

Longitudinal impact analysis of EEF Key Stage 3 and 4 interventions Evaluation Study Plan



Evaluator (institution): Behavioural Insights Team

Principal investigator(s): Dr. Patrick Taylor

PROJECT TITLE	Longitudinal impact analysis of EEF Key Stage 3 and 4 interventions
EVALUATOR (INSTITUTION)	Behavioural Insights Team
PRINCIPAL INVESTIGATOR(S)	Dr. Patrick Taylor
STUDY PLAN AUTHOR(S)	Dr. Laure Bokobza, Trisha Harjani, Dr. Patrick Taylor
STUDY DESIGN	Longitudinal Secondary Data Analysis
PUPIL AGE RANGE AND KEY STAGE	Aged 12 to 18 (KS3 to KS5)
NUMBER OF INTERVENTIONS INCLUDED	18
INTERVENTION SAMPLE SIZE INFORMATION (TOTAL)	Minimum: 557 (Catch Up Literacy) Maximum: 165,151 (Tutor Trust) Mean: 16,505 Median: 6,785
PRIMARY OUTCOMES MEASURE AND SOURCE	Primary outcome 1: Post-16 educational attainment. Measure: Proportion achieving a Level 2 qualification by the age of 19. Primary outcome 2: Post-16 educational progression. Measure: Proportion progressing to a Further Education (FE) destination by the age of 22. Sources: National Pupil Database, Higher Education Statistics Agency, Individual Learner Records.
SECONDARY OUTCOMES MEASURE AND SOURCE	Alternative measures of post-16 educational attainment: - Proportion achieving a Level 3 qualification - Proportion completing a university degree - Proportion passing Maths and English GCSEs by the age of 19 - Proportion enrolled in an HE institution Alternative measures of post-16 educational progression: - Progressing to academic Further Education (academic KS5) - Progressing to non-academic Further Education (non-academic KS5) - Not enrolled in any post-16 educational setting between the ages of 16 and 18 Sources: National Pupil Database, Higher Education Statistics Agency, Individual Learner Records.

Study Plan version history

VERSION	DATE	REASON FOR REVISION
1.3 [<i>latest</i>]	15/07/2025	The progression outcomes (primary outcome 2, secondary outcomes 2.2 and 2.3) cannot be constructed as detailed in v1.2. This is because we are unable to capture progression to a Further Education 'destination' with the available data. As a result, we have modified the definition of Primary Outcome 2, and subsequently of Secondary Outcome 2.2 and Secondary Outcome 2.3.
1.2	24/01/2025	We have removed Secondary Outcome 2.4 on NEET status after confirming that it is not possible to reliably infer it from DfE data.
1.1	14/11/2024	This update responded to peer review comments. The main changes are: • We updated the list of eligible trials (reducing the total number from 20 to 18) based on availability in EEF archive • We added an explanation for using clustered standard errors rather than multilevel models in our analytical specifications
1.0 [<i>original</i>]	19/09/2024	

Table of Contents

[Background and Study Rationale](#)

[Policy Context](#)

[Rationale](#)

[Evaluation design overview](#)

[Interventions](#)

[Systematic trial review](#)

[Theory of change review](#)

[Summary: Characteristics of interventions selected for longitudinal analysis](#)

[Impact Evaluation](#)

[Research questions](#)

[Design overview](#)

[Data sources](#)

[Outcome measures and other data](#)

[Primary outcome](#)

[Secondary outcomes](#)

[Participants](#)

[Sample description and size](#)

[Statistical Analysis](#)

[Intention-to-treat framework](#)

[Accounting for clustering](#)

[Primary outcome analysis](#)

[Robustness checks](#)

[Further analyses](#)

[Ethics](#)

[Data Protection](#)

[Lawful basis for processing data](#)

[Data security](#)

[Personnel](#)

[Risks](#)

[Timeline](#)

[References](#)

[Acknowledgments](#)

[APPENDIX A](#)

[APPENDIX B](#)

[APPENDIX C](#)

[Achieve Together](#)

[Challenge the Gap](#)

[EFA](#)

[Generation STEM Work Experience](#)

[Maths in Context](#)

[Increasing Pupil Motivation](#)
[Texting Parents](#)
[Texting Students and Study Supporters](#)
[The RISE Project](#)
[Tutor Trust](#)
[Best Practice in Setting](#)
[Catch up Literacy \(2014\)](#)
[British Science Association CREST Silver Award](#)
[Ark Mathematics Mastery](#)
[Peer Tutoring in Secondary Schools](#)
[Quest](#)
[Sci-napse](#)
[Youth United](#)

Background and Study Rationale

Policy Context

Improving social mobility and closing the attainment gap between pupils of more and less advantaged backgrounds has been a stated policy priority for the past few Governments in England, in part through investment and targeted intervention programmes. Despite this, the gap remains significant and has widened considerably across educational phases after the Covid-19 pandemic (Education Policy Institute, 2023). Before the pandemic, the gap tended to worsen throughout secondary school, to reach its widest when pupils reach the post-16 stage (Andrews et al. 2017, Smith and Smith, 2014).

Educational research focusing on post-16 students and sectors suffers from a limited evidence base, challenges related to data collection and access, and difficulties to track students across a complex sector with many different types of providers and qualifications (Machin and McNally, 2020). The Education Endowment Foundation (EEF) also notes that there are fewer high-quality interventions that are ready to be trialled in the field at the post-16 stage, which further contributes to a weaker evidence base than in earlier stages.

However, there is a wealth of education interventions targeting pupils in Key Stages 3 and 4 (KS3 and KS4) that are part of a much more established evidence base, including but not limited to interventions targeting GCSE attainment. Crucially, these interventions take place before the juncture between pre- and post-16. GCSEs are a critical juncture: Failure to achieve a Level 4 in Maths and/or English affects students' life trajectories, including likelihood of enrolling into Higher Education (HE) and lifetime earnings (Jerrim, 2022; Lupton et al., 2021, Social Mobility Commission, 2023) These 'upstream' interventions therefore have the potential to have positive long-term effects on a range of post-16 outcomes, including educational attainment and progression. They also have the potential to improve longer-term labour market outcomes for the students who benefited from them. After scoping out the possibility to include those outcomes in the analysis, the availability of linked data made it beyond the scope of this specific project. We therefore focus on post-16 educational attainment and progression.

Rationale

This EEF Data Archive Secondary Research is a project developed in collaboration by the Behavioural Insights Team (BIT) and the EEF, as part of an effort to leverage the EEF archive data to investigate the long-term effects of EEF interventions on pupils' outcomes.

This longitudinal analysis focuses specifically on the post-16 outcomes of selected trials commissioned by the EEF, that were tested and delivered in KS3 and KS4. BIT has been commissioned by the EEF to explore any longer-term impact of these interventions on post-16 educational attainment and progression. To do so, we will link EEF trial data to administrative datasets held by the Department of Education, for the sample of pupils covered by the EEF trial data. We will then re-analyse the original data to examine any long-term impact on educational attainment and progression from KS4 to KS5 (Further Education) and to university (Higher Education).

This project is part of a larger body of research aiming to make the EEF data archive accessible to research teams and to encourage its use to support future research for public good. The EEF archive holds data from approximately 130 trials involving more than 1.8 million children. It includes data on primary outcomes (educational attainment, often captured by standardised test data from the NPD), secondary outcomes, treatment allocation and any other variables used in impact evaluation analysis. The EEF seeks to facilitate access to researchers and encourage the use of the EEF Archive as a time- and cost-efficient method to estimate the long-term impact of the interventions by linking it to longitudinal administrative data. Other objectives include improving the understanding of impact across evaluations and interventions, and improving the methodological approaches used to estimate this impact.

Evaluation design overview

The two overarching research questions we seek to address are:

1. For each of the implemented programmes, what is the impact of educational programmes delivered in KS3 and KS4 on post-16 and higher education (HE) **attainment outcomes** for all pupils?
2. For each of the implemented programmes, what is the impact of educational programmes delivered in KS3 and KS4 on **educational progression** for all pupils?

We start by selecting, among a longlist of EEF-funded programmes delivered in KS3 and KS4, those that are suitable for longitudinal analysis based on their relevance to the overarching research questions, data availability, and trial design. We detail this process in the section on Interventions.

Analyses will then seek to identify any longer-term impact of the selected interventions, with a focus on the following research questions:

1. Was there a difference in the percentage of pupils who **achieved at least a Level 2 qualification** by the age of 19, between pupils allocated to receive an intervention and those allocated to the control group?
 - 1.1. Was there a difference in the percentage of pupils who **achieved a Level 3 qualification** between pupils allocated to receive an intervention and those allocated to the control group?
 - 1.2. Was there a difference in the percentage of pupils who **completed a university degree** between pupils allocated to receive an intervention and those allocated to the control group?
 - 1.3. Was there a difference in the percentage of pupils **passing GCSEs in Maths and English** by the age of 19 between pupils allocated to receive an intervention and those allocated to the control group?
2. Was there a difference in the percentage of pupils who **progressed to a Further Education (FE) destination by the age of 22** between pupils allocated to receive an intervention and those allocated to the control group?
 - 2.1. Was there a difference in the percentage of pupils **enrolled in a HE institution** between pupils allocated to receive an intervention and those allocated to the control group?

- 2.2. Was there a difference in the percentage of pupils **progressing to academic Further Education (academic KS5)** between pupils allocated to receive an intervention and those allocated to the control group?
- 2.3. Was there a difference in the percentage of pupils **progressing to non-academic Further Education (non-academic KS5)** between pupils allocated to receive an intervention and those allocated to the control group?
- 2.4. Was there a difference in the percentage of pupils **not enrolled in any post-16 educational setting between the ages of 16 and 18** between pupils allocated to receive an intervention and those allocated to the control group?

We will also seek to provide evidence of any differential effect for pupils who are classed as Free School Meals (FSM) for the main research questions (1.1 and 2.1).

Once we have successfully linked the EEF archive data with external sources of outcome data from the Department of Education, we will conduct a 'follow-up' analysis on the effectiveness of the intervention on selected long-term outcomes. We will estimate the effect of being allocated to receive an intervention on the outcomes of Research Questions 1.1 to 2.3 respectively. We will conduct this analysis separately for each trial, due to the varied nature of the interventions. We expand on the considerations behind this decision in the Impact Evaluation section.

Interventions

This longitudinal impact evaluation focuses on a total of 18 interventions, each of which have been evaluated in EEF-funded trials.¹ We have selected interventions from a longlist of EEF programmes delivered in KS3 and KS4, based on their relevance to the research questions, data availability, and trial design. In the rest of this section, we outline the process of trial selection and present the resulting group of interventions included in this study.

Systematic trial review

The EEF had pre-identified a longlist of 39 interventions delivered in KS3 and KS4. We started by systematically reviewing this longlist with the aim of selecting trials that were fit for the longitudinal analysis.

We systematically selected trials based on their relevance to the project aims, data availability, and trial design. All trials were originally designed in line with EEF's core mission of narrowing the attainment gap between advantaged and less advantaged pupils. However, the interventions tested are highly variable in terms of subject area and delivery. The interventions include, but are not limited to, CPD programmes, work experience for pupils, catch-up programmes, and text messages to pupils re-sitting GCSEs. We include a fuller explanation of the inclusion and exclusion criteria below and provide an overview of key design features by trial in Table 4.

We started by assessing whether the longlisted interventions were suitable for a longitudinal analysis of their impact on post-16 outcomes. All trials either follow a randomised control trial

¹ By 'trial' we mean any empirical evaluation design, not only Randomised Controlled Trials (RCTs).

(RCT) design, or a quasi-experimental design (QED) where a control group is sampled through non-random processes.² We then applied the following inclusion criteria:³

1. The data from the trial must be available for access in the EEF Archive.
2. The trial should not have a waitlist control design.
3. The trial evaluation should not find evidence of spillover effects.
4. Pupils in the trial should be at least 16 years old by 2021.

Given the outcomes in the longitudinal evaluation are measured with a delay, we set criteria 2 and 3 to ensure that the control group has not received the treatment at the time that the outcome data is captured. However, we can only measure whether pupils received the treatment during the trial. A limitation of our methodology is that the control group may receive the intervention post-trial in another manner e.g., their school implements the intervention as a whole-school practice.

We set criteria 4 based on the predicted availability of DfE data. Specifically, as we will apply to access DfE data for the most up to date available academic year (2022-23), pupils aged 16 in 2021-22 will be 18 years old in 2022-23 allowing for some analysis of their post-16 outcomes. For all 18 included pupils across all trials are expected to be at least 19 years old. Table A1 in Appendix A presents the breakdown of pupil ages across the trial years and up to 2022-23.

This initial screening resulted in a shortlist of 18 eligible trials for analysis, presented in Table 1 below 21 out of 39 trials are not feasible to include for the following reasons:⁴

- 4 trials are not in the EEF archive
- 15 followed waitlist control group designs, where the control schools or pupils were meant to receive the intervention or could request access to it after the trial ended .
- 2 had cohorts that are too young to have post-16 outcomes.
- 1 trial had poor data collection which made it not possible to determine if control schools conducted business as usual or used the intervention (with very low robustness of the evaluation overall).⁵

² 3 were QEDs and 36 were RCTs.

³ See Appendix B for the spreadsheet which displays the granular details of trial eligibility and reasons for screening trials out.

⁴ A trial may have been excluded for multiple reasons and as such, these bullet points will add up to higher than the total number of 19 excluded trials.

⁵ This data for this trial is also unavailable in the EEF Archive.

