Individual participant data meta-analysis of the impact of EEF trials on the educational attainment of pupils on Free School Meals: 2011 - 2019

May 2021

Authors:

Bilal Ashraf, Akansha Singh, Germaine Uwimpuhwe, Tahani Coolen-Maturi, Jochen Einbeck, Steve Higgins and Adetayo Kasim

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 the 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

Durham Research Methods Centre (DRMC) was established in 2018 as part of Durham University's strategic investment in methodological innovations spanning both quantitative and qualitative methods. Through engagement with University colleagues and external partners, we conduct research and training to advance understanding of real-world challenges.

This work on the IPD meta-analysis of the impact of EEF trials on the educational attainment of pupils on Free School Meals was conducted by DRMC fellows and researchers from the School of Education and Department of Anthropology with track records in evaluation of educational interventions, meta-analysis of evidence in education and advanced quantitative methods. We have partnered with 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. This work is part of the Durham-EEF Methods Scheme, a partnership between the DRMC and the Department of Mathematical Sciences at Durham University, UK.

For more information about DRMC or this report please contact:

Prof Adetayo Kasim
Durham Research Methods Centre (DRMC)
NEDTC Hub
1st Floor Arthur Holmes Building
Mountjoy
South Road
Durham
DH1 3LE

0191 334 6209

a.s.kasim@durham.ac.uk

www.dur.ac.uk/researchmethodscentre

Contents

1.0 Executive summary

Overview

The main objective of this report is to estimate the impact of EEF-funded trials on pupils eligible for Free School Meals (FSM) and the attainment gap between FSM and non-FSM pupils based on an analysis of 88 EEF trials and data from over half a million pupils. This report presents findings from the intention to treat (ITT) analysis using two-stage models and Individual Participant Data (IPD) meta-analyses for EEF trials evaluations published from 2011 to 2019. For the meta-analyses, frequentist and Bayesian multilevel models were used to estimate the individual and pooled effect sizes across the categories of explanatory variables. The explanatory variables used in this study were type of outcome, study design, Key Stages (KS) and the type of interventions as used in the EEF evaluation studies.

In this study we investigated the immediate impact of EEF-funded interventions on attainment in literacy and mathematics based on the following research questions:

1. Do EEF trials improve literacy/mathematics attainment of pupils eligible for FSM?
2. What broad types of interventions are more beneficial for mathematics and literacy outcomes of FSM pupils?
3. Do FSM pupils improve their literacy/mathematics attainment more or less from EEF-funded interventions compared to their non-FSM peers?

Key findings

These findings in this report indicate that EEF-funded interventions improved FSM pupils' literacy outcome by an equivalent of about one month's progress. Moreover, the attainment gap in literacy between FSM and non-FSM pupils appears to be dependent on Key Stages. More highlights are provided below.

The overall impact of EEF interventions on the literacy outcomes of FSM pupils was positive with an effect size of 0.06 (0.03, 0.08). The overall effect on the mathematics outcomes of FSM pupils was 0.00 (-0.03, 0.04). These findings answered our first research question and suggested that EEF trials had improved FSM pupils' literacy outcome by an equivalent of about one month's progress.

In order to answer the second research question, we assessed both the outcomes by primary/secondary status, Key Stages (KS), type of intervention and study design. When

literacy outcomes impact was assessed by KS, the greatest effect was observed on the FSM pupils in KS1 with an effect size of 0.09 (0.02, 0.16), followed by those in KS3 with an effect size of 0.08 (0.03, 0.13), KS2 with an effect size of 0.03 (-0.01, 0.07), and the least impact was observed for pupils in KS4 with an effect size of 0.02 (-0.05, 0.08). Overall, a similar impact was observed whether literacy was either a primary or secondary outcome with effect sizes of 0.06 (0.03, 0.08) and 0.06 (-0.04, 0.16), respectively. Interventions designed for a small group of pupils had the highest impact on literacy outcomes with an effect size of 0.14 (0.06, 0.22) followed by that for one-to-one pupils with an effect size of 0.08 (0.04, 0.13), whole school approaches with an effect size of 0.02 (-0.02, 0.06) and whole class interventions with an effect size of 0.01 (-0.04, 0.05). The overal impact by study design methods indicated that multisite trials (MST) had slightly bigger overall effect sizes than cluster-randomised trials (CRT).

There was a positive effect for mathematics as a primary outcome with an effect size of 0.01 (-0.02, 0.05) and a negative effect as a secondary outcome with an effect size of -0.07 (-0.15, 0.00). There was no clear pattern of the impact of the interventions by Key Stages. Contrary to the results for literacy outcomes, one-to-one and whole school interventions had positive effects on maths outcomes of FSM pupils. The overall impacts of CRT and MST trials on maths outcomes were similar.

The third research question was addressed by the analysis of the attainment gap to assess whether FSM pupils benefited more or less compared to their non-FSM peers. The attainment gap is defined in this report as the difference in gain between the two groups when both are receiving the intervention. This is different to how the attainment gap is often defined in observational or longitudinal studies, which considers the absolute difference between FSM pupils and non-FSM pupils.

The attainment gap observed for the literacy outcomes was improved in favour of FSM pupils with an overall attainment gap of 0.01 (-0.01, 0.04). For mathematics, the overall attainment gap was -0.01 (-0.04, 0.02).

For literacy, the impact on narrowing the gap when used as a primary outcome was greater than when literacy was a secondary outcome with an attainment gap of 0.02 (-0.02, 0.04) and 0.00 (-0.06,0.06) respectively. For mathematics outcomes, it was surprising that mathematics as a secondary outcome had positive attainment gap favouring FSM pupils than non-FSM pupils and negative attainment gap favouring non-FSM pupils when mathematics was the primary outcome, with attainment gaps of 0.02 (-0.04, 0.08) and -0.01 (-0.04, 0.02), respectively.

For literacy, the narrowing of the attainment gap between FSM and non-FSM decreased by KS. The gap narrowed most at KS1 with an attainment gap of 0.07 (0.00, 0.14) and least at KS2 and KS4 with attainment gaps of 0.00 (-0.03, 0.03). For maths, the narrowing was greatest at KS3 with an attainment gap of 0.02 (-0.07, 0.10); for the other three Key Stages there was no evidence of the gap narrowing.

By the type of intervention, the attainment gap between FSM and non-FSM literacy outcomes was positive for one-to-one and small group interventions. This indicates that on an average, FSM pupils performed better than the non-FSM pupils in these two subgroups of intervention. Small group interventions (pooled attainment gap = 0.05 (-0.04, 0.14)) benefitted FSM pupils the most followed by one-to-one interventions (pooled attainment gap = 0.02 (-0.04, 0.07)). However, in the case of whole-class and whole-school interventions, pooled attainment gaps were zero. For maths outcomes, the highest pooled attainment gap was observed for whole-class interventions (pooled attainment gap 0.02 (-0.03, 0.06)); however, for the rest of the interventions, the group attainment gap was nearly zero.

Robustness

Reliability of the estimates of overall effect sizes and overall attainment gaps was assessed by sensitivity analysis, which showed that our estimates were consistent across different methodological approaches and even after excluding few trials with fewer than three security padlocks.

2.0 Introduction

Educational attainment has become one of the clearest early indicators of life outcomes such as employment, income and social status, and is a strong predictor of attitudes and wellbeing (Manstead, 2014). Marmot (2010) argued that there are particularly large gaps between extremes of the social hierarchy in the UK with people from the highest social or economic background living longer and with a longer period of their life free from health issues. The impact of low levels of education is not restricted to adulthood; it is also a bigger issue with school-aged children. It is well known that children growing up in poorer families emerge from school with substantially lower levels of educational attainment (Chowdry *et al*., 2010). Since 2011, 60% of children in absolute and relative poverty were eligible for free school meals (FSM) (DWP, 2013), which became mandatory for all pupils in Reception and Years 1 and 2 in England in 2014 (DFE, 2014). Pupils on FSM are reported to make less progress on average compared to their peers (Humprey *et al*., 2013). The gap between disadvantaged pupils and their peers in England is equivalent to one whole General Certificate of Secondary Education (GCSE) grade for mathematics and 0.75 grade in reading. This gap is significantly higher than several other high-income countries in Europe and Asia (Jerrim, Greany & Perera, 2018). The gap between disadvantaged pupils and their peers is evident even when children begin school at age 5 and increases at every stage of education afterwards (EEF, 2017). In Scotland, children living in the most deprived areas are '6 to 13 months behind their peers in problem-solving at age 5; 11 to 18 months behind their peers in expressive vocabulary at age 5; and around two years of schooling behind their peers at age 15' (Scottish Government, 2014). By the time that children leave primary school, those in receipt of FSM are estimated to be significantly behind their more affluent peers (Spencer, 2015). This gap clearly indicates the need to focus on social deprivation to ameliorate the impact of poverty and here schools have a pivotal role to play. High quality education and better teaching methods can be important in reducing this attainment gap (Jerrim *et al*., 2018). Improving educational achievements of pupils on FSM also has the potential to break the cycle of poverty, reduce health inequality, improve lifestyle choices and improve mental health (Hobbs & Vignoles, 2013).

The Education Endowment Foundation (EEF) is an independent charity dedicated to break the link between family income and educational achievement. More than 150 trials have been commissioned by EEF to identify the interventions that can improve the academic attainment of the children and also reduce the attainment gap among deprived pupils as compared to their counterparts (EEF, 2019). Sub-group analyses of pupils on free school meals (FSM) are frequently reported in each trial report, but there is a need to synthesize evidence on impact of EEF interventions on FSM pupils across trials. Analysis of FSM pupils reported for each

trial is useful to complement main findings from individual trials. However, it offers limited insights into how EEF interventions as a whole affect FSM pupils. The trials were also under-powered for the FSM subgroup analysis due to smaller sample size for the group. Are the interventions reducing attainment gaps between FSM pupils and their peers? And what types of interventions are likely to be more beneficial to FSM compared with their peers? These are some of the questions that need answers to provide a better understanding of the progress so far, inform decisions about how to best target specific interventions, and possibly suggest ways to improve the design or implementation of future interventions (Schochet *et al*., 2014).

The current COVID-19 closure of schools is predicted to reverse the progress made to close the attainment gap in the last decade (Coe, 2020). Therefore, it is timely to highlight the characteristics of the most promising interventions that were effective in reducing the attainment gaps between FSM and their peers. This report provides a robust and independent assessment of how EEF-funded interventions benefit FSM pupils and how they impacted on attainment gap by synthesizing evidence from existing trials using individual participant data meta-analysis methods. The traditional meta-analysis approach relies on extracting effect sizes from each trial (Burke *et al*., 2016; Kontopantelis, 2018), but it often suffers from loss of information and lack of consistency in the methods used to calculate individual effect sizes (Debray *et al*. 2015). An individual participant data (IPD) meta-analysis approach can improve reliability of results and it is considered as the gold standard for meta-analysis (MRC, 2020). It is also a more flexible approach to capture variability within and between trials. IPD meta-analysis can also improve standardisation of outcomes; reduce publication, reporting and ecological biases; allow detailed checks of analysis assumptions and consideration of covariates and treatment-covariate interactions which are often lacking in traditional meta-analysis methods (Debray *et al.*, 2015). The delay between research findings and implementation of effective programmes is reduced with the meta-analysis studies. This study will provide a robust and accurate summary of the educational intervention effectiveness in the UK using IPD, a gold-standard of review methods. This robust summary can help education policy makers design appropriate policies and programmes for FSM children in the UK.

This study meta-analysed evidence from randomised controlled trials (RCTs) commissioned by EEF and reported between 2011 and 2019, to assess the impact of EEF-funded interventions on FSM pupils. We defined FSM pupils as pupils ever eligible for FSM in the last six years in schools (EverFSM6). We also aim to identify broad types of interventions (for example, small group versus whole school) which are more likely to improve educational attainment of FSM pupils. This report for the first time provides comparable individual and

global pooled effect sizes for FSM pupils and the estimated attainment gaps in their educational outcomes in literacy and mathematics.

2.1 Research questions

In this study we investigated the immediate impact of EEF-funded interventions on attainment in literacy and mathematics based on the following research questions:

1. Do EEF trials improve literacy/mathematics attainment of pupils eligible for FSM?
2. What broad types of interventions are more beneficial for mathematics and literacy outcomes of FSM pupils?
3. Do FSM pupils improve their literacy/mathematics attainment more or less from EEF-funded interventions compared to their non-FSM peers?

2.2 Study Objectives

The main objective of this research is to estimate the impact of EEF-funded interventions on FSM pupils using individual participant data meta-analysis as specified in the Statistical Analysis Plan (SAP) submitted to EEF.

2.3 Ethics and Data Protection

All the data used in the quantitative analyses were extracted from the EEF Archive generated by the Fischer Family Trust and provided to Durham University as part of the EFF Archive and Database project. The legal basis for processing this data by Durham University is 'Public Task' as defined in Article 6(1e) of the General Data Protection Regulations (GDPR). Ethical approval was granted by the School of Education’s Ethics Committee.

3.0 Methods

3.1 Study Design

EEF has funded over 150 projects, and data from 105 projects were available in the archive. This study initially planned to conduct a meta-analysis using data from all 105 trials in the archive. However, we ended up analysing only 82 projects due to the exclusion of the trials as shown in Figure 1.

For all trials analysed, a pupil's observation was omitted from analysis if the information on any of the outcomes (math/literacy), FSM, intervention, or school was missing, so a complete

case analysis was carried out. Most of the EEF trials are either cluster randomised trials (CRT) or multisite trials (MST). MST refers to multisite trials where randomisation was within school such that pupils in each school are involved in both the intervention and control group (Xiao *et al*., 2016). CRT refers to cluster randomized trial in which clusters, such as schools, classes, or year groups, are randomly assigned to either intervention or control group. It is possible for both designs to be combined in a single trial such as cluster randomisation of classes within schools. Most of the MST and CRT trials in the EEF archive are two-armed trials except few that were three-arms. Few EEF trials used a quasi-experimental design (including regression discontinuity design (RDD)).

Table *1* presents the distribution of the trials by design, including number of schools and pupils. Please note that the trials with more than one treatment (more than two arms) were segregated as separate trials for each treatment (Figure 1).

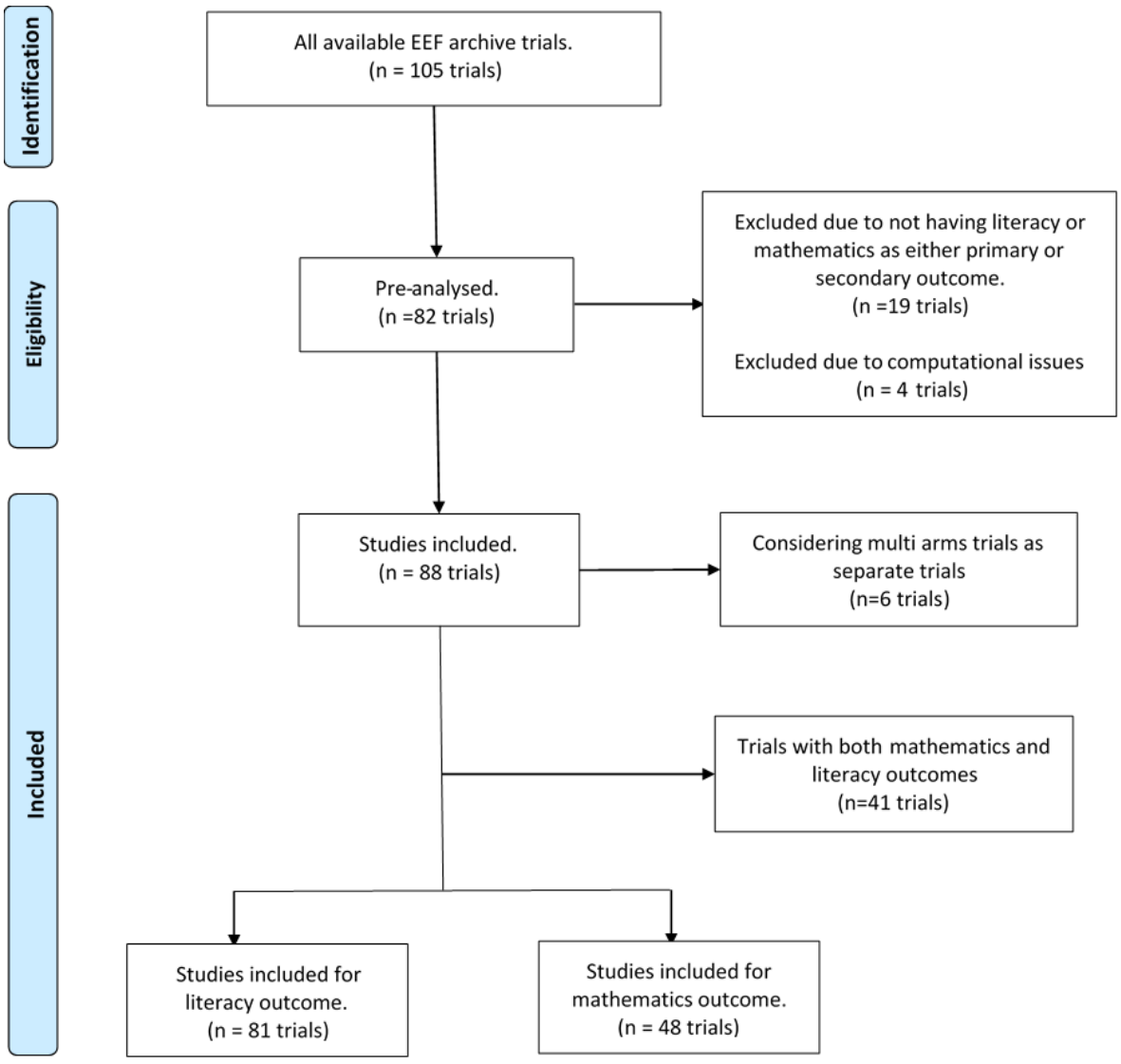

Figure 1: Process from screening to the analysed projects in this study.

Therefore, the total number of trials in each analysis reflects the number of trials considering two or three treatment trials as separate trials. However, 76 trials in the data were one-treatment trials and six trials had more than one treatment; segregating those three-arm trials leads to the total of 88 trials (Figure 1). A list of all the trials used in this study have been provided in the Appendix 1 Table A1. We considered risk of bias using EEF padlock ratings to reflect practice by EEF. Padlock ratings ranged between 0 and 5 with higher padlocks indicating better quality and security ratings for the results from that trial.

Table 1: Distribution of the trials by study design, including number (n) of trials, schools and pupils.

Study Design	Trials (n)	Schools (n)	Pupils (n)
CRT	51	3105	314455
MST	34	985	143451
QUASI	2	294	67021
RDD	1	12	607

3.2 Study outcomes

The outcomes in most trials were literacy and mathematics. These study outcomes were immediate post-test outcomes of the trial. Other outcomes were excluded from meta-analysis due to the nature of the outcomes and their potential underlying constructs. We have used all the primary and secondary outcomes from the trials. Post-test outcomes were mainly continuous data (including ordinal scores). All outcomes were meta-analysed according to pre-defined groups as described in the section 3.3 independent of whether they were primary or secondary outcomes. It is important to note that in the context of evidence synthesis, the false positives are implicitly controlled since the inference is based on pooled evidence across the trials. Hence, adjustment for multiple testing is redundant and not undertaken (Brookes *et al*., 2001). The outcomes in the trials under study were either National Pupil Database (NPD) scores or collected directly by the evaluators preferred measures of literacy and mathematics.

3.3 Grouping variables

Four major groups of variables (primary or secondary outcomes, Key Stages, type of intervention and study design) were considered for the meta-analysis. The effect of EEF interventions was assessed across the Key Stages (KS1, KS2, KS3 and KS4) separately for each Key Stage. Please note that KS1 also includes the data for one trial with the Early years study outcome. The outcomes were also meta-analysed by subgroups of type of intervention in order to determine which group of interventions is more beneficial for FSM pupils. Type of interventions were classified as one-to-one, small group, whole class or whole school. This

classification was adopted from the EEF Evidence database project. Evidence from previous meta-analyses suggests that targeted small group and individual interventions are beneficial for children's educational outcomes (Lou *et al*., 2001). We also meta-analysed the outcomes by study design, where two most common study designs, Clustered randomised trials (CRT) and Multisite trials (MST), were considered. The aim of including study design is to assess whether the choice of study design impacted the outcome of an intervention. Further subgroup analysis was undertaken by cross-classification of the study outcomes by Key Stages, type of intervention and study design. Ideally, these cross-classifications should produce at least 32 subgroups for each outcome. However, we ended up with 10 subgroups for literacy and 7 for maths due to the lack of sufficient trials in most of the resulting strata.

3.4 Two-Stage Meta-analysis Method

A traditional meta-analysis approach mainly aggregates effect sizes from different studies by weighting them proportionally to study-specific variability and the variability between trials. The major drawback of this approach is the loss of information, which is typical of any summarised data. Another limitation is that sometimes the different effect sizes were calculated differently using different statistical approaches and scaling factors. For example, the use of conditional or unconditional variance may result in different estimates of the magnitude of the effect. Retaining the same framework for traditional meta-analysis methods, we proposed to re-estimate an effect size for all trials using the same, consistent methods. Although this approach will not correct for the loss of information, it will reduce variability between effect sizes attributable to analytical approach. Our proposed two-stage meta-analysis involves two steps.

Stage 1: Calculating effect size per trial

Individual trials were analysed independently using the multilevel model (MLM) specified in equation 1. Let Y_{ijk} be the outcome data for pupil i from school j in trial k, the two-level model for each trial is formulated as:

$$Y_{ijk} = \beta_{0\mathrm{k}} + \beta_{1\mathrm{k}} Pret_{ijk} + \beta_{2\mathrm{k}} T_{ijk} + b_{jk} + \varepsilon_{ijk} \quad (1)$$

where, $\beta_{0\mathrm{k}}$ is the overall intercept, $\beta_{1\mathrm{k}}$ is the gradient between post- and pre-test scores, $\beta_{2\mathrm{k}}$ is the adjusted difference between the intervention and control groups based on the indicator for intervention T_{ijk} , defined as $T_{ijk} = 1$ for intervention (treatment) group and $T_{ijk} = 0$ for comparison group for a two-arm trial. $b_{jk} \sim N(0, \omega_{\mathrm{k}} * \omega_{\mathrm{k}})$ captures between-school variability and $\varepsilon_{ijk} \sim N(0, \sigma_{\mathrm{k}} * \sigma_{\mathrm{k}}))$ denotes residual variance. Furthermore, the effect size and its confidence intervals for each trial were calculated as

$$ES_k = \frac{\beta_{2\mathrm{k}}}{\sqrt{{\omega_k}^2 + {\sigma_\mathrm{k}}^2}}, \qquad CI\_lower_k = \frac{Lower(\beta_{2\mathrm{k}})}{\sqrt{{\omega_k}^2 + {\sigma_\mathrm{k}}^2}}, \qquad CI\_upper_k = \frac{Upper(\beta_{2\mathrm{k}})}{\sqrt{{\omega_k}^2 + {\sigma_\mathrm{k}}^2}}.$$

Where $Lower(\beta_2)$ and $Upper(\beta_2)$ are 95% confidence intervals for the adjusted difference between the intervention and comparison group (β_2). Also note that the post-test scores were standardised pre-analysis by subtracting the mean and then divided by their standard deviation, $ES_k = \beta_{2\mathrm{k}}$, $CI_{lower_k} = Lower(\beta_2)$ and $CI_{upper_k} = Upper(\beta_2)$. lme4 package in R was used to fit the multilevel model and to estimate all the parameters.

Stage 2: Weighted Average

The standard error of effect size from trial k (SE_k) was calculated from the confidence interval (CI_upper_k, CI_lower_k) of ES_k as shown in equation 2 and adapted from the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins *et al*., 2019).

$$SE_k = \frac{CI\_upper_k - CI\_lower_k}{3.92} \tag{2}$$

Given that all EEF interventions were not implemented in similar settings, both fixed-effect and random-effects meta-analysis was used to summarise the impact of EEF interventions. The random-effects approach assumes that there is not only one true effect size but a distribution of effects due to differing interventions. In this case, between-trial heterogeneity (τ^2) has to be taken into account (Borenstein *et al.,* 2011), whilst the trials are assumed to be homogenous in fixed effect meta-analysis.

Based on the estimated effect size (ES_k) in stage 1 and τ^2, the weighted average effect size or pooled effect size was calculated as,

$$Pooled\ ES = \frac{\sum_{k=1}^{K} W_k ES_k}{\sum_{k=1}^{K} W_k} \tag{3}$$

Where $W_k = ({SE_k}^2 + \tau^2)^{-1}$ is the weight for the individual trial based on variability for each effect size and the heterogeneity between trials (Hedges & Olkin, 1985; Pigott, 2012). Specific to education trials, the SE_k also accounted for between-school variability when a multilevel model is used. Although this approach provides the global impact of the interventions, it suffers from loss due to the two-stage approach for obtaining the pooled effect size. This type of bias is called the ecological fallacy (Ess & Sudweeks, 2001; Reade *et al.*, 2008) as it does not account for heterogeneity at the individual level (Debray *et al.*, 2015).

3.5 IPD meta-analysis

An IPD meta-analysis method offers a more flexible and pragmatic way to synthesise evidence from existing interventions (Burke *et al*., 2016; Kontopantelis, 2018). It is a more powerful approach than traditional meta-analysis or a two-stage approach because of its ability to pool information across multiple trials, while also accounting for the different sources of variation (Debray *et al*., 2015; Smith, 2016). IPD meta-analysis allows important baseline data and trial-specific characteristics to be accounted for in the same model. Although IPD is more attractive because it fully exploits the available data of individual participants without having to perform additional transition steps (Fanshawe & Perera, 2019), in some cases it produces similar results as a two-stage approach.

IPD meta-analysis can be considered as an extension of a multilevel model where two-level models are extended to incorporate a third level to capture heterogeneity between trials. Within a Bayesian framework (Burke *et al.,* 2017), pupils (level 1) are nested within schools (level 2) and schools are nested within trials (level 3). Let Y_{ijk} be the outcome data for pupil i from school j who participated in trial k as previously defined, a full IPD meta-analysis model can be formulated as:

$$Y_{ijk} = (b_{0k} + \varphi_0) + (b_{1k} + \varphi_1)Pret_{ijk} + (b_{2k} + \varphi_2)T_{ijk} + S_{jk} + \varepsilon_{ijk} \quad (4)$$

where φ_0, $\varphi_1$and φ_2 are the pooled intercept, gradient between pre-test and post-test, and treatment effect across trials. Whilst $b_{0k} \sim N(0, \tau_{\mathrm{k}} * \tau_{\mathrm{k}})$, $b_{1k} \sim N(0, \vartheta_{\mathrm{k}} * \vartheta_{\mathrm{k}})$ and $b_{2k} \sim N(0, \delta_{\mathrm{k}} * \delta_{\mathrm{k}})$ are the trial-specific deviations from the pooled intercept, gradient between pre-test and post-test, and the treatment effects. The additional sources of variation within each trial are captured by $S_{jk} \sim N(0, \omega_{\mathrm{sk}} * \omega_{\mathrm{sk}})$ and $\varepsilon_{ijk} \sim N(0, \sigma_{\mathrm{k}} * \sigma_{\mathrm{k}})$. ω_{sk} denotes heterogeneity between schools in trial k and σ_{k} captures between-pupil variability in trial k.

This model formulation highlights the first challenge with an IPD meta-analysis of evidence from educational trials. The pooled effect of the intervention (φ_2) is only meaningful if the outcomes in each trial are on the same scale, which is often not the case in educational trials. A further challenge is that there is no single measure of heterogeneity between schools (ω_{sk}^2) and within pupils (σ_{k}) per trial except if one is willing to make unrealistic assumptions that $\omega_{sk}^2 = \omega_k^2$ and $\sigma_k^2 = \sigma^2$. Outcome measures in education trials are generally very variable between trials even when measuring the same outcome, due to the fact that each education trial is based on a convenience sample of schools willing to take part in the trial. An even more complicated issue is that the outcome in each trial can be from any of the Key Stages or may use a bespoke test. Additional sources of variability typical in education trials are the nature

of the pre-test scores and how strongly they are correlated with the outcome data. A further challenge is that one cannot safely assume that effect sizes from each trial are from a single distribution or even driven by common underlying factors. This is partly the reason that IPD meta-analysis is not a common approach in education trials despite the methodological advancement in health and clinical trials.

