	0.12	0.18	0.17	0.17	0.08	0.14	0.14	0.14
	2015	0.11	0.11	0.11	0.11	0.11	0.12	0.12	0.12
	2016	0.12	0.14	0.13	0.13	0.07	0.07	0.07	0.07
	2017	0.10	0.10	0.10	0.10	0.12	0.18	0.18	0.18
	2018	0.17	0.21	0.21	0.21	0.09	0.06	0.06	0.06
	2019	-	-	-	-	-	-	-	-

Notes: NULL refers to the model with no covariates. M1 refers to the model with pupil-level raw pre-test score as a covariate. M7 refers to the model with pupil-level raw and school-level mean pre-test scores as covariates. M11 refers to the model with pupil-level pre-test centred around the school mean and school-level pre-test centred around the grand mean of aggregated pre-test scores as covariates. Cells with a '-' sign indicate that no data is available for those years.

Table 16: **Unconditional and conditional ICC estimates for Key Stage 4 English and maths**

	Year	KS1 English				KS1 Maths			
		Uncond-itional	Conditional			Uncond-itional	Conditional		
		NULL	M1	M7	M11	NULL	M1	M7	M11
NPD whole	2012	0.11	0.09	0.08	0.08	0.11	0.10	0.09	0.09
	2013	0.18	0.09	0.08	0.08	0.11	0.09	0.08	0.08
	2014	0.14	0.14	0.13	0.13	0.11	0.10	0.08	0.08
	2015	0.18	0.19	0.18	0.18	0.10	0.11	0.10	0.10
	2016	0.19	0.20	0.19	0.19	0.10	0.09	0.09	0.09
	2017	0.11	0.09	0.07	0.07	0.10	0.10	0.09	0.09
	2018	0.11	0.08	0.07	0.07	0.10	0.10	0.10	0.10
	2019	0.11	0.08	0.07	0.07	0.10	0.10	0.09	0.09

NPD sample	2012	0.10	0.10	0.09	0.09		0.11	0.11	0.10	0.10
	2013	0.10	0.09	0.09	0.09		0.10	0.10	0.08	0.08
	2014	0.10	0.10	0.09	0.09		0.10	0.10	0.08	0.08
	2015	0.11	0.11	0.11	0.11		0.09	0.10	0.10	0.10
	2016	0.11	0.09	0.08	0.08		0.09	0.09	0.08	0.08
	2017	0.11	0.08	0.05	0.05		0.10	0.10	0.08	0.08
	2018	0.10	0.07	0.06	0.06		0.10	0.10	0.10	0.10
	2019	0.10	0.07	0.05	0.05		0.10	0.10	0.09	0.09
EEF studies	2012	-	-	-	-		-	-	-	-
	2013	0.10	0.13	0.12	0.12		0.13	0.14	0.14	0.14
	2014	0.39	0.35	0.32	0.32		0.33	0.24	0.22	0.22
	2015	-	-	-	-		-	-	-	-
	2016	0.06	0.06	0.05	0.05		0.07	0.06	0.06	0.06
	2017	0.09	0.06	0.03	0.03		0.12	0.07	0.05	0.05
	2018	-	-	-	-		-	-	-	-
	2019	-	-	-	-		-	-	-	-

Notes: NULL refers to the model with no covariates. M1 refers to the model with pupil-level raw pre-test score as a covariate. M7 refers to the model with pupil-level raw and school-level mean pre-test scores as covariates. M11 refers to the model with pupil-level pre-test centred around the school mean and school-level pre-test centred around the grand mean of aggregated pre-test scores as covariates. Cells with a '-' sign indicate that no data is available for those years.

Reflections on conditional ICC estimates

NPD (whole and sample)

In KS1 (Table 14), ICC estimates for conditional and unconditional models were similar for all years except 2016, where estimates for conditional models were higher than those for the null/unconditional models. 2016 was the first year of the new KS1 assessment which may account for this inconsistency. For KS1 English, since 2017, unconditional ICC estimates ranged between 0.03 and 0.05 and conditional ICC estimates between 0.05 and 0.07. For KS1 maths, unconditional ICC estimates ranged between 0.02 and 0.05 and conditional ICC estimates between 0.04 and 0.07.

In KS2 (Table 15) ICC estimates for conditional and unconditional models were similar for all years in English, but a greater difference was seen in maths. For KS2 English, since 2017, unconditional ICC estimates ranged between 0.09 and 0.13 and conditional ICC estimates between 0.11 and 0.13. For KS2 maths, unconditional ICC estimates ranged between 0.11 and 0.13, whereas conditional ICC estimates ranged between 0.18 and 0.22.

In KS4 (Table 16) ICC estimates for conditional and unconditional models were similar for all years. For KS4 English, since 2017, unconditional ICC estimates ranged between 0.10 to 0.11 and conditional ICC estimates between 0.05 and 0.09. For KS4 maths, the unconditional ICC estimate was 0.10 and conditional ICC estimates ranged between 0.08 and 0.10.

EEF studies

ICCs for KS1 varied substantially between models for 2013 and other available years (Table 14). For KS1 English, since 2017, unconditional ICC estimates ranged between 0.08 and 0.17 and conditional ICC estimates between 0.05 and 0.16. For KS1 maths, unconditional ICC estimates ranged between 0.12 and 0.20 and conditional ICC estimates between 0.12 and 0.17.

In KS2 (Table 15), ICC estimates for conditional and unconditional models were similar for all available years. For KS2 English, since 2017, unconditional ICC estimates ranged between 0.10 and 0.17 and

conditional ICC estimates between 0.10 and 0.21. For KS2 maths, unconditional ICC estimates ranged between 0.09 and 0.12 and conditional ICC estimates between 0.06 and 0.18.

In KS4 (Table 16), ICC estimates for conditional and unconditional models were similar for all available years. For KS2 English, since 2017, the unconditional ICC estimate was 0.09 and conditional ICC estimates ranged between 0.03 and 0.06. For KS2 maths, the unconditional ICC estimate was 0.12 and conditional ICC estimates ranged between 0.05 and 0.07.

Objective II: Pre/post-test correlations

This section provides estimates of correlations between subsequent Key Stage scores for NPD and EEF Archive data (Table 17 to Table 20). Correlations for three pairs of subsequent Key Stages (EYFS-KS1, KS1-KS2, KS2-KS4) are provided for English and maths outcomes. For EEF studies, pre- and post-test scores for subsequent Key Stages were used to obtain correlations. Correlation estimates were obtained both at the pupil and school levels.

Table 17: **Correlations between Early Years Foundation Stage and Key Stage 1 (EYFS-KS1)**

Pupil-level correlations

Year	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	Maths	English	Maths	English	Maths	English	Maths
2012	0.43	0.11	0.44	0.32	-	-	-	-
2013	0.43	0.13	0.43	0.31	0.88	0.88	0.66	0.59
2014	0.44	0.15	0.45	0.33	0.79	0.74	0.67	0.64
2015	0.44	0.22	0.44	0.32	0.54	-	0.19	0.14
2016	0.59	0.53	0.44	0.37	-	-	-	-
2017	0.61	0.56	0.47	0.42	0.63	0.50	0.52	0.51
2018	0.61	0.57	0.47	0.43	-	0.54	0.58	0.52
2019	0.61	0.57	0.48	0.44	0.75	-	0.44	0.48
Max	0.61	0.57	0.48	0.44	0.88	0.88	0.67	0.64
Min	0.43	0.11	0.43	0.31	0.54	0.50	0.19	0.14
Median	0.52	0.38	0.45	0.35	0.75	0.64	0.55	0.52

Notes: A new EYFS Profile was introduced in 2013. A new KS1 curriculum introduced in 2014; KS1 SATs changed from 2016. Cells with a '-' sign indicate that no data is available for those years.

School-level correlations

Year	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	Maths	English	Maths	English	Maths	English	Maths
2012	0.44	0.17	0.50	0.40	-	-	-	-
2013	0.44	0.20	0.45	0.36	0.95	0.89	0.61	0.65
2014	0.45	0.20	0.51	0.42	0.78	0.66	0.58	0.51
2015	0.41	0.21	0.49	0.38	0.38	-	-0.36	-0.34
2016	0.39	0.32	0.30	0.28	-	-	-	-
2017	0.48	0.38	0.33	0.25	0.46	0.27	0.60	0.52
2018	0.49	0.41	0.34	0.37	-	0.68	0.72	0.59
2019	0.49	0.42	0.33	0.30	0.81	-	0.50	0.46

Max	0.49	0.42	0.51	0.47	0.95	0.89	0.72	0.65
Min	0.39	0.17	0.30	0.30	0.38	0.27	-0.36	-0.34
Median	0.45	0.27	0.40	0.37	0.78	0.67	0.59	0.52

Notes: A new EYFS Profile was introduced in 2013. A new KS1 curriculum was introduced in 2014; KS1 SATs changed from 2016. Cells with a '-' sign indicate that no data is available for those years.

Table 18: Correlations between Key Stage 1 and Key Stage 2 (KS1-KS2)

Pupil-level correlations

Year	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	Maths	English	Maths	English	Maths	English	Maths
2012	0.66	0.71	0.66	0.71	-	-	-	-
2013	0.66	0.71	0.66	0.71	0.54	0.54	0.62	0.67
2014	0.66	0.71	0.66	0.71	0.52	0.91	0.56	0.65
2015	0.67	0.70	0.67	0.70	0.64	0.60	0.59	0.68
2016	0.65	0.67	0.65	0.67	0.64	0.14	0.65	0.68
2017	0.67	0.69	0.67	0.69	0.63	0.67	0.66	0.68
2018	0.65	0.69	0.65	0.69	0.57	0.51	0.66	0.69
2019	0.66	0.69	0.66	0.69	0.26	-	0.65	0.69
Max	0.67	0.71	0.67	0.71	0.64	0.91	0.66	0.69
Min	0.65	0.67	0.65	0.67	0.26	0.14	0.56	0.65
Median	0.66	0.70	0.66	0.70	0.57	0.57	0.65	0.68

Notes: New KS1 and KS2 curricula were introduced in 2014; KS1 and KS2 SATs changed from 2016. Cells with a '-' sign indicate that no data is available for those years.

School-level correlation

Year	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	Maths	English	Maths	English	Maths	English	Maths
2012	0.59	0.55	0.59	0.55	-	-	-	-
2013	0.62	0.53	0.62	0.52	0.55	0.50	0.54	0.44
2014	0.62	0.55	0.61	0.54	0.40	0.36	0.56	0.67
2015	0.62	0.52	0.61	0.52	0.64	0.52	0.47	0.47
2016	0.60	0.45	0.59	0.44	0.60	0.34	0.59	0.45
2017	0.60	0.47	0.60	0.46	0.61	0.55	0.62	0.54
2018	0.57	0.46	0.57	0.46	0.38	0.65	0.63	0.55
2019	0.58	0.48	0.58	0.48	0.32	-	0.58	0.49
Max	0.62	0.54	0.62	0.55	0.64	0.65	0.63	0.67
Min	0.57	0.45	0.57	0.44	0.32	0.34	0.47	0.44
Median	0.60	0.50	0.60	0.50	0.55	0.51	0.58	0.49

Notes: New KS1 and KS2 curricula were introduced in 2014; KS1 and KS2 SATs changed from 2016. Cells with a '-' sign indicate that no data is available for those years.