Table 1. Trial Selection Process.⁶

No.	Programme name	Data in archive	Level of Randomisation	Waistlist Design?	Evidence of spillovers	First Year of intervention	Last year of intervention	Age at intervention	Feasible
1	Accelerated Reader (2015)	yes	Pupil	Yes	no	2013-14	2013-14	11-12	no
2	Achieve together	yes	-	no	no	2014-15	2015-16	14 - 16	yes
3	Ark Mathematics Mastery	yes	School	no	no	2013-14	2013-14	11-12	yes
4	Best Practice in Setting	yes	School	no	no	2015-16	2016-17	11-12	yes
5	British Science Association CREST Silver Award	yes	School	no	yes	2017-18	2017-18	13-14	yes
6	Butterfly Phonics	yes	Pupil	yes	no	2013-14	2013-14	11-12	no
7	Catch Up Literacy (2014)	yes	Pupil	no	no	2013-14	2014-15	10-11	yes
8	Challenge the gap	yes	-	No	no	2013-14	2014-15	12-16	yes
9	Chatterbooks	yes	School	yes	no	2012-13	2012-13	11-12	no
10	EFA	yes	School	no*	no	2015-16	2016-17	14-15	yes
11	Fit to Study	yes	School, class	no	no	2017-18	2017-18	12-13	no
12	Fresh Start	yes		yes	no	2013-14	2013-14	12-13	no
13	Generation STEM Work Experience	yes	School	No	no	2017-18	2017-18	14-15	yes

⁶ Colour coding. Green; before the trial began. Blue; pupils under 19, Red: pupils aged 19 and above

No.	Programme name	Data in archive	Level of Randomisation	Waistlist Design?	Evidence of spillovers	First Year of intervention	Last year of intervention	Age at intervention	Feasible
14	Graduate Coaching Programme	yes	Pupil	yes	yes	2013-14	2013-14	11-12	no
15	Healthy minds	no	School	yes	no	2013-14	2017-18	11 - 15	no
16	ICCAMS	no	School	no	no	2015-16	2017-18	11-13	no
17	INCLUSIVE (Learning Together)	no	School	no	yes	2013-14	2015-16	11-16	no
18	Increasing Pupil Motivation	yes	School	no	no	2012-13	2012-13	14-16	yes
19	Let's Think Secondary Science	yes	School	yes	no	2013-15	2013-15	11-12	no
20	LIT Programme	yes	School	yes	no	2012-13	2012-13	11-12	no
21	Teacher Observation	yes	School	no	yes	2014-15	2015-16	14-16	no
22	Maths in Context	yes	School	no	no	2017-18	2017-18	14-15	yes
23	Project-Based Learning	yes	School	yes	yes	2013-14	2016-17	11-12	no
24	Peer Tutoring in Secondary Schools (Paired Reading)	yes	Class	no	no	2013-14	2013-14	11-12, 13-14	yes
25	REACH	yes	Pupil	yes	yes	2013-14	2013-14	11-13	no
26	Realistic Maths Education	yes	School	no	yes	2018-19	2020-21	11-13	no
27	Quest	yes	School	no	no	2013-14	2013-14	11-12	yes
28	Switch-on Reading	yes	Pupil	yes	no	2012-2013	2012-2014	11-12	no

No.	Programme name	Data in archive	Level of Randomisation	Waistlist Design?	Evidence of spillovers	First Year of intervention	Last year of intervention	Age at intervention	Feasible
29	Talk for Literacy	yes	Pupil	yes	yes	2013-14	2013-15	11-12	no
30	Teacher Effectiveness Enhancement Programme	yes	School	Yes	no	2012-13	2013-14	11-12	no
31	Sci-napse	yes	Class	no	yes	2016-17	2016-17	12-13	yes
32	Texting Parents	yes	Key Stage	No	no	2014-15	2014-15	11-16	yes
33	Texting Students and Study Supporters	yes	Pupil	no	yes	2017-18	2017-18	16-18	yes
34	The RISE Project	yes	School	no	no	2015-16	2016-17	14-16	yes
35	Tutor Trust	yes	-	no	no	2011-12	2013-14	14-16	yes
36	Units of Sound	yes	Pupil	yes	no	2013-14	2013-15	11-12	no
37	Vocabulary Enrichment Intervention	yes	Pupil	yes	yes	2013-14	2013-15	11-12	no
38	Writing about Values	no	Pupil	no	no	2016-17	2017-18	14-16	no
39	Youth United	yes	School	no	no	2014-15	2014-16	13-14	yes

Theory of change review

We then checked that for each of these trials, there is a reasonable causal pathway between the intervention and the long-term outcomes of interest. This review was done prior to the drafting of this Study Plan and using our approximate, working definitions of the primary outcomes which have been slightly modified since. More details on this can be found in Appendix C.⁷

We systematically reviewed the original trial results with two aims: (1) to hypothesise any potential long-term effects (on post-16 outcomes) of the intervention, and (2) to evaluate the quality of implementation. To do so, we drew on the final reports from the original trial, focusing on the Theory of Change (ToC), impact evaluation, and implementation and process evaluation results.⁸

We analysed each trial individually and considered that all plausible causal pathways between the intervention and the post-16 outcomes could be mediated by the intervention's original primary and secondary outcomes. The original outcomes for each trial are listed in Appendix C for each trial. All trials were then reviewed following the same process and considering the same information.

To maintain consistency in how we generate each hypothesis, we weighed available evidence from each trial using some guiding principles:

1. *A non-significant result from the impact evaluation was not considered sufficient to generate a hypothesis about longer-term effects, but the direction and magnitude of the effect were taken into consideration.*
 - For example, if the observed impact was sizeable (e.g. Hedges' g greater than 0.1), a non-significant result was considered sufficient evidence for a hypothesis. Similarly, if the direction of a non-significant result was supported by observations in the process evaluation, this was taken to be a strong indication of a possible long-term effect and the inclusion of a hypothesis.
2. *A significant result from the impact evaluation was considered to be strong evidence to form a hypothesis about longer-term effects unless there were several contradictory observations in the process evaluation.*
 - Overall, while a significant result was generally given more weight than qualitative observations, it was not used to exclude any evidence provided in the process evaluation.
3. *In some cases, coming up with a definitive conclusion about the direction of the supposed effect is challenging* (e.g. when we have identified potentially contradictory mechanisms or there is insufficient clarity of evidence). This is reflected in mixed hypotheses and any missing hypotheses.

Appendix C presents the full review of the ToC for each trial. Table 2 below summarises the direction of the hypothesised mechanism for each trial and main research question via colour

⁷ Note that at the time this review was conducted, we were still hoping to be able to include labour market outcomes in the longitudinal analysis.

⁸ We do not explicitly test these hypotheses, instead we rely on available evidence from the original trials to specify plausible mechanisms between the intervention and longitudinal outcomes of interest.

coding. We will use this to inform our interpretation of results and provide appropriate caveats.

Table 2: Overview of hypotheses based on Theory of Change

Project Name	Original Study Outcomes*	Research Question	
		Post-16 Attainment Outcomes	Post-16 Educational Progression
Achieve together			
Ark Mathematics Mastery			
Best Practice in Setting			
CREST			
Catch Up Literacy (2014)			
Challenge the Gap		*Mixed for FSM	*Mixed for FSM
EFA			
Generation STEM Work Experience			
Increasing Pupil Motivation			
Maths in Context			
Peer Tutoring			
Quest			
Sci-napse			
Texting Parents			
Texting Students and Study Supporters			
The RISE Project			
Tutor Trust			

		Research Question	
Project Name	Original Study Outcomes*	Post-16 Attainment Outcomes	Post-16 Educational Progression
Youth United			

*Note that the direction of the effect does not imply its statistical significance

Colour codes:

Negative	Positive
Mixed Directions / Insufficient Evidence	No plausible relationship identified (not applicable for 1st column, which indicates results)

Summary: Characteristics of interventions selected for longitudinal analysis

Although the 20 selected interventions were all run on pupils aged between 12 to 18; from KS3 to KS5, they are highly variable in terms of their timing and type of intervention. The earliest trial started in the academic year of 2011/12 and the latest trial started in 2017/18. Table A1 in Appendix A shows an overview of pupils' age by trial in the academic year of 2022/23, the latest data point at which the outcome data were collected.

The interventions range in subject area focus with some trials focusing on a particular subject (e.g., [Quest](#) on literacy, and [Ark Mathematics Mastery](#) on mathematics) while others targeted teachers (e.g., [RISE](#)), or enabling pupils with opportunities for early work experience (e.g., [Generation STEM](#)). Overall, the trials can be broadly broken down into:

Academic (10 trials)

- 4 focused on Literacy:
 - [Quest](#); [Catch up Literacy](#); [Peer Tutoring](#)
- 2 focused on Mathematics:
 - [Maths in Context](#); [Ark Mathematics Mastery](#)
- 2 focused on Science:
 - [Sci-napse](#); [British Science CREST Silver Award](#)
- 2 focused on academic support for pupils across multiple subjects:
 - [Texting Students and Study Supporters](#); [Tutor Trust](#)

Non Academic (4 trials)

- 2 focused on non-academic outcomes e.g. motivation, values:
 - [Increasing Pupil Motivation](#)
- 2 focused on targeting pupils' extra-curriculars (e.g. work experience / youth groups):
 - [Youth United](#); [Generation STEM Work Experience](#)

Targeting Parents & Teachers (6 trials)

- 5 focused on Teacher Development / Teacher Training:
 - [Achieve Together](#); [Challenge the Gap](#); [Embedding Formative Assessment \(EFA\)](#); [The RISE project](#); [Best Practice in Setting](#)

- 1 focused on parental support:
 - [Texting Parents](#)

Of all 18 trials, 14 were run as efficacy trials and 4 were run as effectiveness trials.⁹

⁹ See <https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/process-and-people/pipeline-of-eef-trials>

Table 3: Breakdown of Key Details on the original Trial Design

Trial	N			80% Power ¹⁰	Groups	Stage	Age ¹¹	Design	Years
	Treatment	Control	Total ¹²						
<i>Achieve together</i>	2,353	6,045	8,398	Yes	2	KS4	24	QED	2014-16
<i>Ark Mathematics Mastery</i>	3,251	2,687	5,938	Yes	2	KS3	20	RCT	2013-14
<i>Best Practice in Setting</i>	1,351	1,971	3,322	Yes	3	KS3	19	RCT	2015-17
<i>British Science Association</i>	1,067	1,244	2,311	Yes	2	KS3	19	RCT	2017-18
<i>Catch Up Literacy</i>	286	271	557	No	2	KS2, KS3	19	RCT	2013-15
<i>Challenge the Gap</i> ¹³	7,466	7,728	15,194	No	2	KS3, KS4	25	QED	2013-15
<i>EFA</i>	12,358	13,035	25,393	Yes	2	KS4	22	RCT	2015-17
<i>Generation STEM</i>	781	743	1524	Yes	2	KS4	20	RCT	2017-18
<i>Increasing Pupil Motivation</i>	5,096	5,553	10,649	No	3	KS4	26	RCT	2012-13
<i>Maths in Context</i>	4,898	5,017	9,915	Yes	2	KS4	20	RCT	2017-18
<i>Peer Tutoring</i>	1,239	1,336	2,575	Yes	2	KS3	23	RCT	2013-14
<i>Quest</i>	931	1,152	2,083	No	2	KS3	21	RCT	2013-14
<i>Scinapse</i>	2,589	1,275	3,864	No	3	KS3	19	RCT	2016-17
<i>Texting Parents</i>	7,436	7,919	15,355	Yes	2	KS3, KS4	24	RCT	2014-15
<i>Texting Students</i>	2,808	938	3,746	Yes	4	KS5	23	RCT	2017-18
<i>The RISE Project</i> ¹⁴	6,585	6,904	13,489	No	2	KS4	23	RCT	2015-17
<i>Tutor Trust</i> ¹⁵	1,190	163,961	165,151	Yes	2	KS4	26	QED	2011-14
<i>Youth United</i>	3,946	3,686	7,632	Yes	2	KS3	22	RCT	2014-16
Totals	65,631	231,465	297,096	14					

¹⁰ This column indicates whether the trial was powered at 80% to detect an effect at the time of analysis.

¹¹ Here, we note the pupil's age at the time that the outcome data were collected i.e., in 2022/23.

¹² Sample numbers were approximated based on the *n* at analysis stated in the Evaluation Reports. Numbers were taken from the Participant Flows where available and otherwise, from analysis tables.

¹³ Secondary School Sample only.

¹⁴ Note that this trial had 2 cohorts; Cohort A was allocated the intervention for 1 year and Cohort B for 2 years.

¹⁵ Note that the Maths & English tutoring in this trial was run with separate control groups

Impact Evaluation

Design overview

Table 4: Design

Design			Longitudinal Secondary Data Analysis
Unit of analysis (school, pupils)			Pupil-level
Number of interventions to be included in analysis			18
Research Question 1 Attainment	Primary outcome	variable	Post-16 attainment (1)
		measure (instrument, scale, source)	Percentage of pupils with a Level 2 qualification by the age of 19 (1)
	Secondary outcomes	variables	Post-16 and Higher Education (HE) attainment
		measures (instrument, scale, source)	Percentage of pupils <i>achieving</i> a Level 3 qualification by the age of 19 (1.1)
			Percentage of pupils completing a university degree (1.2)
			Percentage of pupils passing GCSEs in English and Maths by the age of 19 (1.3)
Research Question 2 Progression	Primary outcome	variable(s)	Post-16 educational progression (2)
		measure(s) (instrument, scale, source)	Percentage of pupils progressing to any Further Education (FE) destination by the age of 22 (2) ¹⁶
	Secondary outcomes	variable(s)	Post-16 and Higher Educational progression
		measure(s) (instrument, scale, source)	Percentage of pupils enrolled in a HE institution (2.1) Percentage of pupils progressing to academic Further Education (academic KS5) (2.2)

¹⁶ Where progressing to Further Education is defined as pursuing a Level 2 qualification or higher at an institution funded by the Education and Skills Funding Agency (which may therefore submit data to the Individualised Learner Record).