3.6 Simplified IPD Meta-analysis model (sIPD)

The IPD meta-analysis model cannot be directly applied to educational trials without further considerations. We propose to first eliminate heterogeneity between trials by scaling the post-test and pre-test outcome data to a unit variance of 1 per trial. This scaling approach is statistically not the most desirable approach, but it offers the best trade-off in balancing between the challenges of the model and ensuring meaningful results. The scaling of the raw outcome data in each trial was performed separately using mean and standard deviation of the scores within each trial.

The other issue that needs to be addressed is relaxing the assumption that the effects of the interventions are from a single distribution with common mean (φ_2) because the trial-specific impact ($b_{2k} + \varphi_2$) will shrink toward the pooled effects (Kruschke, 2015, Lesaffre & Lawson, 2012; Duchateau, Janssen & Rowlands, 1998). Depending on the shrinkage factor, these estimates may differ from the corresponding estimates from a two-stage meta-analysis approach and the individual effect size in the evaluation report of the different trials. The amount of shrinkage will depend on the extent of the variability (the between-trial variability (τ_k^2), the within-trial variability ($\omega_{sk}^2 + \sigma_k^2$), and the number of schools and pupils in each trial (Laird, 2004). Although the scaling of the post-test and pre-test outcome data removes the between-trial variability, within-trial variability may remain substantially different between the trials. Due to this within-trial variability, a less heterogeneous trial will be disadvantaged, because the lower the between-trial variance, the greater the shrinkage effect (Duchateau, Janssen, & Rowlands, 1998).

To retain the power of an IPD meta-analysis and to ensure meaning of the results in the context of educational interventions, we proposed a simplified IPD meta-analysis model as

$$Y_{ijk}^{s} = \beta_{0k} + \beta_{1k} Pret_{ijk}^{s} + \beta_{2k} T_{ijk} + S_{jk} + \varepsilon_{ijk} \quad (5)$$

Where, Y_{ijk}^{s} and $Pret_{ijk}^{s}$ are standardised post-test and pre-test scores. β_{0k} is a fixed intercept, β_{1k} is a fixed gradient between the standardised post-test and pre-test scores and β_{2k} is the

average effect of the intervention in trial k. However, $S_{jk} \sim N(0, \omega_{\mathrm{sk}} * \omega_{\mathrm{sk}})$ and $\varepsilon_{ijk} \sim N(0, \sigma_{\mathrm{k}} * \sigma_{\mathrm{k}})$ remained as random effects in the model. In order to obtain the pooled effect size, we use,

$$\varphi_2 = \frac{\sum_{k=1}^{K} W_k \beta_{2k}}{\sum_{k=1}^{K} W_k} \quad (6)$$

Where, $W_k = ({\omega_{sk}}^2 + {\sigma_k}^2)^{-1}$ captures within-trial variability given that between-trial variability was pre-scaled to 1. This simplified IPD model is expected to produce results consistent with the two-stage meta-analysis approach and the effect size from the evaluation report for each trial where a multilevel model was used to estimate effect size using conditional variance. Two-stage and IPD meta-analysis methods may produce different results when some studies have unbalanced sample sizes between the treatment and control groups (Danielle *et al.,* 2017).

The proposed IPD meta-analysis method for educational trials was implemented within the Bayesian framework assuming vague normal priors for all fixed effects and vague inverse-gamma priors for all the variance parameters and R2jags R software package. The credible intervals for the pooled effect size and the trial-specific effect size were obtained as 2.5% and 97.5% quantiles from their posterior distributions. To ensure convergence of the parameters, we used three chains with 200,000 Markov Chain Monte Carlo (MCMC) iterations. The first half of each chain was discarded as 'burn-in' part. All results were reported after checking for convergence had been reached using Rhat and trace plots. The separate meta-analysis models were fitted for literacy and maths outcomes using all available data. However, further meta-analyses were performed by different factors such as Key Stage, intervention types, and study design.

3.7 Attainment Gaps

The meta-analysis of effect sizes for only FSM pupils does not provide insight into whether EEF interventions have reduced attainment gaps between them and their peers. It is possible that an intervention will have the same effect on FSM and non-FSM pupils and in such a situation there may be a positive effect in FSM pupils, but the attainment gap will be zero for the specific trial. Another possibility is that an intervention may have no effect on FSM pupils, but may have a positive effect on non-FSM pupils. In such a situation, the intervention is likely to widen the attainment gap. Lastly, an intervention may have a positive effect on FSM pupils and no effect on non-FSM pupils. Such an intervention is likely to reduce the attainment gap as more FSM pupils have improved their educational outcomes. Although this illustration is for individual trials, it is also a possibility for a pooled evidence of impact of EEF interventions. To estimate the attainment gap between FSM and non-FSM pupils, the model specified in

equation 5 was extended with an interaction term between FSM and intervention groups (Kontopantelis, 2018) and using data for all pupils as follows:

$$Y_{ijk}^{s} = \beta_{0k} + \beta_{1k} Pret_{ijk}^{s} + \beta_{2k} T_{ijk} + \gamma_{1k} FSM_{ijk} + \gamma_{2k} T_{ijk} * FSM_{ijk} + S_{jk} + \varepsilon_{ijk} \quad (7)$$

Parameter γ_{2k} is the attainment gap i.e. difference in average effect of the interventions between FSM pupils and their peers in trial k and the impact of the intervention on FSM pupils in trial k, β_{2k} is the impact of the intervention on non-FSM pupils in trial k, and the impact of the intervention on FSM pupils in trial k is $\beta_{2k} + \gamma_{2k}$.

In order to estimate the pooled effect of the EEF interventions on attainment gap, the model is further specified as

$$\text{Attainment Gap } (\eta) = \frac{\sum_{k=1}^{K} V_k \gamma_{2k}}{\sum_{k=1}^{K} V_k} \quad (8)$$

Where $V_k = (\omega_{sk}^2 + \sigma_k^2)^{-1}$. The model was fitted within a Bayesian framework using the same sets of priors as previously defined. The attainment gap was also estimated using the sIPD meta-analytic approach by simply adding an interaction between treatment and FSM variables in the model defined in equation 5 and estimating the attainment gap from each trial and pooling the attainment gap estimate together using the methods provided in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins *et al*., 2019).

3.8 Heterogeneity in the FSM and attainment gap sIPD meta-analysis

We measured the statistical heterogeneity using the statistical test usually applied in meta-analysis for determining whether there is true heterogeneity among the studies' effects adopting the Q test proposed by Cochran (1954) and also described in Bowden *et al.* (2011). The Q-statistic used in this study is defined as

$$Q = \begin{cases} \sum_{k=1}^{K} W_k (\varphi_2 - \beta_{2k})^2 & for\ FSM\ subgroup \\ \sum_{k=1}^{K} V_k (\eta - \gamma_{2k})^2 & for\ \text{Attainment Gap} \end{cases}$$

Further, I^2 index proposed by Higgins and Thompson (2002) was also estimated. This index quantifies the extent of heterogeneity from a collection of effect sizes by comparing the Q value to its expected value assuming homogeneity, that is, to its degrees of freedom (df = k – 1):

$$I^2 = max\left(\frac{Q-(k-1)}{Q} * 100\%, \quad 0\right)$$

4.0 Results

Table 2 and Table 3 provide a summary of the trial outcomes and the number of pupils, schools, and FSM (and non-FSM) eligible pupils by primary/secondary, Key Stages, type of intervention and study design.

The outcomes for meta-analysis in this report were literacy and maths outcomes from 81 and 48 trials, respectively. 72 out of the 81 trials considered literacy as a primary outcome, whilst 9 trials reported literacy as a secondary outcome. Similarly, 42 out of the 48 trials reported maths as a primary outcome, whilst 6 trials reported it as a secondary outcome. Among the 81 trials that assessed literacy as primary or secondary outcomes, 13 trials assessed Key Stage 1, 33 trials assessed Key Stage 2, 29 trials assessed Key Stage 3 and 6 trials assessed Key Stage 4 literacy scores. Similarly, 9, 24, 9 and 6 trials assessed maths Key Stage 1, Key Stage 2, Key Stage 3, and Key Stage 4 scores, respectively. Furthermore, for literacy, 24, 17, 30 and 10 trials were assessed as one-to-one, small group, whole-class and whole-school intervention types, respectively. There were also 10, 7, 23 and 8 one-to-one, small group, whole-class and whole-school types of interventions for maths outcomes, respectively.

Table 2: Overview of literacy trials by outcome types, study design, and types of intervention

		Trials (n)	Schools (n)	Pupils (n)	FSM pupils (n)	Non-FSM pupils (n)
	Overall	81	4000	302138	90218	211920
Outcome Types	Primary	72	3427	262321	78261	184060
	Secondary	9	573	39817	11957	27860
Key Stage Outcomes	KS1	13	529	19905	4444	15461
	KS2	33	2265	102835	34085	68750
	KS3	29	552	39297	10108	29189
	KS4	6	654	140101	41581	98520
Type of interventions	One-to-one	24	1358	97368	28194	69174
	Small group	17	503	22451	6914	15537
	Whole class	30	1339	83550	29774	53776
	Whole school	10	800	98769	25336	73433
Study Design	CRT	46	3011	205928	63358	142570
	MST	32	688	31456	8914	22542

In terms of study design, 46 and 32 trials assessing literacy used cluster-randomised trials (CRT) and multisite trial design (MST), respectively. Similarly, 36 and 11 trials assessing

maths used CRT and MST, respectively. Overall, there were 211,920 instances of FSM pupils from 4000 instances of schools with literacy outcomes and 217,728 instances of FSM pupils from 3178 instances of schools with maths outcomes. We have reported on instances of pupils and schools because there was no indicator to uniquely identify the schools and pupils across the trials.

Table 3: Overview of maths trials by outcome types, study design, types of intervention

		Trials (n)	Schools (n)	Pupils (n)	FSM pupils (n)	Non-FSM pupils (n)
	Overall	48	3178	306975	89247	217728
Outcome Types	Primary	42	2686	275461	79838	195623
	Secondary	6	492	31514	9409	22105
Key Stage Outcomes	KS1	9	639	18718	4394	14324
	KS2	24	1577	79671	25946	53725
	KS3	9	269	30434	6667	23767
	KS4	6	693	178152	52240	125912
Types of intervention	One-to-one	10	857	117290	33754	83536
	Small group	7	496	18391	5032	13359
	Whole class	23	1210	75525	26632	48893
	Whole school	8	615	95769	23829	71940
Study design	CRT	36	2584	186257	56293	129964
	MST	11	582	119955	32519	87436

4.1 Heterogeneity between trials

An important consideration in the meta-analysis of existing evidence is how comparable are the measures of treatment or intervention effects. Variability between trials due to different participating populations, different outcomes with respect to scale or underlying constructs, differences in methods of calculating the effect sizes, and differences in quality of the trials plays a significant role in estimating pooled effects across trials. There is a consensus that variable measures of intervention effects are likely to produce unreliable evidence of the average effects of the interventions across trials, although some of the variability between trials can be accounted for in a random effects meta-analysis.

The level of variability between trials is particularly important in IPD meta-analysis because the individual's data are analysed, which is likely to be different between trials. It is also well known that schools and pupils participating in educational trials are rarely representative of the wider population of schools and pupils. The percentage of variability between trials, schools and residual variance (pupils) for literacy and maths outcomes were presented in Table 4 and Table 5, respectively. The differences between trials accounted for 86% of the

variability in literacy outcomes across trials and 87% of the variability in the maths outcomes when raw data were used. However, using standardised scores of post-test and pre-test outcomes were showing consistent patterns as normally observed in education trials.

Table 4: Estimates and percentage of variability in literacy outcomes explained by differences between trials, differences between schools and residual variance (pupils).

	Pupils variance (%)	School variance (%)	Trial variance (%)
Literacy			
Raw	156.66(12%)	21.47(2%)	1125.80(86%)
Standardised	0.59(82%)	0.09(13%)	0.03(5%)

The majority of the variability in the outcomes was due to the differences between pupils and then due to differences between schools. The difference in effect sizes between trials is negligible. We share the view that IPD meta-analysis of educational trials without properly accounting for the huge heterogeneity between trials will be prone to misleading conclusions because of weighting in the overall effect size in favour of the least variable trial on the original scale of the outcome. The re-scaling of post-test and pre-test scores in each trial will reduce the variability between the trials as shown in Table 4 and Table 5. This approach is not without its own limitations as it may distort the distributions of the outcomes, particularly if the outcomes do not come from a common underlying construct.

Table 5: Estimates and percentage of variability in maths outcomes explained by differences between trials, differences between schools and residual variance (pupils).

	Pupils variance (%)	School variance (%)	Trial variance (%)
Mathematics			
Raw	119.90(11%)	20.79(2%)	946.29(87%)
Standardised	0.52(75%)	0.09(13%)	0.08(12%)

4.2 Simplified IPD model versus Two-stage models

We present the comparison of our proposed simplified IPD meta-analysis model and two-stage methods from Table A1 to Table A4 in Appendix 2. Most of the two-stage and one-stage IPD individual trial and pooled estimates correspond well in terms of direction and magnitude. However, the IPD model produced a greater effect size for literacy outcomes than the two-stage model. One of the reasons why the IPD model resulted in greater effect than a two-stage model may be because of how the weights are defined. The weights in the two-stage

models were defined using standard approximations from confidence intervals, whilst the IPD model directly used estimated variance from the data. Future work using synthetic data will aim to establish the superiority of the method against two stage meta-analysis methods. Table 6 provides an overview of the pooled effect size for FSM pupils' literacy outcomes from IPD meta-analysis and two-stage fixed-effect (FE) and random-effect (RE) meta-analysis using standardised outcome data.

Table 6: Overview of pooled effect size from IPD meta-analysis and two-stage fixed-effect (FE) and random-effect (RE) models using standardised outcome data.

	IPD	Two-stage Fixed-effect	Two-stage Random-effect
Outcome	Pooled ES	Pooled ES	Pooled ES
Literacy	0.06 (0.03, 0.08)	0.02 (0.00, 0.04)	0.03 (0.01, 0.06)
Maths	0.00 (-0.03,0.04)	0.01 (-0.01, 0.03)	0.00 (-0.02, 0.02)

4.3 Meta-analysis of intervention effects on FSM pupils

The pooled effect size for literacy as either a primary or secondary outcome across 81 trials is 0.06 (0.03, 0.08) (Table 7). This means on average EEF-funded interventions had positive benefits on the literacy outcomes of the FSM pupils who participated in the trials, an equivalent of about one month's progress. However, there is no evidence from the 48 trials analysed that EEF interventions had positive effects on the mathematics outcomes of FSM pupils with an effect size of 0.00 (-0.03, 0.04) (Table 7). It is important to note that there is also no evidence that the interventions on average were worsening their mathematics outcomes.

Table 7: Pooled ES and credible intervals for FSM subgroup literacy and maths outcomes.

Outcome	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
Literacy	81	3804	90218	0.06 (0.03, 0.08)
Maths	48	3006	89247	0.00 (-0.03,0.04)

Figure *2* illustrates the individual trial and pooled effect sizes with their credible intervals. The most beneficial interventions for FSM pupils with positive effects on their literacy outcomes were Shared Maths, Graduate Coaching Programme, Accelerated Reader, Online Reading Programme (ABRA), Butterfly Phonics, Response to Intervention, and Nuffield Early Language Intervention 1. The effect sizes for literacy outcomes ranged from -0.20 to 0.42. However, it is surprising that Shared Maths was one of the most effective interventions for

literacy since it was primarily intended to improve attainment in mathematics. Durham Shared Maths was a cross-age peer tutoring pedagogy which pairs older Year 5 pupils (tutors) with younger Year 3 pupils (tutees) to discuss and work through maths problems using a structured stepped approach. The intervention was delivered in the classroom and fits within the existing and ongoing maths teaching (Lloyd *et al*., 2015). This finding needs further evaluation of the process to understand that whether the shared nature of that particular intervention had a positive side effect, or some other factor might have triggered the reaction.

Although there was no evidence of overall effects on maths outcomes, there were promising interventions such as Dialogue Teaching, 'Powerful Learning Conversations', 'Improving Numeracy and Literacy', and 'Act, Sing, Play 1'. The individual trial effect sizes for maths outcomes ranged from -0.18 to 0.31.

Figure 2: Forest plot of effect sizes for literacy and maths outcomes from FSM pupils.

4.3.1 Meta-analysis by primary and secondary outcome

Since interventions in educational trials are mainly powered to improve primary outcomes, we investigated whether analysing primary and secondary outcomes separately would substantially change the pooled effect size. **Error! Reference source not found.** provides t he estimates of pooled effect sizes for literacy and maths outcomes by primary and secondary outcomes. The pooled effect sizes for primary and secondary literacy outcomes were 0.06

(0.03, 0.08) and 0.06 (-0.04, 0.16), respectively. The point estimate for the literacy pooled effect sizes were the same, but literacy as a secondary outcome had wider credible intervals due to fewer trials.

Table 8: Pooled ES and credible intervals for FSM subgroup by primary and secondary outcome.

Outcome	Group	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
Literacy	Primary	72	3250	78261	0.06 (0.03, 008)
	Secondary	9	554	11957	0.06 (-0.04, 0.16)
Maths	Primary	42	2568	79838	0.01 (-0.02,0.05)
	Secondary	6	438	9409	-0.07 (-0.15, 0.00)

The pooled effect sizes for maths as primary and secondary outcomes were 0.01 (-0.02, 0.05) and -0.07(-0.15, 0.00), respectively. Maths as a secondary outcome had on average negative effects on FSM pupils, though there were just six trials with maths as a secondary outcome.

Figure 3 provides the forest plot of effect sizes for literacy outcomes by primary and secondary outcome status. The effect sizes for literacy as a primary outcome ranged from -0.16 to 0.38, while for literacy as a secondary outcome ranged from -0.20 to 0.41. It is interesting that 'Shared Maths' and 'Flipped Learning' had a greater effect on literacy as a secondary outcome than the effect they had on maths as a primary outcome. As shown before, the most promising interventions for literacy were 'Graduate Coaching Programme', 'Accelerated Reader' and 'online reading programme (ABRA)'.

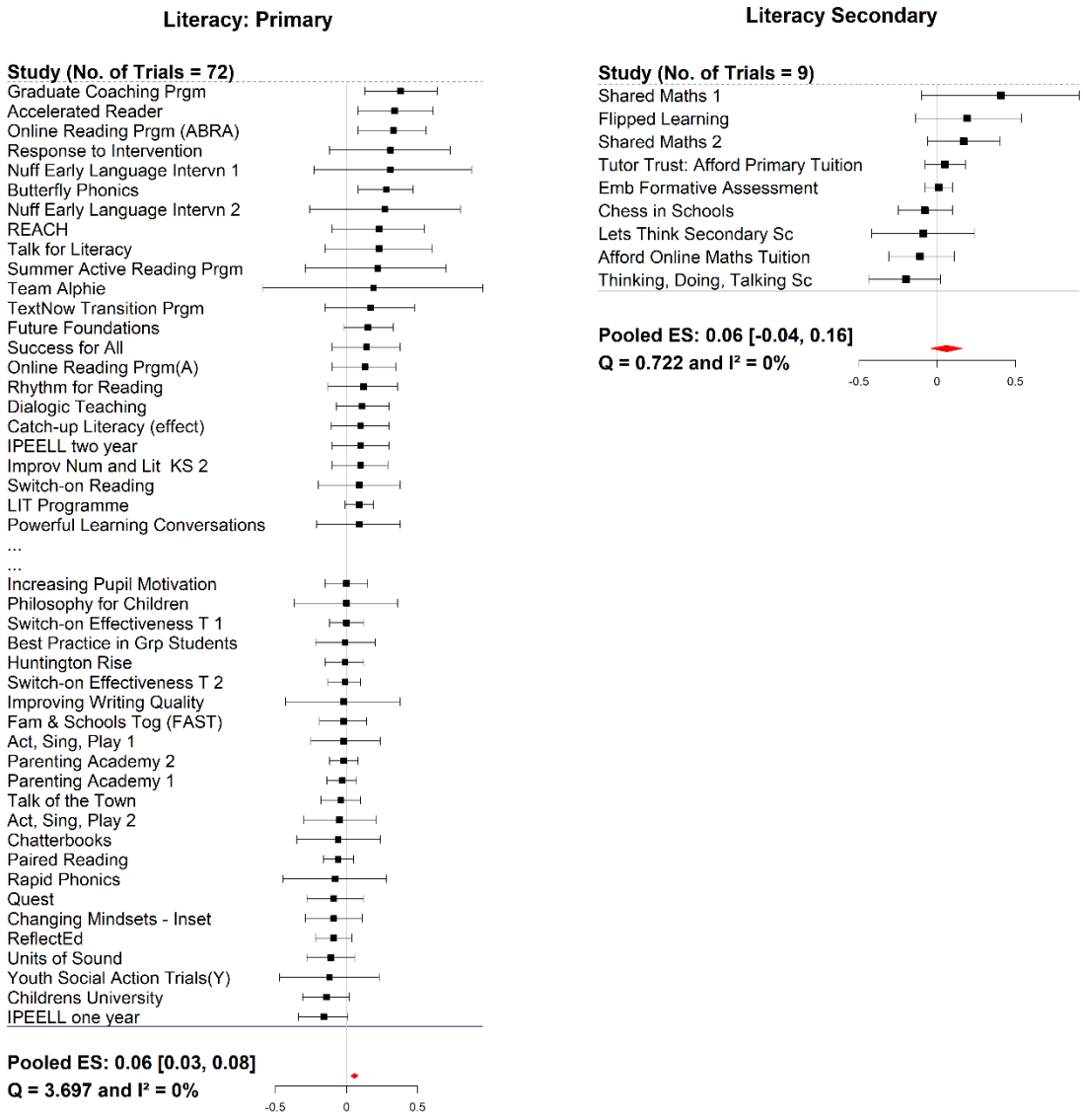

Figure 3: Forest plot of effect sizes for literacy outcomes by primary and secondary outcome status.

Both the primary and secondary literacy outcomes in Figure 3 corresponded well with the pattern observed for overall literacy outcomes in Figure 1. Individual estimates of the primary and secondary maths outcomes shown in Figure 4 suggests that the trials with outcomes other than literacy or mathematics were the major focus of such trials. For example, science score was the primary outcome for the 'Thinking, Doing and Talking Science' trial and GCSE overall attainment for the 'Embedded Formative Assessment' trial.

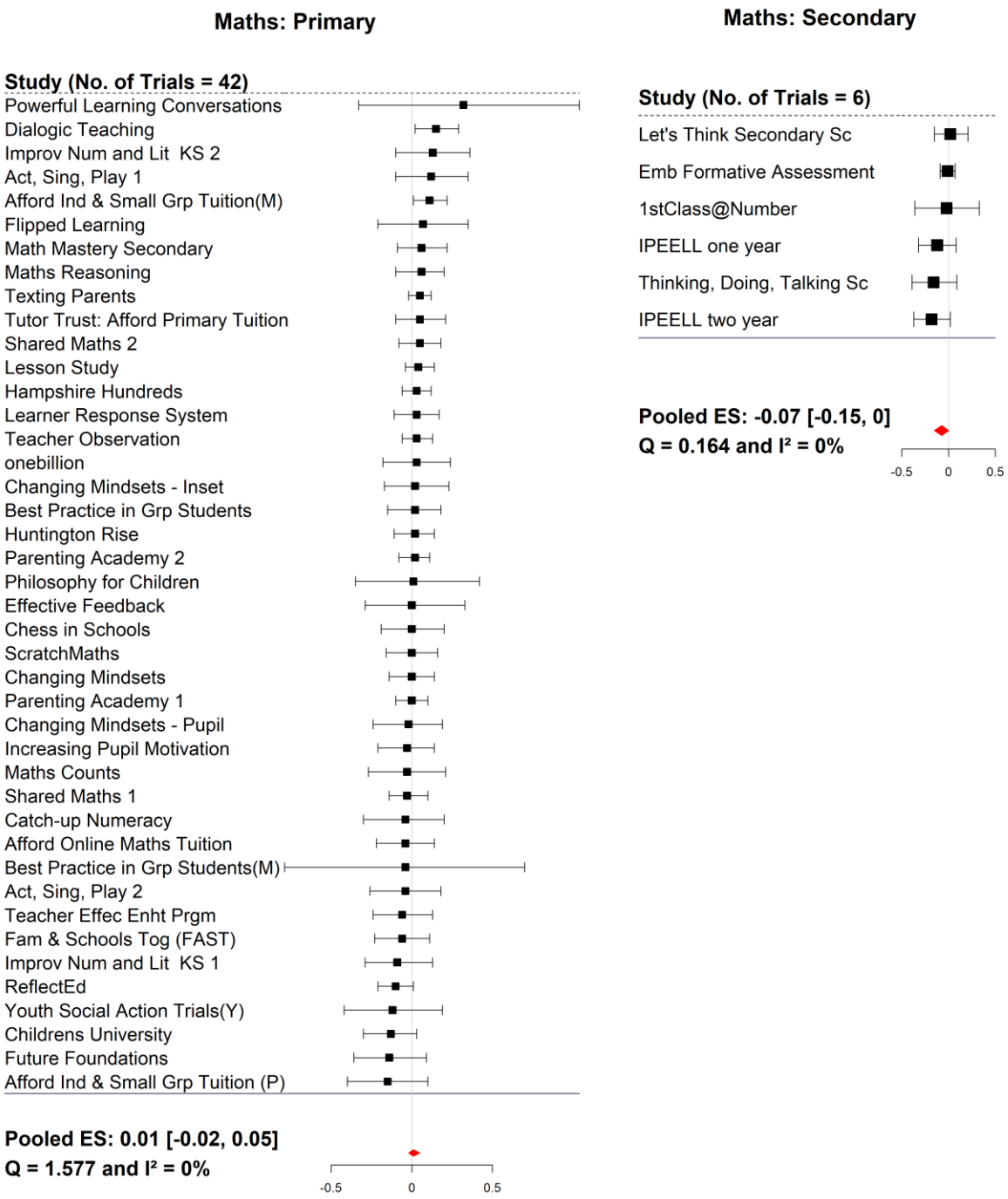

Figure 4: Forest plot of effect sizes for maths outcomes by primary and secondary outcome status.

4.3.2 Meta-analysis by Key Stages

Table 9 provides the estimates of pooled effect sizes for the literacy outcome by four Key Stages (KS). The maximum pooled effect size was observed for KS1 literacy outcome (pooled ES= 0.09 (0.02, 0.16)) followed by the KS3 literacy outcome (pooled ES =0.08 (0.03, 0.13)).

These results clearly suggest that EEF interventions had been most beneficial for FSM pupils in KS1 and KS3.

Table 9: Pooled ES and credible intervals for FSM subgroup literacy outcome by Key Stages (KS).

Key Stage	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
KS1	13	481	4444	0.09 (0.02, 0.16)
KS2	33	2175	34085	0.03 (-0.01, 0.07)
KS3	29	507	10108	0.08 (0.03, 0.13)
KS4	6	641	41581	0.02 (-0.05, 0.08)

Figure 5 provides the individual and pooled effect size estimates with their credible intervals for literacy by Key Stages (KS). The effect sizes for individual trials in KS1 were mostly positive (ten out of thirteen) with the maximum estimate of 0.34 (0.09, 0.57) for the 'Online reading program ABRA'. In KS2, 'Shared Maths' estimate (0.42 (-0.07, 0.94)) was the highest, followed by 'Response to Intervention' (0.32 (-0.09. 0.72)). Most trials in KS3 also had positive effects. 'Butterfly Phonics', 'Accelerated Reader', and 'Graduate Coaching program' were the trials most beneficial for FSM pupils in KS3. Five out of the six trials in KS4 had benefitted FSM pupils.