Table 19: **Correlations between Key Stage 2 and Key Stage 4 (KS2-KS4)**

Pupil-level correlations

Year	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	Maths	English	Maths	English	Maths	English	Maths
2012	0.64	0.70	0.64	0.69	-	-	-	-
2013	0.57	0.65	0.58	0.63	0.69	0.65	0.56	0.60
2014	0.56	0.66	0.57	0.63	0.68	0.78	0.25	0.26
2015	0.17	0.21	0.16	0.18	-	-	-	-
2016	0.52	0.63	0.56	0.61	0.56	0.68	0.31	0.32
2017	0.57	0.66	0.58	0.67	0.60	0.72	0.26	0.33
2018	0.63	0.66	0.63	0.66	-	-	-	-
2019	0.63	0.67	0.64	0.68	-	-	-	-
Max	0.64	0.70	0.64	0.69	0.69	0.78	0.56	0.60
Min	0.17	0.21	0.16	0.18	0.56	0.65	0.25	0.26
Median	0.57	0.66	0.58	0.65	0.64	0.70	0.29	0.33

Notes: A new KS2 curriculum was introduced in 2014; KS2 SATs change from 2016. A new 0 to 9 scale was used in KS4 (GCSE English and maths) from 2017. Cells with a '-' sign indicate that no data is available for those years.

School-level correlations

Year	NPD whole		NPD sample		EEF studies		EEF NPD	
	English	Maths	English	Maths	English	Maths	English	Maths
2012	0.78	0.77	0.71	0.74	-	-	-	-
2013	0.73	0.77	0.65	0.73	0.67	0.69	0.73	0.63
2014	0.57	0.71	0.61	0.74	0.72	0.86	0.44	0.39
2015	0.11	0.18	0.19	0.18	-	-	-	-
2016	0.52	0.65	0.71	0.69	0.67	0.65	0.11	0.11
2017	0.69	0.64	0.74	0.76	0.90	0.88	0.84	0.82
2018	0.76	0.68	0.78	0.66	-	-	-	-
2019	0.77	0.71	0.75	0.73	-	-	-	-
Max	0.78	0.77	0.78	0.76	0.90	0.88	0.84	0.82
Min	0.11	0.18	0.19	0.18	0.67	0.65	0.11	0.11
Median	0.71	0.70	0.71	0.73	0.70	0.78	0.59	0.51

Notes: A new KS2 curriculum was introduced in 2014; KS2 SATs change from 2016. A new 0 to 9 scale was used in KS4 (GCSE English and maths) from 2017. Cells with a '-' sign indicate that no data is available for those years.

Reflections on correlation analyses

NPD (whole and sample)

There is reasonable consistency between estimates from the complete and sampled NPD datasets, most evidently for the correlations between outcomes at KS1 and KS2. Between EYFS and KS1 (Table 17), correlations for English outcomes at the pupil level were slightly (but consistently) higher (in the

range of 0.43 to 0.61) compared to maths outcomes (in the range of 0.11 to 0.57) for most time points. Correlations between EYFS and KS1 increased over time (most clearly for English outcomes).

Between KS1 and KS2 (Table 18), correlations for maths outcomes (in the range of 0.67 to 0.71) were reasonably similar to those of English outcomes (in the range of 0.65 to 0.67). There was also less variability in the correlation estimates for KS1-KS2.

Between KS2 and KS4 (Table 19), correlations for English (in the range of 0.17 to 0.64) and maths outcomes (in the range of 0.21 to 0.70) were much more similar than between KS1 and KS2 or EYFS and KS1. Correlations between KS1 and KS2 were relatively stable over time. Correlations between KS2 and KS4 showed some minor fluctuations but were overall quite stable, except for the 2015 estimates, which were much lower than in other years. Interestingly, correlations at pupil and school levels were reasonably similar for both English and maths in all Key Stages.

EEF studies

Estimates for correlations between EYFS and KS1 were notably higher as compared to the NPD data, but similar for both English (in the range of 0.54 to 0.88) and maths (in the range of 0.50 to 0.88). Correlation estimates for EYFS and KS1 were much higher for earlier years compared to later years, with the largest estimates being for 2013. Although these estimates initially seem to start high and fall through time, the estimates for English in 2019 do not fit this pattern.

The correlations between KS1 and KS2 scores were similar for both maths and English (mostly in the range of 0.50 to 0.70), except for 2014 and 2016 where we see substantial differences (for maths, the correlation for 2014 was 0.91 and for 2016 was 0.14). Estimates for KS1 and KS2 imply a lot of variation for pupil-level correlations, particularly for maths; although school-level correlations also vary, they seem slightly more consistent.

Correlations were, on average, higher for KS2 and KS4 compared to other Key Stage pairs, but mostly in the range of 0.56-0.78. Estimates for KS2 and KS4 were stable through time.

Compared to EEF studies in general, estimates for EEF studies for EYFS and KS1 with equivalent NPD data were more variable. The same variation was observed for English and maths EYFS and KS1 time-specific EEF NPD estimates. On the other hand, KS1 and KS2 correlations were much more consistent over time. Estimates for KS2 and KS4 were generally lower, with the discrepancy between time-specific EEF studies and EEF NPD estimates persisting even more.

Comparison between NPD sample and EEF studies

In general, estimates for EYFS and KS1 were higher for EEF studies than for the NPD sample in most years. There was slightly less difference between these correlation estimates for KS1 and KS2. The correlation estimates were, on average, higher for KS2 and KS4 compared to EYFS-KS1 and KS1-KS2 for both the NPD sample and EEF studies. However, the correlation between KS2 and KS4 scores from the NPD sample matched that of EEF studies. Overall, it is evident from the results that the correlation estimates for the NPD and EEF samples converged better for KS2 and KS4.

FSM subgroup analysis

Table 20 provides correlation estimates for subsequent Key Stage scores for FSM-eligible pupils using NPD and EEF Archive data. Since there was no significant difference between the correlation estimates obtained from the NPD whole and NPD samples for the overall analysis, we reported NPD sample estimates only for the FSM subgroup analysis.

For FSM-eligible pupils, the correlation estimates between KS1 and KS2 obtained from the NPD sample are very similar to those obtained from the EEF studies sample for most years. A similar pattern was

observed for the KS2 and KS4 correlation estimates. There are larger differences for EYFS and KS1, similar to what has been observed for data on all pupils.

Table 20: **Correlation between Key Stages (EYFS-KS1, KS1-KS2, KS2-KS4) for English and maths, FSM-eligible pupils**

	Year	EYFS-KS1 English	EYFS-KS1 Maths	KS1-KS2 English	KS1-KS2 Maths	KS2-KS4 English	KS2 -KS4 Maths
NPD sample	2012	0.40	0.30	0.60	0.67	0.60	0.65
	2013	0.37	0.24	0.61	0.66	0.55	0.60
	2014	0.39	0.29	0.61	0.67	0.54	0.57
	2015	0.37	0.24	0.61	0.66	0.14	0.14
	2016	0.36	0.31	0.61	0.64	0.52	0.56
	2017	0.38	0.34	0.63	0.65	0.52	0.59
	2018	0.40	0.37	0.62	0.66	0.59	0.59
	2019	0.41	0.37	0.64	0.66	0.59	0.61
	Max	0.41	0.37	0.64	0.67	0.60	0.65
	Min	0.36	0.24	0.60	0.64	0.14	0.14
	Median	0.39	0.31	0.61	0.66	0.55	0.59
EEF studies	2013	0.31	0.22	-	-	-	-
	2014	0.78	0.73	0.30	0.52	0.61	0.74
	2015	0.38	-	0.60	0.59	-	-
	2016	-	-	0.58	0.12	0.49	0.64
	2017	0.59	0.49	0.60	0.65	0.49	0.64
	2018	-	0.56	-	0.51	-	-
	2019	0.74	-	0.32	-	-	-
	Max	0.78	0.73	0.60	0.65	0.61	0.74
	Min	0.31	0.22	0.60	0.12	0.49	0.64
	Median	0.59	0.53	0.58	0.52	0.49	0.64

Notes: New KS1 and KS2 curricula were introduced in 2014; KS1 and KS2 SATs changed from 2016. A new 0 to 9 scale was used in KS4 (GCSE English and maths) from 2017. Cells with a '-' sign indicate that no data was available for those years.

Explanatory power

The KS1 to KS4 estimates of explanatory power were obtained from the conditional models that included a pre-test (1, 7 and 11, see Table 4) using the NPD whole, NPD sample and EEF studies datasets. These estimates were compared for the most recent three years in the analyses (2017 to 2019; details on other years are available in the Excel files). Table 21 and Table 22 provide estimates for the explanatory power for the NPD whole and NPD sample datasets, while Table 23 provides the same information for the EEF studies dataset.

Table 21: **Comparing explanatory power estimates for pre-test; whole NPD analysis**

Year	From conditional models								
	Between-school (R_C^2)			Within-schools (Residual R_R^2)			Total (R_T^2)		
	M1	M7	M11	M1	M7	M11	M1	M7	M11
KS1 English									
2017	0.04	0.11	0.11	0.39	0.39	0.39	0.37	0.38	0.38
2018	0.11	0.16	0.16	0.39	0.39	0.39	0.38	0.38	0.38
2019	0.13	0.17	0.17	0.39	0.39	0.39	0.38	0.38	0.38
KS1 Maths									
2017	-0.15	0.00	0.00	0.33	0.33	0.33	0.31	0.32	0.32
2018	-0.10	0.06	0.06	0.34	0.34	0.34	0.32	0.33	0.33
2019	-0.04	0.09	0.09	0.35	0.35	0.35	0.33	0.34	0.34
KS2 English									
2017	0.39	0.39	0.39	0.47	0.47	0.47	0.46	0.46	0.46
2018	0.35	0.35	0.35	0.45	0.45	0.45	0.44	0.44	0.44
2019	0.33	0.33	0.33	0.47	0.47	0.47	0.46	0.46	0.46
KS2 Maths									
2017	0.12	0.13	0.13	0.52	0.52	0.52	0.47	0.47	0.47
2018	0.12	0.14	0.14	0.52	0.52	0.52	0.47	0.48	0.48
2019	0.15	0.15	0.15	0.52	0.52	0.52	0.48	0.48	0.48
KS4 English									
2017	0.44	0.56	0.56	0.31	0.31	0.31	0.33	0.34	0.34
2018	0.54	0.64	0.64	0.38	0.38	0.38	0.40	0.41	0.41
2019	0.53	0.62	0.62	0.39	0.39	0.39	0.40	0.41	0.41
KS4 Maths									
2017	0.44	0.51	0.51	0.44	0.44	0.44	0.44	0.45	0.45
2018	0.41	0.45	0.45	0.44	0.44	0.44	0.44	0.44	0.44
2019	0.44	0.50	0.50	0.45	0.45	0.45	0.45	0.46	0.46

Notes on the negative between-school explanatory power: For KS1 maths, very low school-level variance was observed to increase when a pre-test is added. The largest absolute negative R_C^2 is in 2017 for M1 (-0.15,), While the unconditional school-level variance was 0.02 (unconditional ICC of 0.05), the conditional model with pupil-level raw pre-test score(M1) had a school-level variance of 0.02 (conditional ICC ofof0.08). The overall explanatory power is estimated as 0.31.