		Percentage pupils progressing to non-academic Further Education (non-academic KS5) (2.3)
Baseline for primary outcomes	variable	N/A
	measure (instrument, scale, source)	Original baseline score (trial-specific)
Baseline for secondary outcomes	variable	N/A
	measure (instrument, scale, source)	Original baseline score (trial-specific)

Data sources

This impact evaluation will involve combining two sources of data to conduct a secondary longitudinal re-analysis of the outcomes of selected EEF trials. We will source the outcome data from individual-level administrative datasets held and managed by the Department of Education (DfE) including the NPD, HESA, and ILR. We will therefore request access to:

1. 18 completed EEF trials from the EEF Archive, managed by the Fisher Family Trust (FFT). This dataset contains all variables that were collected at the time that each trial was run and should include all the variables that were included in the original analysis at the time of evaluation. While the longlist of variables will vary by trial, we expect the data to contain common variables across all trials such as randomisation status, baseline, and endline outcome measurements.¹⁷
2. Three datasets held and managed by the Department of Education (DfE), including the NPD, ILR and HESA. We will use this data for the outcomes and covariates in the analysis.

EEF Archive Trial Data

The EEF archive is a collection of all EEF-funded impact evaluations for which the data have been archived. The archive is maintained by FFT and accessed in the Office for National Statistics' Secure Research Service (ONS SRS). In May 2023, the archive held data from approximately 130 interventions, involving more than 1.8 million children.

One key goal of EEF in archiving this data is to encourage further research and analysis using data from previously conducted trials, including longitudinal analysis. As of June 2019, of the 110 published EEF reports, 63 were potentially eligible for reanalysis excluding pilot studies and waitlist-designs (EEF, 2019). Across the trials selected for analysis in this research, we expect this data to contain, among other variables, pupils' pupil matching reference (PMR) number, treatment assignment variable, cluster ID (if relevant), outcome data, and covariate data. The PMR allows linking the archive with the DfE datasets.

¹⁷ The metadata for all available and archived EEF trials will be available [here](#)

National Pupil Database (NPD)

The NPD is a longitudinal database linking pupil/student characteristics to school and college learning aims and attainment information for children in schools in England. The NPD spans over 40 individual datasets and is populated by the DfE using data from schools and local authorities via statutory data collections such as the school census. We will apply to access the following NPD datasets: (1) school census, (2) alternative provision census, (3) Key Stage 4, (4) Key Stage 5, (5) National Client Caseload, (6) Post-16 Learning Aims, (7) Absences, and (8) the Young Persons Matched Administrative Dataset¹⁸.

This data will cover attainment outcomes (RQ1 and RQ 1.1 to 1.3) as well as progression outcomes (RQ2 and RQ 2.1 to 2.3). For example, the NPD Key Stage 4 dataset will contain variables to indicate whether a pupil (re)sat their GCSE exams (RQ 1.3) while the Key Stage 5 dataset will contain variables to indicate whether a pupil progressed to a Further Education (FE) destination (RQ2) as well as the specific type of destination (RQ 2.3).

Individualised Learner Record (ILR)

The ILR is a dataset managed by the Education and Skills Funding Agency (ESFA), which sits within the DfE. The ILR dataset aims to capture all the learning taking place in the Further Education (FE) and Skills sector which includes: Further Education Colleges, former External Institutions, Sixth Form Colleges, Training Organisations, Local Authorities, Academies, and Voluntary and Community Organisations. In addition to the data from the listed sources, the ILR may also contain data from other educational providers whom the EFSA funds including some Higher Educational Institutions or training organisations.¹⁹ Overall, these sources encompass the diverse range of educational, vocational, and specialised post-16 institutions available to pupils in England after the age of 16.

We will request access to the ILR Learner dataset and ILR Aims datasets which cover data on pupil enrolment and attainment in post-16 institutions. The data will then be used to calculate the percentage of pupils who have achieved a level 2 qualification (RQ1), level 3 qualification (RQ1.1), and the percentage of pupils progressing to other Level 3 qualifications such as vocational or specialised education (RQ2.3) for each original treatment and control groups. More details on how we construct these variables can be found in the [Outcome measures and other data](#) section below.

¹⁸ Fuller details on each dataset are available on the Department of Education's "Find and Explore" tool for the National Pupil Database [linked here](#).

¹⁹ For fuller details on where the ILR data is collected from, [visit the linked GOVUK page](#)

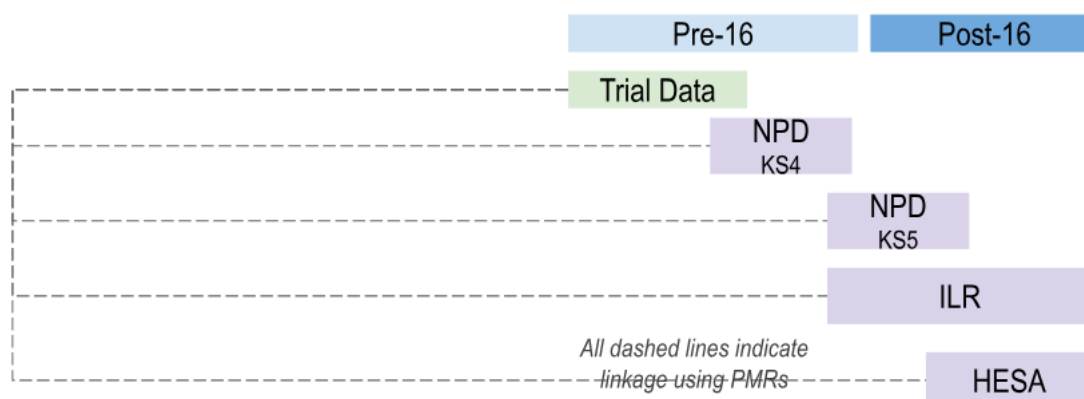
Higher Education and Statistics Agency (HESA)

The Higher Education and Statistics Agency collects data from Higher Education providers (HEPs) across the United Kingdom and English higher education alternative providers²⁰. The dataset covers variables on pupil characteristics, qualifications upon enrolment to higher education (HE), and progression through HE amongst others. We will apply for access to variables indicating whether a pupil is enrolled in any HE institution (RQ 2.1) and whether they have completed a university degree (RQ 1.2).

Linkage

We will access, link, and analyse all the data in a secure virtual environment created by the ONS SRS. In order to identify the relevant pupils required, we will first access the EEF trial data and extract all pupil's PMRs (Pupil Matching Reference) number. The PMRs will then be sent to the DfE who will use this to identify the relevant pupils' data. As the HESA and ILR datasets can be matched to the NPD, we will first link the three outcome datasets (NPD, HESA, and ILR) and then match the outcome data to the EEF trial data using the PMR. Figure 1 represents the datasets and their linkage.

Figure 1: Conceptual Illustration of the data sources and linkage process



Outcome measures and other data

Primary outcome

This secondary analysis will have dual primary outcomes, relating to the two primary research questions:

1. Was there a difference in the **percentage of pupils who achieved a Level 2 qualification by the age of 19**, between pupils allocated to receive an intervention and those allocated to the control group?

²⁰ Higher Education Alternative providers are defined as those “higher education providers who do not receive recurrent funding from the Office for Students (previously HEFCE) or another public body and who are not further education colleges” (HEPI, 2023)

2. Was there a difference in the **percentage of pupils who progressed to a Further Education (FE) destination by the age of 22** between pupils allocated to receive an intervention and those allocated to the control group?

We will also explore any differential effect for pupils who are classed as Free School Meals (FSM) for each of the primary research questions above.

We opt for dual primary outcomes because the focus of each outcome (attainment and progression) maps onto separate research objectives and there is no clear hierarchy between the two in terms of hypothesised causal mechanisms.²¹ Having dual primary outcomes allows us to capture the complexity and diversity of post-16 pathways for individual learners. This has a cost in terms of statistical power for the primary analysis as we adjust for multiple comparisons.

While attainment at post-16 may influence progression to KS5, pupils are not legally required to continue pursuing Level 2 qualifications if they do not achieve this by the age of 16. After the age of 16, young people may either continue education, start an apprenticeship, or split their time between working/ volunteering and being in part-time education or training. (GOVUK, 2024). In England, at the age of 17, around 45% of students begin their A-levels (a type of Level 3 qualification) with the remainder of students pursuing alternative specialised or vocational education qualifications²² or employment (Hupkau et al., 2017).

RQ2 aims to capture all possible Further Education destinations, which may or may not require a Level 2 qualification. While we expect some overlap between the two measures, they track distinctive pathways that a young person may follow. For example, if a young person does not pass their GCSEs at 16 and re-sits at age 17, then pursues their A-levels at 18, they will show up as having achieved Level 2 qualification before the age of 19 (RQ1) and progressing to a Further Education (FE) destination (RQ2). However, if a young person continues to re-sit the exams but only achieves a Level 2 by the age of 19, and chooses not to progress to a Further Education (FE) destination, this will be captured as a “yes” under RQ1 but not under RQ2. This example is a simple illustration and in reality, there are many different FE pathways that a young person may follow in England, which motivates RQ2 (Buttar et al., 2023).

RQ1 - Level 2 qualification by the age of 19

Sample: all 18 interventions.

This measure aims to capture the proportion of pupils that have achieved the lowest bar of attainment that most pupils reach by age 19 or earlier. In 2022/23 in England, 85.5% of pupils achieved a Level 2 (or Level 3) by the age of 19 (Department of Education, 2024b). Our measure allows pupils to re-sit their GCSE (or equivalent assessment) for up to two more academic years.

Based on guidance we received from the DfE, we expect that there is a variable in the ILR Learner dataset that captures whether a pupil has achieved a Level 2 qualification. This is

²¹ We note that this is a departure from EEF statistical guidance and results in a loss of statistical power.

²² See all qualifications split by Level [here](#)

the “ILL_L_FULLEVEL2_ACHIEVE” in the ILR *Learner* dataset. This is expected to be available from the academic year 2007/08 until 2022/23 and hence, should cover all pupils across all 20 EEF trials. See Table 3 below for the variables that we plan to use to calculate this outcome measure if the ILR data is more than 5% missing. See also Table A1 for a breakdown of trial years and pupil’s estimated age by 2022/23.

The stated ILR variables will be coded as 0 if the pupil does not have a Level 2 qualification and 1 if the pupil has achieved a Level 2 qualification. We use the date of birth variable to calculate a pupil’s age. As such, we interpret “by the age of 19” as the academic year in which a pupil turns 19.

RQ2 - Progression to a Further Education (FE) destination by the age of 22

Sample: 10 interventions for which the cohort is at least 22 in 2022/23

This measure aims to capture the lowest bar of academic, technical, or vocational progression available to pupils in England. The pursuit of FE in England is highly variable by age; someone may pursue a FE qualification at the age of 18 or at the age of 30 to change careers (Buttar et al., 2023). However, the FE system is fairly fragmented which implies that there is a lack of reliable data tracking the age breakdown of those enrolled across any FE destination (Hupkau et al., 2017).

For our study, a limitation of this measure arises from variability in trial implementation timing. For example, a pupil aged 16 in 2012-13 participating in the *Increasing Pupil Motivation* versus a pupil aged 16 in 2016-17 participating in the *Writing about Values* trial will be 26 and 22 in 2022-23, respectively. Hence, the pupil in the former trial has four more years of opportunity to progress to a FE destination compared to the latter. As such, we set an age cut-off for this outcome by making a trade-off between providing pupils sufficient years post-16 to enter FE, and the data cut-off. Specifically, the data cut-off refers to the fact that data are only available up to the academic year of 2022/23 where the estimated pupil age ranges from 19 to 26. See Table A1 for the breakdown of age by year and trial.

We select 22 as the age cutoff as pupils in 11 out of 18 trials are estimated to be at least 22 years old by the academic year of 2022/23 (see Table A1 in Appendix A). We expect that most people who intend to go to a Further Education (FE) destination will have done so already by 22. If a pupil passes their GCSEs at 16, this gives them 6 years to enter FE. We also consider the finding that approximately 80% those pursuing a Level 3 qualification at the age of 17 have achieved this by the age of 20 and only 8.4% of those pursuing a Level 2 qualification at the age of 17 have achieved a Level 3 by the age of 20 (Hupkau et al., 2017).

To calculate whether an individual has progressed to a Further Education (FE) destination, we will use data from the ILR Learner and Aims data tables.²³ We have identified the following relevant variables which are available across all years relevant to the trials (2012/13 up to 2022/23):

From the ILR:

²³ Given the complexity, we will approach the DfE for confirmation of this and seek their guidance on additional variables.

- ILL_L_ACTIVE (whether the learner is active for the academic year)
- ILA_PROGTYPE (type of programme e.g., “apprenticeship”, “entry to employment” etc.)
- ILL_L_NLM1 and ILL_L_NLM2 (indicating the learner’s path pursued with several options including a graduate apprenticeship, enrolment towards a Skills for Life qualification; an entry level qualification, amongst others)
- ILA_COMPSTATUS (completion status towards a particular aim)
- ILA_OUTCOME (learning outcome being pursued)
- ILA_D_EMPSTATFDL (employment status on the first day of learning)

We are only interested in whether a pupil has progressed to a Further Education (FE) destination, irrespective of whether they go on to attend for the full duration of the degree or achieve a qualification as a result.²⁴ As such, we code an individual that shows up as active (L_ACTIVE) and pursuing any learning aim (ILA_PROGTYPE) as 1 and all others as 0. In order to verify whether an individual coded as 0 is indeed active elsewhere, we confirm that their data is consistent across the rest of the variables i.e., that the variables ILL_L_NLM1, COMPSTATUS, ILA_OUTCOME are either missing or not applicable. We further include ILA_D_EMPSTATFDL as a measure to verify those that may be partially pursuing Further Education alongside employment. As above, we use the date of birth variable to calculate a pupil’s age. As such, we interpret “by the age of 22” as the academic year in which a pupil turns 22.

The primary outcomes above set the lowest relevant bar to capture the longitudinal outcome of pupils. They balance the tradeoffs between our key considerations for outcome selection in this study: i) maximally capturing all possible routes of attainment and progression, ii) achieving as much consistency as possible with variable interventions, and iii) measuring a longitudinal outcome that is predictive of the young person’s key life outcomes.