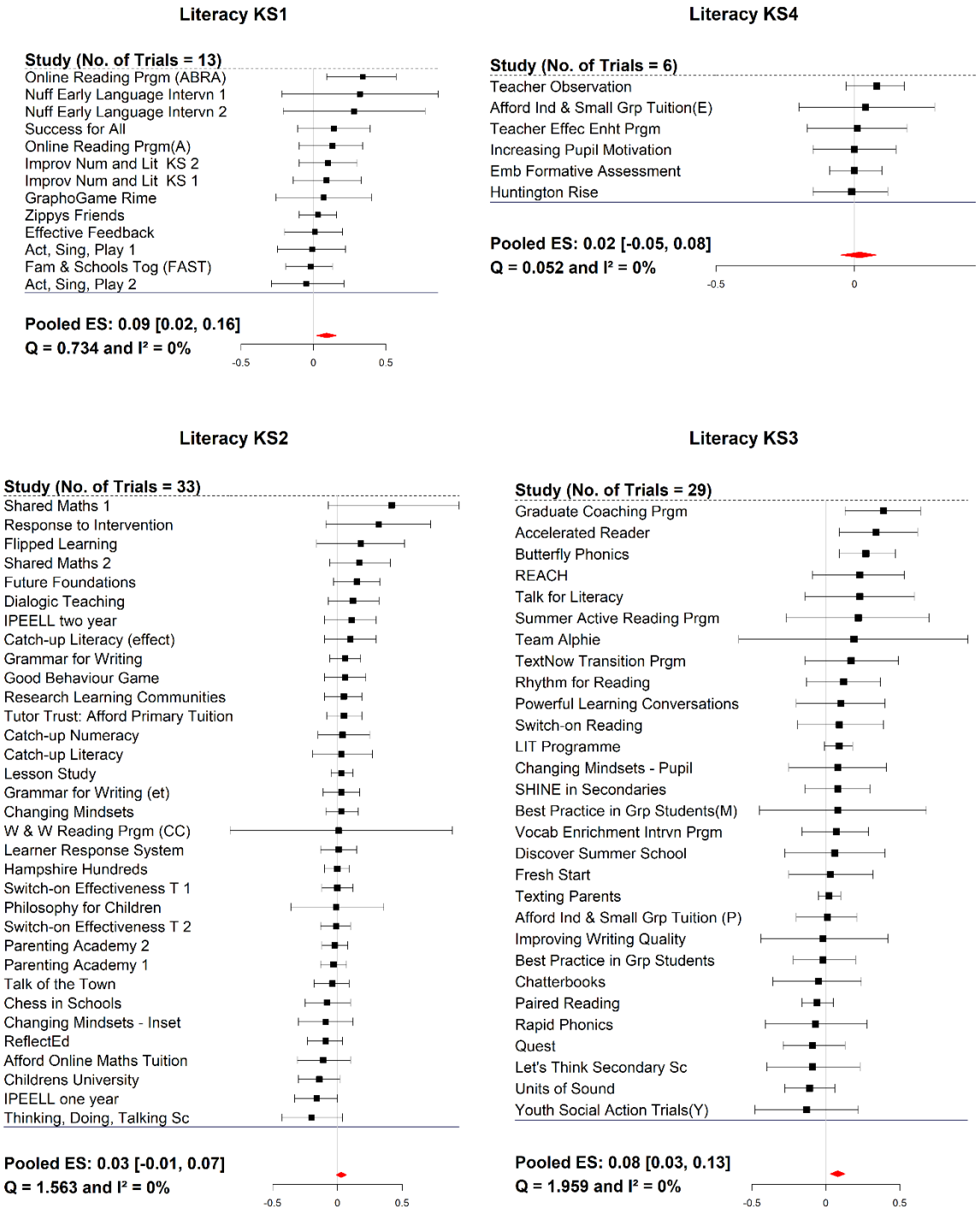

Figure 5: Forest plot of effect sizes for literacy outcome by Key Stages.

As evident from

Table *10*, the pooled estimate of effect size for the maths outcome was about 0.02 SD for KS1 and KS4. From both the literacy and maths outcome analysis, it was evident that EEF interventions had improved the literacy scores in all the Key Stages and maths scores in all

the Key Stages except Key stage 2 . However, the average effect on maths scores for the Key Stages 1, 3 and 4 was smaller than the average score for literacy.

Table 10: Pooled ES and credible intervals for FSM subgroup maths outcome by Key Stages.

Key Stages	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
KS1	9	540	4394	0.02 (-0.07, 0.11)
KS2	24	1524	25946	-0.01 (-0.04, 0.03)
KS3	9	261	6667	0.01 (-0.09, 0.12)
KS4	6	681	52240	0.02 (-0.03, 0.07)

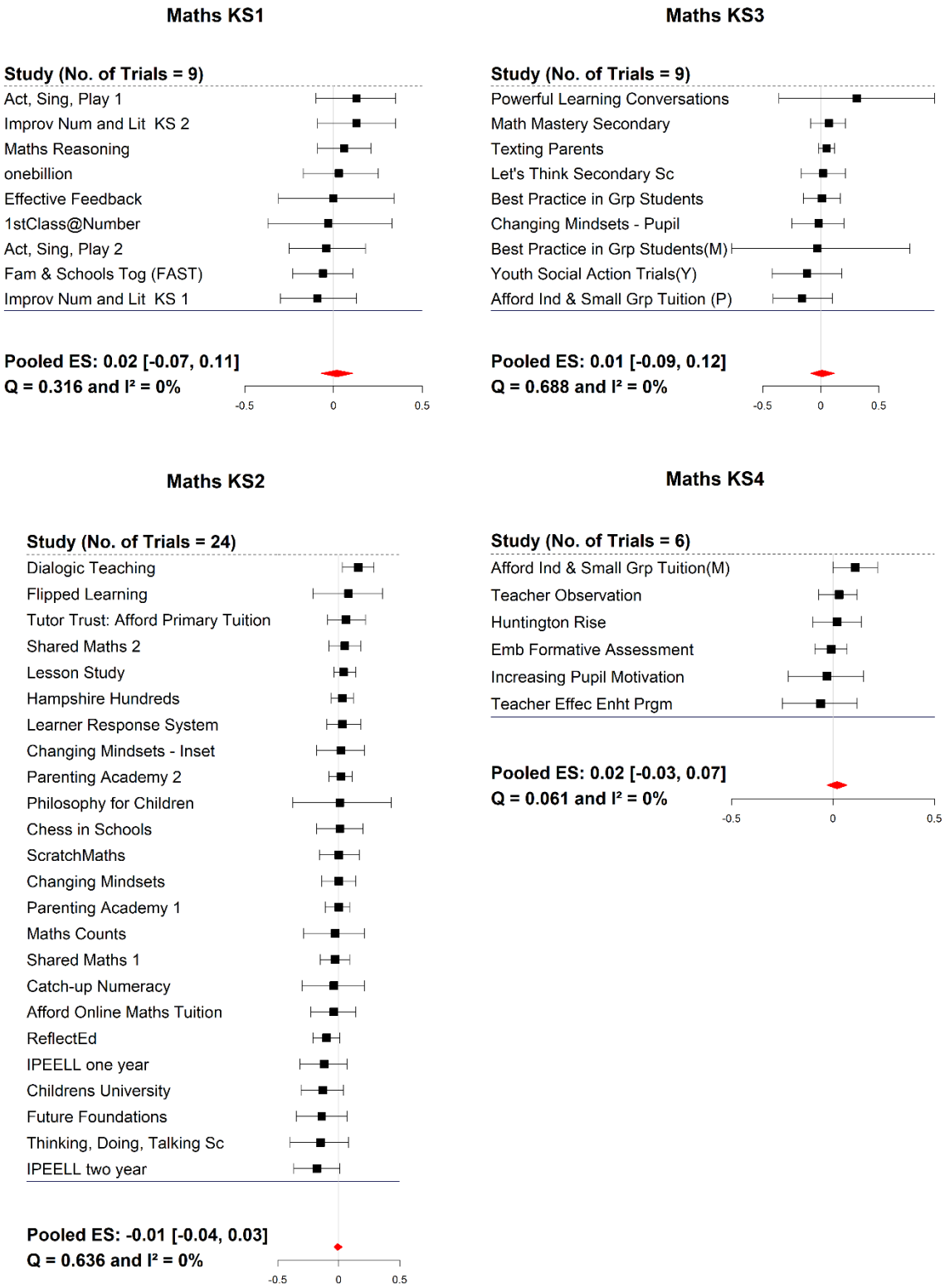

Figure *6* shows the effect size estimates for maths outcomes in KS1 to KS4. In KS1, 'Act, Sing and Play' and 'Improving Numeracy and literacy' were the two trials where the effect size was more than 0.10 SD. However, it is interesting to note that 'Act, Sing, Play (ASP)' offered music and drama tuition to Year 2 pupils and the impact was assessed in term of pupils' maths attainment. In KS2, there were few trials that had a positive impact on the FSM pupils' scores

and ‘Dialogue Teaching’ trial had a maximum impact (pooled ES = 0.16 (0.03, 0.29)). Half of the trials in KS2 had negative effect sizes and the other half had positive effect sizes, though it is worth noting that the larger trials in KS2 had mostly positive effect sizes. ‘Powerful Learning Conversations’ trial in KS3 was the most beneficial intervention for the pupils, followed by ‘Math Mastery Secondary’ and ‘Texting Parents’. It is important to highlight that the ‘Powerful Learning Conversations’ intervention involved a training programme for Year 9 English and Maths teachers, with the aim of improving feedback practices by applying techniques used in sport (Reinzo *et al*., 2016). ‘Affordable Individual Small Groups and Tuition’ trial had the biggest impact on KS4 maths outcomes.

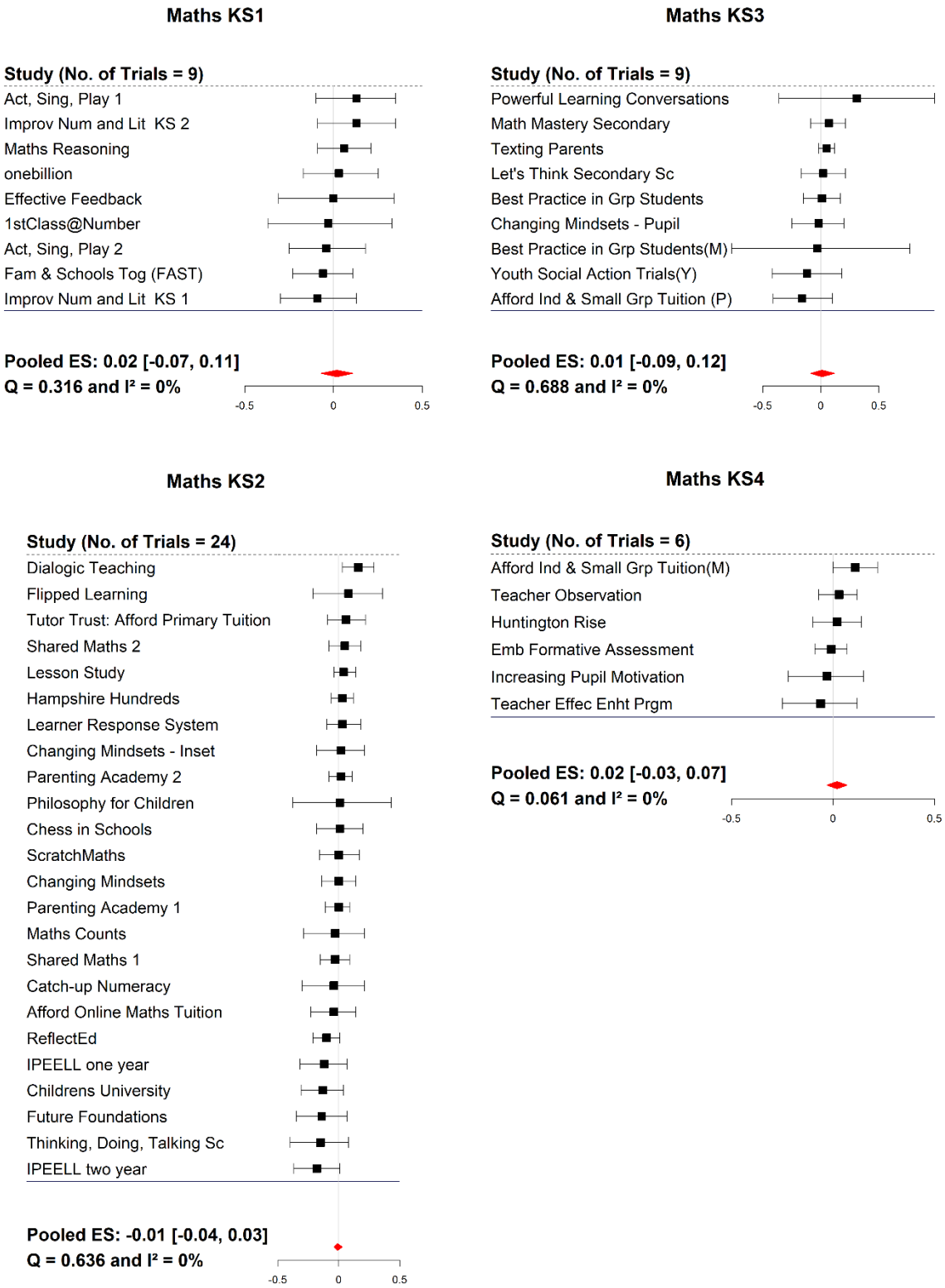

Figure 6: Forest plot of effect sizes for maths outcome by Key Stages.

4.3.3 Meta-analysis by types of intervention

The effects of one-to-one and small group interventions on literacy outcomes were greater than the whole class or whole school interventions. Small group interventions had the pooled effect size of 0.14 (0.06, 0.22), whilst one-to-one had an effect size of 0.08 (0.04, 0.13). Both types of intervention improved the literacy of FSM pupils by more than 0.10 SD, an equivalent of more than one month's progress. The results are presented in Table 11.

Table 11: Pooled ES and credible intervals for FSM subgroup literacy outcome by type of intervention.

Types of Interventions	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
One-to-one	24	1260	28194	0.08 (0.04, 0.13)
Small group	17	463	6914	0.14 (0.06, 0.22)
Whole class	30	1286	29774	0.01 (-0.04, 0.05)
Whole school	10	795	25336	0.02 (-0.02, 0.06)

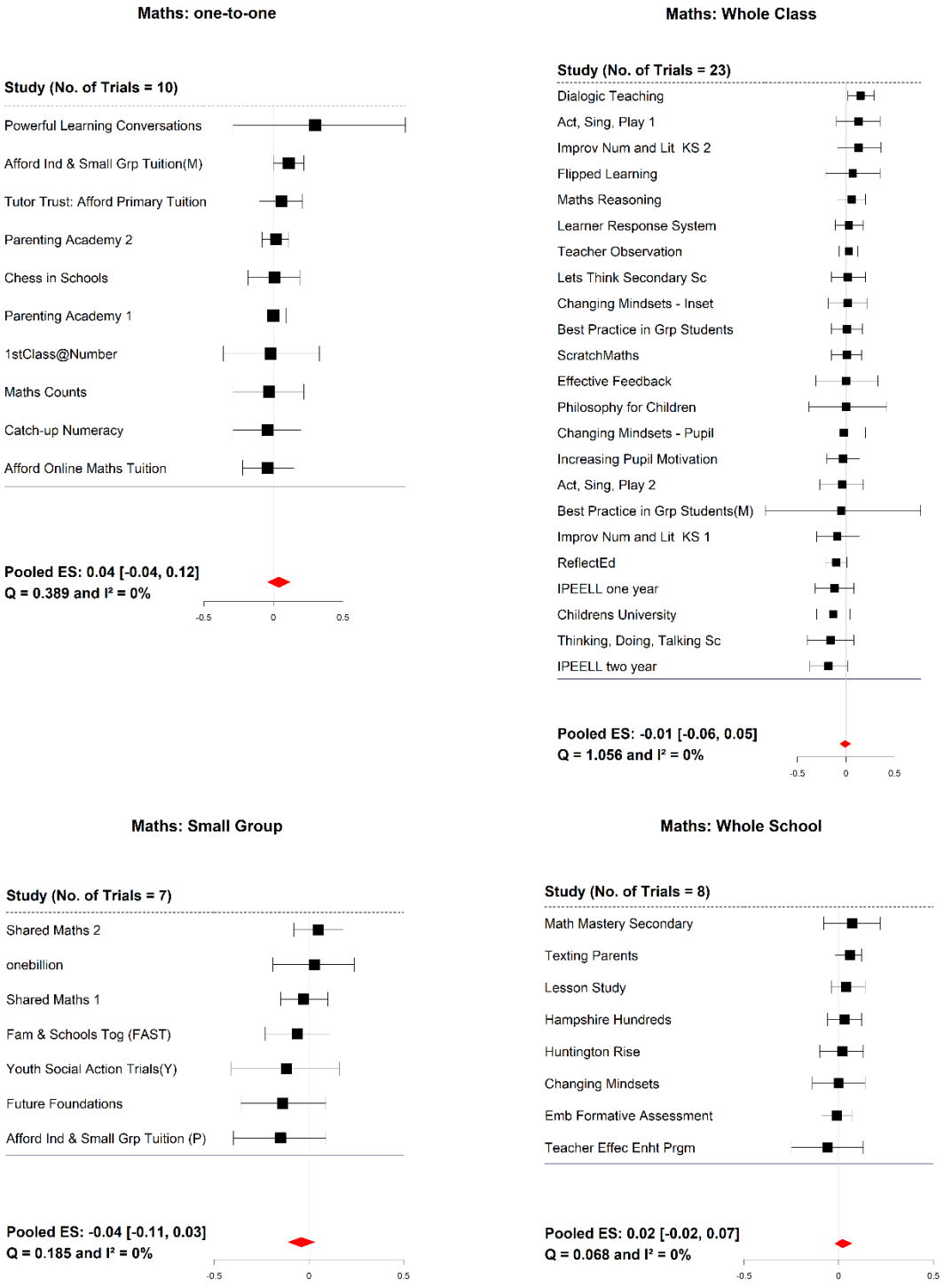

Figure *8* shows the effect sizes for the literacy outcome by the type of intervention. One-to-one interventions, namely, 'Graduate Coaching Programme', 'Accelerated Readers', 'Online Reading Programme (ABRA)' had educationally important effects on literacy. The pooled effect size of one-to-one trials was 0.08 SD. Nearly all small group trials had a positive effect size; the trial 'Shared Maths' had the biggest effect size, followed by 'Butterfly Phonics'. The pooled effect size for all small group trials was 0.14 SD.

Table 12 shows that one-to-one and whole-school interventions had positive effects on maths outcomes of FSM pupils. On the other side, small group and whole-class interventions had a negative impact. However, it should be noted that the number of FSM pupils in small group interventions was much smaller than the other types of interventions.

Table 12: Pooled ES and credible intervals for FSM subgroup maths outcome by type of intervention.

Types of Intervention	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
One-to-one	10	777	33754	0.04 (-0.04, 0.12)
Small group	7	452	5032	-0.04 (-0.11, 0.03)
Whole class	23	1163	26632	-0.01 (-0.06, 0.05)
Whole school	8	614	23829	0.02 (-0.02, 0.07)

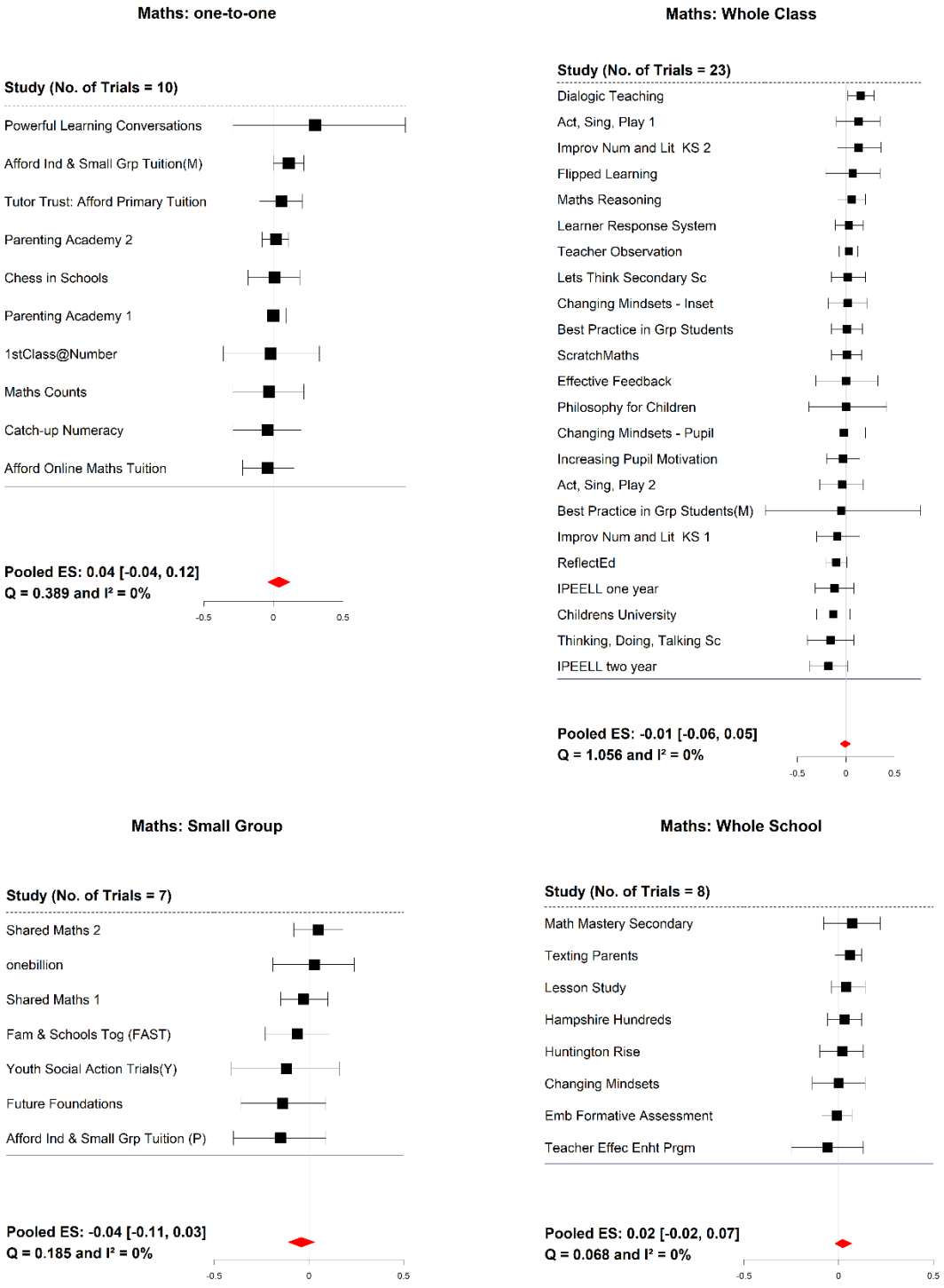

Figure *8* shows the effect sizes of each intervention type for the maths outcome. 'Powerful Learning Conversations' and 'Affordable Tuitions' projects were the most beneficial one-to-one intervention interventions. 'Shared maths' and 'onebillion' were the most beneficial small group interventions. Even though the pooled effect of the class level intervention was negative, several projects such as 'Dialogue Teaching', 'Act, Sing and Play', and 'Improving Numeracy and Literacy' trial had improved the FSM pupils' scores by more than 0.10 SD.

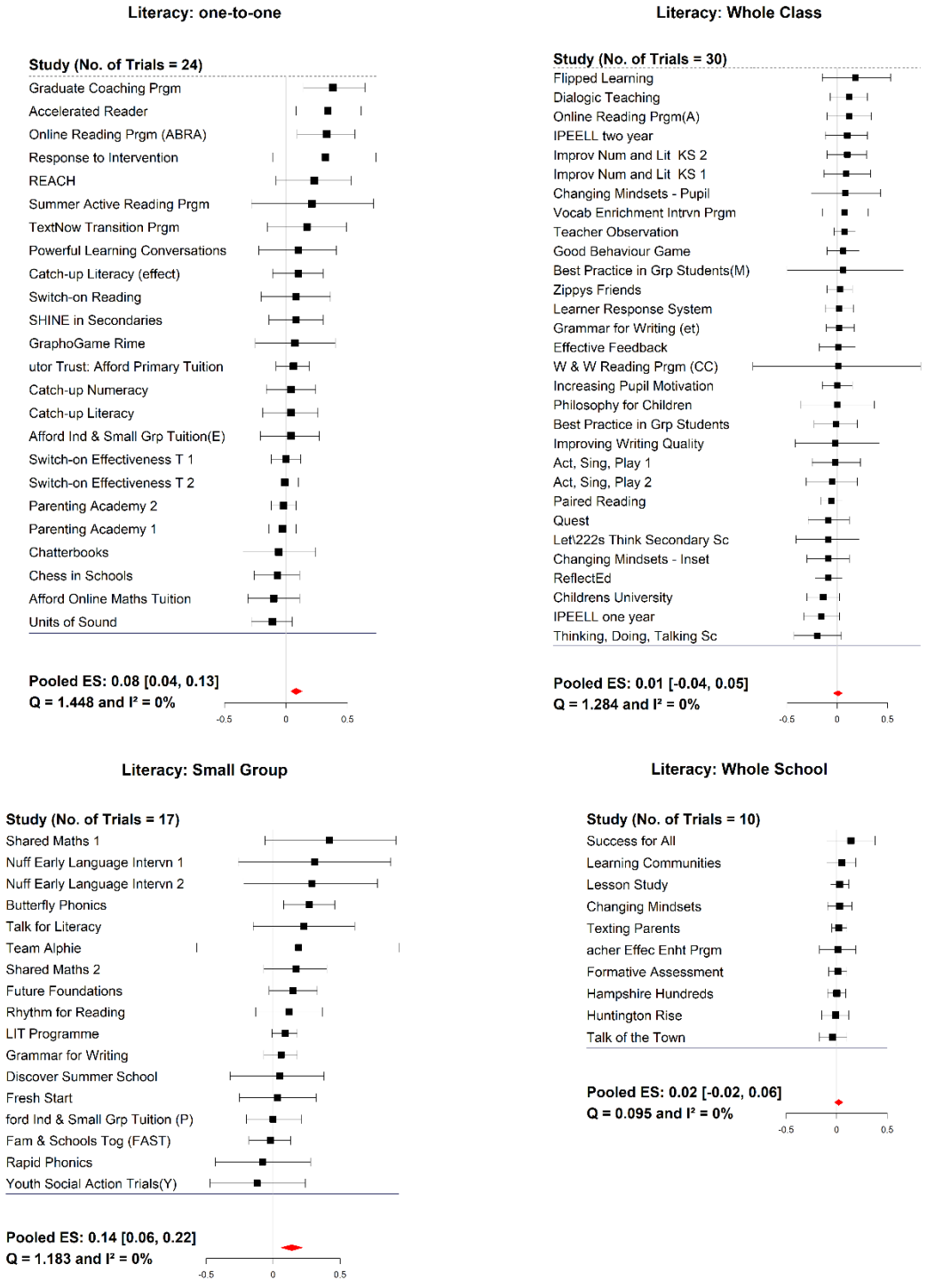

Figure 7: Forest plot of effect sizes for FSM subgroup literacy outcome by type of intervention.