Table 22: Comparing explanatory power estimates for pre-test; NPD sample analysis

Year	From conditional models								
	Between-school (R_C^2)			Within-schools (Residual R_R^2)			Total (R_T^2)		
	M1	M7	M11	M1	M7	M11	M1	M7	M11
KS1 English									
2017	-0.35	-0.24	-0.24	0.24	0.24	0.24	0.21	0.22	0.22
2018	-0.26	-0.17	-0.17	0.24	0.24	0.24	0.22	0.22	0.22
2019	-0.16	-0.13	-0.13	0.24	0.24	0.24	0.23	0.23	0.23
KS1 Maths									
2017	-0.40	-0.21	-0.21	0.19	0.19	0.19	0.17	0.18	0.18
2018	-0.46	-0.26	-0.26	0.20	0.20	0.20	0.18	0.19	0.19
2019	-0.59	-0.36	-0.36	0.21	0.21	0.21	0.19	0.20	0.20
KS2 English									
2017	0.38	0.38	0.38	0.47	0.47	0.47	0.46	0.46	0.46
2018	0.35	0.35	0.35	0.45	0.45	0.45	0.44	0.44	0.44

2019	0.33	0.33	0.33	0.47	0.47	0.47	0.46	0.46	0.46
KS2 Maths									
2017	0.11	0.13	0.13	0.52	0.52	0.52	0.47	0.47	0.47
2018	0.12	0.14	0.14	0.52	0.52	0.52	0.47	0.48	0.48
2019	0.13	0.14	0.14	0.52	0.52	0.52	0.47	0.48	0.48
KS4 English									
2017	0.51	0.70	0.70	0.32	0.32	0.32	0.34	0.36	0.36
2018	0.57	0.65	0.65	0.38	0.38	0.38	0.40	0.41	0.41
2019	0.59	0.69	0.69	0.39	0.39	0.39	0.41	0.42	0.42
KS4 Maths									
2017	0.44	0.44	0.44	0.47	0.58	0.58	0.45	0.46	0.46
2018	0.44	0.44	0.44	0.40	0.43	0.43	0.43	0.44	0.44
2019	0.46	0.46	0.46	0.46	0.52	0.52	0.46	0.47	0.47

Notes on the negative between-school explanatory power: For KS1 English, very low school-level variance was observed to increase when a pre-test is added. The largest negative absolute R_C^2 is in 2017 for M1 (-0.35.). While the unconditional school-level variance was 0.02 (conditional ICC of 0.05), the conditional model with pupil-level raw pre-test score (M1) had a school-level variance of 0.02 (conditional ICC of 0.07). Overall explanatory power is estimated as 0.21. For KS1 maths, very low school-level variance is observed to increase when a pre-test is added. The largest negative R^2 is in 2019 for M1 (-0.59.). While the unconditional school-level variance was 0.004 (unconditional ICC of 0.02), the conditional model with pupil-level raw pre-test score (M1) had a school-level variance of 0.01 (conditional ICC of 0.04). Overall explanatory power is estimated as 0.19.

Table 23: Comparing estimates of explanatory power for pre-test; EEF studies analysis

Year	From conditional models								
	Between-school (R_C^2)			Within-schools (Residual R_R^2)			Total (R_T^2)		
	M1	M7	M11	M1	M7	M11	M1	M7	M11
KS1 English									
2017	0.22	0.23	0.23	0.42	0.42	0.42	0.40	0.40	0.40
2018	-	-	-	-	-	-	-	-	-
2019	0.85	0.85	0.85	0.49	0.50	0.50	0.56	0.56	0.56
KS1 Maths									
2017	-0.07	0.00	0.00	0.28	0.28	0.28	0.24	0.25	0.25
2018	0.52	0.57	0.57	0.24	0.24	0.24	0.29	0.30	0.30
2019	-	-	-	-	-	-	-	-	-
KS2 English									
2017	0.37	0.38	0.38	0.40	0.40	0.40	0.40	0.40	0.40
2018	0.16	0.17	0.17	0.36	0.36	0.36	0.32	0.32	0.32
2019	-	-	-	-	-	-	-	-	-
KS2 Maths									
2017	0.19	0.23	0.23	0.49	0.49	0.49	0.45	0.46	0.46
2018	0.46	0.46	0.46	0.24	0.24	0.24	0.26	0.26	0.26
2019	-	-	-	-	-	-	-	-	-
KS4 English									
2017	0.64	0.79	0.79	0.37	0.37	0.37	0.40	0.41	0.41
2018	-	-	-	-	-	-	-	-	-
2019	-	-	-	-	-	-	-	-	-
KS4 Maths									
2017	0.64	0.73	0.73	0.53	0.53	0.53	0.54	0.54	0.54
2018	-	-	-	-	-	-	-	-	-
2019	-	-	-	-	-	-	-	-	-

Notes on the negative between-school explanatory power: For KS1 maths, very low school-level variance was observed to increase when a pre-test is added. The largest absolute negative R_C^2 is in 2017 for M1 (-0.15,), While the unconditional school-level variance was 0.02 (unconditional ICC of 0.05), the conditional model with pupil-level raw pre-test score(M1) had a school-level variance of 0.02 (conditional ICC ofof0.08). The overall explanatory power is estimated as 0.31.

This subsection provides reflections on the explanatory power analyses for 2017 to 2019 (Table 21 to Table 23) and its synthesis with ICC (Table 9 to Table 16) and correlation analyses (Table 17 to Table 19) in the previous subsections. The estimates shown in this synthesis serve as the basis for the applied example in the following section.

NPD (whole and sample)

KS1 English

- According to Table 8, a very small proportion of variance in KS1 English was observed between-schools (unconditional ICC between 0.03 and 0.05).
- At the pupil level, the bivariate correlation between KS1 and EYFS English was observed as 0.61 (squared correlation estimates=0.37) for the NPD whole sample and ranging between 0.47 and 0.48 (squared correlation estimates=0.22 to 0.23) for the NPD sample.
- At the school level, the bivariate correlation between KS1 and EYFS English was observed as ranging between 0.48 and 0.49 (squared correlation estimates=0.23 to 0.24) for the NPD whole sample and between 0.33 and 0.34 (squared correlation estimates=0.11 to 0.12) for the NPD sample.
- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.05 and 0.07.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.11 and 0.17 for the NPD whole sample and was negative (-0.24 to -0.13) for the NPD sample.
  - The total explanatory power was observed as (R_T^2 =) 0.38 for the NPD whole and between 0.22 and 0.23 for the NPD sample.
  - The explanatory power for the residual variance (within-schools or between-pupils) was observed as (R_R^2 =) 0.39 for the NPD whole sample and 0.24 for the NPD sample.
- The inclusion of pre-tests at both pupil and school levels resulted in greater explanatory power at the school level, but not at the residual level. This is most clearly seen with the NPD whole sample: R_C^2 = 0.11 to 0.17, compared with R_C^2 = 0.04 to 0.13 when only including a pre-test at the pupil level. This can also be seen with the NPD sample: R_C^2 = -0.24 to -0.13 when including pre-tests at both the pupil and school levels, compared with R_C^2 = -0.35 to -0.16 when only including a pre-test at the pupil level.

Drawing on the above synthesis of NPD whole and sample analyses, the following assumptions have been used for the applied example calculation of MDES for KS1 English presented below:

- School-level ICC = 0.05
- R_C^2 = 0.16
- R_R^2 = 0.39
- Inclusion of pre-test covariates at both pupil and school levels

KS1 maths

- A very small proportion of variance in KS1 maths was observed between schools (unconditional ICC between 0.02 and 0.05).
- At the pupil level, the bivariate correlation between KS1 and EYFS maths was observed as 0.56 to 0.57 (squared correlation estimates =0.31 to 0.33) for the NPD whole sample and ranging between 0.42 and 0.44 (squared correlation estimates =0.17 to 0.20) for the NPD sample.

- At the school level, the bivariate correlation between KS1 and EYFS maths was observed as ranging between 0.38 and 0.42 (squared correlation estimates =0.15 to 0.18) for the NPD whole sample and between 0.25 and 0.37 (squared correlation estimates =0.06 to 0.14) for the NPD sample.
- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.04 and 0.07.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.00 and 0.09 for the NPD whole sample and was negative for the NPD sample, ranging between -0.36 and -0.21.
  - The total explanatory power was observed as (R_T^2 =) 0.32 to 0.34 for the NPD whole sample and between 0.18 and 0.20 for the NPD sample.
  - The explanatory power for the residual variance (within-schools, between-pupils) was observed as (R_R^2 =) 0.33 to 0.35 for the NPD whole sample and between 0.19 and 0.21 for the NPD sample.
- The inclusion of pre-tests at both pupil and school levels was observed to result in greater explanatory power at the school level, but not at the residual level. This is most clearly seen with the NPD whole sample: R_C^2 = 0.00 to 0.09, compared with negative R_C^2 = -0.15 to -0.04 when only including a pre-test at the pupil level. This can also be seen with the NPD sample: R_C^2 = -0.36 to -0.21 when including pre-tests at both the pupil and school levels compared with R_C^2 = -0.59 to -0.40 when only including a pre-test at the pupil level.

Drawing on the above synthesis of NPD whole and sample analyses, the following estimates have been used on for the applied example calculating the MDES for KS1 maths presented below:

- School-level ICC = 0.05
- R_C^2 = 0.05
- R_R^2 = 0.33
- Inclusion of pre-test covariates at both pupil and school levels

KS2 English

- A small proportion of variance in KS2 English was observed to be between-schools (unconditional ICC between 0.09 and 0.12).
- At the pupil level, the bivariate correlation between KS2 and KS1 English was observed as 0.65 to 0.67 (squared correlation estimates =0.42 to 0.45) for both the NPD whole and NPD sample.
- At the school level, the bivariate correlation between KS2 and KS1 English was observed as ranging between 0.57 and 0.60 (squared correlation estimates =0.32 to 0.37) for the NPD whole and the NPD sample.
- From the conditional models that included pre-test covariates at both pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.11 and 0.13.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.33 and 0.39 for the NPD whole and the NPD sample.
  - The total explanatory power was observed as (R_T^2 =) 0.44 to 0.46 for the NPD whole and the NPD sample.
  - The explanatory power for the residual variance (within-schools, between-pupils) was observed as (R_R^2 =) 0.45 to 0.47 for the NPD whole and the NPD sample.
- The inclusion of pre-tests at both pupil and school levels was observed to result in no explanatory power gains at the school or residual level for the NPD whole and NPD sample:R_C^2 = 0.33 to 0.39 when including pre-tests at both the pupil and school levels compared with R_C^2 = 0.33 to 0.39 when only including a pre-test at the pupil level.

Drawing from the above synthesis of NPD whole and sample analyses, the following estimates have been used for the applied example calculating the MDES for KS2 English presented below:

- School-level ICC = 0.12

- R_C^2 = 0.35
- R_R^2 = 0.46
- Inclusion of pre-test covariates at both pupil and school levels OR just at the pupil level

KS2 maths

- A small proportion of variance in KS2 maths was observed to be between schools (unconditional ICC between 0.11 and 0.13).
- At the pupil level, the bivariate correlation between KS2 and KS1 maths was observed as 0.69 (squared correlation estimates =0.47) for the NPD whole and the NPD sample.
- At the school level, the bivariate correlation between KS2 and KS1 maths was observed as ranging between 0.46 and 0.48 (squared correlation estimates =0.21 to 0.23) for the NPD whole and the NPD sample.
- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.18 and 0.21.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.13 and 0.15 for the NPD whole and the NPD sample.
  - The total explanatory power was observed as (R_T^2 =) 0.47 to 0.48 for the NPD whole and the NPD sample.
  - The explanatory power for the residual variance (within-schools, between-pupils) was observed as (R_R^2 =) 0.52 for the NPD whole and the NPD sample.
- The inclusion of pre-tests at both pupil and school levels was observed to result in little gains in explanatory power at the school level and no gains at the residual level. This was observed for the NPD whole and NPD sample: R_C^2 = 0.13 to 0.15 when including pre-tests at both the pupil and school levels compared with R_C^2 = 0.11 to 0.15 only including a pre-test at the pupil level.