Secondary outcomes

The secondary outcomes below aim to provide more granular detail on attainment and progression. They cannot always be measured consistently across trials. For example, programmes focused on pupils undergoing GCSE resits e.g., *Texting Students* and *Study Supporters*, are likely to have a more direct impact on pupils’ likelihood of achieving a pass in their GCSEs in Maths and English by the age of 19 (outcome for RQ1.3) than e.g. an intervention focused on providing pupils with work experience such as *Generation STEM*

We categorise the research questions into those pertaining to attainment and fall under primary RQ1: RQ1.1 to RQ1.3 and those pertaining to progression and fall under primary RQ2: RQ2.1 to RQ2.3.

For all outcomes with no set age cut-off, we recognise that some pupils will have a greater or lesser chance to obtain a qualification or progress further through the educational system. At the time of DfE data collection, in the academic year of 2022/23, we have the following estimated pupil ages across trials: four trials with pupils aged 19, three trials with pupils aged 20, one trial with pupils aged 21, two trials with pupils aged 22, three trials with pupils aged

²⁴ Again, this is because theoretically we aim to capture the lowest bar that constitutes progression to Further Education, given that these outcomes are already remote from the time of the intervention.

23, two trials with pupils aged 24, one trial with pupils aged 25, and two trials with pupils aged 26 (see table A1 in Appendix A).

Attainment

The secondary outcomes below measuring Level 3 (RQ 1.1) and university attainment (RQ 1.2, RQ2.1) and progression were chosen to represent significant points of attainment or progression that have been shown to be predictive of an individual's later life outcomes, including lifetime earnings (Jerrim, 2022; Lupton et al., 2021, Social Mobility Commission, 2023). While the impact of the selected interventions on these labour market outcomes is outside the scope of this study due to the unavailability of linked data, they are of clear policy relevance. We have therefore chosen our measures of post-16 attainment with this link in mind.

RQ1.1 - Level 3 qualification by the age of 19

Sample: All 18 interventions.

Whether or not a pupil has achieved a Level 3 qualification by the age of 19 will be taken from the ILR dataset. Specifically, we will use the variable "L_FULLLEVEL3_ACHIEVE" coded as 1, 0 or -1 to signify whether the level was achieved, not achieved, or the data were missing, respectively.

As above, we use the date of birth variable to calculate a pupil's age. As such, we interpret "by the age of 19" as the academic year in which a pupil turns 19.

RQ1.2 - Completed a university degree

Sample: 10 interventions for which the cohort is at least 22 in 2022/23.

Whether or not a pupil has completed a university degree will rely on the HESA dataset and specifically focus on the variable "HE_XQOBTN01" and "HE_XMODE01" described as the "highest qualification obtained" and "mode of study to identify full time/ sandwich/ part-time students", respectively. Other relevant variables will also be cross-referenced to ensure accuracy e.g., the "expected length of study on the programme", the "year of student on this programme", and whether the pupil has suspended their study.

We will only conduct this analysis for trials where individuals are estimated to be at least 22 years old in 2022/23, to allow individuals sufficient time to complete a university degree. Typically, individuals in England are 18 or older when they begin their undergraduate study and graduate at 21 or 22.

RQ1.3 - Passed Maths and English GCSEs by the age of 19

Sample: All 18 interventions.

We selected to measure whether a pupil passed their Maths and English GCSE exams by the age of 19 to identify the subset of pupils who may not have been captured in RQ1 or RQ2. We use this to identify students who may not have met other educational benchmarks (from RQ1 and RQ2), but have still achieved these essential qualifications. This would capture for instance a situation where a pupil did not pass enough GCSEs to achieve a Level

2 but still passed their Maths and English GCSEs. This is noteworthy because the latter qualifications are essential for entering the workforce..We choose to define a passing grade as achieving a grade C or above in Maths and English, including resits, for two reasons. Firstly, pupils in England are required to resit the exams for these two subjects if they do not pass by the age of 18 ([National Careers Service, 2024](#)) and as such pupils at the time of analysis should have had *at least* two opportunities to sit their Maths and English GCSEs and achieve a grade C or above. Secondly, a pupils' GCSE grade only contributes to a Level 2 qualification if they achieve at least a grade of 4 or higher (C or higher as per the alphabetical grading system pre-2017²⁵) ([GOV.UK, 2024](#)).

The DfE has clarified that the KS4 dataset contains several variables to measure whether a pupil has sat an English and Maths GCSE exam in a particular academic year as well as their grade. As such, we will be able to use these variables to understand whether a pupil has sat their exam and whether they have achieved a passing grade. The variable names differ across some years but we consulted with the DfE who confirmed that these are nominal differences and do not present substantial measurement differences.

We will supplement our measurement with variables from the ILR Learner dataset as well using variables such as "ILL_MATHGRADE" and "ILL_ENGGRADE".

As above, we use the date of birth variable to calculate a pupil's age. As such, we interpret "by the age of 19" as the academic year in which a pupil turns 19.

Progression

RQ2.1 - Enrolled in a HE institution

Sample: All 18 interventions.

Whether or not a pupil is enrolled in a HE institution will rely on whether the pupil is *present* in the HESA dataset and have basic details on their programme. Specifically, we will use the variables "HE_XMODE01", "HE_COMDATE", "HE_ENDDATE" which are defined as "mode of study", "Date of Commencement of Programme", "End date", respectively. We will also double-check that the pupil is not classified as having suspended their studies i.e. using "NOTACT".

RQ2.2 and RQ2.3 - Progressing to academic Further Education (academic KS5); Progressing to non-academic Further Education (non-academic KS5)

Sample: All 18 interventions.

RQ2.2 and RQ2.3 were selected in the interest of providing a breakdown of primary RQ2. Specifically, while RQ2 captures progression to any KS5 destination, RQ2.2 focuses on whether pupils were more likely to pursue an academic or non-academic route.

The possible destinations for FE in England are fragmented and complex. They have also undergone significant reform over the past decade (Buttar et al., 2023). We therefore measure whether pupils follow the mainstream academic pathway in England at KS5, or

²⁵ The structure of GCSE grading changed in 2017 from alphabetical (ranging from A* to G) to numeric (1 to 9). See here for details: <https://ofqual.blog.gov.uk/2018/03/02/gcse-9-to-1-grades-a-brief-guide-for-parents/>

another KS5 qualification, to capture this split. The most common form of academic Further Education at KS5 level is A levels: according to the latest DfE data, 66.6% of 18-year olds attending FE locations in England took 3 A level exams in the 2022/23 academic year (GOVUK, 2023).

To calculate the outcome measure for RQ2.2, we will use data from the ILR Aims dataset. Specifically, we will first establish whether an individual is active and engaged in a programme towards a learning aim (ILA_P_ACTIVE), the stage of this learning aim e.g., whether the study is ongoing or complete (ILA_OUTCOME), and what the specific aim is, and whether it is an A or AS level (ILA_LDM_1 , ILA_LDM_2, ILA_LDM_3, ILA_LEARNAIMREF).

To calculate the outcome measure for RQ2.3, we will use similar variables to RQ2.2 from the ILR Aims dataset. Specifically, we will first establish whether an individual is active and engaged in a programme towards a learning aim (ILA_P_ACTIVE), the stage of this learning aim e.g., whether the study is ongoing or complete (ILA_OUTCOME), and what the specific aim is (ILA_LDM_1 , ILA_LDM_2, ILA_LDM_3, ILA_LEARNAIMREF). Here we deviate from RQ2.2 and only count all non-academic KS5 pathways. As we do not have visibility into the coding of these variables, it is unclear whether these include all apprenticeships. At the time of data cleaning, if these variables do not include apprenticeships, we will check this using ILA_PROGTYPE and ILA_D_ACCTOAPPEMPD which capture this.

Variable Specification and Interpretation

Table B1 in Appendix B provides a list of the datasets and variables we plan to use to measure each outcome. We have selected these based on the availability of variables as specified by the DfE.²⁶ The Choice column indicates our order of preference, based on two key considerations: (1) whether the selected variable(s) are consistently available across all years of interest, and (2) the relevance of the variable's description to our outcome measure.

There is some uncertainty over the quality of the data and rates of missingness for the variables used to construct the outcomes. We pre-specify the following rules:

1. For all outcomes, we will opt for the first choice of variable(s) and data source. If less than 5% of pupils successfully matched within a trial have missing outcome data, we will proceed with this measure.
2. If more than 5% of pupils successfully matched within a trial have missing outcome data using the first choice variable(s), we will use the second choice of variable(s) and data source, provided the rate of missingness is less than 5%.
3. If it is not, we will turn to the third choice, provided the rate of missingness is less than 5%.
4. If none of the variables specified allow to reach a rate of missingness less than to 5%, we will investigate patterns of missingness based on observable characteristics we expect to be present across trials, including age and treatment status. We will run a logistic regression with outcome missingness as dependent variable and pupil observable characteristics including age and treatment status as predictors.

²⁶ The tables with the variable names, the years available, and their definition are publicly available [here](#).

- a. If we can identify a pattern, we may i) consult DfE and request additional data; ii) combine variables from different sources, if we are confident that this does not introduce bias.
- b. If we cannot identify a pattern, we will choose the variable(s) and data source with the lowest rate of missing outcome data.

Where the resulting rate of outcome missingness is larger than 5%, we will i) flag this in the report, ii) report missingness by treatment status to assess evidence of differential attrition, and iii) caveat results accordingly. We will not conduct further missing data analysis as it would add considerable time and resources to the analysis and falls beyond the scope and aims of this study. More detail on how missingness will be reported is specified in the subsection *Missing Data*.

For example, for RQ1, we will first attempt to use the ILL_L_FULLEVEL2 and DATEOFBIRTH variables from the ILR dataset. However, if this variable has a higher than 5% rate of missingness, we will calculate whether a pupil has achieved a Level 2 using a series of variables from the NPD (KS4) data. These variables are not our first choice because the variable names change across years, which may reflect changes in measurement and coding across years.

Participants

As this impact evaluation focuses on a secondary data analysis of 20 trials, participants are not included or excluded at the individual level. For all selected trials, all pupils assigned to a treatment or comparison group for that trial are considered eligible for inclusion and analysis (see the subsections *Treatment effects in the presence of non-compliance* and *Missing data* below for details on how we will deal with non-compliance and missing data). Table 3 presents key characteristics of each trial, including sample size and participant characteristics.

Sample description and size

As this study is a secondary data analysis exercise, we do not conduct additional power calculations. Table 3 above shows whether the original trial was powered at the evaluation (analysis) stage. We find that 14 out of 20 trials were powered at 80% to detect an effect at their respective minimum detectable effect size for the primary analysis at the time of the trial's original evaluation.

Statistical Analysis

While we follow [EEF's statistical analysis guidance \(2022\)](#) to a large extent, we depart from it where it is not well-suited for the purpose of this longitudinal, multi-trial analysis using secondary data. We note where that is the case.

Intention-to-treat framework

To draw causal inference about the effect of the intervention, we maintain an “intention-to-treat” analysis framework, building on the initial random assignment of pupils (or other unit of randomisation) to the treatment and control group(s). We will enforce consistency with the original trial analysis specifications in this longitudinal analysis, referring back to the original Statistical Analysis Plans where needed.

For each trial, we will run at most 9 regressions (two primary outcomes, and seven secondary outcomes).²⁷ The regressions will be run separately for each trial, with each intervention being considered as a separate comparison against the corresponding control group. For each trial, each regression will include the same covariates and be clustered at the same level as in the original trial analysis.²⁸

Accounting for clustering

If the trial followed a clustered randomisation and/or clustered analysis, we will cluster the standard errors of the model at the level at which randomisation occurred, in line with EEF statistical analysis guidance (2022). We deviate from the EEF statistical guidance which typically recommends using a multilevel model (MM), as it does not align well with the aims of the study.

First, we do not aim at making inference claims about the general population of year-groups/schools in England. Instead, we aim to identify any long-term impacts of the selected interventions for the sample of pupils who initially took part. Schools in the sample cannot be considered to be representative of all schools, and claims generalising findings beyond that population would lack credibility.

Second, a MM typically provides additional information on (i) the variance between and across clusters which may arise due to variation in fidelity and implementation, and (ii) allows for cluster-level estimates if we were focussing on one trial. The primary motivation behind clustering is to account for the way in which the trial design may influence treatment allocation (i.e., correlations within clusters) and power. However, as we are interested in studying the long-term effect of an intervention, we will not analyse the cluster-level variance and as such, a model with clustered standard errors would still sufficiently reflect the design.

In summary, we choose to use a simpler modelisation that still accounts for clustering here because:

1. The objective is to estimate the effect at the individual level, with an adjusted standard error to account for clustering.
2. There is no need for estimates of cluster-specific variation.
3. We seek to avoid the assumptions required by FE or RE models, particularly if clusters are not truly random samples from a larger population (an issue for RE models in particular).

²⁷ This is in the best-case scenario where all outcome data is available for the trial sample.

²⁸ Our understanding is that these variables may not have been systematically submitted to the archive prior to October 2020, so we will likely have to match with NPD to construct these variables. [EEF note](#), November 2020.

For clustered trials where the cluster ID variable is missing from the archive data, we will rely on the trial protocol and analysis plan to establish the level of clustering. We will then check that this matches treatment assignment in the data by inferring the level of clustering based on whether treatment assignment is the same for all pupils in at least half of the relevant unit (school, school-year, classroom). If there are cases where the trial was clustered at a certain level (e.g. classroom) and the cluster ID is missing from the data, we will cluster at the closest above level for which we have information and the pupil's treatment assignment (e.g. school-treatment or year-group treatment).

Analysis will be conducted in Stata or RStudio, using 2-sided significance tests, at the 5% significance level (before any adjustment for multiple comparisons).

Primary outcome analysis

The analysis for equations 1.1 to 2.2 will include all randomised pupils for whom we have complete data across both data sets (EEF and DfE). Specifically, it will include cases for whom we have data on:

- Treatment allocation (EEF Archive)
- Cluster ID, if applicable (EEF Archive)
- Covariates, if applicable (EEF Archive *and* DfE data)
- Outcome variable (DfE data)

This is in line with EEF's statistical analysis guidance which proposes a complete case analysis.²⁹

Primary outcomes for RQ1 (Proportion with Level 2 qualification by age 19) and RQ2 (Proportion of pupils progressing to Further Education (FE) by the age of 22)

We will run the following analysis for all eligible trials that did not entail clustering at randomisation and/or analysis. As the outcomes are binary, we will use a logistic regression model.