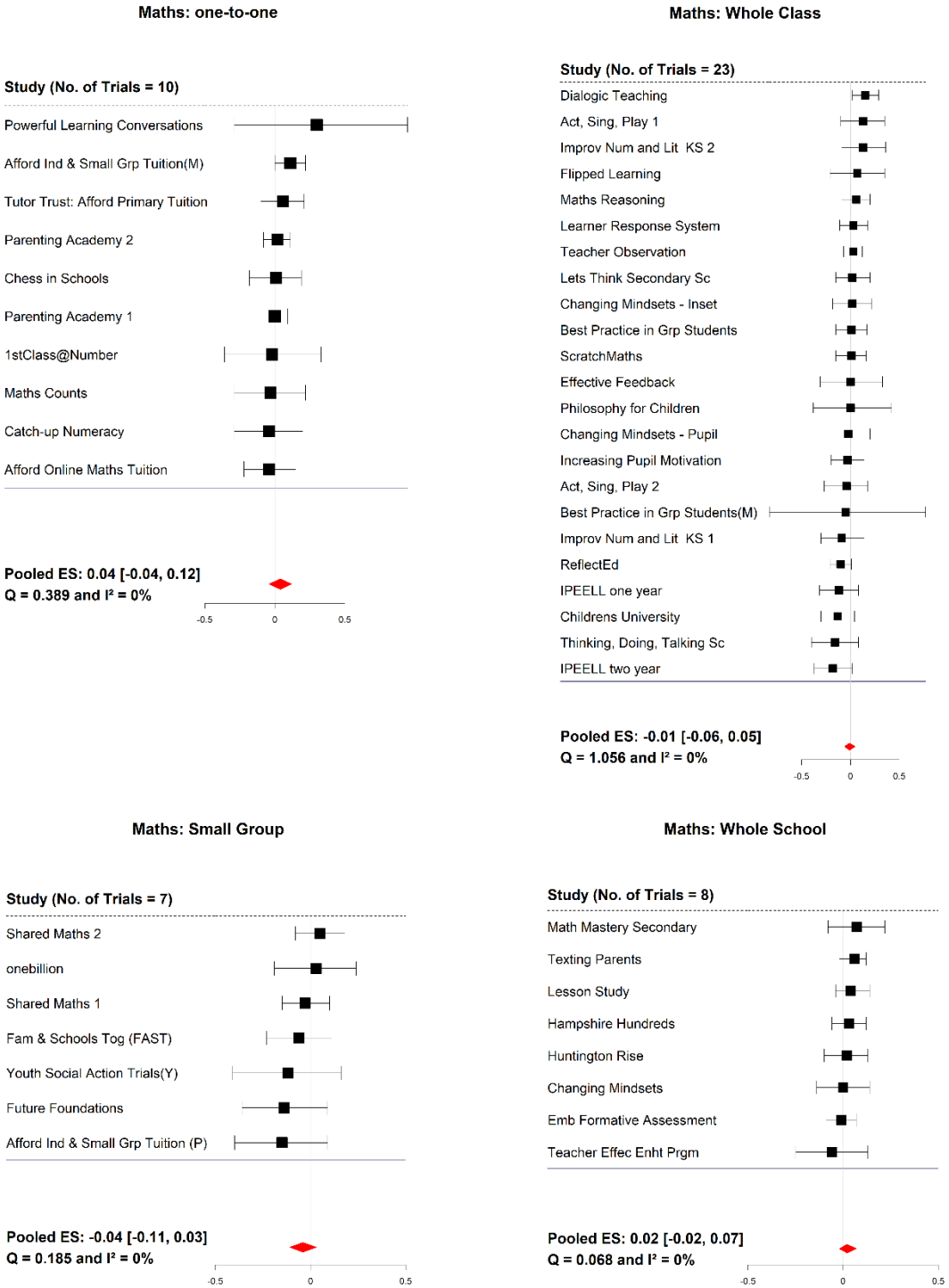

Figure 8: Forest plot of effect sizes for FSM subgroup maths outcome by type of intervention.

4.3.4 Meta-analysis by study design

The choice of study design can sometimes induce variability between trials because of issues like contamination. Clustered randomised controlled trials (CRT) were more common in EEF trials. However, several EEF trials have also used multisite trial designs. Since the trials with other types of study design such as regression discontinuity do not have enough numbers of trials for meta-analysis, they have been excluded from the analysis. Table 13 provides the estimate of the pooled effect size for CRT and MST trials literacy outcomes. The average effects of CRT and MST trials on literacy outcomes were positive. However, the pooled estimate of the effect size for MST trials was slightly higher than the pooled effect size for CRT trials.

Table 13. Pooled ES and credible interval for FSM subgroup literacy outcome by study design.

Study design	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
CRT	46	2905	63358	0.04 (0.00, 0.08)
MST	32	613	8914	0.08 (0.04, 0.13)

Figure 9 provides the effect sizes for the literacy outcomes by study design. CRT trials such as 'Response to Intervention', 'Shared Maths', 'Online Reading Programmes' were the most beneficial trials. For MST design, most of the trials had positive impacts on the literacy outcomes. MST trials such as 'Accelerated Reader', 'Graduate Coaching Programe' and 'Butterfly Phonics' had significant and positive effect sizes as compared to the other trials in this analysis group.

Table 14: Pooled ES and credible interval for FSM subgroup maths outcome by study design.

Study Design	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
CRT	36	2461	56293	0.01 (-0.03, 0.05)
MST	11	535	32519	0.00 (-0.06, 0.06)

The average effect of CRT and MST trials on maths outcomes was more or less similar (Table 14). The individual trial effect sizes for the maths outcome varied from -0.18 SD to 0.30 SD in CRT and -0.13 SD to 0.12 SD for the MST trials. Very few MST trials had reported a positive effect size, as shown in Figure 10, but most of the MST trials had a bigger sample size.

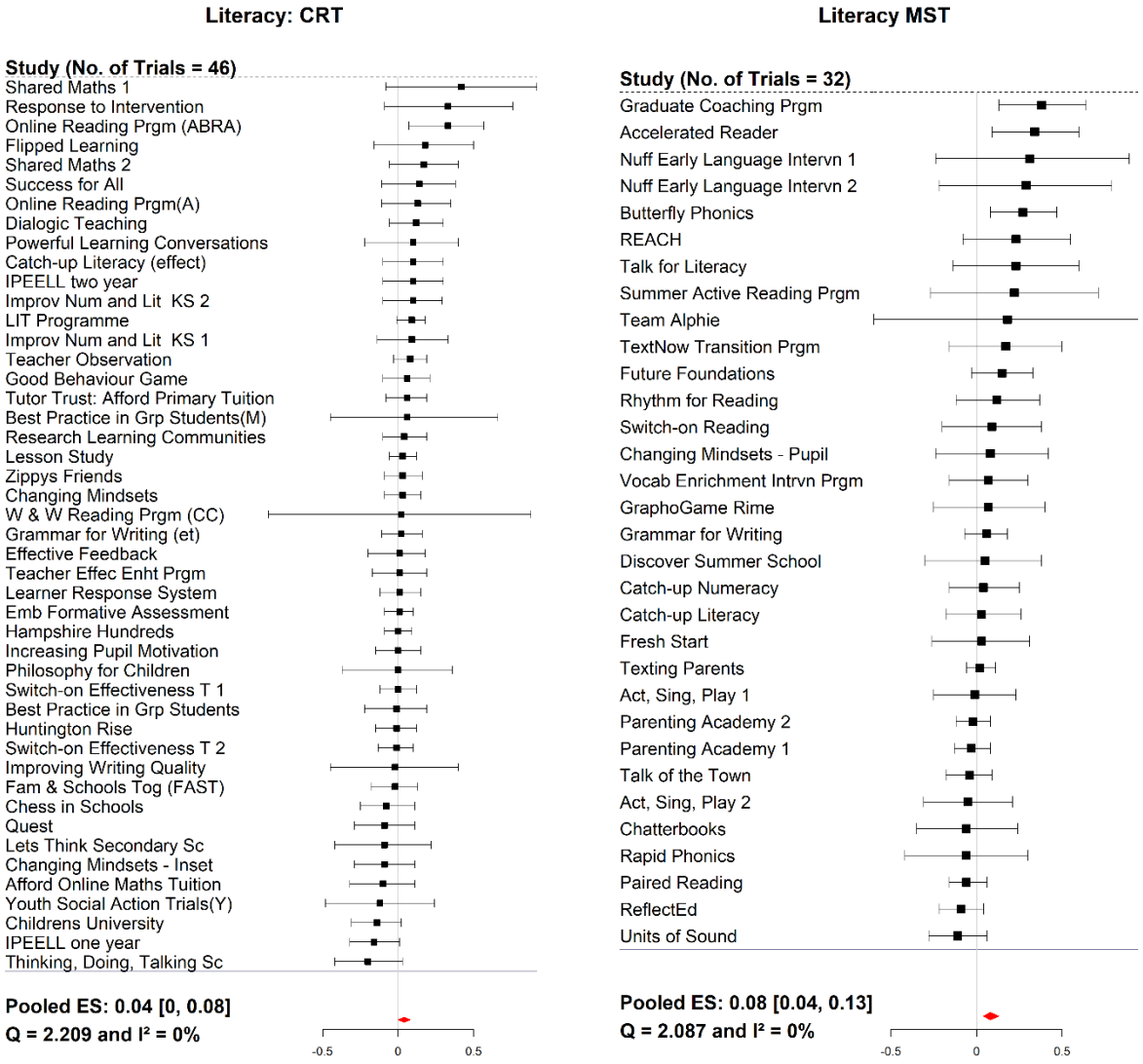

Figure 9: Forest plot with effect sizes for FSM subgroup literacy outcome by study design.

Figure 10: Forest plot with effect sizes for FSM subgroup maths outcomes by study design.

4.3.5 Meta-analysis by study design, Key Stages and types of interventions.

This analysis was conducted by cross-classifying the three grouping variables, i.e., Key Stages, study design, and type of intervention. However, the results presented in

Table *15* provided the pooled estimate for only 10 combinations of the variables due to the few numbers of trials in other strata.

Among the MST trials, KS3 literacy outcomes from one-to-one and small group interventions had pooled effect sizes of 0.16 (0.05, 0.26) and 0.13 (-0.01, 0.26), respectively. Among the CRT trials, KS1 literacy outcomes from whole class interventions had the biggest effect size of 0.06 (-0.02, 0.15).

Table 15: Pooled ES and credible interval for FSM subgroup literacy outcomes by study design, Key Stages and intervention types.

Study Design	Key Stages	Intervention Types	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
CRT	KS1	whole class	5	194	1719	0.06 (-0.02, 0.15)
CRT	KS2	one-to-one	7	661	7011	0.04 (-0.04, 0.11)
CRT	KS2	whole class	12	761	12655	-0.01 (-0.09, 0.07)
CRT	KS2	whole school	4	434	9755	0.02 (-0.04, 0.07)
CRT	KS3	whole class	5	100	2040	-0.04 (-0.18, 0.10)
CRT	KS4	whole school	3	224	12530	0.00 (-0.07, 0.08)
MST	KS2	one-to-one	4	84	2012	0.02 (-0.08, 0.12)
MST	KS3	one-to-one	8	167	979	0.16 (0.05, 0.26)
MST	KS3	small group	7	58	670	0.13 (-0.01, 0.26)
MST	KS3	whole class	3	27	721	0.02 (-0.11, 0.16)

Meta-analysis of maths outcomes by study design, Key Stages, and intervention types were presented in Table 16**Error! Reference source not found.**. Most of the combinations had too few or no trials to enable meta-analysis. The biggest pooled effect size where meta-analysis was done was 0.02 SD for KS1 maths outcomes from whole class interventions and Key Stage 2 maths outcomes from whole school interventions.

Table 16: Pooled ES and credible interval for FSM subgroup maths outcomes by study design, Key Stages and intervention types.

Study Design	Key Stages	Intervention Types	Trials (n)	Schools (n)	FSM pupils (n)	Pooled ES
CRT	KS 1	whole class	4	222	2179	0.02 (-0.22, 0.28)
CRT	KS 2	one-to-one	3	266	3711	0.00 (-0.10, 0.10)
CRT	KS 2	whole class	10	632	10448	-0.02 (-0.09, 0.05)
CRT	KS 2	whole school	3	318	7923	0.02 (-0.04, 0.09)
CRT	KS 3	whole class	3	100	1266	0.00 (-0.22, 0.23)
CRT	KS 4	whole school	3	224	12614	-0.01 (-0.07, 0.06)
MST	KS 2	one-to-one	4	103	2064	-0.01 (-0.10, 0.07)

4.4 Meta-analysis of attainment gaps between FSM and non-FSM pupils

The attainment gap in literacy between FSM and non-FSM pupils was close to zero, but positive. This seems to suggest that on average, EEF interventions had similar effects on FSM and non-FSM pupils across all trials. It can therefore be argued that although EEF interventions were evidenced to reduce the attainment gaps, no evidence was found to suggest that EEF interventions widen attainment gaps between FSM and non-FSM pupils. The results were provided in Table 17. Similarly, the attainment gap in maths was negative. It can also be argued that there was no evidence of the attainment gaps widening.

Table 17. Pooled attainment gap and credible interval for the study outcomes.

Outcome	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
Literacy	81	4000	302138	0.01 (-0.01, 0.04)
Maths	48	3178	306975	-0.01 (-0.04, 0.02)

Figure *11* shows the individual trial and the average attainment gap for both the literacy and maths outcomes. More than half of the trials had a positive attainment gap in literacy scores, which means that on average FSM pupils were more likely to benefit than their peers. The attainment gap in literacy scores between FSM and non-FSM pupils was more than 0.20 SD for trials such as 'Text Now Transition Programme', 'Affordable Individual and Small Group Tuitions Programme', 'Nuffield Early Language Intervention', 'Improving Numeracy and

Literacy', and 'Best Practice in Grouping Students'. However, the attainment gap in maths between FSM and non-FSM pupils was closer to 0.0 SD.

Figure 11: Forest plot with attainment gap between FSM and their peers by study outcomes.

4.4.1 Attainment gaps by primary and secondary outcomes

The attainment gap did not vary much by type of outcome. Overall, the attainment gap between FSM and non-FSM pupils for trials with literacy as the primary outcome was slightly higher than trials with literacy as secondary outcomes (Table 18). This finding also suggests that FSM pupils were not worse off than their peers. The secondary maths attainment gap was positive, whilst the gap from maths primary outcome was negative. Figure 12 and Figure 13 provide the individual and pooled attainment gap between FSM and their peers' estimates. 'Nuffield Early Language Intervention', 'TextNow Transition Programme', 'Best Practice in Grouping Students' and 'Improving Numeracy and Literacy' were the interventions most likely to benefit FSM pupils more than their peers. Similarly, 'Act, Sing and Play' and 'Improving Numeracy and Literacy' were most likely to benefit FSM pupils than their peers.

Table 18: Pooled attainment gap and credible interval by primary and secondary outcome.

Study outcome	Outcome Types	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
Literacy	Primary	72	3427	262321	0.02 (-0.02, 0.04)
Literacy	Secondary	9	573	39817	0.00 (-0.06, 0.06)
Maths	Primary	42	2686	275461	-0.01 (-0.04, 0.02)
Maths	Secondary	6	492	31514	0.02 (-0.04, 0.08)

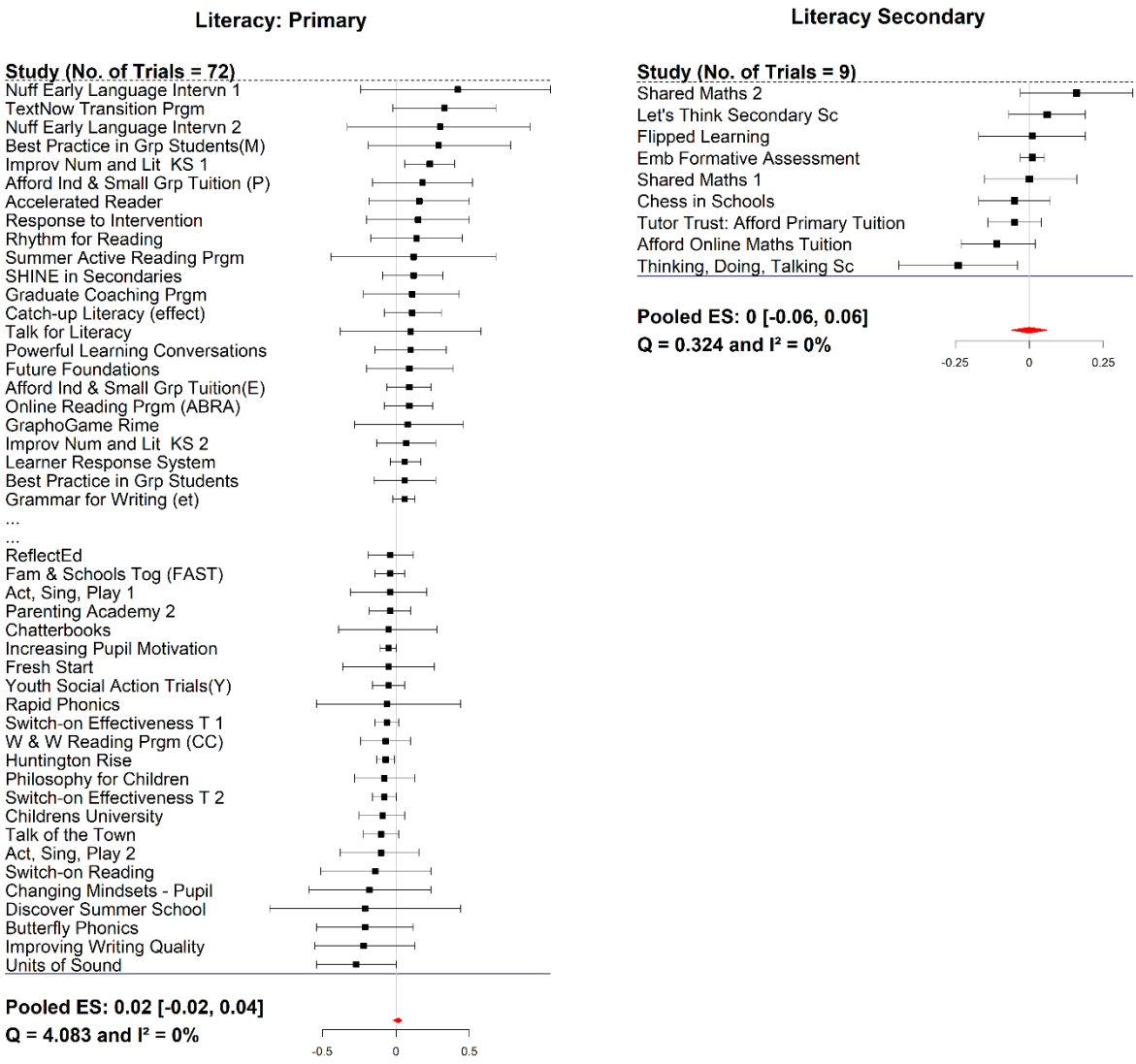

Figure 12: Forest plot with attainment gap between FSM and their peers for literacy as a primary and secondary outcome.

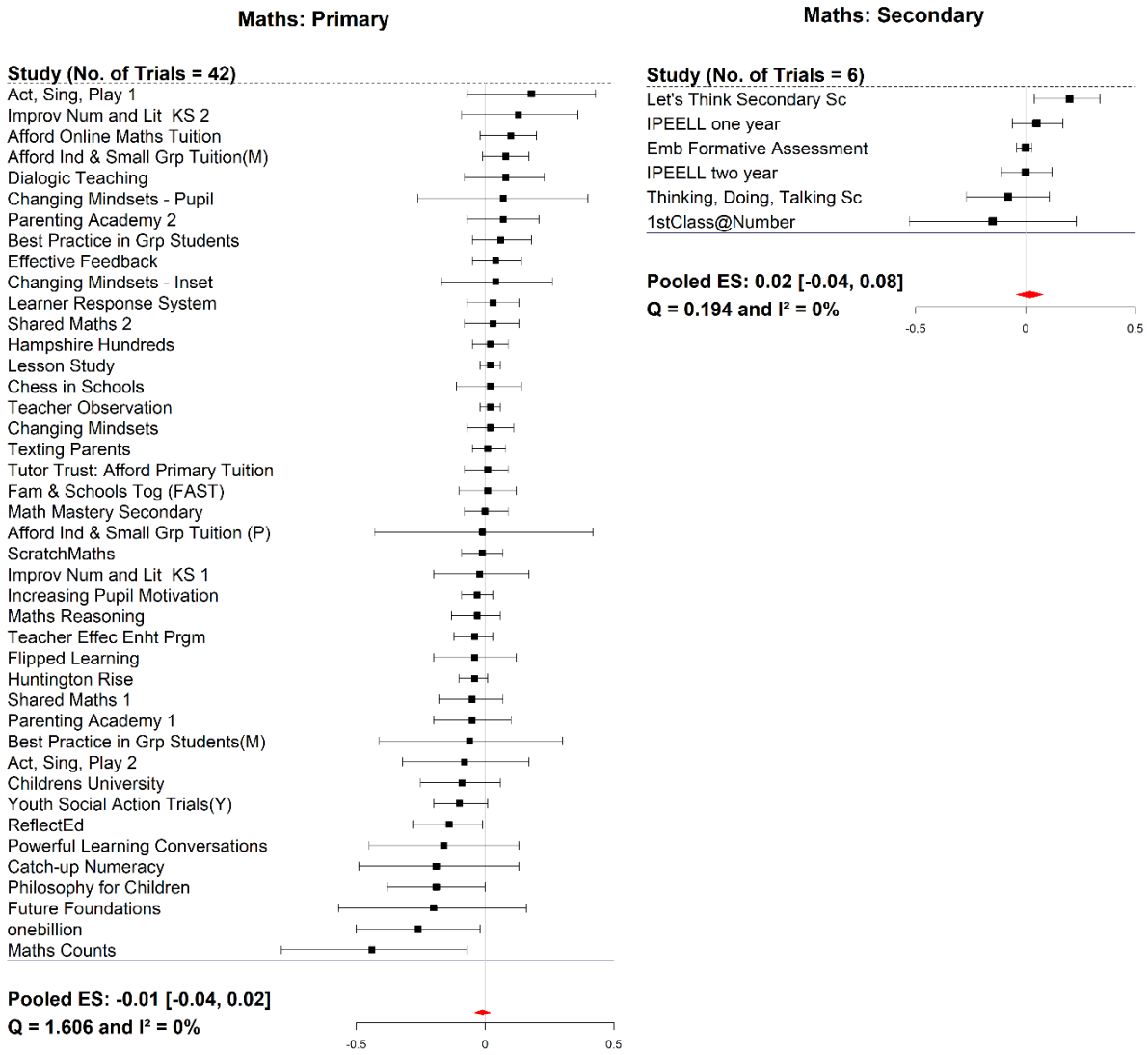

Figure 13: Forest plot with attainment gap between FSM and their peers for maths as a primary and secondary outcome.

4.4.2 Attainment gaps by Key Stages

The attainment gap in literacy between FSM and non-FSM pupils appears to be dependent on Key Stages (Table 19). The pooled attainment gap for KS1 was 0.07 (0.00, 0.14), whilst KS2 and KS4 had 0.00 (-0.03, 0.03) pooled attainment gap. There was an evidence that interventions in KS1 helped FSM pupils, while for other key stages, there was no clear evidence that the intervention favour FSM pupils than non-FSM pupils.

Table 19: Pooled attainment gap and credible interval for literacy outcome by Key Stages.

Key Stages	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
KS1	13	529	19905	0.07 (0.00, 0.14)
KS2	33	2265	102835	0.00 (-0.03, 0.03)
KS3	29	552	39297	0.01 (-0.05, 0.07)
KS4	6	654	140101	0.00 (-0.03, 0.03)

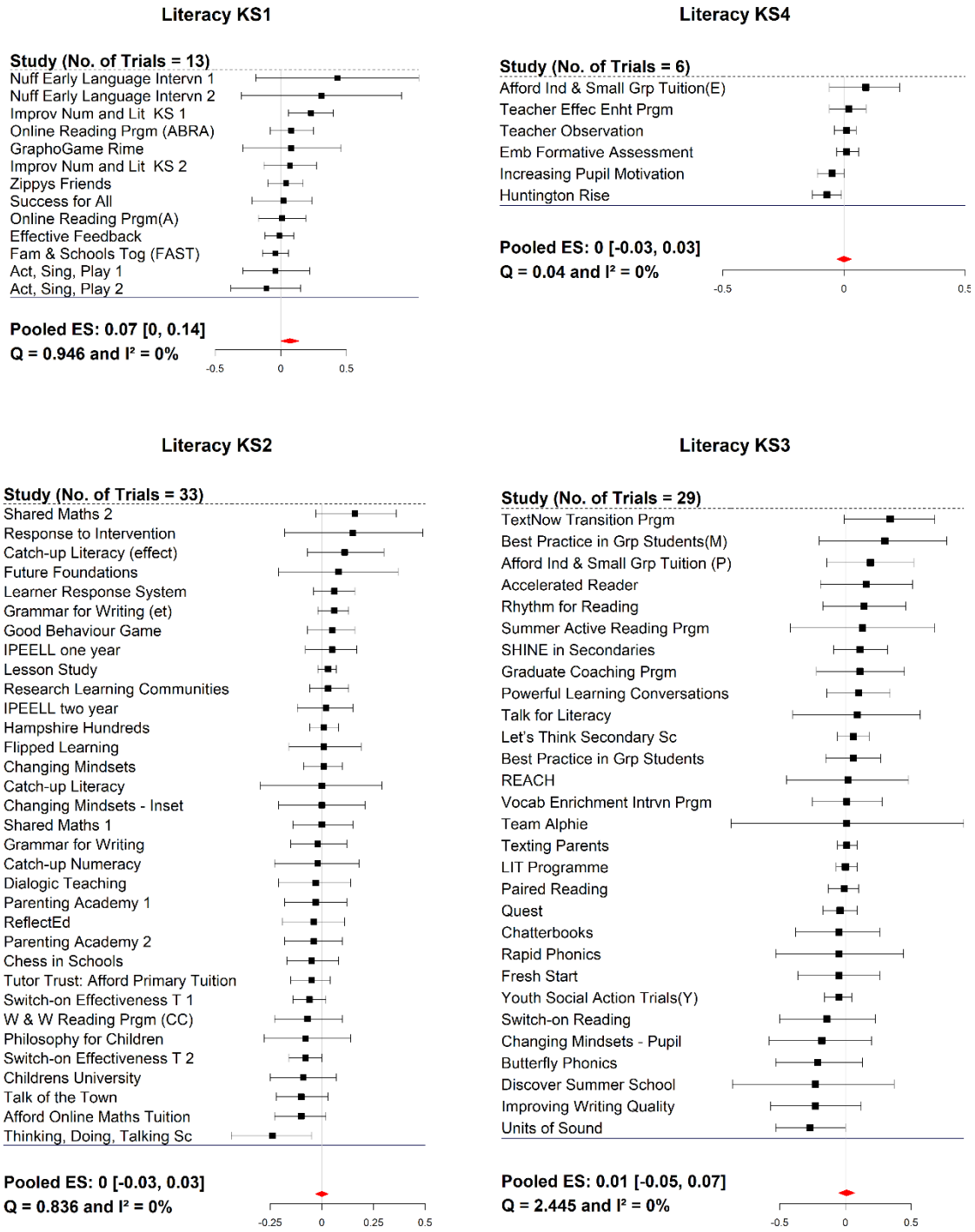

Figure 14 provides the individual and pooled attainment gap estimates for the literacy outcome by Key Stages. In KS1, most of the trials had a positive attainment gap in literacy scores and few had negative attainment gap. The trial 'Nuffield Early Language Intervention 1' had highest attainment gap of 0.42. However, trials with a larger sample size, such as 'Families & School Together (FAST)' and 'Effective Feedback' had negative attainment gap which resulted in a pooled attainment gap of 0.07 (0.00, 0.14) for KS1. Nearly, half of the trials in KS2 had positive attainment gaps and tended to benefit FSM pupils more than non-FSM pupils. The trials include 'Shared Maths 2', 'Response to Intervention', 'Catch up Literacy' effectiveness trials

with an attainment gap more than 0.10 SD. More than two-thirds of the trials in KS3 also had a positive attainment gap. In KS4, four trials had a positive attainment gap. Overall, these trials showed no differential benefit between FSM and non-FSM pupils.

Table 20: Pooled attainment gap and credible interval for maths outcome by Key Stages.

Key Stages	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
KS1	9	639	18718	0.00 (-0.06, 0.07)
KS2	24	1577	79671	-0.02 (-0.06, 0.00)
KS3	9	269	30434	0.02 (-0.07, 0.10)
KS4	6	693	178152	0.00 (-0.02, 0.02)

The pooled attainment gaps for all Key Stages was zero, except for KS2 maths (Table 20).