Drawing from the above synthesis of NPD whole and sample analyses, the following estimates have been used for the applied example calculating the MDES for KS2 maths presented below:

- School-level ICC = 0.13
- R_C^2 = 0.14
- R_R^2 = 0.52
- Inclusion of pre-test covariates at both pupil and school levels OR just at the pupil level

KS4 English

- A small proportion of variance in KS4 English was observed to be between-schools (unconditional ICC between 0.10 and 0.11).
- At the pupil level, the bivariate correlation between KS4 and KS2 English was observed as ranging between 0.57 and 0.64 (squared correlation estimates =0.32 to 0.41) for both the NPD whole and NPD sample.
- At the school level, the bivariate correlation between KS4 and KS2 English was observed as ranging between 0.69 and 0.78 (squared correlation estimates =0.48 to 0.61) for the NPD whole and the NPD sample.
- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.05 and 0.07.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.56 and 0.70 for the NPD whole and the NPD sample.
  - The total explanatory power was observed as (R_T^2 =) 0.34 to 0.42 for the NPD whole and the NPD sample.
  - The explanatory power for the residual variance (within-schools, between-pupils) was observed as (R_R^2 =) 0.31 to 0.39 for the NPD whole and the NPD sample.

- The inclusion of pre-tests at both pupil and school levels was observed to result in greater explanatory power at the school level, but not at the residual level. This was observed for the NPD whole and NPD sample: R_C^2 = 0.56 to 0.70 when including pre-tests at both the pupil and school levels compared with R_C^2 = 0.44 to 0.59 when only including a pre-test at the pupil level.

Drawing from the above synthesis of NPD whole and sample analyses, the following estimates have been used for the applied example calculation of MDES estimates for KS4 English presented below:

- School-level ICC = 0.11
- R_C^2 = 0.64
- R_R^2 = 0.38
- Inclusion of pre-test covariates at both pupil and school levels

KS4 maths

- A small proportion of variance in KS4 maths was observed to be between-schools (unconditional ICC = 0.10).
- At the pupil level, the bivariate correlation between KS4 and KS2 maths was observed as ranging between 0.66 and 0.68 (squared correlation estimates =0.43 to 0.46) for both the NPD whole and NPD sample.
- At the school level, the bivariate correlation between KS4 and KS2 maths ranged from 0.64 to 0.76 (squared correlation estimates =0.42 to 0.58) for the NPD whole and the NPD sample.
- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.08 and 0.10.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.43 and 0.58 for the NPD whole and the NPD sample.
  - The total explanatory power was observed as (R_T^2 =) 0.44 to 0.47 for the NPD whole and the NPD sample.
  - The explanatory power for the residual variance (within-schools, between-pupils) was observed as (R_R^2 =) 0.44 to 0.46 for the NPD whole and the NPD sample.
- The inclusion of pre-tests at both pupil and school levels was observed to result in slightly greater explanatory power at the school level, but not at the residual level. This was observed for the NPD whole and NPD sample: R_C^2 = 0.43 to 0.58 when including pre-tests at both the pupil and school levels compared with R_C^2 = 0.40 to 0.47 when only including a pre-test at the pupil level.

Drawing on the above synthesis of NPD whole and sample analyses, the following estimates have been used for the applied example calculating the MDES estimates for KS4 maths presented below:

- School-level ICC = 0.10
- R_C^2 = 0.51
- R_R^2 = 0.45
- Inclusion of pre-test covariates at both pupil and school levels

EEF studies

In line with the NPD analyses, a synthesis of the ICC, correlation and explanatory power is also provided for the EEF studies dataset. However, the applied MDES example in the section further below, only leverages parameters estimated using the NPD datasets. Similar procedures can be followed to obtain relevant MDES estimates using parameters derived from EEF studies.

KS1 English

- A small proportion of variance in KS1 English was observed to be between-schools and much higher than that observed in NPD datasets (unconditional ICC between 0.08 and 0.17).

- At the pupil level, the bivariate correlation between KS1 and EYFS English was observed as ranging between 0.63 and 0.75 (squared correlation estimates =0.40 to 0.56).
- At the school level, the bivariate correlation between KS1 and EYFS English was observed as ranging between 0.46 and 0.81 (squared correlation estimates =0.21 to 0.66).
- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.06 and 0.10.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.23 and 0.85.
  - The total explanatory power ranged between (R_T^2 =) 0.40 and 0.56.
  - The explanatory power for the residual variance (within-schools, between-pupils) ranged between (R_R^2 =) 0.42 and 0.50.
- Considering the average estimates for explanatory power, the inclusion of pre-tests at both pupil and school levels was observed to result in greater explanatory power at the school level, but not at the residual level.

Drawing from the above synthesis of EEF studies, the following estimates can be used for calculating the MDES for KS1 English:

- School-level ICC = 0.13
- R_C^2 = 0.54
- R_R^2 = 0.46
- Inclusion of pre-test covariates at both pupil and school levels

KS1 maths

- A small proportion of variance in KS1 maths was observed to be between-schools (unconditional ICC between 0.12 to 0.20).
- At the pupil level, the bivariate correlation between KS1 and EYFS maths was observed as ranging between 0.50 and 0.54 (squared correlation estimates =0.25 to 0.29).
- At the school level, the bivariate correlation between KS1 and EYFS maths was observed as ranging between 0.27 and 0.68 (squared correlation estimates =0.07 to 0.46).
- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.12 and 0.17.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.00 and 0.57.
  - The total explanatory power ranged between (R_T^2 =) 0.25 and 0.30.
  - The explanatory power for the residual variance (within-schools, between-pupils) ranged between (R_R^2 =) 0.24 and 0.28.
- The inclusion of pre-tests at both pupil and school levels was observed to result in greater explanatory power at the school level, but not at the residual level.

Drawing from the above synthesis of EEF studies, the following estimates can be used for calculating the MDES for KS1 maths:

- School-level ICC = 0.16
- R_C^2 = 0.29
- R_R^2 = 0.26
- Inclusion of pre-test covariates at both pupil and school levels

KS2 English

- Unconditional ICC varied between 0.03 and 0.17 for KS2 English.
- At the pupil level, the bivariate correlation between KS2 and KS1 English was observed as ranging between 0.26 and 0.63 (squared correlation estimates =0.07 to 0.40).

- At the school level, the bivariate correlation between KS2 and KS1 English was observed as ranging between 0.32 and 0.61 (squared correlation estimates =0.10 to 0.37).
- From the conditional models that included pre-test covariates at both pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.10 and 0.21.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.17 and 0.38.
  - The total explanatory power ranged between (R_T^2 =) 0.32 and 0.40.
  - The explanatory power for the residual variance (within-schools, between-pupils) ranged between (R_R^2 =) 0.36 and 0.40.
- The inclusion of pre-tests at both pupil and school levels was observed to result in no explanatory power gains at the school or residual level compared to the inclusion of pre-test at the pupil level only.

Drawing from the above synthesis of EEF studies, the following estimates can be used for calculating the MDES for KS2 English:

- School-level ICC = 0.10
- R_C^2 = 0.28
- R_R^2 = 0.38
- Inclusion of pre-test covariates at both pupil and school levels OR just at the pupil level

KS2 maths

- A small proportion of variance in KS2 maths was observed to be between schools and much higher than that observed in NPD datasets (unconditional ICC between 0.09 and 0.12).
- At the pupil level, the bivariate correlation between KS2 and KS1 maths was observed as ranging between 0.51 and 0.67 (squared correlation estimates =0.26 to 0.45).
- At the school level, the bivariate correlation between KS2 and KS1 maths was observed as ranging between 0.55 and 0.65 (squared correlation estimates =0.30 to 0.42).
- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimates ranged between 0.06 and 0.18.
  - At the school level, explanatory power ranged between (R_C^2 =) 0.23 and 0.46.
  - The total explanatory power ranged between (R_T^2 =) 0.26 and 0.46.
  - The explanatory power for the residual variance (within-schools, between-pupils) ranged between (R_R^2 =) 0.24 and 0.49.
- The inclusion of pre-tests at both pupil and school levels was observed to result in little gains in explanatory power at the school level and none at the residual level.

Drawing from the above synthesis of EEF studies, the following estimates can be used for calculating the MDES for KS2 maths:

- School-level ICC = 0.11
- R_C^2 = 0.35
- R_R^2 = 0.37
- Inclusion of pre-test covariates at both pupil and school levels OR just at the pupil level

KS4 English

- A small proportion of variance in KS4 English was observed to be between schools and much higher than that observed in NPD datasets (unconditional ICC 0.09).
- At the pupil level, the bivariate correlation between KS4 and KS2 English was observed as 0.60 (squared correlation estimates =0.36).
- At the school level, the bivariate correlation between KS4 and KS2 English was observed as 0.90 (squared correlation estimates =0.81).

- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimate was 0.03.
  - At the school level, explanatory power was observed as (R_C^2 =) 0.79.
  - The total explanatory power was observed as (R_T^2 =) 0.41.
  - The explanatory power for the residual variance (within-schools, between-pupils) was observed as (R_R^2 =) 0.37.
- The inclusion of pre-tests at both pupil and school levels was observed to result in greater explanatory power at the school level, but not at the residual level. This was R_C^2 = 0.56 to 0.70 when including pre-tests at both the pupil and school levels compared with R_C^2 = 0.44 to 0.59 when only including a pre-test at the pupil level.

Drawing from the above synthesis of EEF studies, the following estimates can be used for calculation the MDES for KS4 English:

- School-level ICC = 0.09
- R_C^2 = 0.79
- R_R^2 = 0.37
- Inclusion of pre-test covariates at both pupil and school levels

KS4 maths

- A small proportion of variance in KS4 maths was observed to be between schools and much higher than that observed in NPD datasets (Unconditional ICC 0.09).
- At the pupil level, the bivariate correlation between KS4 and KS2 maths was observed as 0.72 (squared correlation estimates =0.52).
- At the school level, the bivariate correlation between KS4 and KS2 maths was observed as 0.88 (squared correlation estimates =0.77).
- From the conditional models that included pre-test covariates at pupil and school levels (M7):
  - Conditional ICC estimate was 0.05.
  - At the school level, explanatory power was observed as (R_C^2 =) 0.73.
  - The total explanatory power was observed as (R_T^2 =) 0.54.
  - The explanatory power for the residual variance (within-schools, between-pupils) was observed as (R_R^2 =) 0.53.
- The inclusion of pre-tests at both pupil and school levels was observed to result in greater explanatory power at the school level, but not at the residual level. This was clearly observed with the explanatory power comparison across the three models (M1, M7, M11) for school and pupil levels.

Drawing from the above synthesis of EEF studies, the following estimates can be used for calculating the MDES for KS4 maths:

- School-level ICC = 0.09
- R_C^2 = 0.73
- R_R^2 = 0.53
- Inclusion of pre-test covariates at both pupil and school levels

Overall, this analysis from NPD and EEF datasets showed that the pre-test stands out as the most significant covariate, accounting for a substantial portion of the school- or pupil-level variation for maths and English outcomes across different Key Stages. However, it is important to note that the extent of their explanatory power may vary for different Key Stages.

There were a few instances where negative explanatory power was observed. Hox et al. (2017) says that using these formulas may lead to a conclusion that the specific explanatory variable has a negative contribution to the explained variance. This will lead to a negative R_C^2 *or* R_R^2, which is an impossible value.

This always happen when a predictor variable with lowest-level variation, such as a group mean centred predictor, or a measurement occasion in a longitudinal model with fixed occasions, is added to the model. The reason is that the decomposition of the total variance into the first level and second-level variance in the empty model assumes random sampling at each level, and a variable with only lowest-level variance violates that assumption.

Objective III: Using ICC and pre-test explanatory power estimates for the design of 2-level CRTs in educational settings (applied example)

The following section draws on estimates from the NPD analyses for the most recent three academic years (2017, 2018 and 2019).

The MDES calculations are obtained using the Bloom et al. (2007) formula. More details for the parameters and the equation used are provided in the Methods section above.