Specifically, we will estimate a model of the form:

Equation 1

$$Y_i \sim \text{bernoulli}(p_i); \text{logit}(p_i) = \alpha + \beta_1 \text{Treatment}_i + \beta_2 X_{i,j}$$

Where:

- Y_i is a binary outcome variable which is equal to 1 if individual i meets the definition of the outcome variable, and 0 otherwise
- α is a constant
- Treatment_i is a vector of binary variables that will be equal to 1 if the individual i is allocated to that treatment group, and 0 if they were allocated to the control or comparison group, or another treatment group.
- $X_{i,j}$ is a vector of covariates included in the original primary trial analysis, if relevant. For most trials, we would expect this to be limited to the baseline value of the outcome or prior attainment measure.

²⁹ We expect to have both the complete and non complete cases in the EEF Archive based on the latest archiving guidance (appendix 3).

Where the original evaluation report notes clustered randomisation and/or analysis, we will add ε_c , the heteroskedasticity-robust Huber-White error term for pupil i from cluster c to the above model.

Multiple Comparisons Adjustments

Our overall analytical approach considers each trial as a ‘family’ and asks, for each outcome of interest, whether each of the implemented programmes had an impact, respectively. This implies that we will run each of the regressions specified below per trial, evaluating each trial independently as a follow-up. We will be adjusting for multiple comparisons for the primary and secondary outcomes separately.

Treating each trial as a ‘family’ makes sense theoretically for a number of reasons. First, the interventions are highly variable ranging from those providing work experience (e.g. *Generation STEM*) to those providing support on specific subjects (e.g. *Ark Mathematics Mastery*), or those targeting teachers (e.g. *Achieve Together*). Second, the proximal findings for each trial were also highly heterogeneous and we cannot hypothesise a uniform direction for any of the outcomes across trials (see Table B1). Lastly, our review of the interventions’ respective theory of change (see Appendix C) shows that the trials were different in aims and delivered to varying degrees of fidelity and quality.³⁰

We will apply the Benjamini-Hochberg correction where the standard 5% threshold for significance is adjusted based on the number of comparisons being made. See Table 5 below for how we will calculate the number of comparisons per analysis.

Table 5: Number of Comparisons (m)

		Number of Treatment Arms		
Number of Outcomes		1	2	3
	2	2	4	6
	3	3	6	9
	4	4	8	12
	5	5	10	15
	6	6	12	18
	7	7	14	21

³⁰ An alternative approach would have been to treat all trials taken together as a ‘family’ and adjust for multiple comparisons based on the number of trials included in each follow-up analysis, in addition to the number of primary outcomes in the analysis. However, even though we specify the same outcomes, the set of data used to answer the research questions will be different for each trial, and we deem this approach overly penalising.

For example, if we are studying the long-term effect of a trial with 1 treatment arm using 7 outcome measures, we would be making a total of 7 comparisons (see Table 5 above). In order to correct the p-values, we would follow these steps:

1. Rank the p-values for all 7 outcomes from smallest to largest
2. Calculate the *new* threshold values per outcome using the equation below

$$\alpha_{adjusted} = \frac{rank}{m}$$

where m is the number of comparisons and $rank$ represents the position of the p-value after ranking in step 1.³¹

We will be adjusting for multiple comparisons for the primary and secondary outcomes separately. Given there are 2 primary and 6 secondary outcomes, and the maximum number of treatment arms we will consider is 3 (see Table 4), the maximum number of comparisons is 6 for primary analyses and 18 for secondary analyses.

Inference

Logistic Regressions (Equation 1)

For each intervention, the measure of whether the trial had positive longitudinal effects would be if the treatment group(s) had:

- A significantly higher probability of achieving a Level 2 qualification by the age of 19 compared to the comparison group (RQ1).
- A significantly *higher* probability of progressing to a Further Education (FE) destination compared to the comparison group (RQ2).

β_1 from the model equations will be expressed in terms of log-odds in the raw output of the logistic regression, we will need to convert this into a probability. To do this we will:

1. Estimate the mean for the control or comparison group: We will estimate the mean proportion of individuals who achieve a Level 2 qualification by age 19 (RQ1) or progress to a Further Education destination (RQ2) in the control or comparison group. This provides the baseline probability for the control group.
2. Calculate the log-odds of the baseline probability: Using the baseline probability p from the control or comparison group, we calculate the log-odds:

$$logodds = \log\left(\frac{p}{1-p}\right)$$

3. Adjust for the treatment effect: We will incorporate the treatment effect from the logistic regression by adding the coefficient β_1 to the log-odds: $y = logodds + \beta_1$
4. Convert y back to a probability, using the inverse logit function:

³¹ Although EEF statistical guidance recommends the Bonferroni procedure, it states its openness to other methods e.g. the Hochberg (1988), a similar one to what we have stated above.

$$\text{invlogit}(y) = \frac{\exp(y)}{1 + \exp(y)}$$

If the analysis employed the specifications set out in Equations 1.2 and 2.2, and thus takes the form of a logistic regression with clustered errors, the interpretation of the effect of treatment allocation, β_1 coefficient, on the outcome of interest, will not change.

Further analyses

Secondary outcomes analyses

The following analysis will be run for the secondary outcomes listed above.

The analysis for equation 2 will include all randomised pupils for whom we have complete data across both data sets. Specifically, it will include cases for whom we have data on:

- Treatment allocation (EEF Archive)
- Cluster ID, *if applicable* (EEF Archive)
- Covariates (EEF Archive *and* DfE data)
- Outcome variable (DfE data)

This is in line with EEF's statistical analysis guidance which proposes a complete case analysis.

As with the primary analysis, analysis will be carried out using a logistic regression and follow clustering where clustering took place at the time of the intervention's initial evaluation:

Equation 2

$$Y_i \sim \text{bernoulli}(p_i); \logit(p_i) = \alpha + \beta_1 \text{Treatment}_i + \beta_2 X_i$$

Where:

- Y_i is a binary outcome variable which takes the following value of 1 if:
 - RQ 1.1: If individual i has achieved a Level 3 qualification
 - RQ 1.2: If individual i completed a university degree
 - RQ 1.3: If individual i passed their GCSEs in Maths and English
 - RQ 2.1: If individual i is or was enrolled in a HE institution
 - RQ 2.2: If individual i progressed to A levels
 - RQ 2.3: If individual i progressed to a Level 3 qualifications
 and takes the value of 0 otherwise
- α is a constant
- Treatment_i is a vector of binary variables that will be equal to 1 if the individual i is allocated to that treatment group, and 0 if they were allocated to the control or comparison group, or another treatment group.
- $X_{i,j}$ is a vector of covariates included in the original primary trial analysis, if relevant. For most trials, we would expect this to be limited to the baseline value of the outcome or prior attainment measure.
- For trials with clustered randomisation and/or analysis, we will add ε_c the heteroskedasticity-robust Huber-White error term for pupil i from cluster c to the model above.

Subgroup outcome analyses

For each trial, we will run one subgroup analysis for FSM-eligible pupils to study any heterogeneous treatment effects in that group, for the primary outcomes only. To do so, we will re-estimate the model specified in Equations 1 and 2, including in addition an interaction term indicating FSM classification, using the entire sample (see Equation 3).³² The analysis for equation 3 will include all randomised pupils for whom we have complete data across both data sets. Specifically, it will only include cases for whom we have data on:

- Treatment allocation (EEF Archive)
- Cluster ID, if applicable (EEF Archive)
- Covariates, if applicable (EEF Archive *and* DfE data)
- Outcome variable (DfE data)

This is in line with EEF's statistical analysis guidance which proposes a complete case analysis.

Equation 3

$$Y_{ic} \sim \text{bernoulli}(p_i); \text{logit}(p_i) = \alpha + \beta_1 \text{Treatment}_i \times \text{FSM}_i + \beta_2 X_i + \beta_3 \text{Treatment}_i \beta_4 \text{FSM}_i + e_i$$

Where:

- Y_i is a binary outcome variable which will take different values depending on the outcome:
 - For RQ1, Y_i will be equal to 1 if an individual has a Level 2 qualification by the age of 19, and 0 otherwise.
 - For RQ2, Y_i will be equal to 1 if an individual progressed to *any* Further Education (FE) destination by the age of 22, and 0 otherwise.
- α is a constant.
- Treatment_i is a vector of binary variables that will be equal to 1 if the individual i is allocated to that treatment group, and 0 if they were allocated to the control or comparison group, or another treatment group.
- FSM_i is a binary variable to indicate whether a pupil was eligible for free school meals at the time that the trial was run
- $X_{i,j}$ is a vector of covariates included in the original primary trial analysis, if relevant. For most trials, we would expect this to be limited to the baseline value of the outcome or prior attainment measure.
- For trials with clustered randomisation and/or analysis, we will add ϵ_c the heteroskedasticity-robust Huber-White error term for pupil i from cluster c to the model above.

³² This constitutes a minor departure from EEF statistical analysis guidance, which recommends running in addition the analysis on the subgroup of FSM and non-FSM eligible pupils respectively. These two approaches ultimately estimate the same thing, however the interaction model restricts the other covariates to have the same effect across the values of FSM status, while the subgroup analyses allow these other coefficients to vary (if we included all interactions in the interacted model, these two alternatives would produce the same results). Given the large number of pre-specified analyses, we privilege one approach.

To maximise completeness, FSM status will be taken from the School Census data table in the NDP database. Specifically, the variable “FSM” will be used for the first year that the trial was run i.e., FSM status at baseline.

Treatment effects in the presence of non-compliance

Non-compliance in this context refers to non-compliance with the original treatment assignment at the time that the trial was run. Generally, non-compliance is likely to cover instances where (i) pupils in the control or comparison group received treatment, and/or (ii) pupils in the treatment group(s) did not receive the treatment.

We will not conduct any further analyses to measure the effect of compliance on the outcomes for two reasons. Firstly, we do not expect to receive compliance data as the EEF guidelines did not mandate evaluators to submit data on compliance until 2020 (EEF, 2020). As the latest trial in our dataset is from 2016/17, we are unlikely to receive this variable for most trials. Secondly, even if we are able to access most or all of the compliance data, this project’s primary focus is the longitudinal effect of EEF trials on pupil outcomes and the project constraints do not allow for an additional compiler analysis per trial.

Missing data

There are two potential sources of missing data that may result in differences between the re-analysis sample and the original analysis sample; (i) outcome data and/or (ii) other variables included in the analysis i.e., covariate data. We describe below how we handle each case.

Outcome data

We anticipate that missingness in the outcome data can reflect three scenarios:

- 1. The individual is no longer in the English education system (for example, if a pupil is deceased or has moved out of the country);
- 2. The individual is still in the English education system but their PMR is missing or lost,
- 3. The individual is still in the English education system with a valid PMR but their outcome data are missing due to genuine data loss or human error.

Ideally, we would be able to distinguish between these three scenarios. However, the DfE datasets (NPD, HESA, and ILR) do not contain sufficiently clear information indicating whether an individual has left the English education system and/or the possible reason for a data loss. We know that 95% of datasets in the archive include a Pupil Matching Reference (PMR), and that 95% of records within linkable datasets have PMRs, which suggests a high potential matching rate.³³

We will report the rate of missing outcome data by treatment arm and outcome. As this impact evaluation has 2 primary and 7 secondary outcomes, we will report at least 18 rates of missing outcome data per trial. See Table 6 below for an example of this filled with arbitrary numbers.

Table 6: Template for reporting Missing Outcome Data

	% of Data Missing
--	-------------------

³³ EEF archive conference, June 2023.

	Treatment Arm	Control Arm
Level 2 qualification by 19 (P)	2	3
Progressed to any Further Education (FE) destination by 22 (P)	2	4
Level 3 qualification by the age of 19	1	5
Completed a university degree	0.6	7
Passed GCSEs in English, Maths by 19	2	2
Enrolled in a HE institution	1	1
Progressed to academic Further Education (academic KS5)	0.6	0.6
Progressed to non-academic Further Education (non-academic KS5)	2.3	3

If the rate of missingness is higher than 5% for any of the primary outcomes, EEF statistical guidelines would ordinarily encourage researchers to take further steps to determine whether multiple imputation and further sensitivity analyses are necessary. Specifically, the guidance first proposes a check to determine whether patterns of missingness in the data are; missing at random (MAR), where missingness is correlated with other observed variables in the data, missing not at random (MNAR), where missingness depends on some unobserved variable, or missing completely at random (MCAR), where missingness is not correlated with any observed, or unobserved variables. This information is then used to make a decision about whether multiple imputation or further sensitivity analyses are necessary.

For this study, given the number of trials, analyses conducted, and the primary aims of the research, we will only report the % of missingness as seen in Table 6 above and continue the analysis in the same way as specified above for all trials.

Covariate data

Missing covariate data is likely to be due to this data being missing at the time the trial was run. We will first supplement covariate data from the trial with covariate data from the National Pupil Database. This is to ensure that we remain as close as possible to the initial randomisation in the longitudinal analysis, to maximise the sample size and statistical power, and to use the most reliable source of data for covariates.

After we supplement the data using the NPD, we will then report the % of missing data by each covariate and treatment arm for covariates included in the primary analysis, similar to the approach taken in the case of missing outcome data in Table 6 above. If any of the covariates exhibit higher than 5% missingness, we will refer back to the output of Equation 3 as specified in the [Robustness Checks](#) section. This model provides the treatment estimate per trial without the inclusion of any covariates.

In case of missing covariate and/or outcome data, we will not conduct further tests to investigate the reasons for missingness. We deviate from EEF statistical guidance as this analysis is outside of the project scope. The potential resource and time burden of repeating robustness analyses 20 times per trial is not proportionate to the aims of this project.