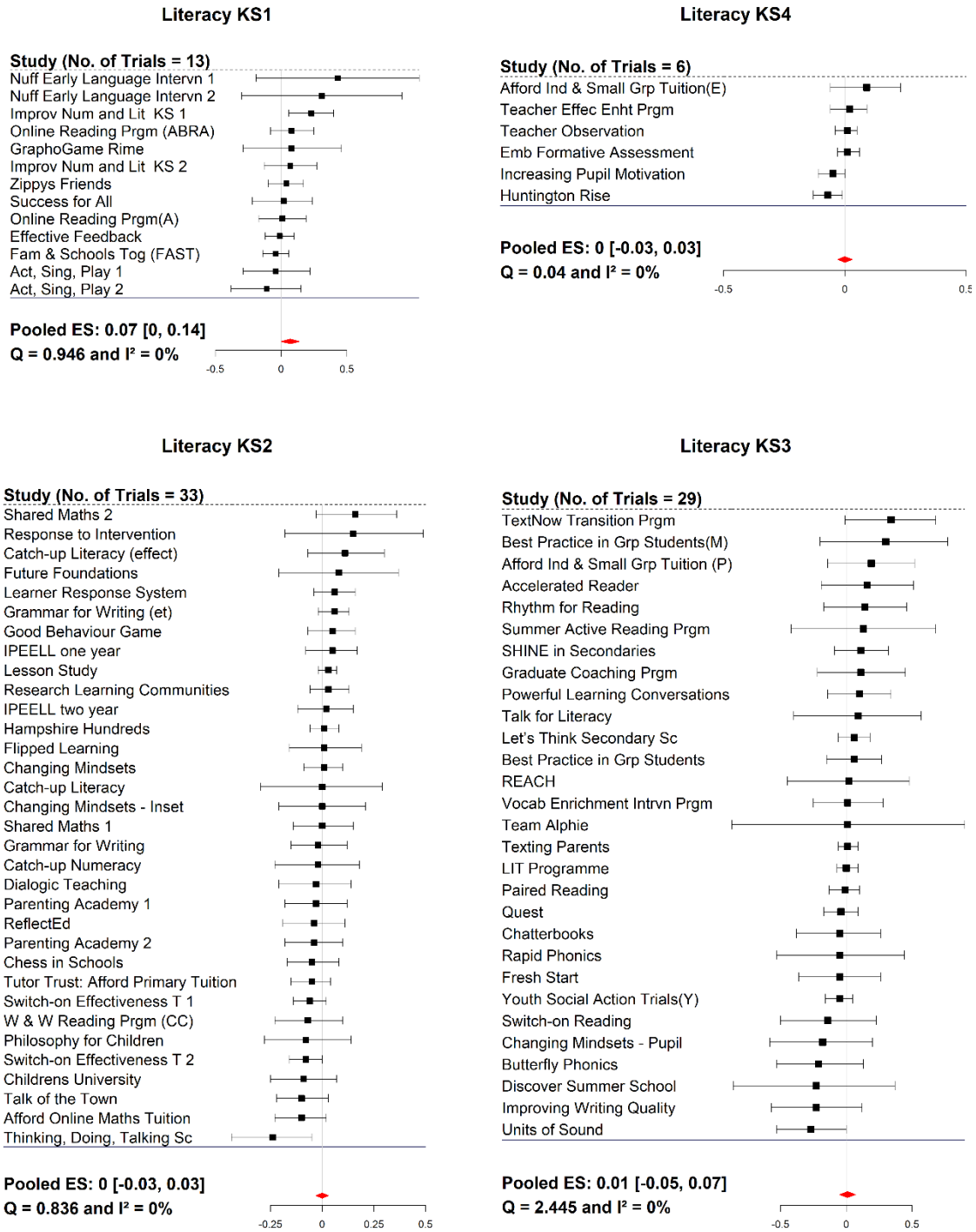

Figure 14: Forest plot with attainment gap between FSM and their peers for literacy outcomes by Key Stages.

Figure 15 provides the individual trial and pooled attainment gap estimates for the maths outcome by Key Stages. In KS1, two trials ('Act, Sing and Play' and 'Improving Numeracy and Literacy') benefitted the FSM pupils the most, with an attainment gap of more than 0.10 SD. Most of the trials in KS2 had a positive attainment gap. The attainment gap was maximum for 'Affordable Maths' trial in KS2. Five trials in KS3 also had positive attainment gaps and 'Let's Think Secondary Science' trial was the most beneficial for FSM. There were also several trials

in KS4 with positive attainment gaps in favour of FSM pupils, though the overall pooled attainment gap was zero.

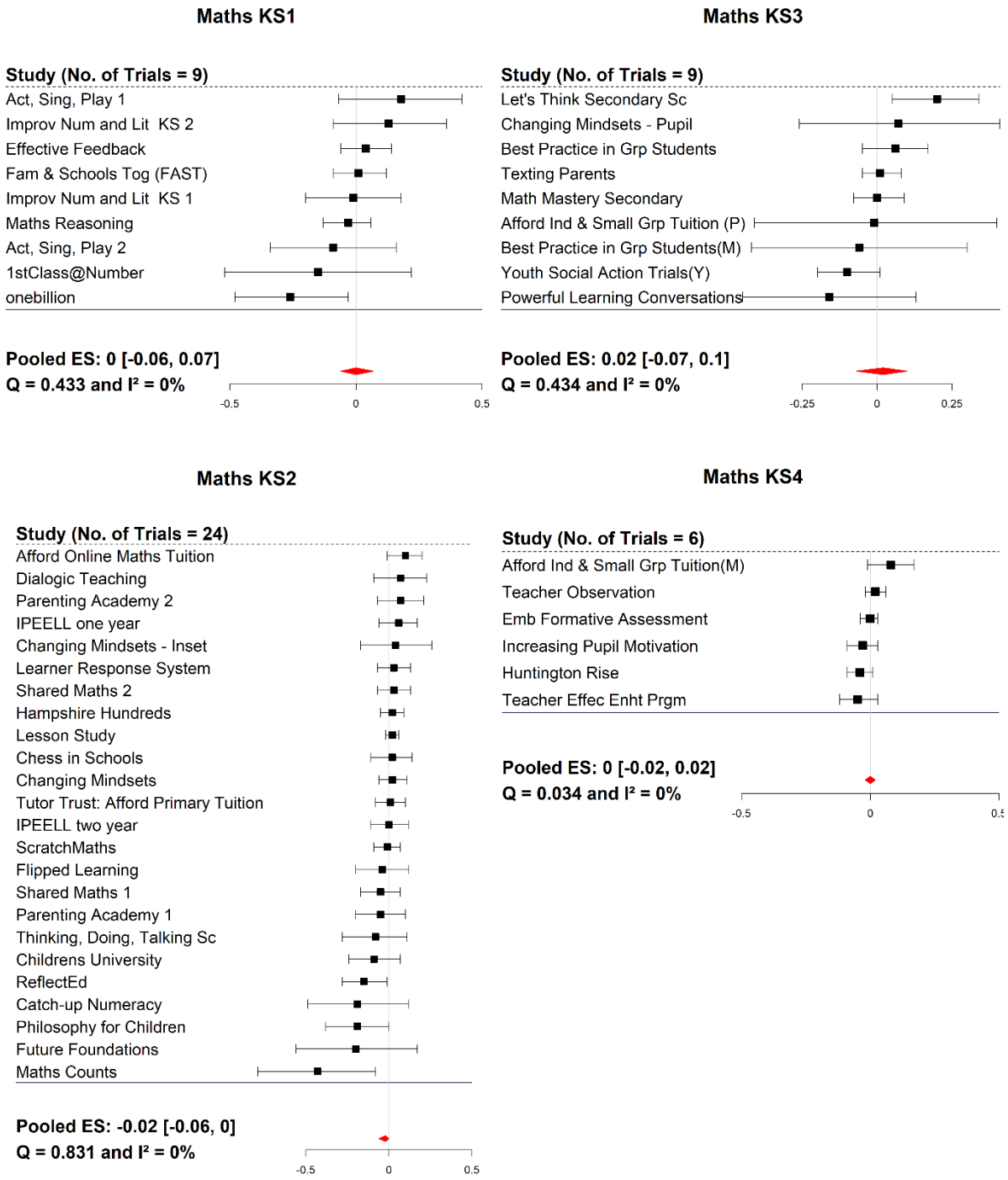

Figure 15: Forest plot with attainment gap between FSM and their peers for maths outcomes by Key Stages.

Overall, comparison of the attainment gaps in Figure 16 showed that the gaps across the Key Stages were positive for literacy outcomes and mostly negative for the maths outcome. KS3

was the only subgroup where the attainment gap was positive for both literacy and maths outcomes. This indicates that FSM pupils in KS3 benefited more than the non-FSM pupils.

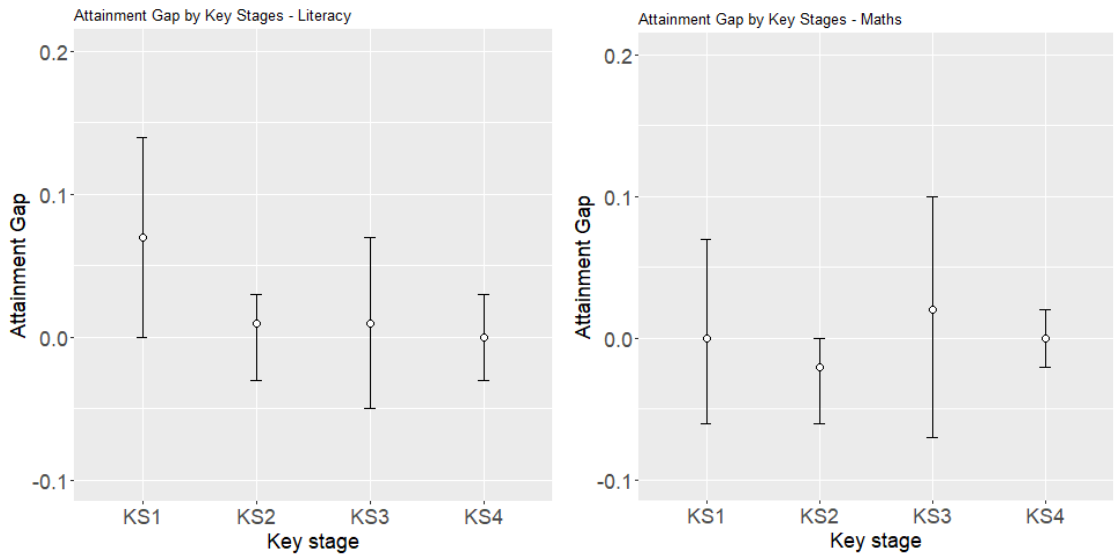

Figure 16: Pooled attainment gap between FSM and their peers for literacy and maths outcomes by Key Stages.

4.4.3 Attainment gaps by types of intervention

The narrowing of the attainment gap in literacy outcomes between FSM and non-FSM pupils was maximum for one-to-one and small group interventions (Table 21). On average, FSM pupils were 0.02 SD and 0.05 SD better than the non-FSM pupils for one-to-one and small group interventions, respectively.

Table 21: Pooled attainment gaps and credible interval for literacy outcome by types of intervention.

Intervention type	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
One-to-one	24	1358	97368	0.02 (-0.04, 0.07)
small group	17	503	22451	0.05 (-0.04, 0.14)
whole class	30	1339	83550	0.00 (-0.04, 0.04)
whole school	10	800	98769	0.00 (-0.03, 0.03)

Figure *17* provides the forest plot with the individual trial and pooled attainment gaps for the literacy outcome by type of intervention. Half of one-to-one and small group interventions had positive attainment gaps, suggesting that FSM pupils were more likely to benefit from these interventions. 'Text now Transition Programme' (one-to-one) intervention and 'Nuffield Early Language Intervention' (small group), 'Best Practice in Grouping Students' (whole class) and 'Improving Literacy and Numeracy' trial (whole class) interventions were the most beneficial interventions for the FSM pupils.

The pooled attainment gaps in maths scores between the FSM and non-FSM pupils was positive for the whole class interventions (Table 22). One-to-one and small group interventions were the least beneficial for FSM pupils. This was contradictory to the pattern for literacy outcomes where one-to-one and small group interventions were the most beneficial for FSM pupils.

Table 22: Pooled attainment gap and credible interval for maths outcome by type of intervention.

Intervention type	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
One-to-one	10	857	117290	-0.05 (-0.12, 0.01)
small group	7	496	18391	-0.06 (-0.14, 0.02)
whole class	23	1210	75525	0.02 (-0.03, 0.06)
whole school	8	615	95769	0.00 (-0.02, 0.03)

Figure *18* shows the individual and pooled attainment gap estimates for the maths outcome by types of intervention. The trial-specific attainment gap in one-to-one interventions varied from -0.44 to 0.10, small group varied from -0.26 to 0.03, whole class varied from -0.19 to 0.20 and for whole school varied from -0.04 to 0.02. ‘Affordable Online Maths Tuition’ was the one-to-one intervention with an attainment gap bigger than 0.10 SD.

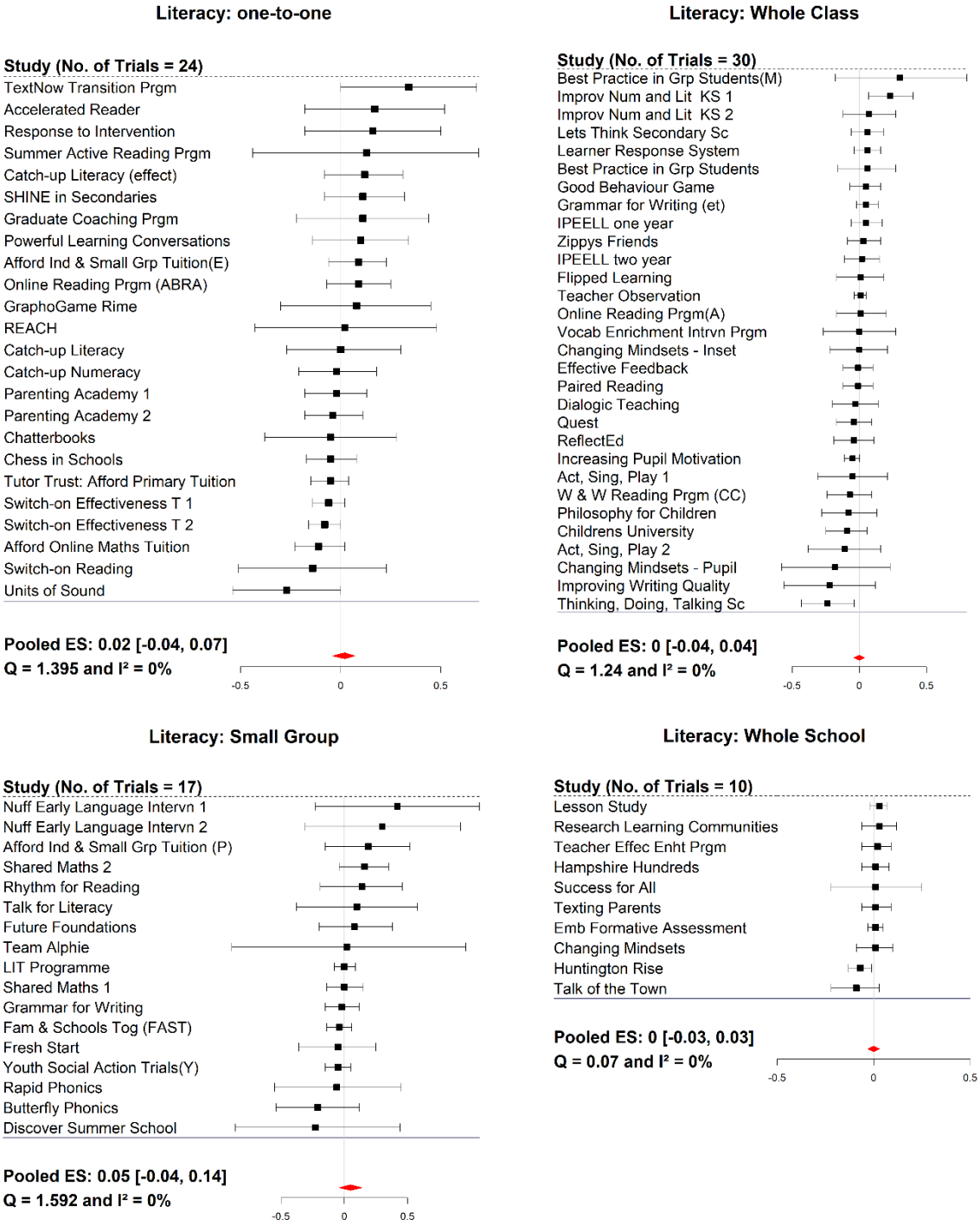

Figure 17: Forest plot with attainment gap between FSM and their peers for literacy outcomes by type of intervention.

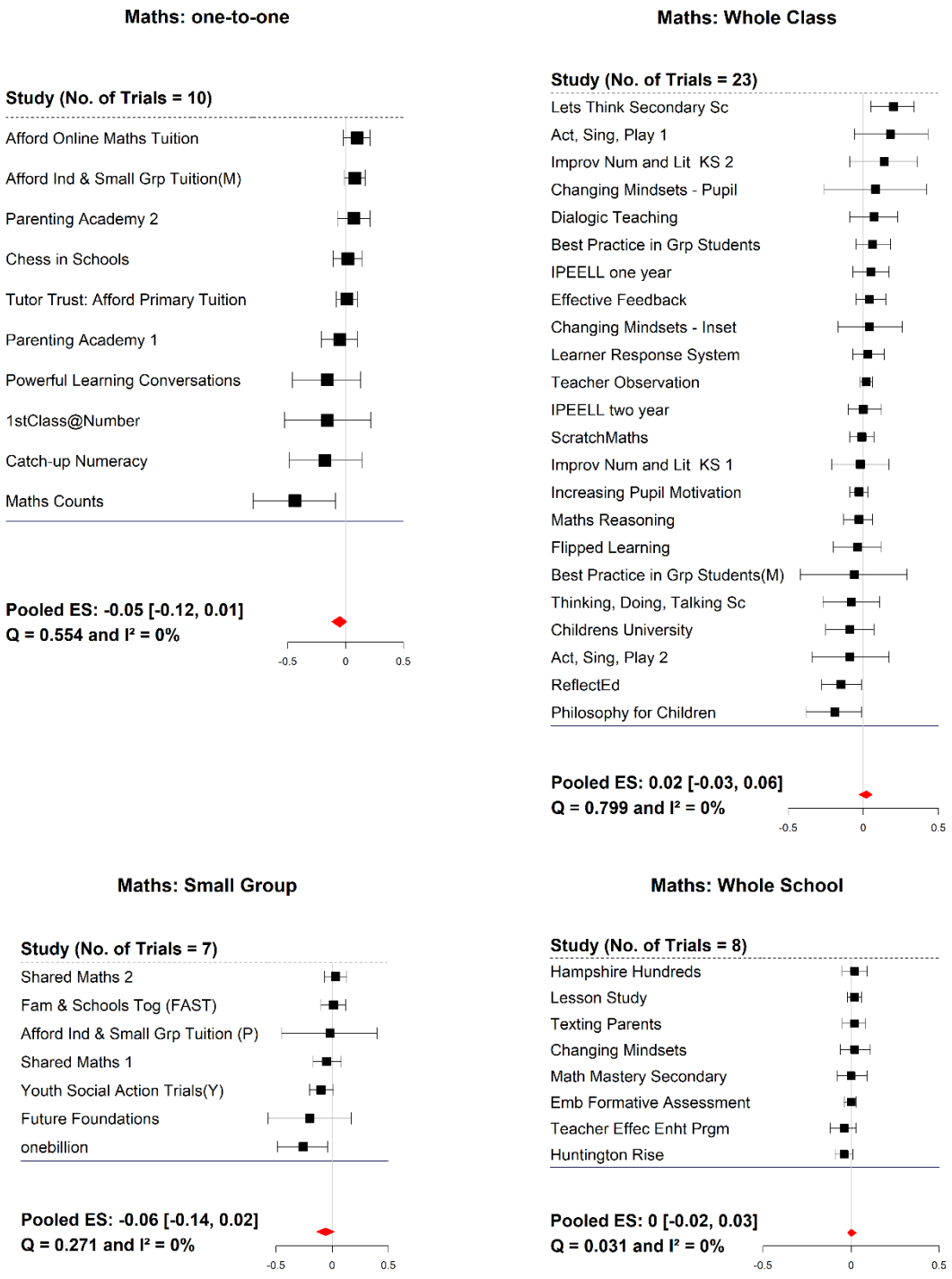

Figure 18: Forest plot with attainment gaps between FSM and their peers for maths outcomes by type of intervention.

Figure 19 provides a comparative overview of attainment gaps for the literacy and maths outcomes by type of intervention. One-to-one or small groups interventions were the better interventions for literacy, whilst whole-class and whole-school interventions were the better interventions and reduced attainment gaps in maths.

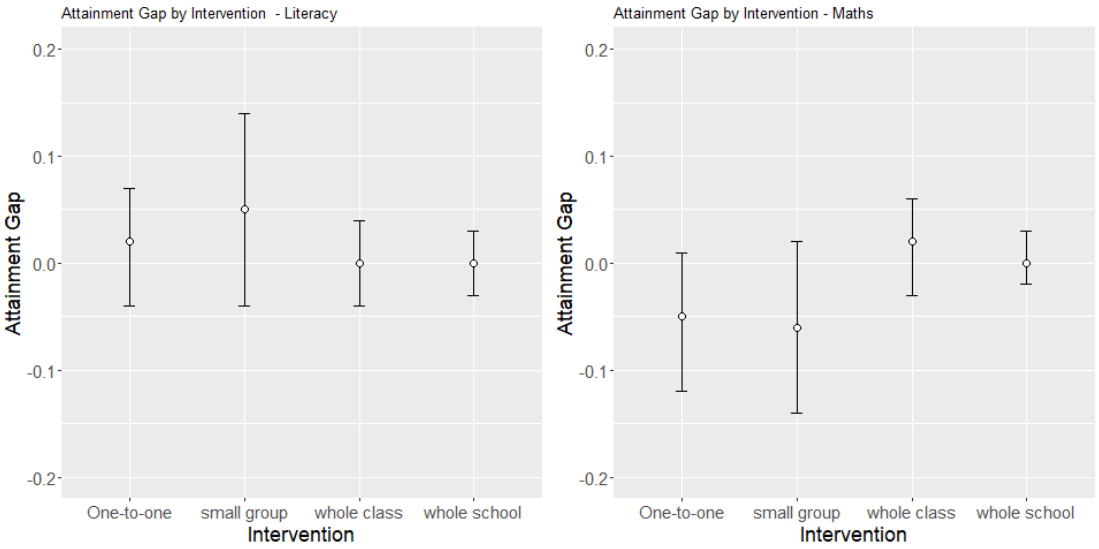

Figure 19: Pooled attainment gap between FSM and their peers for literacy and maths outcomes by type of intervention.

4.4.4 Attainment gaps by study design

Table 23 shows that the narrowing of the attainment gap in literacy was slightly more for CRT than MST with pooled attainment gaps of 0.02 (-0.01, 0.04) and 0.00 (-0.06, 0.05), respectively.

Table 23: Pooled attainment gap and credible intervals for literacy outcome by study design.

Study design	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
CRT	46	3011	205928	0.02 (-0.01, 0.04)
MST	32	688	31456	0.00 (-0.06, 0.05)

Figure 20 provides individual trial and pooled attainment gaps for the literacy outcome by study design. CRT Trials such as 'Powerful Learning Conversations', 'Improving Numeracy and Literacy', 'Best practice in Grouping Students', 'Catch Up Literacy' were more likely to have bigger attainment gaps in favour of FSM pupils. However, there were 15 MST trials with positive attainment gaps and seven trials with attainment gaps of more than 0.10 SD. The attainment gap in maths outcome for CRT trials was zero and negative for MST trials (Table 24).

Table 24: Pooled attainment gap and credible intervals for maths outcome by study design.

Study design	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
CRT	36	2584	186257	0.00 (-0.03, 0.02)
MST	11	582	119955	-0.03 (-0.11, 0.04)

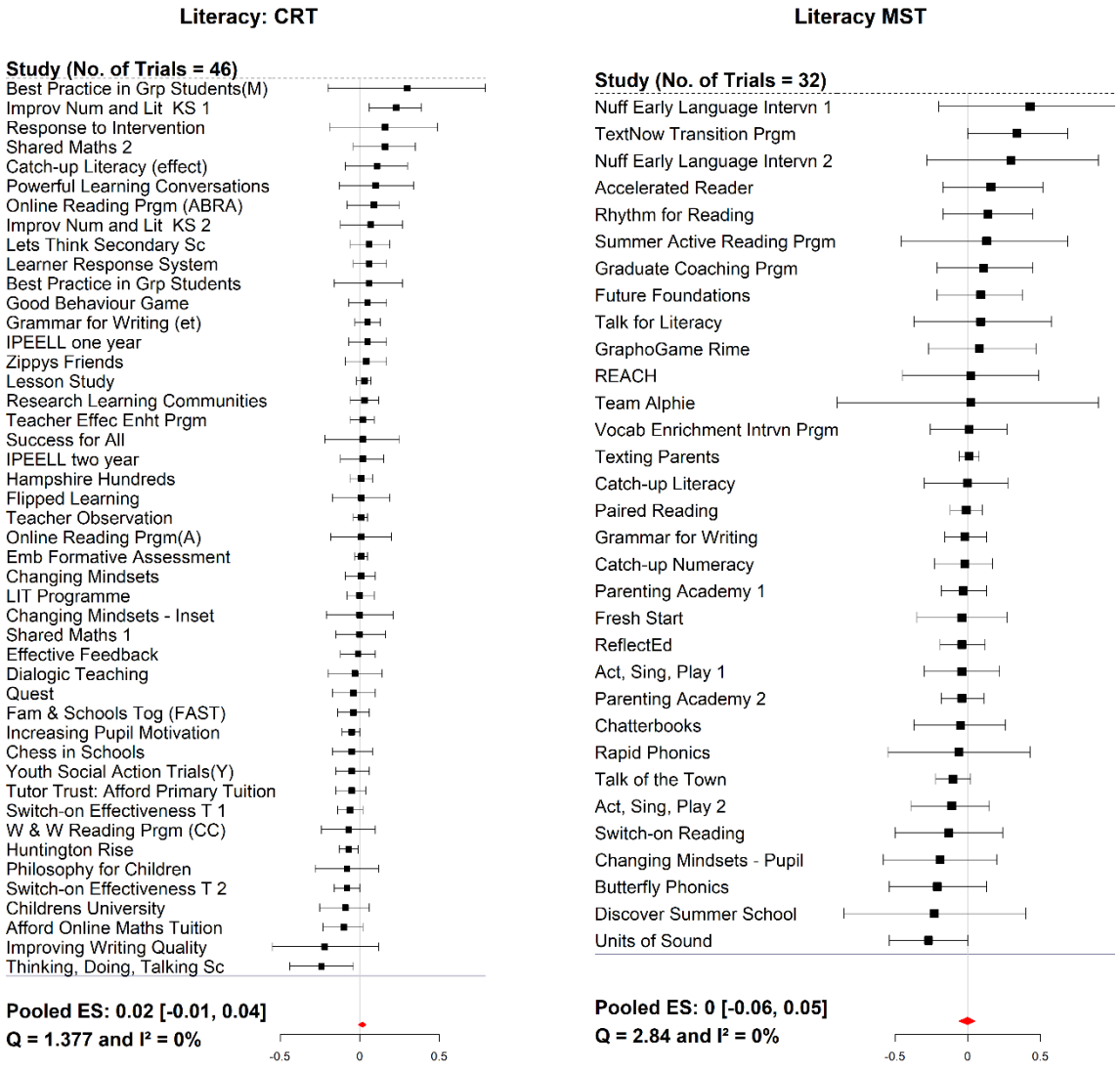

Figure 20: Forest plot with attainment gap between FSM and their peers for literacy outcome by study design.