For all the MDES estimates, the following values are assumed for all Key Stage outcomes:

- P = 0.50 (half of schools randomly allocated to intervention and control groups).
- m = 2 cluster-level covariates (group membership and school-level pre-test).
- The number of schools (J) ~ allowed to vary between 50 and 150.
- The number of pupils per school (n) ~ allowed to vary between 5 and 30.

EYFS English

- Unconditional ICC ~ 0.05 to 0.06 [fixed at 0.06]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = zero
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) = zero

Our analyses provide estimates for the unconditional ICC, but not the explanatory power, because there are no pre-tests available from the NPD prior to EYFS at the end of Y0. Where available, trials that collect a commercial baseline pre-test whilst using EYFS as a test outcome might be used to provide explanatory power details. This means that for the purpose of the MDES estimates below, explanatory power has been set at zero (i.e., this is an outcome only 2-level CRT design).

Table 24: MDES estimates by number of schools and pupils, EYFS English

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per school	5	0.40	0.32	0.28	0.23
	10	0.32	0.25	0.22	0.18
	20	0.26	0.21	0.19	0.15
	30	0.24	0.19	0.17	0.14

Notes: Shaded cells indicate MDES of 0.20 or less.

EYFS maths

- Unconditional ICC ~ 0.07 to 0.08 [fixed at 0.08]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = zero
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) = zero

As with EYFS English, the power analysis for maths draws on estimates for the unconditional ICC but not on the explanatory power. Therefore, this is also an outcome only 2-level CRT design.

Table 25: **MDES estimates by number of schools and pupils, EYFS maths**

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per school	5	0.42	0.33	0.29	0.24
	10	0.34	0.26	0.23	0.19
	20	0.29	0.23	0.20	0.16
	30	0.27	0.21	0.19	0.15

Notes: Shaded cells indicate MDES of 0.20 or less.

For EYFS English, with an outcome-only design, detecting a MDES of 0.20 requires 80+ schools with 30+ pupils per school; 100+ schools with 20+ pupils per school; or 150+ schools with 10+ pupils per school. Whilst to detect a MDES of 0.10 requires 250+ schools with 50+ pupils per school; or 300+ schools with 30+ pupils per school.

For EYFS maths, with an outcome-only design, detecting a MDES of 0.20 requires 100+ schools with 20+ pupils per school; or 150+ schools with 10+ pupils per school. Whilst to detect a MDES of 0.10 requires 290+ schools with 50+ pupils per school; or 320+ schools with 30+ pupils per school.

KS1 English

- Unconditional ICC ~ 0.03 to 0.05 [fixed at 0.05]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.11 to 0.17 [fixed at 0.14]
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) = 0.39

Table 26: **MDES estimates by number of schools and pupils, KS1 English**

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per school	5	0.32	0.25	0.23	0.18
	10	0.26	0.20	0.18	0.15
	20	0.22	0.17	0.15	0.12
	30	0.20	0.16	0.14	0.11

Notes: Shaded cells indicate MDES of 0.20 or less.

KS1 maths

- Unconditional ICC ~ 0.02 to 0.05 [fixed at 0.05]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.00 to 0.09 [fixed at 0.05]
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) ~ 0.33 to 0.35 [fixed at 0.34]

Table 27: **MDES estimates by number of schools and pupils, KS1 maths**

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per school	5	0.34	0.26	0.24	0.19
	10	0.27	0.21	0.19	0.15
	20	0.23	0.18	0.16	0.13
	30	0.21	0.17	0.15	0.12

Notes: Shaded cells indicate MDES of 0.20 or less.

For KS1 English, detecting a MDES of 0.20 with an EYFS pre-test requires 50+ schools with 30+ pupils per school; 80+ schools with 20+ pupils per school; 100+ schools with 10+ pupils per school; or 150+ schools with 5+ pupils per school. Whilst to detect a MDES of 0.10 requires at least 180+ schools with 50+pupils per school; or 200+ schools with 30+ pupils per school.

For KS1 maths, detecting a MDES of 0.20 with an EYFS pre-test requires 80+ schools with 20+ pupils per school; 100+ schools with 10+ pupils per school; or 150+ schools with 5+ pupils per school. Whilst to detect a MDES of 0.10 requires at least 180+ schools with 50+pupils per school; or 200+ schools with 30+ pupils per school.

KS2 English

- Unconditional ICC ~ 0.09 to 0.12 [fixed at 0.12]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.33 to 0.39 [fixed at 0.35]
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) ~ 0.45 to 0.47 [fixed at 0.46]

Table 28: **MDES estimates by number of schools and pupils, KS2 English**

MDES estimates		Number of schools			
		50	80	100	150
Pupils per school	5	0.34	0.26	0.24	0.19
	10	0.29	0.22	0.20	0.16
	20	0.26	0.20	0.18	0.15
	30	0.25	0.19	0.17	0.14

Notes: Shaded cells indicate MDES of 0.20 or less.

KS2 maths

- Unconditional ICC ~ 0.11 to 0.13 [fixed at 0.13]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.13 to 0.15 [fixed at 0.14]
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) =0.52

Table 29: MDES estimates by number of schools and pupils, KS2 maths

MDES estimates		Number of schools			
		50	80	100	150
Pupils per school	5	0.36	0.28	0.25	0.20
	10	0.32	0.25	0.22	0.18
	20	0.29	0.23	0.21	0.17
	30	0.29	0.23	0.20	0.16

Notes: Shaded cells indicate MDES of 0.20 or less.

For KS2 English, detecting a MDES of 0.20 with a KS1 pre-test requires 80+ schools with 20+ pupils per school; 100+ schools with 10+ pupils per school; or 150 schools with 5+ pupils per school. Whilst to detect a MDES of 0.10 requires at least 260+ schools with 50+ pupils per school; or 270+ schools with 30+ pupils per school.

For KS2 maths, detecting a MDES of 0.20 with a KS1 pre-test requires 100+ schools with 30+ pupils per school; or 150+ schools with 5+ pupils per school. Detecting a MDES of 0.10 requires at least 350+ schools with 50+ pupils per school; or 360+ schools with 30+ pupils per school.

KS4 English

- Unconditional ICC ~ 0.10 to 0.11 [fixed at 0.11]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.56 to 0.70 [fixed at 0.63]
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) ~ 0.31 to 0.39 [fixed at 0.35]

Table 30: **MDES estimates by number of schools and pupils, KS4 English**

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per school	5	0.32	0.25	0.22	0.18
	10	0.25	0.20	0.18	0.14
	20	0.21	0.17	0.15	0.12
	30	0.20	0.16	0.14	0.11

Notes: Shaded cells indicate MDES of 0.20 or less.

KS4 maths

- Unconditional ICC ~ 0.10
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.43 to 0.58 [fixed at 0.50]
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) ~ 0.41 to 0.51 [fixed at 0.45]

Table 31: **MDES estimates by number of schools and pupils, KS4 maths**

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per school	5	0.31	0.24	0.22	0.18
	10	0.26	0.20	0.18	0.15
	20	0.22	0.17	0.15	0.13
	30	0.21	0.16	0.15	0.12

Notes: Shaded cells indicate MDES of 0.20 or less.

For KS4 English, detecting a MDES of 0.20 with a KS2 pre-test requires 50+ schools with 30+ pupils per school; 80+ schools with 20+ pupils per school; 100+ schools with 10+ pupils per school; or 150+ schools with 5+ pupils per school. Whilst to detect a MDES of 0.10 requires at least 160+ schools with 50+ pupils per school; or 180+ schools with 30+ pupils per school.

For KS4 maths detecting a MDES of 0.20 with a KS2 pre-test requires 80+ schools with 20+ pupils per school; 100+ schools with 10+ pupils per school; or 150+ schools with 5+ pupils per school. Whilst to detect a MDES of 0.10 requires at least 180+ schools with 50+ pupils per school; or 230+ schools with 30+ pupils per school.

FSM subgroup analysis

For FSM-eligible pupils, a similar analysis was conducted for KS2 and KS4 using NPD sample data.

KS2 English

- Unconditional ICC ~ 0.09 to 0.10 [fixed at 0.10]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.32 to 0.36 [fixed at 0.35]

- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) ~ 0.51 to 0.53 [fixed at 0.52]

Table 32: **MDES estimates by number of schools and pupils, KS2 English**

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per School	5	0.30	0.24	0.21	0.17
	10	0.25	0.20	0.18	0.15
	20	0.22	0.18	0.16	0.13
	30	0.21	0.17	0.15	0.12

Notes: Shaded cells indicate MDES of 0.20 or less.

KS2 maths

- Unconditional ICC ~ 0.11 to 0.14 [fixed at 0.12]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.24 to 0.28 [fixed at 0.26]
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) ~ 0.56 to 0.58 [fixed at 0.57]

Table 33: **MDES estimates by number of schools and pupils, KS2 maths**

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per School	5	0.32	0.25	0.23	0.19
	10	0.28	0.22	0.20	0.16
	20	0.26	0.21	0.18	0.15
	30	0.25	0.20	0.18	0.15

Notes: Shaded cells indicate MDES of 0.20 or less.

For KS2 English, detecting a MDES of 0.20 with a KS1 pre-test requires 80+ schools with 10+ FSM-eligible pupils per school; or 100+ schools with 5+ FSM-eligible pupils per school. However, detecting a MDES of 0.10 requires at least 200+ schools with 30+ FSM-eligible pupils per school.

For KS2 maths, detecting a MDES of 0.20 with a KS1 pre-test requires 80+ schools with 30+ FSM-eligible pupils per school; 100+ schools with 10+ FSM-eligible pupils per school; or 150+ schools with 5+ FSM-eligible pupils per school. However, detecting a MDES of 0.10 requires at least 300+ schools with 30+ FSM-eligible pupils per school.

KS4 English

- Unconditional ICC ~ 0.06 to 0.09 [fixed at 0.08]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.45 to 0.57 [fixed at 0.50]
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) ~ 0.25 to 0.34 [fixed at 0.30]

Table 34: **MDES estimates by number of schools and pupils, KS4 English**

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per School	5	0.33	0.26	0.23	0.19
	10	0.26	0.20	0.18	0.15
	20	0.21	0.17	0.15	0.12

	30	0.20	0.16	0.14	0.11

Notes: Shaded cells indicate MDES of 0.20 or less.

KS4 maths

- Unconditional ICC ~ 0.06 to 0.07 [fixed at 0.07]
- Covariate explanatory power for cluster/school-level variance (R_C^2) = 0.05 to 0.30 [fixed at 0.20]
- Covariate explanatory power for residual/within-school, between-pupil variance (R_R^2) ~ 0.38 to 0.39 [fixed at 0.38]

Table 35: **MDES estimates by number of schools and pupils, KS4 maths**

MDES estimates		Number of schools			
		50	**80**	**100**	**150**
Pupils per School	5	0.33	0.26	0.23	0.19
	10	0.27	0.21	0.19	0.15
	20	0.23	0.18	0.16	0.13
	30	0.22	0.17	0.15	0.13

Notes: Shaded cells indicate MDES of 0.20 or less.

For KS4 English, detecting a MDES of 0.20 with a KS2 pre-test requires 80+ schools with 10+ FSM-eligible pupils per school; or 100+ schools with 5+ FSM-eligible pupils per school. However, detecting a MDES of 0.10 requires at least 150+ schools with 30+ FSM-eligible pupils per school; or 180+ schools with 20+ pupils per school.

For KS4 maths, detecting a MDES of 0.20 with a KS2 pre-test requires 80+ schools with 10+ FSM-eligible pupils per school; or 100+ schools with 5+ FSM-eligible pupils per school. However, detecting a MDES of 0.10 requires at least 200+ schools with 30+ FSM-eligible pupils per school.