Ethics

We conducted an ethical review in line with BIT's internal research ethics policy, which follows the guidance and principles developed by the Government Social Research Profession (GSRP). Key principles include: 1) sound application and conduct of social research methods and appropriate dissemination and utilisation of the findings; 2) participation based on valid informed consent; 3) enabling participation; 4) avoidance of personal and social harm; and 5) non-disclosure of identity and personal information.

The level of ethical review and process depends on the estimated risk level of the research. This project was deemed medium risk because while it exclusively uses secondary data, it analyses large amounts of data on pupils, including potentially sensitive data. Ethical approval has been conducted by the BIT internal ethics committee and was approved on 1st February 2024. The internal review was conducted by non-project member(s) of staff who are part of BIT's ethics reviewers panel and have been trained in BIT's ethical review process. They assess the research plan and data collection materials to ensure the project has considered all the potential risks and have appropriate procedures in place to mitigate these. The committee worked with the project team until they were satisfied that all relevant risks had been identified and that the appropriate mitigations were in place.

Data Protection

BIT will act as the independent data controller of the linked data from the two sources specified above.

For the constituent parts:

- FFT will act as the data controller of the EEF trials
- The Department of Education (DfE) acts as the data controller of the NPD-HESA-ILR data

As such, all requests to update or remove data will be dealt with by FFT and DfE. Their privacy notices are [here](#) and [here](#), respectively. The DfE further publishes specific privacy notices for the National Pupil Database (NPD) [here](#), for the Individualised Learner Record (ILR) [here](#), and for the Higher Education and Statistics Agency database [here](#). The SRS also has a privacy policy [here](#).

All data will only be accessed via the Official of National Statistics' Secure Research Service (ONS-SRS). This is a virtual environment that will host the EEF Archive data, all DfE data, and any necessary applications required to clean and analyse the data. In this case, we will conduct all the analysis using RStudio. Only ONS-SRS accredited researchers will access the data and adhere to the secure research service guidelines and data protection principles such as the "Five Safes" explained [here](#). All research outputs will undergo clearance by the SRS to ensure that there are no disclosures of personal or identifiable information. The process is outlined [here](#).

BIT may proceed to archive the data from this project in EEF's data archive. This will be decided at a later stage, in collaboration with EEF.

Lawful basis for processing data

The legal basis for processing this data is legitimate interest. It is necessary in BIT's 'legitimate interests' to process personal data in order to produce public-purpose research in this area. The research will benefit the Department of Education, and other social impact organisations such as the EEF, to understand the feasibility, challenges, and general process of linking administrative datasets such as the LEO dataset with EEF and other trial data. This can help guide future researchers aiming to pursue similar data linkage. It will also provide practitioners with an evidence base to assess what interventions may or may not have long-term impacts on pupils in KS2, KS3, KS4 and KS5 and aid the DfE to decide which interventions and research to fund next to boost educational progression, employability, and attainment. The research project fulfils BIT's core business aims including undertaking research, evaluation and information activities in sectors that will deliver social impact.

Data security

As mentioned above, we will only access and analyse the data in the ONS-SRS. While the chances of data leakage or insecurity are low, we will take steps to protect personal information and follow procedures designed to minimise unauthorised access, alteration, loss or disclosure of personal information. Those who have permanent or regular access to personal data, or that are involved in the processing of personal data, are trained and informed of their rights and responsibilities when processing personal data. We have put in place procedures to deal with any suspected personal data breach and will notify staff and parents and any applicable regulator of a breach where we are legally required to do so.

We will also rely on the measures taken by the SRS to secure the virtual environment, or project space that we will access. The ONS-SRS adopts the *Five Safes Framework*. More information on the SRS is detailed [here](#).

Data retention

We will be granted access to the SRS virtual environment, and therefore, all the raw data until September 2025.

Personnel

Table 7: Overview of Evaluation Team

Researcher Name	Role	Responsibilities	Institutional Affiliation
Dr. Laure Bokobza	Senior Research Advisor	Project Lead; end-to-end project management; oversight of research activities.	The Behavioural Insights Team

Dr. Patrick Taylor	Principal Research Advisor	Overseeing project; ensuring design and reporting meet research aims; technical support; management support for Senior Research Advisor(s).	The Behavioural Insights Team
Neus Torres-Blas	Research Advisor	Supporting research design, and data cleaning and analysis.	The Behavioural Insights Team
Dr. Tom O’Keeffe	Senior Research Advisor	Quality assurance of the study plan and data cleaning and analysis.	The Behavioural Insights Team
Trisha Harjani	Associate Quantitative Research Advisor	Supporting trial selection; data access; leading on data cleaning and analysis with the support of senior researchers.	The Behavioural Insights Team

Risks

Table 8: Risks

Risk	Impact	Likelihood	Mitigations
Poor archive data quality	High	Low	Accessing the archive data as early as possible, and planning for time to rigorously assess data quality and missingness. Selecting trials that have high-quality archive data.
Delays/difficulties in accessing administrative data from NPD, IRL and HESA	Medium	Medium	Submitting data applications ahead of the release of unamended datasets. Using high-level contacts at NPD to monitor progress and troubleshoot issues. Agile resourcing within BIT, with several ONS certified quantitative researchers.
Difficulties to identify the relevant individuals in the administrative data from DfE	High	Low	Agreeing on a process with DfE, EEF and FFT to extract the list of PMRs for the selected trials from the archive and share it with DfE. Aim to have one or more points of contact within the Department of Education to assist with difficulties and/or queries.
Missing data in the administrative datasets from NPD, ILR and HESA resulting in missing outcome data and/or covariate data in the main analysis	Medium	Medium	Create contingency plans for defining outcome data and pre-specify the variables used for each outcome. Set thresholds for missing data and set actions to implement in case the threshold is crossed.
Personal data breach	High	Low	Only accredited researchers will access the ONS-SRS and therefore engage with any personal data. This implies that they have been trained (or re-trained every 5 years) and assessed on the Five Safes Framework. Only deidentified data is used i.e., no names.

Timeline

Date	Activity	Staff responsible/leading
------	----------	---------------------------

October 2023 to December 2023	Selection of trials	Dr. Patrick Taylor, Dr. Laure Bokobza, Trisha Harjani, Neus Torres-Blas
November 2023 to February 2024	Scoping data access	Dr. Patrick Taylor, Dr. Laure Bokobza, Trisha Harjani, Neus Torres-Blas
March 2024 to October 2024	Data Access Request and Linkage in ONS SRS	Dr. Laure Bokobza, Trisha Harjani, Neus Torres-Blas, Dr Tom O' Keefe
October 2024 to March 2025	Analysis and reporting	Dr. Laure Bokobza, Trisha Harjani, Neus Torres-Blas, Dr Tom O' Keefe, Dr Patrick Taylor

References

- Andrews, J., Hutchinson, J., & Robinson, D. (2017). Closing the gap? Trends in educational attainment and disadvantage. In <https://epi.org.uk/publications-and-research/closing-gap-trends-educational-attainment-disadvantage/>. Education Policy Institute. Retrieved August 7, 2024, from https://epi.org.uk/wp-content/uploads/2017/08/Closing-the-Gap_EPI-.pdf
- Department for Education. (2024a, March 28). *Statistics: NEET and participation*. GOV.UK. Retrieved August 7, 2024, from <https://www.gov.uk/government/collections/statistics-neet>
- Department of Education. (2024b, April 25). *Level 2 and 3 attainment age 16 to 25*. Explore Education Statistics - GOV.UK. Retrieved August 7, 2024, from <https://explore-education-statistics.service.gov.uk/find-statistics/level-2-and-3-attainment-by-young-people-aged-19#releaseHeadlines-summary>
- EEF. (2020). Archiving evaluation data analysed using the ONS Secure Research Service. In EEF. EEF. Retrieved August 7, 2024, from https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/archiving-evaluation-data/Archiving-evaluation-data-analysed-in-the-SRS-Project-completion-and-archiving-form_v2.pdf?v=1720712834
- EEF. (2022). Statistical analysis guidance for EEF evaluations. In EEF (p. 1) [Report]. Retrieved August 7, 2024, from <https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1722968052>
- EEF. (2019). *Longitudinal analysis of EEF trials*. EEF. Retrieved August 7, 2024, from https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/longitudinal_guidance.pdf?v=1721031784
- Education Policy Institute, *Annual Report 2023*, from <https://epi.org.uk/annual-report-2023/>
- Jerrim, J. (2022). The benefits of meeting key grade thresholds in high-stakes examinations. New evidence from England. *British Journal of Educational Studies*, 71(1), 5–28. <https://doi.org/10.1080/00071005.2022.2033692>
- GOV.UK. (2024, April 25). *Level 2 and 3 attainment age 16 to 25*. GOV.UK. Retrieved September 19, 2024, from <https://explore-education-statistics.service.gov.uk/methodology/level-2-and-3-attainment-age-16-to-25#content-section-3-content-1>
- GOV.UK. (2024). *School leaving age*. Retrieved August 7, 2024, from <https://www.gov.uk/know-when-you-can-leave-school>
- Hupkau, C., McNally, S., Ruiz-Valenzuela, J., & Ventura, G. (2017). Post-Compulsory education in England: choices and implications. *National Institute Economic Review*, 240, R42–R57. <https://doi.org/10.1177/002795011724000113>
- Lupton, R., Thomson, S., Velthuis, S., & Unwin, L. Moving on from initial GCSE ‘failure’: Post-16 transitions for ‘lower attainers’ and why the English education system must do better. From <https://www.nuffieldfoundation.org/wp-content/uploads/2021/02/Post16-transitions-for-lower-attainers-Final-report.pdf>

Machin, S., & McNally, S. (2012). The evaluation of English education Policies. *National Institute Economic Review*, 219, R15–R25. <https://doi.org/10.1177/002795011221900103>

National Careers Service (2024). *Exam resits*. National Careers Service. Retrieved September 19, 2024, from <https://nationalcareers.service.gov.uk/exam-results/resits#:~:text=Subjects%20you%20can%20retake,in%20English%20or%20maths%20instead>.

Ofqual. (2023, August 17). *Infographics for A level results, 2023 (accessible)*. GOV.UK. Retrieved August 7, 2024, from <https://www.gov.uk/government/publications/infographic-a-level-results-2023/infographics-for-a-level-results-2023-accessible>

Smith, G., & Smith, T. (2014). Targeting educational disadvantage by area: continuity and change in urban areas in England, 1968–2014. *Oxford Review of Education*, 40(6), 715–738. <https://doi.org/10.1080/03054985.2014.981436>

Social Mobility Commission, Buttar, J., Alonso, R., Martin, K., Donkor, E., & Guy, J. (2023). Labour market value of higher and further education qualifications. In L. Sibiet, S. McNally, M. Dickson, & Institute for Fiscal Studies (Eds.), *Social Mobility Commission* [Research report]. Retrieved August 7, 2024, from <https://www.gov.uk/government/organisations/social-mobility-commission>

APPENDIX A

Table A1. EEF Trials selected for analysis and maximum age of pupils by academic year³⁴

	Programme name	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22	2022-23
1	Best Practice in Setting				12	13	14	15	16	17	18	19
2	British Science Association: CREST Silver Award						14	15	16	17	18	19
3	Scinapse					13	14	15	16	17	18	19
4	Catch Up Literacy (2014)			11	12	13	14	15	16	17	18	19
5	Generation STEM						15	16	17	18	19	20
6	Maths in Context						15	16	17	18	19	20
7	Ark Mathematics Mastery		11	12	13	14	15	16	17	18	19	20
8	Quest		12	13	14	15	16	17	18	19	20	21
9	EFA				15	16	17	18	19	20	21	22
10	Youth United			14	15	16	17	18	19	20	21	22
12	Peer Tutoring in Secondary Schools		14	15	16	17	18	19	20	21	22	23
11												
12	Texting Students						18	19	20	21	22	23
13	The RISE Project				16	17	18	19	20	21	22	23
	Texting Parents (Oldest)			16	17	18	19	20	21	22	23	24
14	Texting Parents (Youngest)			11	12	13	14	15	16	17	18	19
15	Achieve together			16	17	18	19	20	21	22	23	24
16	Challenge the Gap (KS3 and KS4 only)		16	17	18	19	20	21	22	23	24	25
17	Increasing Pupil Motivation	16	17	18	19	20	21	22	23	24	25	26

³⁴ Colour coding. **Green**; before the trial began. **Blue**; pupils under 19, **Red**: pupils aged 19 and above

	Programme name	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22	2022-23
1	Best Practice in Setting				12	13	14	15	16	17	18	19
2	British Science Association: CREST Silver Award						14	15	16	17	18	19
3	Scinapse					13	14	15	16	17	18	19
4	Catch Up Literacy (2014)			11	12	13	14	15	16	17	18	19
5	Generation STEM						15	16	17	18	19	20
6	Maths in Context						15	16	17	18	19	20
7	Ark Mathematics Mastery		11	12	13	14	15	16	17	18	19	20
8	Quest		12	13	14	15	16	17	18	19	20	21
18	Tutor Trust	16	17	18	19	20	21	22	23	24	25	26

APPENDIX B

Table B1: Summary of outcome measures for attainment outcomes

RQ	Outcome Primary in bold	Choice	Data	Requires calculation ³⁵	All years Available ³⁶	Variable(s)
1	Percentage of pupils with a Level 2 qualification by the age of 19	1st	ILR <i>Learner</i>	No	Yes	ILL_L_FULLEVEL2, DATEOFBIRTH
		2nd	NPD <i>KS4</i>	Yes	Yes	KS4_PASS_1AC, KS4_PASS_1AC_PTQ, KS4_LEVEL2_EM_PTQ_EE, KS4_LEVEL2_PTQ, KS4_DOB
		3rd	NPD <i>YPMAD</i>	No	No ³⁷	YPMAD_L2_OVERALL, YPMAD_DOB
1.1	Percentage of pupils achieving a Level 3 qualification by the age of 19	1st	ILR <i>Learner</i>	No	Yes	ILL_L_FULLEVEL3, DATEOFBIRTH
		2nd	NPD <i>KS5</i>	Yes	Yes	KS5_PASS1L3, KS5_PASS2LV3, KS5_PTQ_INCLUDE, KS5_LEV3, KS5_DOB
		3rd	NPD <i>YPMAD</i>	No	No	YPMAD_L3_ACADEMIC, YPMAD_L3_FE_ILR, YPMAD_L3_OVERALL, YPMAD_DOB
1.2	Percentage of pupils completing a university degree	1st	HESA	No	Yes	HE_XQOBTN01, HE_XMODE01
		2nd	HESA	No	Yes	XQLEV301
1.3	Percentage of pupils passing GCSEs in	1st	NPD <i>KS4</i>	Yes	Yes	KS4_PASS_LEV2EM, KS4_PASS_LEV2EM_PTQ, KS4_PASS_LEV2EM_PTQ_EE KS4_DOB

³⁵ This column indicates whether calculating this variable will require calculation and subsequently, additional data checks. This may be for several reasons including changes in the coding, labelling, or definition of a variable over time. For example, the 2nd choice of variables for RQ1 are available across all years but will be sourced for from KS4_PASS_1AC for 2012/13, from KS4_PASS_1AC_PTQ for 2013/14, and from KS4_LEVEL2_EM_PTQ_EE for 2014/15 up to 2022/23.