Figure 21 shows the individual trial and pooled attainment gap estimates for the maths outcomes by study design. The trial-specific attainment gap for maths outcome varied from -0.26 SD to 0.20 SD for CRT trials and from -0.44 SD to 0.18 SD for MST trials. Half of the CRT and MST design trials with maths outcome had a positive attainment gap in favour of FSM pupils. 'Let's Think Secondary Science' (CRT trial) and 'Act, Sing and Play' (MST trial)

had bigger attainment gap in favour of FSM pupils. ‘Maths Count’ (MST trial), ‘onebillion’ and ‘Powerful Learning Conversations’ (CRT trial) had negative attainment gaps suggesting that the FSM pupils lagged behind the non-FSM pupils in these trials.

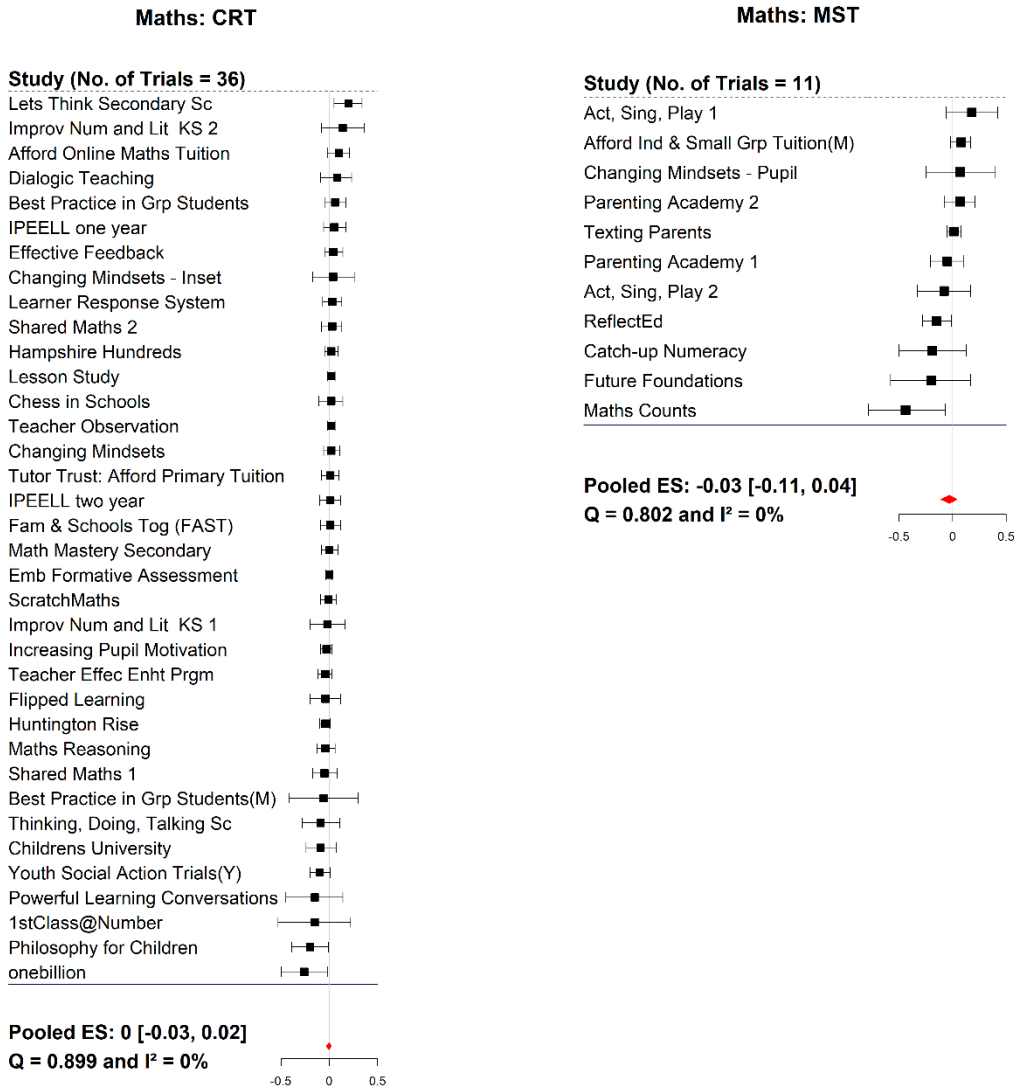

Figure 21: Forest plot with attainment gap between FSM and their peers for maths outcomes by study design.

Figure 22 provides a comparative overview of the pooled attainment gaps between FSM pupils and their peers for literacy and maths outcomes by study design. The attainment gaps for literacy were positive for both study designs, but in opposite directions for maths. Further, variability in attainment gaps as evident from the credible intervals was higher in the MST trials than the CRT trials. This may be due to an extra source of variation in the MST trials.

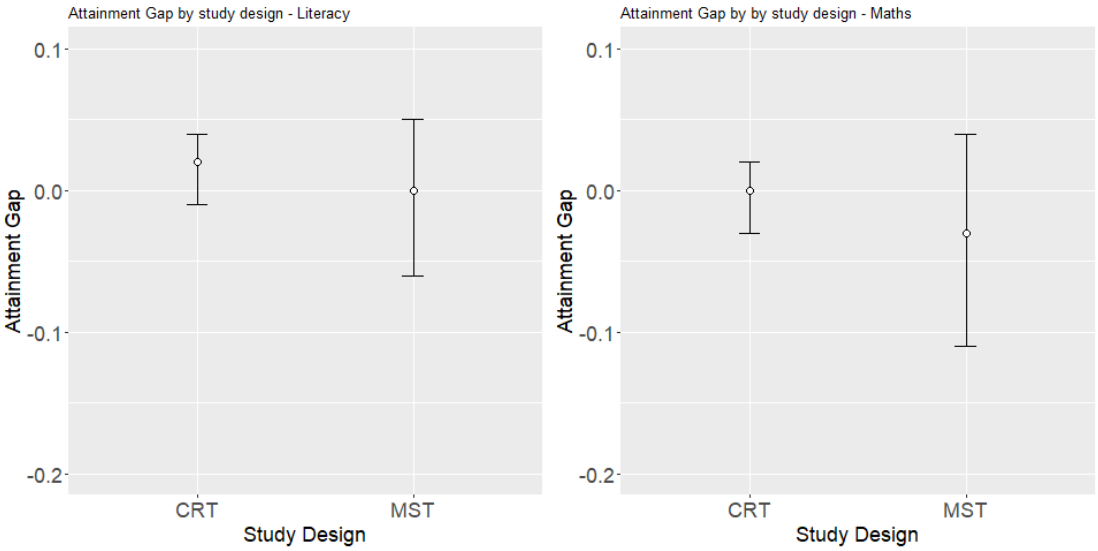

Figure 22: Pooled attainment gap between FSM and their peers for literacy and maths outcomes by study design.

4.4.5 Attainment gap by study design, Key Stages and types of intervention

Table 25 provides pooled attainment gap estimates for combinations of study design, Key Stage and types of interventions. Among the CRT trials, whole class interventions in Key Stage 1 and Key Stage 3 had been more beneficial for FSM pupils. Among the MST trials, one-to-one interventions in Key Stage 3 had been most beneficial for FSM pupils with the pooled effect size of 0.02 (-0.12, 0.15).

Table 25: Pooled attainment gap and credible intervals for literacy outcomes by study design, Key Stages and intervention types.

Study design	Key Stages	Intervention Types	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
CRT	KS1	whole class	5	215	10569	0.07 (0.00, 0.14)
CRT	KS2	one-to-one	7	696	22641	-0.01 (-0.08, 0.05)
CRT	KS2	whole class	12	780	30704	-0.02 (-0.07, 0.03)
CRT	KS2	whole school	4	434	36912	0.01 (-0.03, 0.06)
CRT	KS3	whole class	5	106	6229	0.03 (-0.09, 0.14)
CRT	KS4	whole school	3	224	46435	-0.02 (-0.05, 0.02)
MST	KS2	one-to-one	4	101	4157	-0.02 (-0.14, 0.10)
MST	KS3	one-to-one	8	193	2578	0.02 (-0.12, 0.15)
MST	KS3	small group	7	66	1623	-0.05 (-0.21, 0.12)
MST	KS3	whole class	3	27	3334	-0.06 (-0.22, 0.10)

Table 26 shows the analysis of maths outcome by the combinations of study design, Key Stages, and types of intervention. In CRT, KS3 whole class intervention had highest pooled attainment gap 0.07 (-0.06, 0.19), and KS4 whole class intervention had lowest pooled attainment gap -0.03 (-0.06, 0.00).

Table 26 : Pooled attainment gap and credible intervals for maths outcomes by study design, Key Stages and intervention types.

Study design	Key Stages	Intervention Types	Trials (n)	Schools (n)	Pupils (n)	Pooled Attainment Gap
CRT	KS 1	whole class	4	245	11756	0.04 (0.04, 0.12)
CRT	KS 2	one-to-one	3	268	10609	0.04 (-0.02, 0.11)
CRT	KS 2	whole class	10	647	26391	-0.02 (-0.07, 0.03)
CRT	KS 2	whole school	3	318	31606	0.02 (-0.03, 0.07)
CRT	KS 3	whole class	3	102	4516	0.07 (-0.06, 0.19)
CRT	KS 4	whole school	3	224	46744	-0.03 (-0.06, 0.00)
MST	KS 2	one-to-one	4	125	3947	-0.13 (-0.25, -0.01)

4.5 Sensitivity analysis for padlock ratings

The sensitivity of the pooled effect size for the padlock rating of the trials was checked by excluding the trials with fewer than three padlock ratings. The EEF padlock rating is often used to measure the quality of trial (Lortie-Forgues & Inglis, 2019). Padlock ratings ranged between 0 and 5 with a higher padlocks indicating better quality and security rating for the results from that trial. Table 27 provides the results of the analysis alongside the analysis using all the trials. There was no evidence to suggest that padlock ratings substantially affected the average effect of the interventions or the average attainment gaps between FSM and non-FSM pupils and attainment gaps from the trials.

Table 27: Sensitivity analysis for literacy and maths outcome by excluding trials with less than 3 padlocks.

Outcome	Effect	Schools (n)	Pupils (n)	ALL	Schools (n)	Pupils (n)	Padlock ≥3
Literacy	FSM	3804	90218	0.06 (0.03, 0.08)	2337	48216	0.06 (0.03, 0.10)
	Gap	4000	302138	0.01 (-0.01, 0.04)	2436	156004	0.02 (-0.02, 0.06)
Maths	FSM	3006	89247	0.00 (-0.03, 0.04)	1990	49783	0.01 (-0.02, 0.05)
	Gap	3178	306975	-0.01 (-0.03, 0.02)	2115	165735	-0.00 (-0.04, 0.03)

5.0 Discussion and conclusions

Evidence-based interventions need to be developed for FSM pupils in order to reduce the attainment gap between FSM and their peers. With this aim, an IPD meta-analysis was conducted to synthesise evidence of the overall impact of EEF-funded education interventions on FSM pupils and to quantify the effects of the interventions on the gaps between FSM pupils and their peers. Meta-analysis helps to counteract the risk that individual studies may be underpowered due to the smaller sample size of FSM pupils. There has been no previous attempt in education research to systematically review such a large archive of individual pupils' data in education trials and provide reliable individual and pooled estimates of effect size and attainment gap for the key study outcomes of FSM pupils and describing these outcomes by a range of important factors such as study design, type of intervention and Key Stage of pupils using a gold standard of meta-analysis.

Key findings

The overall impact of EEF interventions on the literacy outcomes of FSM pupils was positive (pooled effect size = 0.06 (0.03, 0.08)). When this impact was assessed by KS, the greatest effect was observed on the FSM pupils of KS1 (pooled effect size 0.09 (0.02, 0.16)) followed by those of KS3 (pooled effect size = 0.08 (0.03, 0.13)), KS2 (pooled effect size = 0.03 (-0.01, 0.07)), and the least impact on those of KS4 (pooled effect size = 0.02 (-0.05, 0.08)). Overall a similar impact was seen for literacy as a primary (pooled effect size = 0.06 (0.03, 0.08)) or secondary (pooled effect size = 0.06 (-0.04, 0.16)) outcome.

Previous evidence showed that one-to-one tuition can be an effective intervention method, although comparison of one-to-one with small group tuition had shown mixed results (EEF, 2020). This study showed that the interventions designed for a small group of pupils had the highest impact on literacy outcomes (pooled effect size = 0.14 (0.06, 0.22)) followed by that for one-to-one pupils (pooled effect size = 0.08 (0.04, 0.13)), then whole school approaches (pooled effect size = 0.02 (-0.02, 0.06)) and whole class interventions (pooled effect size = 0.01 (-0.04, 0.05)). This finding is consistent with existing evidence regarding the benefit of small group interventions (Lou et al., 2001).

There was no overall effect on the mathematics outcomes of FSM pupils (pooled effect size= 0.00 (-0.03, 0.04)). There was a positive effect for mathematics as a primary outcome (pooled effect size = 0.01 (-0.02, 0.05)) and a negative effect as a secondary outcome (pooled effect size = -0.07 (-0.15, 0.00)). By KS there was a similar impact on pupils in KS1 (pooled effect size = 0.02 (-0.07, 0.11)) and KS4 (pooled effect size =0.02 (-0.03, 0.07)), followed by those

in KS3 (pooled effect size = 0.01 (-0.09, 0.12)), and KS2 (pooled effect size = -0.01 (-0.04, 0.03)).

An analysis of the attainment gap indicated that literacy outcomes for FSM pupils were improved by EEF interventions marginally more than for non-FSM pupils (pooled attainment gap =0.01 (-0.01, 0.04)). Mathematics outcome was affected in a similar way for both the FSM pupils and their non-FSM peers.

For literacy, the impact on attainment gap as a primary outcome was greater (pooled attainment gap = 0.02 (-0.02, 0.04)) than when it was a secondary outcome. However, mathematics as a secondary outcome (pooled attainment gap = 0.02 (-0.04, 0.08)) had greater impact than when it was a primary outcome (pooled attainment gap = -0.01 (-0.04, 0.02)).

For literacy, the gap was maximum for KS1 (pooled attainment gap = 0.07 (0.00, 0.14)) and least at KS4 (pooled attainment gap = 0.00 (-0.03, 0.03)). For maths, the attainment gap was maximum at KS3 (pooled attainment gap = 0.02 (-0.07, 0.10)); for the other three Key Stages there was no evidence of the gap narrowing.

By the type of intervention, the attainment gap between FSM and non-FSM literacy outcomes was positive for one-to-one and small group interventions. This indicates that on an average, FSM pupils performed better than the non-FSM pupils in these two subgroups of intervention. Small group interventions (pooled attainment gap = 0.05 (-0.04, 0.14)) benefitted FSM pupils the most followed by one-to-one interventions (pooled attainment gap = 0.02 (-0.04, 0.07)). However, in case of whole class and whole school interventions, the pooled attainment gap estimates were zero. For maths outcomes, the highest pooled attainment gap was observed for whole class interventions (pooled attainment gap = 0.02 (-0.03, 0.06)); however, for the rest of the interventions groups attainment gap was nearly zero. Further analysis of the impacts of the different interventions is contained within the report.

Overall, EEF interventions had beneficial impacts on the literacy outcomes of pupils eligible for free school meals compared to maths outcomes which showed no overall effect. Trials with a literacy focus were more beneficial for the FSM pupils than the trials with a maths focus as evident from the pooled effect sizes and attainment gap estimates. An attainment gap report from EEF showed that the FSM pupil's scores and grades were much lower than the non-FSM pupils and the gap was negative. This finding remains the same in different key stages (EEF, 2017). The attainment gap estimates for KS3 outcomes in this study were positive for both maths and literacy outcomes, from which it can be concluded that the EEF interventions helped FSM pupils to perform better than non-FSM pupils in KS3. Both CRT and MST designed trials had benefited FSM pupils more than non-FSM pupils. This was mainly the case for literacy

outcomes. There was huge variability in attainment gap estimates for mathematics outcomes in MST designed trials. By the type of intervention, individual or small group interventions had improved literacy outcomes of FSM pupils considerably while the intervention with focus on the whole class or school were beneficial for the maths outcomes. Previous evidence suggests that small-group work is a key part of academic learning (Gillies & Ashman, 2003). The sense of identity and belonging that a student can experience in a well-run group cannot be underestimated (Jacques, 2000). Interventions conducted in a small group are the best trade-off between cost and effectiveness of interventions as one-to-one tuition or interventions can be a relatively expensive programme to implement.

Overall, evidence from this report can be used to identify, test and scale successful educational interventions with positive impact which can be implemented in schools to improve educational attainment of FSM children. This project had provided a better understanding of the different interventions' effects, inform decisions about specific interventions to target FSM subgroups, and can be used to suggest ways to improve the design or implementation of the tested interventions among FSM children. It also indicates the extent of the challenge of identifying and scaling possible solutions to reduce educational disadvantage in schools.

6.0 References

Black, A. (2019). A picture of special educational needs in England–an overview. In *Frontiers in Education* (Vol. 4, p. 79). Frontiers.

Burke, D. L., Ensor, J., & Riley, R. D. (2017). Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Statistics in Medicine*, 36(5), 855-875.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). Introduction to meta-analysis. John Wiley & Sons.

Brookes, S. T., Whitley, E., Peters, T. J., Mulheran, P. A., Egger, M., & Davey Smith, G. (2001). Subgroup analysis in randomised controlled trials: quantifying the risks of false-positives and false-negatives.

Bowden, J., Tierney, J. F., Copas, A. J., & Burdett, S. (2011). Quantifying, displaying and accounting for heterogeneity in the meta-analysis of RCTs using standard and generalised Q statistics. BMC medical research methodology, 11(1), 1-12.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, *10*, 101-129.

Chowdry, H., Crawford, C., Dearden, L., Joyce, R., Sibieta, L., Sylva, K., & Washbrook, E. (2010). Poorer children's educational attainment: how important are attitudes and behaviour. *Joseph Rowntree Foundation*, 1-72.

Coe, R., Weidmann, B., Coleman, R., & Kay, J. (2020). Impact of school closures on the attainment gap: rapid evidence assessment. June 2020.

Debray, T.P., Moons, K.G., van Valkenhoef, G., Efthimiou, O., Hummel, N., Groenwold, R.H., Reitsma, J.B., & GetReal Methods Review Group (2015). Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Research synthesis methods*, 6(4), pp.293-309.

DfE (Department for Education). (2014). Children and Families Act London: HMSO(2014).

Debray, T. P., Moons, K. G., van Valkenhoef, G., Efthimiou, O., Hummel, N., Groenwold, R. H., ... & GetReal Methods Review Group. (2015). Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Research synthesis methods*, *6*(4), 293-309.

Department for Work & Pensions (DWP). (2013). *Free school meal entitlement and child poverty in England*. Retrieved from https://dera.ioe.ac.uk/19084/1/free-school-meals-and-poverty.pdf

Duchateau, L., Janssen, P., & Rowlands, J. (1998). *Linear mixed models. An introduction with applications in veterinary research*. ILRI (aka ILCA and ILRAD).

Education Endowment Foundation. (2020). Retrieved from https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit/one-to-one-tuition/

Education Endowment Foundation. (2019). Retrieved from https://educationendowmentfoundation.org.uk/.

Education Endowment Foundation. (2017). *The attainment gap: 2017*. London: Education Endowment Foundation.

Ess, C., & Sudweeks, F. (2001). The problem lies with the 'ecological fallacy' (or fallacy of division) the impulse to apply group or societal level characteristics onto individuals within that group. *Culture, technology, communication: towards an intercultural global village*. SUNY Press, p.90.

Fanshawe, T. R., & Perera, R. (2019). Conducting one-stage IPD meta-analysis: which approach should I choose?. *BMJ evidence-based medicine*, bmjebm-2018.

Gillies, R. M., & Ashman, A. F. (Eds.). (2003). *Co-operative learning: The social and intellectual outcomes of learning in groups*. Psychology Press.

Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a metaanalysis. *Statistics in Medicine, 21*, 1539-1558.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.

Hobbs, G., & Vignoles, A. (2010). Is children's free school meal 'eligibility 'a good proxy for family income? *British Educational Research Journal*, 36(4), 673-690.

Humphrey, N., Wigelsworth, M., Barlow, A., & Squires, G. (2013). The role of school and individual differences in the academic attainment of learners with special educational needs and disabilities: a multi-level analysis. *International Journal of Inclusive Education*, 17(9), 909-931. doi: 10.1080/13603116.2012.718373

Jacques, D. (2000). *Learning in a Group: A Handbook for Improving Group Work*. Routledge.

Jerrim, J., Greany, T., & Perera, N. (2018). *Educational disadvantage: how does England compare?.* London: Education Policy Institute.

Kontopantelis, E. (2018). A comparison of one-stage vs two-stage individual patient data meta-analysis methods: A simulation study. *Research synthesis methods*, 9(3), 417-430.

Kruschke, J. K. (2015). Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan. Waltham, *MA: Academic Press / Elsevier.*

Laird, N. (2004). Random effects and the linear mixed model. In Analysis of Longitudinal and Cluster-Correlated Data (pp. 79-95). *Institute of Mathematical Statistics and the American Statistical Association*.

Lesaffre, E., & Lawson, A. B. (2012). Bayesian biostatistics. John Wiley & Sons.

Lloyd, C., Edovald, T., Morris, S. P., Skipp, A., Kiss, Z., & Haywood, S. (2015). *Durham Shared Maths project. Evaluation report and executive summary*. London: Education Endowment Foundation.

Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned?. *Educational Researcher*, *48*(3), 158-166.

Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. Review of educational research, 71(3), 449-521.

Manstead, A. (2014). *The wellbeing effect of education. Evidence Briefing*. Swindon: Economic and Social Research Council.

Marmot, M. (2010). *Fair society, healthy lives: the marmot review*. London: University College.

Mathew, T., & Nordström, K. (2010). Comparison of one-step and two-step meta-analysis models using individual patient data. Biometrical Journal: *Journal of Mathematical Methods in Biosciences*, 52(2), 271-287.

MRC. (2020). Individual Participant Data Meta-analyses. https://www.ctu.mrc.ac.uk/about-us/our-structure/individual-participant-data-ipd-meta-analyses/

Pigott, T. (2012). Advances in meta-analysis. New York: Springer, p.9.

Plummer, M. (2013). rjags: Bayesian graphical models using MCMC. R package version, 3(10). http://mcmc-jags.sourceforge.net

Reade, M.C., Delaney, A., Bailey, M.J., & Angus, D.C., 2008. Bench-to-bedside review: avoiding pitfalls in critical care meta-analysis–funnel plots, risk estimates, types of heterogeneity, baseline risk and the ecologic fallacy. *Critical Care*, 12(4), p.220.

Rienzo, C., Rolfe, H., & Wilkinson, D. (2016). *Powerful Learning Conversations: Evaluation Report and Executive Summary. London: Education Endowment Foundation*.

Schochet, P. Z., Puma, M., & Deke, J. (2014). *Understanding variation in treatment effects in education impact evaluations: An overview of quantitative methods.* US Department of Education, Washington, DC. Report No. NCEE, 4017.

Scottish Government (SG) (2014). *Raising Attainment for All – Scotland: The Best Place in the World to Go to School.* Edinburgh: Scottish Government.

Smith, C. T., Marcucci, M., Nolan, S. J., Iorio, A., Sudell, M., Riley, R., ... & Williamson, P. R. (2016). Individual participant data meta-analyses compared with meta-analyses based on aggregate data. *Cochrane Database of Systematic Reviews*, (9).

Spencer, S. (2015). *The cost of the school day.* Report, Child Poverty Action Group in Scotland, Glasgow.

Torgerson, D. J., & Torgerson, C. J. (2008). *Designing randomised trials in health, education and the social sciences: An introduction.* London, UK: Palgrave Macmillan.

Xiao, Z., Kasim, A., & Higgins, S. (2016). Same difference? Understanding variation in the estimation of Effect sizes from educational trials. *International Journal of Educational Research*, 77, 1–14. doi: 10.1016/j.ijer.2016.02.00118

Appendix 1: List of EEF trials meta-analysed in this report

Table A1: List of EF trials analysed in this study.

S.No.	Web-link for trials	Trial name
	One treatment trial	
1	Link to EEF Website	Future Foundations
2	Link to EEF Website	Switch-on Reading
3	Link to EEF Website	Grammar for Writing
4	Link to EEF Website	Rhythm for Reading
5	Link to EEF Website	Response to Intervention
6	Link to EEF Website	Effective Feedback
7	Link to EEF Website	Changing Mindsets - Pupil
8	Link to EEF Website	Catch-up Numeracy
9	Link to EEF Website	Chatterbooks
10	Link to EEF Website	Discover Summer School
11	Link to EEF Website	Literacy Programme
12	Link to EEF Website	Rapid Phonics
13	Link to EEF Website	Accelerated Reader
14	Link to EEF Website	Butterfly Phonics
15	Link to EEF Website	Improving Writing Quality
16	Link to EEF Website	Summer Active Reading Programme
17	Link to EEF Website	TextNow Transition Programme
18	Link to EEF Website	Affordable Individual & Small Group Tuition (P)
19	Link to EEF Website	Hampshire Hundreds
20	Link to EEF Website	Units of Sound
21	Link to EEF Website	Vocabulary Enrichment Intervention Programme
22	Link to EEF Website	Increasing Pupil Motivation
23	Link to EEF Website	Word and Word Reading Programme (CC)
24	Link to EEF Website	REACH
25	Link to EEF Website	Catch-up Literacy
26	Link to EEF Website	Fresh Start
27	Link to EEF Website	Talk for Literacy
28	Link to EEF Website	Teacher Effectiveness Enhancement Programme
29	Link to EEF Website	Math Mastery Secondary
30	Link to EEF Website	Tutoring with Alphie
31	Link to EEF Website	Quest
32	Link to EEF Website	Philosophy for Children
33	Link to EEF Website	Affordable Individual & Small Group Tuition (E)
34	Link to EEF Website	Lesson Study
35	Link to EEF Website	SHINE in Secondaries
36	Link to EEF Website	Talk of the Town

37	Link to EEF Website	Thinking, Doing, Talking Science
38	Link to EEF Website	Success for All
39	Link to EEF Website	Chess in Schools
40	Link to EEF Website	Let's Think Secondary Science
41	Link to EEF Website	Powerful Learning Conversations
42	Link to EEF Website	Affordable Online Maths Tuition
43	Link to EEF Website	Texting Parents
44	Link to EEF Website	Online Reading Programme (ABRA)
45	Link to EEF Website	Flipped Learning
46	Link to EEF Website	Graduate Coaching Programme
47	Link to EEF Website	Paired Reading
48	Link to EEF Website	Changing Mindsets - Inset
49	Link to EEF Website	Youth Social Action Trials(Y)
50	Link to EEF Website	Affordable Individual & Small Group Tuition(M)
51	Link to EEF Website	ReflectEd
52	Link to EEF Website	Dialogic Teaching
53	Link to EEF Website	Learner Response System
54	Link to EEF Website	Teacher Observation
55	Link to EEF Website	Research Learning Communities
56	Link to EEF Website	Best Practice in Grouping Students
57	Link to EEF Website	Childrens University
58	Link to EEF Website	ScratchMaths
59	Link to EEF Website	Good Behaviour Game
60	Link to EEF Website	Online Reading Programme (A)
61	Link to EEF Website	GraphoGame Rime
62	Link to EEF Website	Embedding Formative Assessment
63	Link to EEF Website	Zippys Friends
64	Link to EEF Website	1stClass@Number
65	Link to EEF Website	Tutor Trust: Affordable Primary Tuition
66	Link to EEF Website	Huntington Rise
67	Link to EEF Website	Maths Counts
68	Link to EEF Website	Grammar for Writing (et)
69	Link to EEF Website	IPEELL one year
70	Link to EEF Website	Catch-up Literacy (effect)
71	Link to EEF Website	Maths Reasoning
72	Link to EEF Website	Best Practice in Grouping Students (M)
73	Link to EEF Website	IPEELL two year
74	Link to EEF Website	onebillion
75	Link to EEF Website	Families & Schools Together (FAST)
76	Link to EEF Website	Changing Mindsets
Two treatment trials		

77	Link to EEF Website	Shared Maths 1
78	Link to EEF Website	Shared Maths 2
79	Link to EEF Website	Act, Sing, Play 1
80	Link to EEF Website	Act, Sing, Play 2
81	Link to EEF Website	Nuffield Early Language Intervention 1
82	Link to EEF Website	Nuffield Early Language Intervention 2
83	Link to EEF Website	Improving Numeracy and Literacy KS 1
84	Link to EEF Website	Improving Numeracy and Literacy KS 2
85	Link to EEF Website	Parenting Academy 1
86	Link to EEF Website	Parenting Academy 2
87	Link to EEF Website	Switch-on Effectiveness T 1
88	Link to EEF Website	Switch-on Effectiveness T 2

Table A2: Types of Intervention as per toolkit strand classification adopted in EEF Evidence database project.