Objective IV: Value of commercial pre-tests

The following analysis includes all EEF funded trials where both pre- and post-test were commercial assessments. The aim was to understand the relative value of employing commercial pre-tests compared to an NPD derived pre-test.

In total, 14 EEF trials with English/literacy outcomes and seven EEF trials with maths outcomes were eligible for this analysis. Commercial and equivalent NPD pre-test information was utilised. All those pupils whose NPD pre-test information was not available were removed from the analysis. Therefore, for the purpose of this analysis, the sample size for these trials may be different from that found in the original trials data. Furthermore, raw scores from the selected trials were utilised since these specific trials employed consistent outcome measures for both pre- and post-test.

Table 36 provides pre-test and post-test correlations. The second column (Com*) provides commercial pre- and post-test correlation estimates. The third column, NPD** shows NPD pre-test and commercial post-test correlation estimates. In general, the correlation estimates for commercial and NPD pre-tests for English outcomes exhibited minor differences, with a few exceptions considered as outliers. For a few trials, such as, "*Word and World reading*", "*Talk of the Town*" or "*Integrating English*", correlations were nearly the same. However, trials with smaller sample sizes such as "*Response to intervention*", "*SPOKES*", "*Lexia*", displayed greater variability in the correlation estimates while estimates for larger trails remained more stable. This is not surprising given that smaller trials are more sensitive to any change. For maths outcomes, there were slightly more differences as compared to the English outcome. However, correlation estimates were broadly comparable.

Table 36: **Correlation between post-test and pre-test (commercial and NPD) in EEF data for English and maths outcomes**

Project	Com*	NPD**	Sch.	Pup.	NPD pre-test variable
	Correlation		Number		
English/literacy					
Grammar for Writing	0.50	0.40	50	2,065	KS1_READPOINTS
Response to intervention	0.65	0.49	48	348	KS1_READPOINTS
Effective feedback	0.69	0.50	13	1,247	FSP_CLL_TOTAL
Word and World Reading	0.63	0.62	16	1,184	KS1_READPOINTS
Catch up Literacy	0.58	0.46	15	496	Sum of FSP_LIT_G09 and FSP_LIT_G10
Act, Sing and Play	0.76	0.62	19	792	FSP_CLL_TOTAL
Improving numeracy and literacy	0.82	0.62	54	1,789	FSP_CLL_TOTAL
SPOKES	0.23	0.50	67	481	KS1_READPOINTS
Talk of the Town	0.68	0.64	63	2,611	KS1_READPOINTS
Success for all	0.42	0.68	50	1,292	Sum of FSP_LIT_G09 and FSP_LIT_G10
Grapho Game Rime	0.58	0.42	14	344	Sum of FSP_LIT_G09 and FSP_LIT_G10
Zippy's Friends	0.74	0.60	81	3,223	Sum of FSP_LIT_G09 and FSP_LIT_G10
Integrating English	0.56	0.51	80	3,306	KS1_READPOINTS
Lexia	0.74	0.35	57	600	Sum of FSP_LIT_G09 and FSP_LIT_G10
Maths					
Effective feedback	0.67	0.49	13	1,247	FSP_PSRN_TOTAL
Act, Sing and Play	0.73	0.56	19	798	FSP_PSRN_TOTAL
Improving numeracy and literacy	0.75	0.58	54	1,790	FSP_PSRN_TOTAL
1stClass@Number	0.28	0.35	129	457	Sum of FSP_MAT_G11 and FSP_LIT_G12
RISE	0.50	0.74	84	1,347	KS1_MATPOINTS
Onebillion	0.54	0.33	111	1,018	Sum of FSP_MAT_G11 and FSP_LIT_G12
Digital Feedback	0.34	0.67	32	1,227	KS1_MATPOINTS

Notes: Com* refers to the commercial post-test and pre-test correlation; NPD** refers to the commercial post-test and NPD pre-test correlation; Sch. refers to the number of schools while Pup. refers to the number of pupils.

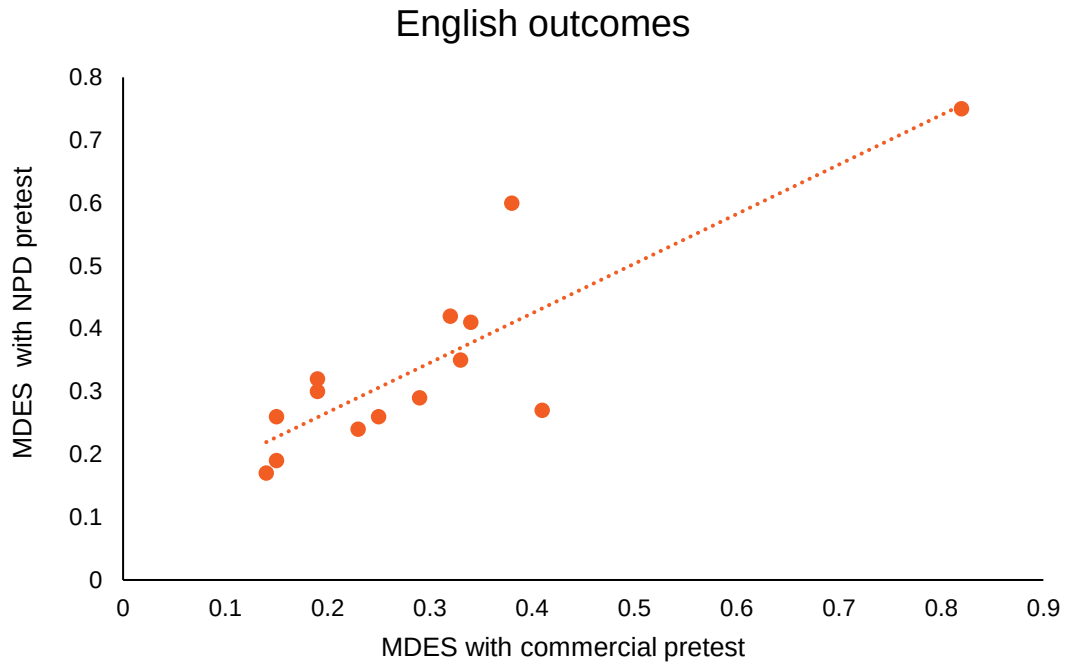
Table 37 and Table 38 provides ICC, variance and MDES estimates for the English and maths outcomes for three different sets of models: one with a commercial pre-test, another with an NPD pre-test only, and a third with both an NPD and a commercial pre-test. The aim for this analysis was to understand the role of NPD and commercial pre-tests on ICC and MDES estimation for the commercial outcomes. It is important to mention here that the provided MDES estimates stem from the values obtained from the analysis stage of these trials.

Including any of the pre-test resulted in significant reduction of the school and residual variance. Additionally, MDES estimates obtained from the commercial or NPD pre-test-based models do not vary

significantly across these trials, except for a few cases. For example, the MDES obtained using estimates from the "*Grammar for Writing*" trial with a commercial pre-test was 0.33, while using an NPD pre-test resulted in a MDES of 0.35, a difference of 0.02 standard deviations. In most cases, MDES estimates obtained using a commercial or NPD pre-test are broadly comparable, except for a few trials where there were larger differences. This suggests that using NPD scores as pre-test does not significantly affect MDES calculations. Similar outcomes were observed in data from several other trials.

Furthermore, Figure 3 also shows that there is a strong positive correlation between MDES estimates derived from NPD or commercial pre-test data from the eligible trials, except for a few outliers. It is worthy to note that for several trials, EYFS literacy or maths scores were used as NPD pre-test scores (see Table 36). While it is well known that there have been significant changes in the EYFS literacy or maths outcomes collected over time in NPD, this analysis did not identify significant deviation in the MDES estimates for commercial and NPD pre-test for larger trials such as "*Zippy friends*" or "*Improving numeracy and literacy*" trial, where different EYFSP scores were used as pre-test.

Figure 3: MDES estimates with commercial pre-test vs NPD pre-test for English and maths outcomes from EEF studies



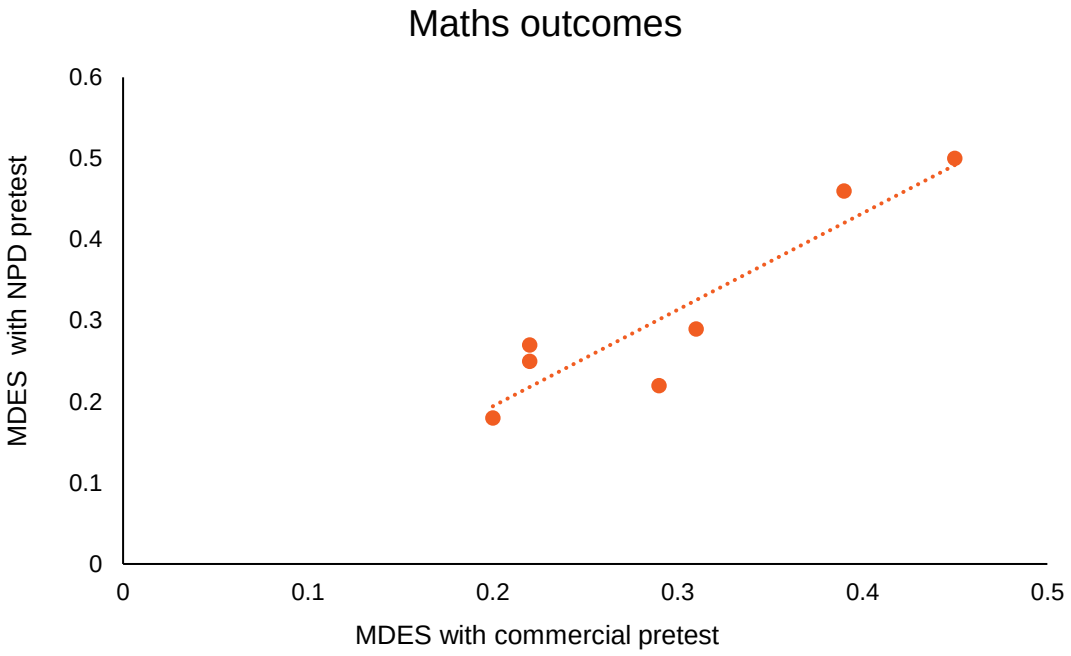


Table 37: **ICC, MDES and variance for trials with commercial English outcome**

Project	Model with no covariate			Model with commercial pre-test				Model with NPD pre-test				Model with commercial and NPD pre-test			
	School Var	Residual Var	ICC	School Var	Residual Var	ICC	MDES	School Var	Residual Var	ICC	MDES	School Var	Residual Var	ICC	MDES
Grammar for Writing	8.24	34.83	0.19	6.99	25.88	0.21	0.33	7.81	28.52	0.21	0.35	6.99	25.66	0.21	0.33
Response to intervention	1023.45	1846.69	0.36	316.47	1375.00	0.19	0.34	516.78	1711.87	0.23	0.41	233.36	1365.24	0.15	0.31
Effective feedback	1.81	13.25	0.12	0.78	7.23	0.10	0.38	2.07	9.43	0.18	0.60	0.79	5.96	0.12	0.38
Word and World Reading	122.23	187.52	0.39	108.07	98.68	0.52	0.82	89.58	107.97	0.45	0.75	96.32	89.87	0.52	0.77
Catch up Literacy	3.41	105.25	0.03	6.73	65.21	0.09	0.41	1.24	84.42	0.01	0.27	5.17	61.82	0.08	0.37
Act, Sing and Play	4.87	64.20	0.07	0.87	29.04	0.03	0.19	3.14	40.12	0.07	0.30	0.89	26.31	0.03	0.18
Improving numeracy and literacy	4.01	61.24	0.06	2.02	19.66	0.09	0.15	6.44	34.69	0.16	0.26	2.07	18.49	0.10	0.15
SPOKES	3.97	57.43	0.06	3.67	54.59	0.06	0.29	5.09	41.02	0.11	0.29	5.28	40.78	0.11	0.29
Talk of the Town	10.70	182.77	0.06	6.38	98.58	0.06	0.15	11.00	102.96	0.10	0.19	7.23	83.96	0.08	0.15
Success for all	75.59	1092.99	0.06	62.39	900.42	0.06	0.23	88.18	551.56	0.14	0.24	83.67	545.98	0.13	0.24
GraphoGame Rime	5.18	56.46	0.08	1.21	39.16	0.03	0.32	2.90	47.89	0.06	0.42	0.96	37.33	0.03	0.30
Zippy's Friends	4.78	60.03	0.07	2.65	27.00	0.09	0.14	4.03	37.13	0.10	0.17	2.57	25.33	0.09	0.14
Integrating English	6.20	30.31	0.17	5.34	19.69	0.21	0.25	5.71	21.44	0.21	0.26	5.26	19.00	0.22	0.25
Lexia	397.77	1935.82	0.17	59.49	1006.69	0.06	0.19	274.19	1772.24	0.13	0.32	56.37	1004.23	0.05	0.19

Note: Post-test outcome is the commercial English/literacy outcome. MDES calculations here assumed that P=0.5 that is units are equally allocated to intervention and control groups and the explanatory power estimates were calculated using residual and school-level variance from the model with no covariate and model with covariate.