³⁶ Years of interest range from the academic years of 2012/13 to 2022/23; so where this column states “All years” it refers to the full range. Where columns

³⁷ The YPMAD is a newer dataset that only covers the academic year of 2022/23. However, the YPMAD data table captures pupil attainment at age 19. The dataset covers anyone aged 19 to 24 in 2022/23. This would provide data for 17 out of 20 trials.

RQ	Outcome Primary in bold	Choice	Data	Requires calculation ³⁵	All years Available ³⁶	Variable(s)
	English and Maths by the age of 19	2nd	NPD <i>YPMAD</i>	No	No	YPMAD_GCSE_GRADE_ENG, YPMAD_GCSE_GRADE_MATHS, QUALCODE, EXAMSEASON, EXAMYEAR, YPMAD_DOB
		3rd	ILR <i>Learner</i>	No	No	ILL_L_EDFMath, ILL_L_EDFEng, ILL_L_MGA, ILL_L_EGA
		4th	NPD <i>Pupil Level School Census</i>	No	No	MathsGCSEHighestPriorAttainment_[term][yy], EnglishGCSEHighestPriorAttainment_[term][yy]
2	Percentage of pupils progressing to any Further Education (FE) destination by the age of 22	1st	ILR <i>Learner & Aims</i>	No	Yes	ILL_L_NLM1, ILL_L_NLM2, ILL_L_ACTIVE, ILA_PROGTYPE, ILA_COMPSTATUS, ILA_OUTCOME, ILA_D_EMPSTATFDL
		2nd	NPD <i>KS5 & PLAMS</i>	Yes	Yes	KS5_NFTYPE, KS5_MMSCH, KS5_MMSCH2, KS5_MSCH, KS5_MSCH2, KS5_DOB, pl_LearningAimStatus, pl_OnRoll
		3rd	NPD <i>YPMAD</i>	No	No	YPMAD_ILR_FE_DATAONPUPIL, YPMAD_L3_FE_ILR_CA, YPMAD_L2_MA_Academic_Study
2.1	Percentage of pupils enrolled in a HE institution	1st	HESA	No	Yes	HE_XMODE01, HE_COMDATE, HE_ENDDATE
		2nd	HESA	No	Yes	HE_COURSEAIM
2.2	Percentage of pupils progressing to academic Further Education (academic KS5)	1st	ILR <i>Learner & Aims</i>	Yes	Yes	ILA_LDM_1, ILA_LDM_2, ILA_LDM_3, ILA_LEARNAIMREF, ILA_OUTCOME, ILA_P_ACTIVE
		2nd	NPD <i>KS5 & PLAMS</i>	Yes	Yes	KS5_ENROLSTS, KS5_NFTYPE, KS5_ENTRY_GA, KS5_ENTRY_GAS, KS5_ENTRY_GDAS, KS5_ENTRY_VGAA, KS5_ENTRY_VGAAS, KS5_ENTRY_K3, KS5_TRIGGER, KS5_TRIGGER_2016, KS5_TRIGGER_2021, pl_LearningAimStatus, pl_OnRoll
2.3	Percentage pupils progressing	1st	ILR <i>Aims</i>	Yes	Yes	ILA_LDM_1, ILA_LDM_2, ILA_LDM_3, ILA_PROGTYPE, ILA_D_ACCTOAPPEMPD

RQ	Outcome Primary in bold	Choice	Data	Requires calculation ³⁵	All years Available ³⁶	Variable(s)
	to non-academic Further Education (non-academic KS5)	2nd	NPD <i>KS5 & PLAMS</i>	Yes	No	KS5_ONAPPREN, pl_LearningAimStatus, pl_OnRoll, KS5_NFTYPE

APPENDIX C

Table C1: Overview of hypotheses based on Theory of Change

Project Name	Original Study Outcomes*	Research Question		
		Post-16 Attainment Outcomes	Post-16 Educational Progression	Post-16 Labour Market Activities
Achieve together				
Ark Mathematics Mastery				
Best Practice in Setting				
CREST				
Catch Up Literacy (2014)				
Challenge the Gap		*Mixed for FSM	*Mixed for FSM	
EFA				
Generation STEM Work Experience				
Increasing Pupil Motivation				
Maths in Context				
Peer Tutoring				
Quest				
Sci-napse				
Texting Parents				*None for FSM
Texting Students and Study Supporters				
The RISE Project				
Tutor Trust				
Youth United				

**Note that the direction of the effect does not imply its statistical significance*

Colour codes:

Negative	Positive
Mixed Directions / Insufficient Evidence	No plausible relationship identified (not applicable for 1st column, which indicates results)

Table C2

Mechanisms	Number of trials for which the mechanism was identified RQ1: Attainment Outcomes
Increased Attainment*	13 ³⁸
Pupil Engagement/ Enjoyment	2
Teaching quality	9
Pupil Confidence*	6 ³⁹
Pupil Absenteeism/ Attendance	3
Pupil Attitudes	2
Pupil Effort	2

Table C3

Mechanisms	Number of trials for which the mechanism was identified RQ2: Educational Progression
Increased Attainment*	8 ⁴⁰
Pupil Engagement/ Enjoyment	6
Pupil Confidence*	2 ⁴¹
Teaching quality	2
Pupil Effort	2
Pupil Motivation	2

**All mechanisms with an asterisk indicate that they were in the top 3 for both research questions.*

³⁸ 4 of these relate to a hypothesised negative relationship

³⁹ 1 of these relates to a hypothesised negative relationship

⁴⁰ 3 of these relate to a hypothesised negative relationship

⁴¹ 1 of these relates to a hypothesised negative relationship

The above tables list all mechanisms that were identified at least twice across all trials, by research question. For simplicity, some similar mechanisms were grouped. Specifically, “teaching quality” may refer also to teaching practices, teacher support, teaching effort, TA quality. Similarly, “pupil attitudes” range from attitudes to school or a specific subject, and “attainment” may refer to general attainment scores being higher, or specific skill improvement e.g. “improved pupil reading ability” or “better quantitative skills”. The mechanisms may relate to a hypothesised positive or negative relationship.

Achieve Together

Background info

- **Population:** KS4
- **Intervention:** Subsidised leadership development training for teachers in middle and senior leadership roles with a specific focus on participants working collaboratively on a school improvement impact initiative (p.6⁴²)
 - **Deliverer:** External Providers
 - **Mode of delivery:** Full classroom (teacher training)
 - **Dosage:** Averaged 83.7 days of staff time over 3 years (p.32, Table 15)

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
KS4 points: 0.014 [-0.063, 0.091]	+	<i>Not significant</i>
(Lower) Absence: -0.025 [-0.072, 0.024]	+	<i>Not significant</i>
KS4 points (FSM): -0.003 [-0.077, 0.083]	N/A	<i>Not significant</i>
KS4 points (after 1 year): -0.067 [-0.147, 0.013]	-	<i>p=0.092</i>

- **Outcomes:** copied from Key Conclusions (p.4 & p.47)
 - ‘Pupils in Achieve Together schools made no additional progress, on average, compared to other children in the trial, as measured by GCSE results. The three padlock security rating means we are moderately confident in this result.’
 - ‘Similar results were found for children eligible, and not eligible, for free school meals, and children with higher and lower prior attainment, although these sub-group results are less secure because of the smaller numbers of pupils.’
 - ‘...Some participants found that the process increased reflective practice and provided impetus for further school improvement. Others felt that it was resource intensive and had minimal additional impact.’
 - ‘Some teachers felt Achieve Together would have a positive impact on pupil attainment once new teaching skills and new initiatives were embedded. The EEF data archive allows analysis of future results for these schools to assess whether this is the case.’
- **Any mention of longer-term outcomes:**
 - ‘Staff anticipated that the leadership skills developed from participating in individual leadership programmes would ultimately feed into improved pupil attainment, but that this was a long-term goal that would not be apparent immediately.’ (p.44)

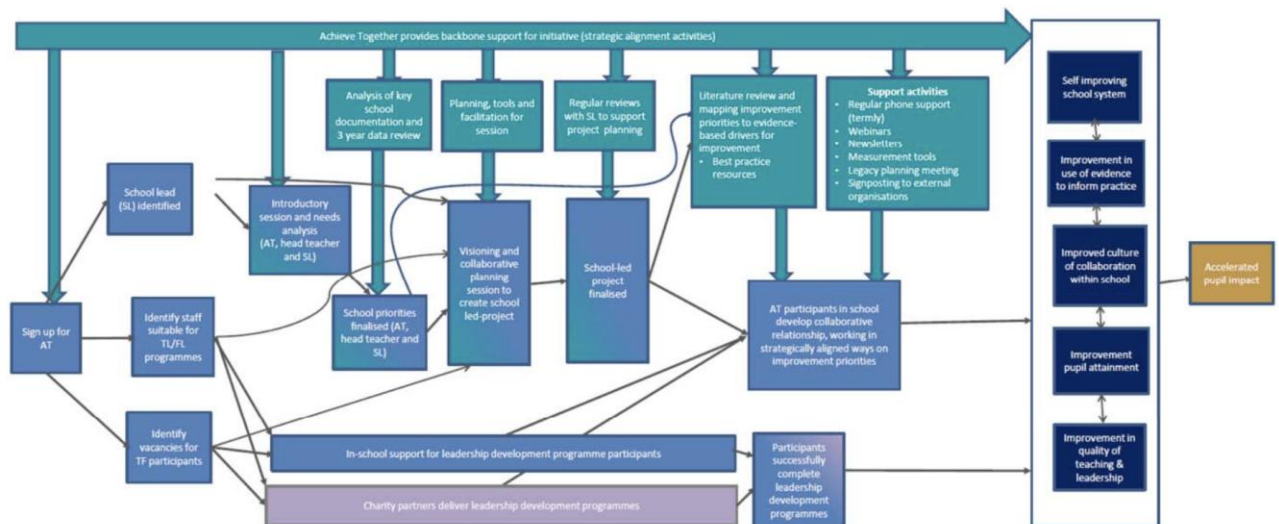
Resource list

- Efficacy trial ([EEF Page](#))

⁴² Unless stated otherwise, page numbers refer to the original evaluation final report, as published on EEF's website.

- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant: N/A
- Suggested further research: pages 9-10 covers background research on the Achieve Together / similar programmes

Programme logic model/ theory of change (see p.54)



Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
 % of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **improved teaching quality and training**
 - ‘Staff described thinking more strategically, developing their confidence, improving their management skills, and using these skills to support other staff and contribute to school improvement’ (p.43)
 - “‘I mean hopefully the things that I’ve picked up from it, I’ll continue to use over the next couple of years, and then that will hopefully mean that I’m doing my job better which will then have a direct impact on the staff in the team which will in turn have an impact on the results, so I think ... could be quite long-term’ (TL Fellows participant, Year 2, School 4).” (p.44)
- May **increase** post-16 attainment through **lowered pupil absence**
 - Reduction (non significant) in absence found; see outcomes table above
 - “In the case of a school that had used its impact initiative to develop and roll out a new approach to attitudes to learning, staff identified positive outcomes for pupil levels of engagement in lessons and there was an expectation that this would ultimately improve outcomes.” (p.45)
- May **decrease** post-16 attainment through **lowered attainment**
 - See outcomes table above, 1 year later, students had lower attainment in the intervention group (significant at the 10% level only).

FSM Pupils

- May **increase FSM pupils’** post-16 attainment through **improved teaching quality and training**

- Same as above (see Table 10, p.30 for evidence of subgroup analysis that found the positive *albeit not statistically significant* results held up for FSM pupils)
- May **increase FSM pupils'** post-16 attainment through **lowered pupil absence**
 - Same as above (see Table 8 for secondary analysis; FSM is included as a control variable here)
- May **decrease FSM pupils'** post-16 attainment through **lowered attainment**
 - Same as above; there is no sub-group analysis that was done at the 1-year mark

RQ2: Educational Progression
% of pupils progressing to KS5

- May **increase** post-16 progression through **improved pupil engagement**
 - Pupil attainment was varied throughout the programme, but many teachers and schools anticipated longer-term positive effects on pupil attainment once programmes were further embedded (see p.44, 47)

FSM Pupils

- May **increase FSM pupils'** post-16 progression through **improved pupil engagement**
 - Same as above

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **improved school practices**
 - The process evaluation notes that “schools were divided on the extent to which they felt the Achieve Together programme had contributed to school improvement (beyond the impact of the individual leadership programmes). For schools that succeeded...positive outcomes included: An increase in **school improvement impetus**...An **increase in reflective practice**” (p.43-44)

FSM Pupils

- May **increase FSM pupils'** participation in labour market/education/ training through **improved school practices**
 - Same as above

Any Additional Evidence [OPTIONAL]

N/A

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?