No.	Types of Intervention
1	Arts participation
2	Aspiration interventions
3	Behaviour interventions
4	Block scheduling
5	Built environment
6	Collaborative learning
7	Digital technology
8	Early years intervention
9	Extending school time
10	Feedback
11	Homework
12	Individualised instruction
13	Learning styles
14	Mastery learning
15	Metacognition and self-regulation
16	Mentoring
17	One to one tuition
18	Oral language interventions
19	Outdoor adventure learning
20	Parental engagement
21	Peer Tutoring
22	Performance pay
23	Phonics
24	Reading comprehension strategies
25	Reducing class size
26	Repeating a year
27	School uniform
28	Setting or streaming
29	Small Group Tuition
30	Social and emotional learning
31	Sports participation
32	Summer schools
33	Teaching assistants

Appendix 2: Comparison of Pooled and individual trial effect sizes in EEF trials for different meta-analysis methods

Table A1. Comparison of Pooled and individual trial effect sizes for FSM subgroup literacy outcome using IPD and two-stage fixed effect (FE) and random effect (RE) meta-analysis methods.

			Standardised Score			Raw Score						Standardised Score					
	Number of		IPD			Two stage (FE)			Two stage (RE)			Two stage (FE)			Two stage (RE)		
Trial	Schools	Pupils	ES	Low	Up	ES	Low	Up	ES	Low	Up	ES	Low	Up	ES	Low	Up
Pooled ES	3804	90218	0.06	0.03	0.08	0.03	0.01	0.05	0.04	0.01	0.07	0.02	0.00	0.04	0.03	0.01	0.06
Shared Maths 1	34	157	0.42	-0.07	0.93	0.55	-0.09	1.18	0.55	-0.09	1.18	0.42	-0.07	0.91	0.42	-0.07	0.91
Graduate Coaching Programme	4	124	0.38	0.13	0.64	0.53	0.18	0.87	0.53	0.18	0.87	0.39	0.13	0.64	0.39	0.13	0.64
Accelerated Reader	4	109	0.34	0.08	0.60	0.50	0.12	0.87	0.50	0.12	0.87	0.34	0.09	0.60	0.34	0.09	0.60
Online Reading Programme (ABRA)	45	431	0.33	0.09	0.56	0.41	0.11	0.70	0.41	0.11	0.70	0.33	0.08	0.56	0.33	0.08	0.56
Response to Intervention	30	94	0.32	-0.11	0.74	0.37	-0.13	0.86	0.37	-0.13	0.86	0.31	-0.11	0.72	0.31	-0.11	0.72
Nuff Early Language Intervn 1	17	38	0.30	-0.26	0.88	0.31	-0.24	0.87	0.31	-0.24	0.87	0.29	-0.22	0.82	0.29	-0.22	0.82
Nuff Early Language Intervn 2	20	46	0.28	-0.24	0.80	0.34	-0.26	0.92	0.34	-0.26	0.92	0.28	-0.21	0.77	0.28	-0.21	0.77
Butterfly Phonics	6	209	0.27	0.08	0.47	0.39	0.12	0.66	0.39	0.12	0.66	0.27	0.09	0.46	0.27	0.09	0.46
REACH	19	88	0.23	-0.08	0.55	0.30	-0.10	0.70	0.30	-0.10	0.70	0.23	-0.08	0.54	0.23	-0.08	0.54
Summer Active Reading Programme	32	61	0.22	-0.26	0.69	0.23	-0.27	0.74	0.23	-0.27	0.74	0.22	-0.25	0.69	0.22	-0.25	0.69
Talk for Literacy	3	61	0.22	-0.14	0.58	0.32	-0.20	0.81	0.32	-0.20	0.81	0.22	-0.14	0.57	0.22	-0.14	0.57
Team Alphie	6	33	0.20	-0.57	1.06	0.14	-0.54	0.88	0.14	-0.54	0.88	0.15	-0.57	0.93	0.15	-0.57	0.93
Flipped Learning	24	430	0.19	-0.15	0.54	0.22	-0.17	0.61	0.22	-0.17	0.61	0.18	-0.14	0.51	0.18	-0.14	0.51
Shared Maths 2	33	154	0.17	-0.06	0.40	0.30	-0.12	0.73	0.30	-0.12	0.73	0.17	-0.07	0.41	0.17	-0.07	0.41
TextNow Transition Programme	45	116	0.16	-0.16	0.49	0.18	-0.16	0.52	0.18	-0.16	0.52	0.17	-0.14	0.48	0.17	-0.14	0.48
Future Foundations	28	170	0.15	-0.04	0.33	0.17	-0.04	0.37	0.17	-0.04	0.37	0.15	-0.03	0.33	0.15	-0.03	0.33
Success for All	46	322	0.13	-0.11	0.38	0.15	-0.11	0.41	0.15	-0.11	0.41	0.15	-0.11	0.40	0.15	-0.11	0.40
Rhythm for Reading	6	137	0.12	-0.13	0.37	0.14	-0.16	0.45	0.14	-0.16	0.45	0.12	-0.13	0.36	0.12	-0.13	0.36
Dialogic Teaching	69	614	0.12	-0.07	0.30	0.14	-0.09	0.38	0.14	-0.09	0.38	0.12	-0.07	0.30	0.12	-0.07	0.30
Online Reading Programme(A)	45	399	0.12	-0.10	0.34	0.16	-0.14	0.46	0.16	-0.14	0.46	0.12	-0.10	0.35	0.12	-0.10	0.35
Switch-on Reading	18	98	0.10	-0.18	0.38	0.11	-0.27	0.51	0.11	-0.27	0.51	0.08	-0.19	0.36	0.08	-0.19	0.36
Catch-up Literacy (effect)	132	495	0.10	-0.09	0.30	0.12	-0.13	0.37	0.12	-0.13	0.37	0.10	-0.10	0.30	0.10	-0.10	0.30
IPEELL two year	75	809	0.10	-0.09	0.30	0.13	-0.12	0.38	0.13	-0.12	0.38	0.11	-0.09	0.30	0.11	-0.09	0.30
Improv Num and Lit KS 2	31	214	0.10	-0.10	0.30	0.17	-0.14	0.49	0.17	-0.14	0.49	0.10	-0.08	0.28	0.10	-0.08	0.28
Changing Mindsets - Pupil	5	64	0.09	-0.25	0.43	0.11	-0.38	0.60	0.11	-0.38	0.60	0.08	-0.25	0.40	0.08	-0.25	0.40
LIT Programme	34	1416	0.09	-0.01	0.18	0.13	-0.02	0.29	0.13	-0.02	0.29	0.09	-0.01	0.18	0.09	-0.01	0.18
Powerful Learning Conversations	15	190	0.09	-0.21	0.39	0.10	-0.28	0.48	0.10	-0.28	0.48	0.08	-0.23	0.40	0.08	-0.23	0.40
Improv Num and Lit KS 1	30	253	0.09	-0.13	0.33	0.15	-0.21	0.52	0.15	-0.21	0.52	0.10	-0.13	0.32	0.10	-0.13	0.32
SHINE in Secondaries	4	332	0.08	-0.14	0.30	0.14	-0.24	0.51	0.14	-0.24	0.51	0.08	-0.14	0.29	0.08	-0.14	0.29
Teacher Observation	82	8157	0.08	-0.03	0.18	0.08	-0.03	0.20	0.08	-0.03	0.20	0.08	-0.03	0.18	0.08	-0.03	0.18

Vocab Enrichment Intrvn Prgm	12	159	0.07	-0.16	0.31	0.10	-0.21	0.42	0.10	-0.21	0.42	0.07	-0.15	0.30	0.07	-0.15	0.30
GraphoGame Rime	12	111	0.07	-0.26	0.41	0.09	-0.29	0.46	0.09	-0.29	0.46	0.08	-0.25	0.40	0.08	-0.25	0.40
Grammar for Writing	50	722	0.06	-0.06	0.18	0.06	-0.07	0.19	0.06	-0.07	0.19	0.06	-0.07	0.18	0.06	-0.07	0.18
Discover Summer School	9	32	0.06	-0.29	0.39	0.06	-0.26	0.36	0.06	-0.26	0.36	0.06	-0.26	0.37	0.06	-0.26	0.37
Good Behaviour Game	75	589	0.06	-0.11	0.22	0.08	-0.14	0.31	0.08	-0.14	0.31	0.06	-0.10	0.22	0.06	-0.10	0.22
Best Practice in Grp Students(M)	8	109	0.06	-0.49	0.65	0.12	-0.53	0.79	0.12	-0.53	0.79	0.10	-0.43	0.65	0.10	-0.43	0.65
Research Learning Communities	116	1725	0.05	-0.10	0.20	0.06	-0.13	0.25	0.06	-0.13	0.25	0.05	-0.10	0.19	0.05	-0.10	0.19
Tutor Trust: Afford Primary Tuition	103	1622	0.05	-0.08	0.19	0.07	-0.10	0.24	0.07	-0.10	0.24	0.05	-0.08	0.19	0.05	-0.08	0.19
Catch-up Numeracy	37	105	0.04	-0.16	0.24	0.07	-0.32	0.46	0.07	-0.32	0.46	0.03	-0.16	0.23	0.03	-0.16	0.23
Catch-up Literacy	15	113	0.03	-0.18	0.25	0.06	-0.31	0.42	0.06	-0.31	0.42	0.04	-0.18	0.25	0.04	-0.18	0.25
Fresh Start	10	104	0.03	-0.25	0.31	0.03	-0.35	0.42	0.03	-0.35	0.42	0.02	-0.25	0.30	0.02	-0.25	0.30
Afford Ind & Small Grp Tuition(E)	272	17128	0.03	-0.21	0.28	0.04	-0.24	0.32	0.04	-0.24	0.32	0.03	-0.21	0.28	0.03	-0.21	0.28
Lesson Study	181	5812	0.03	-0.06	0.12	0.04	-0.07	0.16	0.04	-0.07	0.16	0.03	-0.06	0.12	0.03	-0.06	0.12
Zippys Friends	75	493	0.03	-0.10	0.16	0.05	-0.13	0.23	0.05	-0.13	0.23	0.03	-0.09	0.15	0.03	-0.09	0.15
Grammar for Writing (et)	154	3136	0.03	-0.11	0.17	0.03	-0.12	0.18	0.03	-0.12	0.18	0.03	-0.11	0.16	0.03	-0.11	0.16
Changing Mindsets	101	1573	0.03	-0.09	0.15	0.04	-0.11	0.20	0.04	-0.11	0.20	0.03	-0.09	0.15	0.03	-0.09	0.15
W & W Reading Programme (CC)	14	395	0.02	-0.83	0.84	0.01	-0.78	0.81	0.01	-0.78	0.81	0.01	-0.76	0.78	0.01	-0.76	0.78
Texting Parents	29	1745	0.02	-0.05	0.10	0.03	-0.07	0.14	0.03	-0.07	0.14	0.02	-0.06	0.10	0.02	-0.06	0.10
Effective Feedback	13	360	0.01	-0.20	0.19	0.00	-0.35	0.33	0.00	-0.35	0.33	0.00	-0.18	0.17	0.00	-0.18	0.17
Teacher Effec Enht Programme	45	2524	0.01	-0.17	0.19	0.01	-0.17	0.19	0.01	-0.17	0.19	0.01	-0.17	0.18	0.01	-0.17	0.18
Learner Response System	99	3462	0.01	-0.12	0.15	0.01	-0.12	0.15	0.01	-0.12	0.15	0.01	-0.12	0.15	0.01	-0.12	0.15
Emb Formative Assessment	140	6489	0.01	-0.09	0.10	0.01	-0.11	0.13	0.01	-0.11	0.13	0.01	-0.09	0.10	0.01	-0.09	0.10
Afford & Small Grp Tuition (P)	10	486	0.00	-0.20	0.20	0.01	-0.26	0.27	0.01	-0.26	0.27	0.01	-0.20	0.21	0.01	-0.20	0.21
Hampshire Hundreds	36	645	0.00	-0.09	0.09	0.00	-0.22	0.22	0.00	-0.22	0.22	0.00	-0.09	0.09	0.00	-0.09	0.09
Increasing Pupil Motivation	63	3766	0.00	-0.16	0.15	0.00	-0.20	0.20	0.00	-0.20	0.20	0.00	-0.15	0.15	0.00	-0.15	0.15
Switch-on Effectiveness T 1	116	1372	0.00	-0.12	0.12	0.00	-0.18	0.18	0.00	-0.18	0.18	0.00	-0.12	0.12	0.00	-0.12	0.12
Philosophy for Children	45	774	-0.01	-0.39	0.37	-0.01	-0.37	0.35	-0.01	-0.37	0.35	-0.01	-0.36	0.34	-0.01	-0.36	0.34
Best Practice in Grp Students	37	265	-0.01	-0.23	0.19	-0.01	-0.31	0.29	-0.01	-0.31	0.29	-0.01	-0.23	0.21	-0.01	-0.23	0.21
Huntington Rise	39	3517	-0.01	-0.15	0.12	-0.02	-0.17	0.14	-0.02	-0.17	0.14	-0.01	-0.14	0.12	-0.01	-0.14	0.12
Switch-on Effectiveness T 2	117	1378	-0.01	-0.13	0.11	-0.02	-0.20	0.16	-0.02	-0.20	0.16	-0.01	-0.13	0.10	-0.01	-0.13	0.10
Improving Writing Quality	16	123	-0.02	-0.45	0.40	-0.01	-0.46	0.43	-0.01	-0.46	0.43	-0.01	-0.45	0.42	-0.01	-0.45	0.42
Fam & Schools Tog (FAST)	116	1526	-0.02	-0.18	0.14	-0.02	-0.20	0.15	-0.02	-0.20	0.15	-0.02	-0.18	0.14	-0.02	-0.18	0.14
Act, Sing, Play 1	15	118	-0.02	-0.25	0.22	-0.02	-0.37	0.33	-0.02	-0.37	0.33	-0.01	-0.25	0.22	-0.01	-0.25	0.22
Parenting Academy 2	16	924	-0.02	-0.12	0.08	-0.03	-0.16	0.10	-0.03	-0.16	0.10	-0.02	-0.12	0.08	-0.02	-0.12	0.08
Parenting Academy 1	16	870	-0.03	-0.13	0.08	-0.04	-0.18	0.10	-0.04	-0.18	0.10	-0.03	-0.13	0.08	-0.03	-0.13	0.08
Talk of the Town	62	984	-0.04	-0.17	0.09	-0.05	-0.22	0.12	-0.05	-0.22	0.12	-0.04	-0.18	0.09	-0.04	-0.18	0.09
Chatterbooks	12	128	-0.05	-0.35	0.26	-0.06	-0.42	0.29	-0.06	-0.42	0.29	-0.05	-0.34	0.24	-0.05	-0.34	0.24
Act, Sing, Play 2	16	133	-0.05	-0.32	0.20	-0.06	-0.40	0.27	-0.06	-0.40	0.27	-0.05	-0.30	0.20	-0.05	-0.30	0.20
Paired Reading	10	498	-0.06	-0.17	0.05	-0.09	-0.26	0.08	-0.09	-0.26	0.08	-0.06	-0.16	0.05	-0.06	-0.16	0.05
Rapid Phonics	18	94	-0.08	-0.42	0.27	-0.08	-0.48	0.34	-0.08	-0.48	0.34	-0.07	-0.41	0.28	-0.07	-0.41	0.28
Chess in Schools	100	1288	-0.08	-0.25	0.11	-0.07	-0.25	0.10	-0.07	-0.25	0.10	-0.07	-0.25	0.10	-0.07	-0.25	0.10
Quest	19	826	-0.09	-0.30	0.12	-0.13	-0.40	0.15	-0.13	-0.40	0.15	-0.09	-0.28	0.10	-0.09	-0.28	0.10
Let's Think Secondary Sc	20	717	-0.09	-0.42	0.23	-0.13	-0.55	0.29	-0.13	-0.55	0.29	-0.09	-0.39	0.21	-0.09	-0.39	0.21

Changing Mindsets - Inset	23	176	-0.09	-0.29	0.12	-0.14	-0.44	0.16	-0.14	-0.44	0.16	-0.09	-0.28	0.11	-0.09	-0.28	0.11
ReflectEd	28	465	-0.09	-0.22	0.04	-0.10	-0.25	0.05	-0.10	-0.25	0.05	-0.09	-0.22	0.04	-0.09	-0.22	0.04
Units of Sound	33	255	-0.11	-0.28	0.06	-0.15	-0.40	0.09	-0.15	-0.40	0.09	-0.11	-0.27	0.06	-0.11	-0.27	0.06
Afford Online Maths Tuition	63	762	-0.11	-0.31	0.10	-0.12	-0.37	0.12	-0.12	-0.37	0.12	-0.10	-0.31	0.10	-0.10	-0.31	0.10
Youth Social Action Trials(Y)	63	1529	-0.13	-0.47	0.23	-0.12	-0.44	0.21	-0.12	-0.44	0.21	-0.12	-0.47	0.22	-0.12	-0.47	0.22
Childrens University	66	996	-0.14	-0.30	0.02	-0.15	-0.32	0.02	-0.15	-0.32	0.02	-0.14	-0.31	0.02	-0.14	-0.31	0.02
IPEELL one year	80	936	-0.16	-0.33	0.01	-0.22	-0.44	0.01	-0.22	-0.44	0.01	-0.16	-0.33	0.00	-0.16	-0.33	0.00
Thinking, Doing, Talking Sc	37	338	-0.20	-0.44	0.04	-0.23	-0.50	0.04	-0.23	-0.50	0.04	-0.20	-0.44	0.04	-0.20	-0.44	0.04

Table A2. Comparison of pooled and individual trial effect sizes for FSM subgroup maths outcome using IPD and two-stage fixed effect (FE) and random effect (RE) meta-analysis methods.

	Number of		Standardised Score IPD			Raw Score Two stage (FE)			Raw Score Two stage (RE)			Standardised Score Two stage (FE)			Standardised Score Two stage (RE)		
Trial	Schools	Pupils	ES	Low	Up	ES	Low	Up	ES	Low	Up	ES	Low	Up	ES	Low	Up
Pooled ES	3006	89247	0.00	-0.03	0.04	0.01	-0.02	0.04	0.00	-0.03	0.03	0.01	-0.01	0.03	0.00	-0.02	0.02
Powerful Learning Conversations	11	84	0.31	-0.25	0.98	0.38	-0.30	1.08	0.38	-0.30	1.08	0.32	-0.25	0.91	0.32	-0.25	0.91
Dialogic Teaching	69	637	0.16	0.02	0.29	0.22	0.03	0.42	0.22	0.03	0.42	0.16	0.02	0.29	0.16	0.02	0.29
Improv Num and Lit KS 2	31	220	0.13	-0.10	0.36	0.20	-0.15	0.55	0.20	-0.15	0.55	0.13	-0.10	0.36	0.13	-0.10	0.36
Act, Sing, Play 1	15	119	0.12	-0.10	0.35	0.18	-0.14	0.51	0.18	-0.14	0.51	0.12	-0.10	0.35	0.12	-0.10	0.35
Afford Ind & Small Grp Tuition(M)	312	27746	0.11	0.01	0.22	0.15	0.01	0.30	0.15	0.01	0.30	0.11	0.01	0.22	0.11	0.01	0.22
Flipped Learning	24	427	0.07	-0.20	0.35	0.10	-0.28	0.48	0.10	-0.28	0.48	0.07	-0.20	0.34	0.07	-0.20	0.34
Math Mastery Secondary	43	1609	0.06	-0.08	0.21	0.09	-0.11	0.28	0.09	-0.11	0.28	0.06	-0.08	0.21	0.06	-0.08	0.21
Tutor Trust: Afford Primary Tuition	103	1634	0.06	-0.10	0.21	0.07	-0.13	0.28	0.07	-0.13	0.28	0.06	-0.10	0.21	0.06	-0.10	0.21
Maths Reasoning	148	1342	0.06	-0.09	0.20	0.07	-0.11	0.24	0.07	-0.11	0.24	0.06	-0.09	0.20	0.06	-0.09	0.20
Lesson Study	181	5703	0.05	-0.04	0.14	0.06	-0.06	0.19	0.06	-0.06	0.19	0.04	-0.05	0.14	0.04	-0.05	0.14
Texting Parents	29	1683	0.05	-0.02	0.12	0.08	-0.02	0.19	0.08	-0.02	0.19	0.05	-0.02	0.12	0.05	-0.02	0.12
Shared Maths 2	74	535	0.05	-0.07	0.18	0.10	-0.14	0.35	0.10	-0.14	0.35	0.06	-0.07	0.18	0.06	-0.07	0.18
Hampshire Hundreds	36	645	0.03	-0.06	0.11	0.07	-0.17	0.31	0.07	-0.17	0.31	0.03	-0.06	0.12	0.03	-0.06	0.12
Learner Response System	99	3537	0.03	-0.10	0.17	0.03	-0.11	0.17	0.03	-0.11	0.17	0.03	-0.11	0.18	0.03	-0.11	0.18
Teacher Observation	82	8128	0.03	-0.07	0.12	0.04	-0.08	0.16	0.04	-0.08	0.16	0.03	-0.06	0.12	0.03	-0.06	0.12
onebillion	87	274	0.03	-0.18	0.24	0.04	-0.24	0.32	0.04	-0.24	0.32	0.03	-0.18	0.25	0.03	-0.18	0.25
Let's Think Secondary Sc	16	439	0.02	-0.15	0.20	0.03	-0.24	0.30	0.03	-0.24	0.30	0.02	-0.16	0.20	0.02	-0.16	0.20
Changing Mindsets - Inset	23	178	0.02	-0.18	0.22	0.04	-0.29	0.38	0.04	-0.29	0.38	0.02	-0.18	0.23	0.02	-0.18	0.23
Huntington Rise	39	3547	0.02	-0.10	0.13	0.02	-0.13	0.17	0.02	-0.13	0.17	0.02	-0.10	0.13	0.02	-0.10	0.13
Parenting Academy 2	16	937	0.02	-0.08	0.12	0.02	-0.11	0.15	0.02	-0.11	0.15	0.02	-0.08	0.11	0.02	-0.08	0.11
Philosophy for Children	45	774	0.01	-0.38	0.39	0.01	-0.40	0.44	0.01	-0.40	0.44	0.01	-0.37	0.40	0.01	-0.37	0.40
Chess in Schools	100	1291	0.01	-0.18	0.19	0.00	-0.17	0.18	0.00	-0.17	0.18	0.01	-0.18	0.19	0.01	-0.18	0.19
Best Practice in Grp Students	75	713	0.01	-0.15	0.17	0.02	-0.23	0.26	0.02	-0.23	0.26	0.01	-0.15	0.17	0.01	-0.15	0.17
Effective Feedback	13	361	0.00	-0.30	0.34	0.01	-0.54	0.57	0.01	-0.54	0.57	0.00	-0.27	0.29	0.00	-0.27	0.29
ScratchMaths	109	1636	0.00	-0.15	0.16	0.00	-0.19	0.20	0.00	-0.19	0.20	0.00	-0.15	0.16	0.00	-0.15	0.16
Changing Mindsets	101	1575	0.00	-0.14	0.14	0.00	-0.19	0.19	0.00	-0.19	0.19	0.00	-0.14	0.14	0.00	-0.14	0.14
Emb Formative Assessment	140	6564	-0.01	-0.08	0.08	-0.01	-0.13	0.11	-0.01	-0.13	0.11	-0.01	-0.09	0.07	-0.01	-0.09	0.07

Parenting Academy 1	16	888	-0.01	-0.11	0.10	-0.01	-0.15	0.13	-0.01	-0.15	0.13	-0.01	-0.11	0.10	-0.01	-0.11	0.10
Changing Mindsets - Pupil	5	61	-0.02	-0.24	0.19	-0.04	-0.52	0.44	-0.04	-0.52	0.44	-0.02	-0.24	0.20	-0.02	-0.24	0.20
1stClass@Number	85	149	-0.02	-0.37	0.33	-0.02	-0.37	0.34	-0.02	-0.37	0.34	-0.02	-0.36	0.34	-0.02	-0.36	0.34
Increasing Pupil Motivation	63	3752	-0.03	-0.21	0.14	-0.04	-0.26	0.18	-0.04	-0.26	0.18	-0.03	-0.21	0.14	-0.03	-0.21	0.14
Shared Maths 1	76	554	-0.03	-0.15	0.10	-0.03	-0.22	0.15	-0.03	-0.22	0.15	-0.02	-0.15	0.10	-0.02	-0.15	0.10
Catch-up Numeracy	37	106	-0.04	-0.30	0.21	-0.04	-0.30	0.22	-0.04	-0.30	0.22	-0.04	-0.29	0.20	-0.04	-0.29	0.20
Afford Online Maths Tuition	63	786	-0.04	-0.22	0.15	-0.06	-0.32	0.19	-0.06	-0.32	0.19	-0.04	-0.22	0.14	-0.04	-0.22	0.14
Maths Counts	34	133	-0.04	-0.29	0.21	-0.04	-0.35	0.27	-0.04	-0.35	0.27	-0.03	-0.28	0.21	-0.03	-0.28	0.21
Best Practice in Grp Students(M)	9	114	-0.04	-0.83	0.73	-0.04	-1.00	0.93	-0.04	-1.00	0.93	-0.03	-0.70	0.65	-0.03	-0.70	0.65
Act, Sing, Play 2	16	135	-0.04	-0.27	0.18	-0.06	-0.38	0.25	-0.06	-0.38	0.25	-0.04	-0.26	0.17	-0.04	-0.26	0.17
Teacher Effec Enht Programme	45	2503	-0.06	-0.24	0.13	-0.06	-0.24	0.12	-0.06	-0.24	0.12	-0.06	-0.24	0.13	-0.06	-0.24	0.13
Fam & Schools Tog (FAST)	115	1538	-0.06	-0.23	0.12	-0.07	-0.25	0.12	-0.07	-0.25	0.12	-0.06	-0.23	0.11	-0.06	-0.23	0.11
Improv Num and Lit KS 1	30	256	-0.09	-0.30	0.13	-0.11	-0.42	0.21	-0.11	-0.42	0.21	-0.08	-0.29	0.15	-0.08	-0.29	0.15
ReflectEd	28	544	-0.10	-0.21	0.01	-0.13	-0.28	0.01	-0.13	-0.28	0.01	-0.10	-0.21	0.01	-0.10	-0.21	0.01
Youth Social Action Trials(Y)	63	1529	-0.12	-0.42	0.18	-0.12	-0.40	0.16	-0.12	-0.40	0.16	-0.12	-0.41	0.17	-0.12	-0.41	0.17
Childrens University	66	1002	-0.13	-0.30	0.04	-0.14	-0.32	0.04	-0.14	-0.32	0.04	-0.13	-0.31	0.04	-0.13	-0.31	0.04
IPEELL one year	80	950	-0.13	-0.32	0.08	-0.16	-0.41	0.09	-0.16	-0.41	0.09	-0.12	-0.32	0.07	-0.12	-0.32	0.07
Future Foundations	27	167	-0.14	-0.37	0.08	-0.17	-0.44	0.10	-0.17	-0.44	0.10	-0.14	-0.36	0.08	-0.14	-0.36	0.08
Afford Ind & Small Grp Tuition (P)	10	435	-0.16	-0.40	0.10	-0.21	-0.54	0.11	-0.21	-0.54	0.11	-0.15	-0.40	0.08	-0.15	-0.40	0.08
Thinking, Doing, Talking Sc	37	338	-0.16	-0.39	0.07	-0.22	-0.53	0.10	-0.22	-0.53	0.10	-0.16	-0.39	0.08	-0.16	-0.39	0.08
IPEELL two year	80	969	-0.18	-0.36	0.01	-0.24	-0.48	0.01	-0.24	-0.48	0.01	-0.18	-0.37	0.01	-0.18	-0.37	0.01

Table A3. Comparison of pooled and individual trial attainment gap between FSM and non-FSM pupils' literacy outcome using IPD and two stage fixed effect (FE) and random effect (RE) meta-analysis methods.