Table 38: **ICC, MDES and variance for trials with commercial maths outcome**

Project	Model with no covariate			Model with commercial pre-test				Model with NPD pre-test				Model with both commercial and NPD pre-test			
	School Var	Residual Var	ICC	School Var	Residual Var	ICC	MDES	School Var	Residual Var	ICC	MDES	School Var	Residual Var	ICC	MDES
Effective feedback	1.22	9.62	0.11	0.83	5.15	0.14	0.45	1.02	7.31	0.12	0.50	0.64	4.41	0.13	0.40
Act, Sing and Play	15.90	88.96	0.15	9.27	41.11	0.18	0.39	12.89	59.20	0.18	0.46	8.88	38.98	0.19	0.38
Improving numeracy and literacy	2.96	23.87	0.11	1.87	10.26	0.15	0.22	2.82	15.13	0.16	0.27	1.80	9.61	0.16	0.21
1stClass@Number	4.40	16.07	0.21	4.14	14.73	0.22	0.31	2.81	15.18	0.16	0.29	2.76	14.21	0.16	0.28
RISE	16.51	178.27	0.08	11.94	134.63	0.08	0.20	13.11	73.28	0.15	0.18	11.83	72.97	0.14	0.18
Onebillion	3.56	14.90	0.19	1.82	11.33	0.14	0.22	2.69	13.82	0.16	0.25	1.69	11.12	0.13	0.21
Digital Feedback	3.12	33.87	0.08	2.37	30.40	0.07	0.29	1.41	18.73	0.07	0.22	1.32	18.56	0.07	0.22

Note: Post-test outcome was the commercial maths outcome. MDES calculations here assumed that P=0.5 that is units are equally allocated to intervention and control groups and the explanatory power estimates were calculated using residual and school-level variance from the model with no covariate and model with covariate.

Practical implications of this study

The findings presented in this report carry significant practical implications. This study provides estimates key study design parameters used to determine MDES in educational trials. The pre- and post-test correlations, unconditional and conditional ICCs, and the explanatory power estimates provided in this study can be used to inform the design of 2-level CRTs in England.

The report provides examples for MDES estimates for each Key Stage, drawing from whole NPD dataset estimates for the three most recent years (2017 to 2019). By employing the estimates provided in the Results section, evaluators and researchers can estimate MDES and the required sample sizes for their specific study.

NPD data (2017-2019) analysis from this study shows that for KS1 maths, KS2 English, KS4 English and maths, detecting a MDES of 0.20 with a Key Stage NPD pre-test requires 80+ schools with 20+ pupils per school; 100+ schools with 10+ pupils per school; or 150+ schools with 5+ pupils per school. While for EYFS English/maths, detecting a MDES of 0.20 requires 100+ schools with 20+ pupils per school; 150+ schools with 10+ pupils per school. KS1 English is the only outcome which requires only 50+ schools with 20+ pupils per school to detect MDES of 0.20.

Furthermore, detecting a MDES of 0.10, which is a more commonly observed actual effect in EEF trials across various Key Stages, with a Key Stage NPD pre-test, requires a sample of 180+ schools with 30+ pupils per school for KS4 English; 200+ schools with at least 30+ pupils per school for KS1 English and maths; 230+ schools with 30+ pupils per school for KS4 maths; 270+ schools with 30+ pupils per school for KS2 English; and more than 300+ schools with 30+ pupils per school in EYFS and KS2 maths. It is important to mention here that these estimates are obtained based on certain assumptions for the number of units to be randomised for a fixed sample size. These calculations also make use of the most recent estimates of the ICC and explanatory power from the NPD data. However, researchers who are keen to estimate MDES for their specific study may opt to use different values that align with the feasibility and practical relevance of their research objectives.

Most educational trials require estimation of a MDES not only for all pupils but also for specific subgroups, such as FSM-eligible pupils, in order to conduct appropriate subgroup analyses. The ICC (unconditional/conditional) and variance estimates provided for the EEF or NPD datasets could be used to estimate MDES for FSM-eligible pupils. Based on the analysis of the most recent NPD data, to detect a MDES of 0.20 for the subgroup of FSM-eligible pupils using NPD KS2 as pre-test requires 80+ schools with 10+ FSM-eligible pupils per school for KS4 English or maths.

This study has provided estimates for ICC (unconditional/conditional), pre- and post-test correlation and variance from the NPD whole data, NPD sample data and EEF trial data. This variety of estimates provides researchers with flexibility in selecting the most relevant data source based on their objectives and research questions. For example, estimates derived from the NPD whole data refers to the entire population of pupils in England. Researchers who are keen to utilise population specific estimates for different Key Stages, specifically in relation to English and maths outcomes, may consider estimates derived from the whole NPD dataset to be most suitable. More details for the NPD and EEF specific estimates are also available in the supplementary Excel files.

The aim of this study was to conduct a comparative analysis of estimates from various data sources across recent years. This comprehensive approach allows researchers to evaluate trends and patterns in the estimates. Considering that the study provides estimates from 2012 to 2019, researchers who prefer to use the most current information can use estimates from 2019 or, if preferable, consider using an average estimate from the last 2-3 years.

This study has also provided additional trial specific estimates of the pre- and post-test correlations, unconditional and conditional ICC using pre-test as a covariate including variance parameters for trials stored in the EEF Archive. This provides educational researchers specific insights in scenarios akin to these past educational trials. More details for the trial specific estimates are available in the supplementary Excel files.

Finally, this study also provides some additional analysis to examine the value of commercial pre-tests in educational trials. Overall, findings from this study indicate that the MDES for the English and maths outcomes do not vary much in instances where an NPD pre-test is used in place of a commercial pre-test. Therefore, given the strong correlation between a MDES obtained using NPD pre-test and commercial pre-test, it can be inferred that researchers who are designing trials with commercial outcomes could potentially use historic NPD Key Stage scores to obtain MDES for their study designs.

Conclusions

This study investigated and empirically derived parameters commonly used for statistical power and sample size calculations to better inform future trial design; specifically, through estimating school-level unconditional and conditional ICCs for English and maths attainment outcomes at four educational Key Stages - EYFS, KS1, KS2 and KS4 - by using the NPD and EEF Archive data from 2012-2019. Additionally, correlation coefficients between test scores at pupil and school level for English and maths for three subsequent Key Stages - EYFS to KS1, KS1 to KS2, and KS2 to KS4 - were also estimated from both data sources. Finally, conditional multilevel models were used to provide explanatory power estimates for pre-test covariates. The empirical estimations of ICC, pre- and post-test correlation coefficients, and explanatory power were also conducted for FSM-eligible pupils to examine the variation in these parameters commonly used for statistical power and sample size calculations.

This work also adds to Allen et al. (2018), a study which examined properties of commercial test scores for a few EEF trials. This study not only examined a number of trials with commercial test scores but also contributed to examining the absolute reduction in MDES achieved by commercial tests used in EEF trials relative to NPD data.

Unconditional ICC estimates obtained from the NPD analyses shows that ICC estimates for EYFS English ranged between 0.05 and 0.06, and between 0.07 and 0.08 for maths. In KS1, the ICC estimates for the English and maths estimates are broadly comparable, ranging between 0.03 and 0.05. In KS2, the ICC estimates are larger than those observed in KS1, and were again broadly comparable for English and maths, ranging between 0.09 and 0.13. In KS4, the ICC estimates are slightly lower than those seen in KS2 and remarkably consistent over time, ranging between 0.10 and 0.11, since 2015, for both English and maths. For EEF studies, unconditional estimates for EYFS and KS1 suggest slightly higher ICCs for maths when compared to English. As we move to KS2, ICCs become slightly higher for English than what we would expect to find for educational studies, ranging from 0.10 to 0.25, except for 2019. A comparison between the NPD sample data and EEF studies shows that unconditional ICC estimates converge for KS2 and KS4. However, for KS1 the ICC estimates for EEF studies were much higher than for the NPD sample data.

Then, conditional ICC analyses considering pre-test as a covariate shows that ICC estimates for conditional models were able to explain large amounts of the school- and pupil-level variation (e.g., explanatory power for pre-test at the school level was more than 0.62 and at the residual level more than 0.39 for NPD KS4 English in 2019). In a model with pre-test included at the pupil and school level, additional covariates such as FSM, SEN or EAL do not explain any additional variation in the school- or pupil-level variance once pre-tests at pupil and school levels had been included in the model (e.g., explanatory power for all variables at the school level was more than 0.68 and at the residual level more than 0.40 for NPD KS4 English in 2019). This validates the EEF's preferred analytical model, which stresses the importance of including a pre-test as covariate (EEF, 2022). Furthermore, estimates for conditional ICC models with pre-test as covariate and unconditional models were similar for most years across all three Key Stages (KS1, KS2 and KS4) for both NPD data and EEF studies.

Pre- and post-test correlations between subsequent Key Stages shows that correlations between KS1 and KS2 scores (more than 0.60 for most years) and KS2 and KS4 scores (more than 0.50 for most of the years) were strong, while correlations between EYFS and KS1 scores were moderate but increased over time. Interestingly, correlations at pupil- and school-levels are reasonably similar for both English

and maths across all Key Stages. For EEF studies, correlation estimates between EYFS and KS1 were notably higher than that found in NPD data, but similar for both English (ranging between 0.54 and 0.88) and maths (ranging between 0.50 and 0.88). Correlation for both English and maths for KS1 and KS2 as well as KS2 and KS4, ranged mostly between 0.50 and 0.70 over the years. Correlation estimates for NPD sample data and EEF studies converged much better for KS1 and KS2 as well as for KS2 and KS4. For FSM-eligible pupils, correlation estimates from NPD data and EEF studies were much closer, while for EYFS and KS1, there are larger differences, more akin to what has been observed in pre- and post-test data correlation estimates across data including all pupils.