			Standardised Score			Raw Score						Standardised Score					
	Number of		IPD			Two stage (FE)			Two stage (RE)			Two stage (FE)			Two stage (RE)		
Trial	Schools	Pupils	ES	Low	Up	ES	Low	Up	ES	Low	Up	ES	Low	Up	ES	Low	Up
Overall gap	4000	302138	0.01	-0.01	0.04	0.00	-0.02	0.01	0.00	-0.03	0.02	0.00	-0.02	0.01	0.00	-0.02	0.02
Nuff Early Language Intervn 1	23	159	0.42	-0.22	1.06	0.49	-0.22	1.20	0.49	-0.22	1.20	0.43	-0.20	1.06	0.43	-0.20	1.06
TextNow Transition Programme	54	391	0.34	-0.01	0.68	0.40	-0.01	0.80	0.40	-0.01	0.80	0.34	0.00	0.68	0.34	0.00	0.68
Nuff Early Language Intervn 2	23	161	0.31	-0.32	0.92	0.37	-0.31	1.04	0.37	-0.31	1.04	0.32	-0.27	0.91	0.32	-0.27	0.91
Best Practice in Grp Students(M)	8	328	0.30	-0.21	0.79	0.35	-0.22	0.95	0.35	-0.22	0.95	0.29	-0.18	0.79	0.29	-0.18	0.79
Improv Num and Lit KS 1	37	1345	0.23	0.06	0.40	0.40	0.11	0.69	0.40	0.11	0.69	0.23	0.06	0.40	0.23	0.06	0.40
Afford Ind & Small Grp Tuition (P)	12	814	0.19	-0.14	0.53	0.24	-0.20	0.67	0.24	-0.20	0.67	0.18	-0.16	0.51	0.18	-0.16	0.51
Accelerated Reader	4	326	0.17	-0.19	0.52	0.22	-0.24	0.67	0.22	-0.24	0.67	0.17	-0.18	0.51	0.17	-0.18	0.51
Shared Maths 2	41	888	0.16	-0.04	0.36	0.32	-0.07	0.71	0.32	-0.07	0.71	0.16	-0.03	0.35	0.16	-0.03	0.35
Response to Intervention	48	373	0.15	-0.20	0.52	0.20	-0.26	0.69	0.20	-0.26	0.69	0.15	-0.19	0.50	0.15	-0.19	0.50
Rhythm for Reading	6	355	0.14	-0.17	0.46	0.18	-0.22	0.57	0.18	-0.22	0.57	0.14	-0.17	0.45	0.14	-0.17	0.45
Summer Active Reading Programme	48	182	0.12	-0.43	0.69	0.13	-0.47	0.74	0.13	-0.47	0.74	0.12	-0.42	0.67	0.12	-0.42	0.67
SHINE in Secondaries	4	548	0.12	-0.08	0.31	0.19	-0.14	0.53	0.19	-0.14	0.53	0.11	-0.08	0.31	0.11	-0.08	0.31
Graduate Coaching Programme	4	291	0.11	-0.22	0.45	0.15	-0.30	0.60	0.15	-0.30	0.60	0.11	-0.22	0.44	0.11	-0.22	0.44
Catch-up Literacy (effect)	141	1006	0.11	-0.08	0.30	0.14	-0.10	0.37	0.14	-0.10	0.37	0.11	-0.08	0.30	0.11	-0.08	0.30

Powerful Learning Conversations	15	1723	0.10	-0.15	0.32	0.12	-0.17	0.42	0.12	-0.17	0.42	0.10	-0.40	0.56	0.10	-0.40	0.56
Talk for Literacy	3	213	0.09	-0.38	0.58	0.12	-0.50	0.69	0.12	-0.50	0.69	0.10	-0.14	0.33	0.10	-0.14	0.33
Afford Ind & Small Grp Tuition(E)	285	63392	0.09	-0.06	0.22	0.11	-0.06	0.28	0.11	-0.06	0.28	0.09	-0.21	0.38	0.09	-0.21	0.38
Online Reading Programme (ABRA)	50	1969	0.09	-0.08	0.25	0.11	-0.10	0.32	0.11	-0.10	0.32	0.09	-0.05	0.23	0.09	-0.05	0.23
GraphoGame Rime	14	360	0.09	-0.30	0.45	0.09	-0.35	0.54	0.09	-0.35	0.54	0.09	-0.08	0.25	0.09	-0.08	0.25
Future Foundations	33	310	0.08	-0.21	0.39	0.10	-0.24	0.45	0.10	-0.24	0.45	0.08	-0.29	0.45	0.08	-0.29	0.45
Improv Num and Lit KS 2	35	1279	0.07	-0.13	0.26	0.12	-0.23	0.47	0.12	-0.23	0.47	0.07	-0.13	0.27	0.07	-0.13	0.27
Let's Think Secondary Sc	20	2400	0.06	-0.06	0.18	0.08	-0.09	0.25	0.08	-0.09	0.25	0.06	-0.06	0.18	0.06	-0.06	0.18
Learner Response System	99	5842	0.06	-0.04	0.17	0.06	-0.04	0.16	0.06	-0.04	0.16	0.06	-0.04	0.17	0.06	-0.04	0.17
Best Practice in Grp Students	37	939	0.06	-0.16	0.28	0.07	-0.21	0.36	0.07	-0.21	0.36	0.06	-0.02	0.13	0.06	-0.02	0.13
Grammar for Writing (et)	155	6955	0.06	-0.02	0.13	0.06	-0.03	0.16	0.06	-0.03	0.16	0.05	-0.16	0.27	0.05	-0.16	0.27
Good Behaviour Game	77	2504	0.05	-0.07	0.17	0.08	-0.11	0.26	0.08	-0.11	0.26	0.05	-0.07	0.17	0.05	-0.07	0.17
IPEELL one year	83	2429	0.05	-0.07	0.17	0.07	-0.10	0.23	0.07	-0.10	0.23	0.05	-0.07	0.17	0.05	-0.07	0.17
Zippys Friends	81	3306	0.04	-0.09	0.16	0.05	-0.14	0.25	0.05	-0.14	0.25	0.04	-0.09	0.17	0.04	-0.09	0.17
Lesson Study	181	24592	0.03	-0.02	0.07	0.04	-0.02	0.09	0.04	-0.02	0.09	0.03	-0.43	0.48	0.03	-0.43	0.48
Research Learning Communities	116	4969	0.03	-0.06	0.12	0.04	-0.08	0.16	0.04	-0.08	0.16	0.03	-0.02	0.07	0.03	-0.02	0.07
REACH	19	192	0.02	-0.46	0.49	0.04	-0.55	0.62	0.04	-0.55	0.62	0.03	-0.06	0.12	0.03	-0.06	0.12
Teacher Effec Enht Programme	45	10384	0.02	-0.06	0.09	0.02	-0.07	0.11	0.02	-0.07	0.11	0.02	-0.06	0.09	0.02	-0.06	0.09
IPEELL two year	77	2182	0.02	-0.12	0.15	0.02	-0.15	0.19	0.02	-0.15	0.19	0.02	-0.11	0.15	0.02	-0.11	0.15
Hampshire Hundreds	36	2828	0.01	-0.06	0.08	0.02	-0.16	0.19	0.02	-0.16	0.19	0.01	-0.06	0.08	0.01	-0.06	0.08
Vocab Enrichment Intervention Programme	12	570	0.01	-0.25	0.27	0.01	-0.36	0.38	0.01	-0.36	0.38	0.01	-0.22	0.24	0.01	-0.22	0.24
Success for All	50	1307	0.01	-0.22	0.24	0.02	-0.24	0.27	0.02	-0.24	0.27	0.01	-0.06	0.09	0.01	-0.06	0.09
Texting Parents	29	11414	0.01	-0.06	0.09	0.02	-0.08	0.12	0.02	-0.08	0.12	0.01	-0.17	0.19	0.01	-0.17	0.19
Flipped Learning	24	1133	0.01	-0.17	0.18	0.01	-0.22	0.24	0.01	-0.22	0.24	0.01	-0.04	0.05	0.01	-0.04	0.05
Teacher Observation	82	21002	0.01	-0.04	0.05	0.01	-0.05	0.07	0.01	-0.05	0.07	0.01	-0.17	0.20	0.01	-0.17	0.20
Online Reading Programme(A)	49	1845	0.01	-0.17	0.19	0.01	-0.21	0.24	0.01	-0.21	0.24	0.01	-0.03	0.05	0.01	-0.03	0.05
Emb Formative Assessment	140	22628	0.01	-0.03	0.05	0.01	-0.04	0.07	0.01	-0.04	0.07	0.01	-0.09	0.10	0.01	-0.09	0.10
Changing Mindsets	101	4523	0.01	-0.09	0.11	0.01	-0.12	0.14	0.01	-0.12	0.14	0.00	-0.08	0.09	0.00	-0.08	0.09
LIT Programme	34	4415	0.00	-0.08	0.09	0.01	-0.13	0.14	0.01	-0.13	0.14	0.00	-0.26	0.27	0.00	-0.26	0.27
Catch-up Literacy	15	555	0.00	-0.28	0.29	0.00	-0.40	0.41	0.00	-0.40	0.41	0.00	-0.28	0.28	0.00	-0.28	0.28
Team Alphie	6	72	0.00	-0.92	0.90	-0.02	-0.85	0.84	-0.02	-0.85	0.84	0.00	-0.21	0.21	0.00	-0.21	0.21
Changing Mindsets - Inset	24	914	0.00	-0.22	0.20	0.00	-0.34	0.34	0.00	-0.34	0.34	0.00	-0.15	0.15	0.00	-0.15	0.15
Shared Maths 1	39	950	0.00	-0.14	0.15	0.00	-0.21	0.22	0.00	-0.21	0.22	-0.01	-0.11	0.10	-0.01	-0.11	0.10
Grammar for Writing	50	2219	-0.01	-0.15	0.12	-0.02	-0.18	0.14	-0.02	-0.18	0.14	-0.01	-0.13	0.10	-0.01	-0.13	0.10
Effective Feedback	13	2794	-0.01	-0.12	0.10	-0.02	-0.24	0.21	-0.02	-0.24	0.21	-0.02	-0.15	0.12	-0.02	-0.15	0.12
Paired Reading	10	2580	-0.01	-0.12	0.10	-0.02	-0.21	0.17	-0.02	-0.21	0.17	-0.02	-0.22	0.18	-0.02	-0.22	0.18
Catch-up Numeracy	54	316	-0.02	-0.23	0.18	-0.05	-0.51	0.41	-0.05	-0.51	0.41	-0.02	-0.86	0.85	-0.02	-0.86	0.85
Dialogic Teaching	69	1217	-0.03	-0.20	0.14	-0.04	-0.26	0.18	-0.04	-0.26	0.18	-0.03	-0.20	0.14	-0.03	-0.20	0.14
Parenting Academy 1	16	1589	-0.03	-0.18	0.13	-0.04	-0.24	0.18	-0.04	-0.24	0.18	-0.03	-0.18	0.13	-0.03	-0.18	0.13
Quest	19	2090	-0.04	-0.17	0.09	-0.05	-0.24	0.13	-0.05	-0.24	0.13	-0.04	-0.17	0.09	-0.04	-0.17	0.09
ReflectEd	28	1353	-0.04	-0.19	0.12	-0.05	-0.24	0.15	-0.05	-0.24	0.15	-0.04	-0.19	0.12	-0.04	-0.19	0.12
Fam & Schools Tog (FAST)	116	4293	-0.04	-0.14	0.06	-0.05	-0.17	0.07	-0.05	-0.17	0.07	-0.04	-0.14	0.06	-0.04	-0.14	0.06
Act, Sing, Play 1	19	542	-0.04	-0.30	0.22	-0.07	-0.47	0.33	-0.07	-0.47	0.33	-0.04	-0.30	0.22	-0.04	-0.30	0.22

Parenting Academy 2	16	1697	-0.04	-0.18	0.10	-0.05	-0.25	0.14	-0.05	-0.25	0.14	-0.04	-0.18	0.10	-0.04	-0.18	0.10
Increasing Pupil Motivation	63	9272	-0.05	-0.11	0.00	-0.08	-0.16	0.00	-0.08	-0.16	0.00	-0.05	-0.38	0.27	-0.05	-0.38	0.27
Fresh Start	10	419	-0.05	-0.36	0.25	-0.07	-0.50	0.36	-0.07	-0.50	0.36	-0.05	-0.54	0.43	-0.05	-0.54	0.43
Chess in Schools	100	3694	-0.05	-0.18	0.08	-0.05	-0.18	0.08	-0.05	-0.18	0.08	-0.05	-0.11	0.00	-0.05	-0.11	0.00
Youth Social Action Trials(Y)	66	6619	-0.05	-0.16	0.05	-0.05	-0.15	0.05	-0.05	-0.15	0.05	-0.05	-0.35	0.26	-0.05	-0.35	0.26
Tutor Trust: Afford Primary Tuition	104	3844	-0.05	-0.15	0.04	-0.07	-0.20	0.06	-0.07	-0.20	0.06	-0.05	-0.18	0.08	-0.05	-0.18	0.08
Chatterbooks	12	465	-0.06	-0.39	0.28	-0.07	-0.48	0.34	-0.07	-0.48	0.34	-0.05	-0.16	0.05	-0.05	-0.16	0.05
Rapid Phonics	21	178	-0.06	-0.56	0.44	-0.06	-0.66	0.53	-0.06	-0.66	0.53	-0.05	-0.15	0.04	-0.05	-0.15	0.04
Switch-on Effectiveness T 1	119	5318	-0.06	-0.15	0.02	-0.09	-0.21	0.03	-0.09	-0.21	0.03	-0.06	-0.14	0.02	-0.06	-0.14	0.02
W & W Reading Programme (CC)	16	1223	-0.07	-0.23	0.10	-0.07	-0.25	0.10	-0.07	-0.25	0.10	-0.07	-0.23	0.10	-0.07	-0.23	0.10
Huntington Rise	39	13423	-0.07	-0.13	-0.01	-0.09	-0.17	-0.01	-0.09	-0.17	-0.01	-0.07	-0.13	-0.01	-0.07	-0.13	-0.01
Philosophy for Children	48	1470	-0.08	-0.28	0.12	-0.09	-0.34	0.15	-0.09	-0.34	0.15	-0.08	-0.28	0.13	-0.08	-0.28	0.13
Switch-on Effectiveness T 2	120	5479	-0.08	-0.16	0.00	-0.12	-0.25	0.00	-0.12	-0.25	0.00	-0.08	-0.16	0.00	-0.08	-0.16	0.00
Childrens University	67	3482	-0.09	-0.24	0.06	-0.09	-0.25	0.06	-0.09	-0.25	0.06	-0.09	-0.25	0.06	-0.09	-0.25	0.06
Talk of the Town	63	2701	-0.10	-0.22	0.02	-0.12	-0.28	0.04	-0.12	-0.28	0.04	-0.10	-0.22	0.03	-0.10	-0.22	0.03
Afford Online Maths Tuition	64	2927	-0.10	-0.23	0.02	-0.14	-0.30	0.03	-0.14	-0.30	0.03	-0.10	-0.23	0.02	-0.10	-0.23	0.02
Act, Sing, Play 2	19	545	-0.11	-0.37	0.16	-0.16	-0.56	0.23	-0.16	-0.56	0.23	-0.11	-0.38	0.15	-0.11	-0.38	0.15
Switch-on Reading	19	308	-0.13	-0.50	0.24	-0.18	-0.66	0.30	-0.18	-0.66	0.30	-0.13	-0.50	0.23	-0.13	-0.50	0.23
Changing Mindsets - Pupil	5	184	-0.18	-0.57	0.21	-0.29	-0.88	0.34	-0.29	-0.88	0.34	-0.19	-0.58	0.22	-0.19	-0.58	0.22
Butterfly Phonics	6	307	-0.21	-0.52	0.14	-0.29	-0.78	0.17	-0.29	-0.78	0.17	-0.20	-0.54	0.12	-0.20	-0.54	0.12
Discover Summer School	14	79	-0.23	-0.86	0.38	-0.21	-0.76	0.34	-0.21	-0.76	0.34	-0.23	-0.83	0.37	-0.23	-0.83	0.37
Improving Writing Quality	22	472	-0.23	-0.57	0.12	-0.27	-0.68	0.13	-0.27	-0.68	0.13	-0.23	-0.57	0.11	-0.23	-0.57	0.11
Thinking, Doing, Talking Sc	41	1353	-0.24	-0.44	-0.05	-0.31	-0.56	-0.06	-0.31	-0.56	-0.06	-0.24	-0.44	-0.04	-0.24	-0.44	-0.04
Units of Sound	33	423	-0.27	-0.53	0.00	-0.38	-0.76	0.00	-0.38	-0.76	0.00	-0.27	-0.53	0.00	-0.27	-0.53	0.00

Table A4. Comparison of pooled and individual trial attainment gaps between FSM and non-FSM pupils' maths outcome using IPD and two-stage fixed effect (FE) and random effect (RE) meta-analysis methods.

			Standardised Score			Raw Score						Standardised Score					
	Number of		IPD			Two stage (FE)			Two stage (RE)			Two stage (FE)			Two stage (RE)		
Trial	Schools	Pupils	ES	Low	Up	ES	Low	Up	ES	Low	Up	ES	Low	Up	ES	Low	Up
Overall gap	3178	306975	-0.01	-0.04	0.02	0.00	-0.02	0.02	-0.01	-0.04	0.03	0.00	-0.01	0.01	-0.01	-0.03	0.02
Let's Think Secondary Sc	17	1780	0.20	0.04	0.35	0.30	0.07	0.52	0.30	0.07	0.52	0.20	0.05	0.34	0.20	0.05	0.34
Act, Sing, Play 1	19	548	0.18	-0.06	0.44	0.26	-0.10	0.62	0.26	-0.10	0.62	0.18	-0.07	0.43	0.18	-0.07	0.43
Improv Num and Lit KS 2	35	1282	0.14	-0.08	0.36	0.20	-0.14	0.53	0.20	-0.14	0.53	0.14	-0.09	0.35	0.14	-0.09	0.35
Afford Online Maths Tuition	64	3041	0.10	-0.02	0.20	0.13	-0.03	0.29	0.13	-0.03	0.29	0.10	-0.02	0.21	0.10	-0.02	0.21
Afford Ind & Small Grp Tuition(M)	324	101331	0.08	-0.01	0.18	0.12	-0.02	0.26	0.12	-0.02	0.26	0.08	-0.01	0.17	0.08	-0.01	0.17
Changing Mindsets - Pupil	5	180	0.07	-0.26	0.40	0.13	-0.47	0.73	0.13	-0.47	0.73	0.07	-0.26	0.40	0.07	-0.26	0.40
Dialogic Teaching	69	1258	0.07	-0.08	0.23	0.10	-0.12	0.33	0.10	-0.12	0.33	0.07	-0.08	0.23	0.07	-0.08	0.23
Parenting Academy 2	16	1721	0.07	-0.07	0.22	0.10	-0.10	0.29	0.10	-0.10	0.29	0.07	-0.07	0.21	0.07	-0.07	0.21
Best Practice in Grp Students	76	2383	0.06	-0.05	0.17	0.10	-0.08	0.28	0.10	-0.08	0.28	0.06	-0.05	0.17	0.06	-0.05	0.17
IPEELL one year	83	2441	0.05	-0.07	0.17	0.07	-0.08	0.23	0.07	-0.08	0.23	0.05	-0.06	0.17	0.05	-0.06	0.17
Effective Feedback	13	2796	0.04	-0.06	0.14	0.09	-0.12	0.29	0.09	-0.12	0.29	0.04	-0.06	0.14	0.04	-0.06	0.14

Changing Mindsets - Inset	24	928	0.04	-0.17	0.26	0.07	-0.27	0.40	0.07	-0.27	0.40	0.04	-0.17	0.26	0.04	-0.17	0.26
Learner Response System	99	5964	0.03	-0.07	0.14	0.03	-0.07	0.13	0.03	-0.07	0.13	0.03	-0.07	0.13	0.03	-0.07	0.13
Shared Maths 2	79	2598	0.03	-0.08	0.13	0.05	-0.14	0.24	0.05	-0.14	0.24	0.03	-0.07	0.13	0.03	-0.07	0.13
Hampshire Hundreds	36	2795	0.02	-0.05	0.10	0.05	-0.13	0.23	0.05	-0.13	0.23	0.02	-0.05	0.09	0.02	-0.05	0.09
Lesson Study	181	24283	0.02	-0.02	0.06	0.03	-0.03	0.08	0.03	-0.03	0.08	0.02	-0.02	0.06	0.02	-0.02	0.06
Chess in Schools	100	3705	0.02	-0.12	0.14	0.02	-0.12	0.15	0.02	-0.12	0.15	0.02	-0.11	0.14	0.02	-0.11	0.14
Texting Parents	29	11589	0.02	-0.05	0.08	0.02	-0.08	0.13	0.02	-0.08	0.13	0.01	-0.05	0.08	0.01	-0.05	0.08
Teacher Observation	82	20829	0.02	-0.02	0.06	0.03	-0.02	0.09	0.03	-0.02	0.09	0.02	-0.02	0.06	0.02	-0.02	0.06
Changing Mindsets	101	4528	0.02	-0.06	0.11	0.03	-0.09	0.15	0.03	-0.09	0.15	0.02	-0.06	0.11	0.02	-0.06	0.11
Tutor Trust: Afford Primary Tuition	104	3863	0.01	-0.08	0.10	0.01	-0.11	0.14	0.01	-0.11	0.14	0.01	-0.08	0.10	0.01	-0.08	0.10
Fam & Schools Tog (FAST)	115	4308	0.01	-0.09	0.12	0.01	-0.11	0.13	0.01	-0.11	0.13	0.01	-0.09	0.12	0.01	-0.09	0.12
Math Mastery Secondary	44	5830	0.00	-0.08	0.09	0.01	-0.11	0.12	0.01	-0.11	0.12	0.00	-0.08	0.09	0.00	-0.08	0.09
Emb Formative Assessment	140	22935	0.00	-0.04	0.03	-0.01	-0.06	0.05	-0.01	-0.06	0.05	0.00	-0.04	0.03	0.00	-0.04	0.03
IPEELL two year	82	2539	0.00	-0.10	0.12	0.00	-0.15	0.16	0.00	-0.15	0.16	0.00	-0.11	0.11	0.00	-0.11	0.11
Afford Ind & Small Grp Tuition (P)	12	763	-0.01	-0.41	0.41	-0.02	-0.57	0.50	-0.02	-0.57	0.50	-0.02	-0.43	0.38	-0.02	-0.43	0.38
ScratchMaths	110	5818	-0.01	-0.10	0.07	-0.01	-0.13	0.10	-0.01	-0.13	0.10	-0.01	-0.09	0.07	-0.01	-0.09	0.07
Improv Num and Lit KS 1	37	1344	-0.02	-0.21	0.18	-0.02	-0.31	0.26	-0.02	-0.31	0.26	-0.02	-0.20	0.17	-0.02	-0.20	0.17
Increasing Pupil Motivation	63	9248	-0.03	-0.09	0.03	-0.04	-0.12	0.04	-0.04	-0.12	0.04	-0.03	-0.09	0.03	-0.03	-0.09	0.03
Maths Reasoning	160	6334	-0.03	-0.13	0.07	-0.04	-0.16	0.08	-0.04	-0.16	0.08	-0.03	-0.13	0.06	-0.03	-0.13	0.06
Flipped Learning	24	1129	-0.04	-0.20	0.12	-0.06	-0.29	0.17	-0.06	-0.29	0.17	-0.04	-0.20	0.12	-0.04	-0.20	0.12
Huntington Rise	39	13489	-0.04	-0.10	0.01	-0.06	-0.14	0.02	-0.06	-0.14	0.02	-0.04	-0.10	0.01	-0.04	-0.10	0.01
Teacher Effec Enht Programme	45	10320	-0.05	-0.12	0.03	-0.05	-0.14	0.04	-0.05	-0.14	0.04	-0.05	-0.12	0.03	-0.05	-0.12	0.03
Shared Maths 1	79	2710	-0.05	-0.17	0.07	-0.07	-0.26	0.12	-0.07	-0.26	0.12	-0.05	-0.17	0.08	-0.05	-0.17	0.08
Parenting Academy 1	16	1619	-0.05	-0.20	0.10	-0.07	-0.28	0.14	-0.07	-0.28	0.14	-0.05	-0.20	0.10	-0.05	-0.20	0.10
Best Practice in Grp Students(M)	9	353	-0.06	-0.43	0.30	-0.09	-0.62	0.45	-0.09	-0.62	0.45	-0.06	-0.41	0.30	-0.06	-0.41	0.30
Thinking, Doing, Talking Sc	41	1353	-0.08	-0.28	0.11	-0.11	-0.36	0.14	-0.11	-0.36	0.14	-0.08	-0.28	0.11	-0.08	-0.28	0.11
Act, Sing, Play 2	19	550	-0.08	-0.35	0.16	-0.12	-0.48	0.23	-0.12	-0.48	0.23	-0.09	-0.33	0.16	-0.09	-0.33	0.16
Childrens University	67	3491	-0.09	-0.25	0.07	-0.09	-0.25	0.06	-0.09	-0.25	0.06	-0.09	-0.25	0.06	-0.09	-0.25	0.06
Youth Social Action Trials(Y)	66	6619	-0.10	-0.20	0.01	-0.10	-0.20	0.01	-0.10	-0.20	0.01	-0.10	-0.20	0.01	-0.10	-0.20	0.01
ReflectEd	28	1507	-0.15	-0.28	-0.01	-0.20	-0.38	-0.02	-0.20	-0.38	-0.02	-0.15	-0.28	-0.01	-0.15	-0.28	-0.01
Powerful Learning Conversations	11	937	-0.16	-0.46	0.14	-0.22	-0.62	0.19	-0.22	-0.62	0.19	-0.16	-0.45	0.14	-0.16	-0.45	0.14
1stClass@Number	129	466	-0.16	-0.52	0.22	-0.16	-0.55	0.23	-0.16	-0.55	0.23	-0.15	-0.53	0.22	-0.15	-0.53	0.22
Catch-up Numeracy	54	316	-0.19	-0.50	0.11	-0.20	-0.53	0.13	-0.20	-0.53	0.13	-0.19	-0.50	0.12	-0.19	-0.50	0.12
Philosophy for Children	48	1470	-0.19	-0.38	0.00	-0.25	-0.50	-0.01	-0.25	-0.50	-0.01	-0.19	-0.38	0.00	-0.19	-0.38	0.00
Future Foundations	33	303	-0.20	-0.57	0.18	-0.23	-0.66	0.19	-0.23	-0.66	0.19	-0.20	-0.56	0.16	-0.20	-0.56	0.16
onebillion	112	1090	-0.26	-0.49	-0.03	-0.32	-0.60	-0.04	-0.32	-0.60	-0.04	-0.26	-0.50	-0.03	-0.26	-0.50	-0.03
Maths Counts	39	291	-0.43	-0.78	-0.06	-0.52	-0.95	-0.10	-0.52	-0.95	-0.10	-0.43	-0.78	-0.08	-0.43	-0.78	-0.08

This document is available for download at https://educationendowmentfoundation.org.uk

Education Endowment Foundation

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

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