School-level explanatory power estimates obtained from conditional models that included pre-tests at both pupil and school levels were consistently lower than those obtained from the correlation estimates. Moreover, residual explanatory power estimates (within-school, between-pupils) were slightly greater compared with those obtained from the pupil-level correlation estimates. The reasons for this seem clear. The conditional models were multivariate, which means that the school-level explanatory power estimates drew on pre-test variance to account for school-level variance in an outcome. In contrast, the estimates obtained from the (squared) school-level correlations were bivariate, meaning they did not factor in any covariance between pupil- and school-level pre-tests. The (squared) pupil-level correlation estimates closely aligned with estimates for total explanatory power across all NPD analyses. However, it is important to note that trial sensitivity draws on residual rather than total explanatory power. Total explanatory power is an estimate of the proportion of explained variance at both pupil- and school-levels whilst residual explanatory power is an estimate of the proportion of explained variance that is within-schools, between-pupils – essentially, the variance that remains after accounting for between-school variance.

Lastly, we estimated MDES for commercial or NPD pre-test based models for trials with commercial test scores. These results suggested that MDES estimates for both models do not vary significantly except for a few outliers. There is a strong positive relationship between MDES estimated for commercial and NPD pre-test for both English and maths outcomes.

Study limitations

This study has provided useful insights on the parameters associated with estimation of the sample size and power calculation in educational trials. However, there were a few limitations. First, given that there have been some changes in the literacy/mathematics outcome measurements over time within NPD data, changes over time need to be interpreted with slight caution as the real change in the estimates may get obscured by the changes in measurements. Secondly, EEF trials encompass a wide range of outcome measures to capture English/literacy or maths attainment. Thus, this study standardised all outcome measures to Z-scores due to the need to estimate ICCs and assessment correlation parameters for each year, and therefore might have introduced some level of abstraction that could impact the results' applicability. Thirdly, due to the unavailability of the data for KS3 from NPD, analysis for KS3 was not conducted, even though some EEF trials provides information for the KS3 pupils. Fourthly, conditional models including FSM, SEN and EAL failed to converge in a few cases due to missing data issues. Lastly, it is also important to mention here that EYFS and KS1 scores are no longer available in the same format as used for the analysis presented in this report. Therefore, present-day changes may affect the ability of evaluators to use the provided EYFS and KS1 data as a baseline measure in their trials.

Future research

As a final remark, this study has focused on 2-level CRT designs where pupils are clustered into schools and with randomisation at the school-level. To date, this is the most common form of RCT design funded by the EEF (Demack et al., 2021). However, this 2-level CRT design does not account for any within-school clustering of pupils. For example, policies of setting/streaming may result in very high clustering at the classroom level (Demack et al. 2021; Demack, 2019). Additionally, the composition of

classrooms in terms of both pupils and teachers may change over time which may relate to the formation or re-configuration of classroom sets or streams (and/or to teachers moving between classrooms). With a 2-level CRT design, the strength of within-school clustering of attainment data and shifts pupil/teacher classroom composition are hidden. Therefore, to gain more understanding on these key structural contexts, more 3-level CRTs that acknowledge clustering at both school and classroom levels are needed. A further and closer examination of 3-level CRTs would be valuable complement to this piece of work.

References

Allen, R., Jerrim, J., Parameshwaran, M., & Thompson, D. 2018. Properties of commercial tests in the EEF database. London: EEF Research Paper, (001).

Bloom, H. S., Richburg-Hayes, L., and Black, A. R. 2007. Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. *Educational Evaluation and Policy Analysis*, 29(1), pp. 30-59.

Dong, N., Kelcey, B., Maynard, R. and Spybrook, J. 2015 *PowerUp! Tool for power analysis.*

Demack, S. 2019. Does the classroom level matter in the design of educational trials? A theoretical & empirical review. London: EEF Research Paper (003).

Demack, S., Maxwell, B., Coldwell, M., Stevens, A., Wolstenholme, C., Reaney-Wood, S. & Stiell, B. 2021. Review of EEF Projects. Summary of Key Findings. Education Endowment Foundation.

Demack, S., Culliney, M., Boylan, M. & Wolstenholme, C. 2022. Realistic Maths Education: Evaluation Report.

Education Endowment Foundation (EEF) 2020. Archiving evaluation data. London: EEF. https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/archiving-evaluation-data/Archiving_evaluation_data_analysed_in_the_SRS_-_November_2020.pdf

Education Endowment Foundation (EEF) 2022. Statistical analysis guidance for EEF evaluations. London: EEF. EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf (d2tic4wvo1iusb.cloudfront.net)

Gorard, S., 2018. *Education policy*. Bristol: Policy Press.

Hox, J.J., Moerbeek, M. and Van de Schoot, R., 2017. *Multilevel analysis: Techniques and applications*. Routledge.

Hedges, L.V. and Hedberg, E.C., 2007. Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, *29*(1), pp.60-87.

Hedges, L. V., and Hedberg, E. C. 2013. Intraclass correlations and covariate outcome correlations for planning two-and three-level cluster-randomized experiments in education. *Evaluation review*, 37(6), pp. 445-489.

Hemmings, B., Grootenboer, P. and Kay, R., 2011. Predicting mathematics achievement: The influence of prior achievement and attitudes. *International Journal of Science and Mathematics Education*, *9*, pp.691-705.

Muijs, D. and Dunne, M., 2010. Setting by ability–or is it? A quantitative study of determinants of set placement in English secondary schools. *Educational Research*, *52*(4), pp.391-407.

Singh, J., Liddy, C., Hogg, W., and Taljaard, M. 2015. Intracluster correlation coefficients for sample size calculations related to cardiovascular disease prevention and management in primary care practices. BMC research notes, 8(1), pp. 89.

Spybrook, J., Shi, R. and Kelcey, B. 2016 Progress in the past decade: an examination of the precision of cluster randomised trials funded by the US Institute of Education Studies. International Journal of Research & Method in Education 39(3) pp. 255-267.

Strand, S., Malmberg, L. and Hall, J., 2015. English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database.

Appendix A: EEF Archive data

The EEF has funded several educational trials with the aim to improve outcomes for children and young people in England, particularly those from disadvantaged backgrounds. All projects funded by the EEF are independently evaluated by a number of evaluation teams from different universities and independent research organisations. The data from these projects are deposited in an archive which has become a rich repository of findings from EEF interventions. The Fischer Family Trust (FFT) is the organisation responsible for transferring and validating the data from the different EEF funded evaluations and adding them to the EEF Archive (EEF, 2020). The FFT currently manages this data repository within the Secure Research Service (SRS) of the ONS.

For the purposes of this study, the EEF Archive was made available to the Durham University and Sheffield Hallam University research teams through the SRS. More than 200 trials have been commissioned by the EEF since 2011, involving over 1,000,000 pupils, out of which 109 trials data are currently available in the EEF Archive. The aim of creating this archive is to provide researchers a rich data source of educational trials. This archive can be used to summarise the effect of interventions, track their longer-term impacts, estimate indicators for sample size estimations (as is done in the present study), and to conduct further methodological research that can be of use for researchers in the field of education.

The EEF Archive contains trial data from 2013 to the most recent years. This includes pupils-level data from all Key Stages. The majority of the EEF trials were cluster randomised trials. The outcomes in all the trials are English/literacy, Mathematics, and other outcomes, with attainment data either obtained from the NPD or collected directly by the evaluators' preferred outcome measures given the specifics of the programme that was evaluated. Although this provided a consistent dataset, the differences in assessment across the complex domains of English/literacy and maths need to be borne in mind. This study utilised all those trials from 2013 to 2019 which provides English/literacy or maths outcomes.

Appendix table 1: **EEF trials used in this study**

Trial code	Description
1	Future foundations
3	Grammar for Writing
5	Response to intervention
6	Effective feedback
9	Catch up numeracy
25	Increasing pupil Motivation
26	Word and World Reading
29	Catch up Literacy
35	Act, Sing and Play
41	Improving numeracy and literacy
42	Philosophy for Children
43	SPOKES
44	Tutor Trust Secondary
45	Lesson Study
49	Talk of the Town
51	Success for all
52	Chess in Schools
66	Affordable Online Maths Tuition
68	ABRA
72	Flipped Learning
88	Tutor Trust Secondary
90	Parenting Academy
93	ReflectEd

95	Dialogic Teaching
97	Learner Response System
98	Teacher Observation
101	Switch on Reading
104	Children's University
105	Scratch Maths
106	Good Behaviour Game
109	GraphoGame Rime
117	The Literacy Octopus
121	Zippy's Friends
122	1stClass@Number
124	Improving Working Memory
126	Tutal trust Primary
127	The RISE Project
128	Maths Count
131	Grammar for Writing
132	IPEEL 1 YEAR
134	Mathematical Reasoning
136	IPEEL 2 YEAR
137	RISE
140	Onebillion
141	Families and Schools Together
144	Changing Mindsets
145	Evidence based literacy support
150	Digital Feedback
151	Integrating English
152	Accelerated reader-effectiveness
165	Lexia trial
167	Same Day Intervention

Additional information about each of these archived trials can be found within the evaluation reports accessible on the EEF project website.

A significant number of missing values were observed for FSM, EAL, and SEN variables within the EEF archive data (30%, 93% and 95% respectively for FSM, EAL and SEN). This resulted in the failure of numerous models to converge successfully. To deal with this issue, we updated the missing values of these variables with the appropriate NPD values (refer to Appendix table 2).

Appendix table 2: **Updated variables, the values original missing were updated as 1 or 0, from their corresponding variable in NPD**

Variables	Originally non-missing		Originally missing		
	0	1	0	1	NA
FSM	477,581	217,479	64,844	50,781	178,433
EAL	47,304	20,912	137,094	536,309	247,499
SEN	41,146	12,522	567,618	118,085	249,747

The analysis for unconditional and conditional models was ran again for both EEF trial data and EEF data replaced by NPD outcomes.

Appendix B: NPD variables description

Appendix table 3: NPD categorical variables description used in this study

Y2, KS1	English	2012-15	KS1 Reading	W-6	W = Working towards level 1 1 = Achieved Level 1 2 = Achieved Level 2 3 = Achieved Level 3 4 = Achieved Level 4 4+ = Achieved Level 4 or above 5 = Achieved Level 5 6 = Achieved Level 6
		2016-19	KS1 Reading	BLW-GDS	BLW = Below - corresponds with P-scales or NOTSEN PKF = Pre-Key stage - Foundations for the expected standard, PK1 = Pre-Key stage standard 1 PK2 = Pre-Key stage standard 2 PK3 = Pre-Key stage standard 3 PK4 = Pre-Key stage standard 4 WTS = Working towards the expected standard EXS = Working at the expected standard GDS = Working at a greater depth within the expected standard
	Maths	2012-15	KS1 Maths	W-6	W = Working towards level 1 1 = Achieved Level 1 2 = Achieved Level 2 3 = Achieved Level 3 4 = Achieved Level 4 4+ = Achieved Level 4 or above 5 = Achieved Level 5 6 = Achieved Level 6
		2016-19	KS1 Maths	BLW-GDS	BLW = Below - corresponds with P-scales or NOTSEN PKF = Pre-Key stage - Foundations for the expected standard, PK1 = Pre-Key stage standard 1 PK2 = Pre-Key stage standard 2 PK3 = Pre-Key stage standard 3 PK4 = Pre-Key stage standard 4 WTS = Working towards the expected standard EXS = Working at the expected standard GDS = Working at a greater depth within the expected standard
Y11, KS4	English	2012-16	GCSE English	A*-U	* = A* at GCSE A = A at GCSE B = B at GCSE C = C at GCSE D = D at GCSE E = E at GCSE F = F at GCSE G = G at GCSE U = Ungraded at GCSE
	Maths	2012-16	GCSE Maths	A*-U	* = A* at GCSE A = A at GCSE B = B at GCSE C = C at GCSE D = D at GCSE E = E at GCSE F = F at GCSE G = G at GCSE U = Ungraded at GCSE