2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)

- Some notes on lack of clarity in delivery due to unclear aims, recruitment delays, timing of alignment activities, compatibility of leadership programmes, and school capacity to lead the collaboration. (pp. 34-46)
- Fidelity was difficult to assess given flexibility of programme: *"Given this flexibility it is difficult to make a judgement on fidelity to the programme. However, across the participating schools there was variation both in terms of the number of staff participating in the leadership programmes (from 5 to 23) and in terms of which leadership programmes were taken-up. This variation may impact on the fidelity of the programme by limiting the scope for collaboration in schools with fewer staff participating."* (p.43)

3. Risk of Spillover (high/low): Low

Challenge the Gap

Background info

- **Population:** KS1, KS2, KS3, KS4
- **Intervention:** School to school improvement programme where "lead" schools share best practices with "accelerator" schools through encouraging schools to work together, share practice and engage in joint practice development to improve attainment among disadvantaged pupil groups (p.3)
 - **Deliverer:** Teacher/ Staff and External
 - **Mode of delivery:** 12-15 *disadvantaged* students per school
 - **Dosage:** 6 workshops over two years

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
KS2 points: 0.001 [-0.035, 0.034]	+	<i>Not significant</i>
KS4 points: -0.007 [-0.03, 0.016]	-	<i>Not significant</i>
KS2 points (FSM): 0.101 [0.048, 0.154]	+	<i>p=0.025</i>
KS4 points (FSM): -0.096 [-0.143, -0.049]	-	<i>p=0.006</i>

- **Outcomes:** copied from Key Conclusions (p.1 & p.34)
 - *"The project found no evidence that Challenge the Gap (CtG) **increased average attainment for either primary or secondary school pupils, overall.** The security of the primary school results is low to moderate, and the security for the secondary school results is low."*
 - *"The **findings are different for children eligible for free school meals.** ...[FSM-eligible children in] **primary** schools made **2 months' additional progress** ...while FSM-eligible children in CtG **secondary** schools made **2 months' fewer progress** ... less secure than the overall findings."*
 - Teachers and non-teaching staff from participating schools were **extremely positive about the involvement of their schools** in the CtG programme

- and valued both the professional development opportunities it provided and the opportunity to collaborate with partner schools.”
 - “... **large variation** between the strategies that were adopted...”
- Any mention of longer term outcomes:
 - “Chapman et al. (2014) argue that adopting a collaborative inquiry approach to improvement is an effective strategy for tackling the attainment gap because: it generates a commitment to long-term sustainability of performance improvement, extending beyond the time-frame of a particular project” (p.7)

Resource list:

- Efficacy trial ([EEF Page](#))
- [Final report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant: N/A
- Suggested further research: It might be useful to look at the previously successful programmes of Challenge the Gap and understand their longer-term impacts
 - 2012 version of the CtG programme (Challenge Partners have since developed and continue to develop the approach).

Programme logic model (if found in project documents, skip if not):

- No explicit theory of change but the report mentions: “*In the event, the local variations between schools and implementation strategies meant that the identification of a single ‘theory of change’ was not possible. Consequently, efforts were directed towards identifying common elements that schools associated with effective implementation and impact and common problems encountered through the data collection methods outlined above.*” (p.14)

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
 % of pupils with level 2 as highest qualification

- **May increase post-16 attainment through increased awareness and improvement in teaching practices and effort in teaching.**
 - ‘It seems clear from the above that, to a large extent, the experiences reported by schools taking part in CtG reinforces previous research findings regarding the conditions that lead to effective networking: the majority of the participants believe that working closely with colleagues from other schools has added an additional dimension to their efforts.’ (p.27)
 - ‘... (at that workshop) there was thinking about resilience and resilient’ learners and they talked about an intervention at the lead school there, that we’ve basically pretty much stolen and we’re trying to put into place now, which is having all of the students... giving a talk during their form time.’ (p.28)
- **May increase post-16 attainment through increased pupils’ attendance/engagement with school**
 - ‘...many instances where teachers referred to visible improvements in pupils’ progress, though these were often expressed in terms of changes in attitudes, engagement or attendance, rather than demonstrated in measured attainment.’ (p.28)

FSM Pupils

- May increase FSM pupils' post-16 attainment through **better support from teachers**
 - *'...the TAs understand how to best support PP pupils in all classes and they have developed strong relationships with the pupils in the groups' (p.28)*
- May increase FSM pupils' post-16 attainment through **pupils' increased attainment in KS2**
 - *'...there is a small gain in months' progress (two months) for FSM eligible primary school pupils.' (p.19)*
- May decrease FSM pupils' post-16 attainment through **pupils' decreased attainment in KS4**
 - *'FSM eligible pupils in Key Stage 4 show a statistically significant decline in two months' progress when compared to comparison schools' (p.19)*

RQ2: Educational Progression
 % of pupils progressing to KS5

- May increase students' post-16 progression through **better teaching quality and broadening teacher's network/ exposure**
 - As noted above, teachers found the exposure to other schools useful and noted that time and authority constraints may have prevented the interventions from being effective in the short-run but **learnings may be implemented in the longer-term**
 - *"profound impact that the programme has had on some teachers from single-form entry primary schools, who reported that the trios provided them with the first opportunity available to them in their professional lives to plan teaching jointly with an experienced colleague" (p.28)*
- May increase students' post-16 progression through **perceived increased student progress**
 - *'there was a clear perception among those involved that the programme was accelerating the progress of targeted disadvantaged pupils' (p.28)*

FSM Pupils

- May increase FSM pupils' post-16 progression through **pupils' increased attainment in KS2**
 - *'...there is a small gain in months' progress (two months) for FSM eligible primary school pupils.' (p.19)*
- May decrease students' post-16 progression through **pupils' decreased attainment in KS4**
 - *'FSM eligible pupils in Key Stage 4 show a statistically significant decline in two months' progress when compared to comparison schools' (p.19)*

RQ3: Labour Market Activities

of months not in employment, education or training from August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May increase post-16 participation in labour market/ education/ training through **cross-school contact**
 - *'Practical collaborations between schools have often extended to the target pupil groups, offering them new experiences and through cross-school contact, deepening the feeling of collective responsibility for achieving the intended outcomes for all involved.' (p.27)*

FSM Pupils

- May increase FSM pupils' post-16 participation in labour market/ education/ training through cross-school contact
 - Same as above

Any Additional Evidence [OPTIONAL]

- Common elements (for impact) detailed in the process evaluation: '...we can see that there are a number of **common elements** that seem to run across most participating schools, and these include the **nature and size of the target group**, the **willingness to collaborate** both within and across schools to support this group, deliberate and **regular monitoring of impact**, and **commitment to sustain and spread** those aspects of the intervention that seem to impact most clearly on pupils' progress' (p.28)

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues / violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - The answers to all the questions above were a 'No', but:
 - **The study was not fully powered to detect small changes** for secondary schools. This is particularly important as KS4 / high school students made 2 months less progress and this effect was significant ($p = 0.006$), but needs to be replicated on a larger scale given the power issues (see table at bottom of p.2).
 - **Study had a security rating of 2 for primary and 1 for secondary**

3. Risk of Spillover (high/low): Low

EFA

Background info

- **Population:** KS4
- **Intervention:** Professional development programme which aims to improve pupil outcomes by embedding the use of formative assessment strategies across a school (p.6)
 - **Deliverer:** Teachers
 - **Mode of delivery:** Full classroom
 - **Dosage:** 18 workshops over 2 years and monthly peer observations

Impact and Outcomes

Finding	Direction	Significance
Attainment 8: 0.10 [-0.01, 0.21]	+	<i>Not significant</i>
Attainment 8 (FSM): 0.07 [-0.04, 0.19]	+	<i>Not significant</i>

Attainment 8 (schools not previously exposed to TEEP): 0.13 [0, 0.25]	+	$p=0.04$
---	---	----------

- **Outcomes:** copied from Key Conclusions (p.4 & p.41)
 - “Students in the Embedding Formative Assessment schools made **the equivalent of two additional months’ progress in their Attainment 8 GCSE score**, using the standard EEF conversion from pupil scores to months progress. This result has a very high security rating.”
 - “The project found no evidence that Embedding Formative Assessment improved English or Maths GCSE attainment specifically.”
 - “The **additional progress made by children in the lowest third for prior attainment was greater than that made by children in the highest third**. These results are less robust and have a lower security rating than the overall findings because of the smaller number of pupils.”
 - “**Teachers were positive about the Teacher Learning Communities...** improved their practice by allowing valuable dialogue between teachers, and encouraged experimentation with formative assessment strategies.”
- **Any mention of longer-term outcomes:**
 - “... teachers **often argued that improvements in teaching practices and pupils’ approaches to learning would take time to feed into higher pupil attainment**. This was reinforced by the frequent observation among teachers that younger pupils were more receptive to the intervention than their older and more exam-minded peers. As such, the **intervention may be more suited for a long-term evaluation in which the impact on pupil attainment is assessed a number of years after delivery**, among pupils who were younger at the time of the intervention’ (p.42)

Resource list

- Efficacy trial (EEF Page):
- [Final Report](#)
- Effectiveness trial ([EEF page](#))
- Any cited studies that might be relevant:
- Suggested further research:

Programme logic model/ theory of change

N/A

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes

% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **higher pupil attainment**
 - Significant effects on pupil attainment in the short term (see outcomes table above)
 - A sense that there would be longer-term improvement (p.41,42; see quotes above). Similarly most teachers predicted that the EFA programme would produce some increase pupil attainment (Table 17, p.38)
- May **increase** post-16 attainment through **better teaching quality**
 - ‘In particular, some teachers and school leads focused on the positive impact on teaching practices and teaching quality, including increases in high quality feedback, more awareness of current student attainment and understanding

of topics, the increased use of engaging teaching strategies, and more self-reflection about own teaching practices. It was generally believed that, ultimately, this would result in higher attainment' (p.38)

FSM Pupils

- May **increase FSM pupils'** post-16 attainment through **higher pupil attainment**
 - Same as above (effects held, albeit non-significant, for FSM pupils)
- May **increase FSM pupils'** post-16 attainment through **better teaching quality**
 - Same as above (no specific comments found for FSM pupils in the process evaluation)

RQ2: Educational Progression

% of pupils progressing to KS5

- May **increase** post-16 progression through **better student engagement**
 - *"It was also suggested that the intervention had increased pupils' engagement and enjoyment of lessons. Improvements in behaviour were also noted though sometimes attributed to other developments in the school. Impacts on higher-achieving pupils were also observed..." (p.38)*
 - *"In addition, school staff identified positive improvements on pupils other than attainment, such as non-cognitive outcomes and changes in the school culture around formative assessment and dialogue." (p.39)*
- May **increase** post-16 progression through **higher longer-term pupil attainment**
 - This was noted throughout the report i.e. that embedding these practises are a longer-term process. Specifically:
 - *'...participants explained that the embedding of formative assessment principles and the improvement in pupils' approaches to learning was a long process. In addition, teachers frequently observed that the intervention had had more impact on younger pupils' (p.38)*

FSM Pupils

- May **increase FSM pupils'** post-16 progression through **higher pupil attainment**
 - Same as above
- May **increase FSM pupils'** post-16 progression through **better student engagement**
 - Same as above

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **higher pupil attainment**
 - *'teachers often argued that improvements in teaching practices and pupils' approaches to learning would take time to feed into higher pupil attainment. Regarding potential improvements to GCSE results, this was exacerbated by the perceived tendency of younger pupils to be more responsive to the intervention' (p.39)*

FSM Pupils

- May **increase FSM pupils'** participation in labour market/education/ training through **higher pupil attainment**

- Same as above

Any Additional Evidence [OPTIONAL]

- **NOTE:** *Given the repeated mentions of higher impact on younger students, perhaps a sub-group analysis can be done for this trial by age/ cohort*

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - Fidelity was highly variable & most schools adapted the programme but there are no direct implications for the hypotheses above
 - *"While schools achieved the broader aim of increasing the use of formative assessment and facilitating dialogue and sharing of practices, the actual implementation of the intervention varied significantly across case study schools. This led to variation in the nature of the intervention, often away from the programme specification. This occurred particularly in relation to the format and structure of TLCs and the use and frequency of peer observations" (p.34)*
3. Risk of Spillover (high/low): Low

Generation STEM Work Experience

Background info

- **Population:** KS4
- **Intervention:** Generation STEM is a work experience placement intervention for Year 10 students aiming to develop students' life skills and interest in STEM subjects (science, technology, engineering, and mathematics).
 - **Deliverer:** External Provider (CSW group or one of their delivery partners)
 - **Mode of delivery:** Nominated students with an interest in the internship
 - **Dosage:** 5-days of work experience (p.8)

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
KS4 Maths and Science: 0.04 [-0.07, 0.15]	+	<i>Not significant</i>
KS4 Maths and Science (FSM): 0.08 [-0.1, 0.26]	+	<i>Not significant</i>

- **Outcomes:** *copied from Key Conclusions (p.4 & p.62)*

- “Children in Generation STEM schools **made the equivalent of 0 months’ additional progress** in mathematics and science, on average, compared with children in other schools. This result has a high security rating.”
- “Children **eligible for free school meals in Gen STEM schools made the equivalent of one month’s additional progress in mathematics and science**, on average, compared with children in other schools. This finding should be interpreted with caution, as it is based on a small sub-group of pupils and there is considerable uncertainty around the result.”
- “There was **no evidence** in the secondary outcome data that Gen STEM had an **impact on students’ attendance in Year 11, uptake of STEM-related subjects at A-level, ability to make future decisions, or attitudes to STEM, life and school.**”
- **Any mention of longer-term outcomes:**
 - “All young people who undertook the work experience believed it had confirmed or expanded their interest in STEM and increased their confidence. Additionally, most of the young people felt that the overall experience confirmed their career interests or intentions. Similarly, most employers viewed the quality of the Gen STEM work experience programme as good and believed it had increased young people’s confidence, provided insights into working life, and progressed their career decision-making.” (p.5)

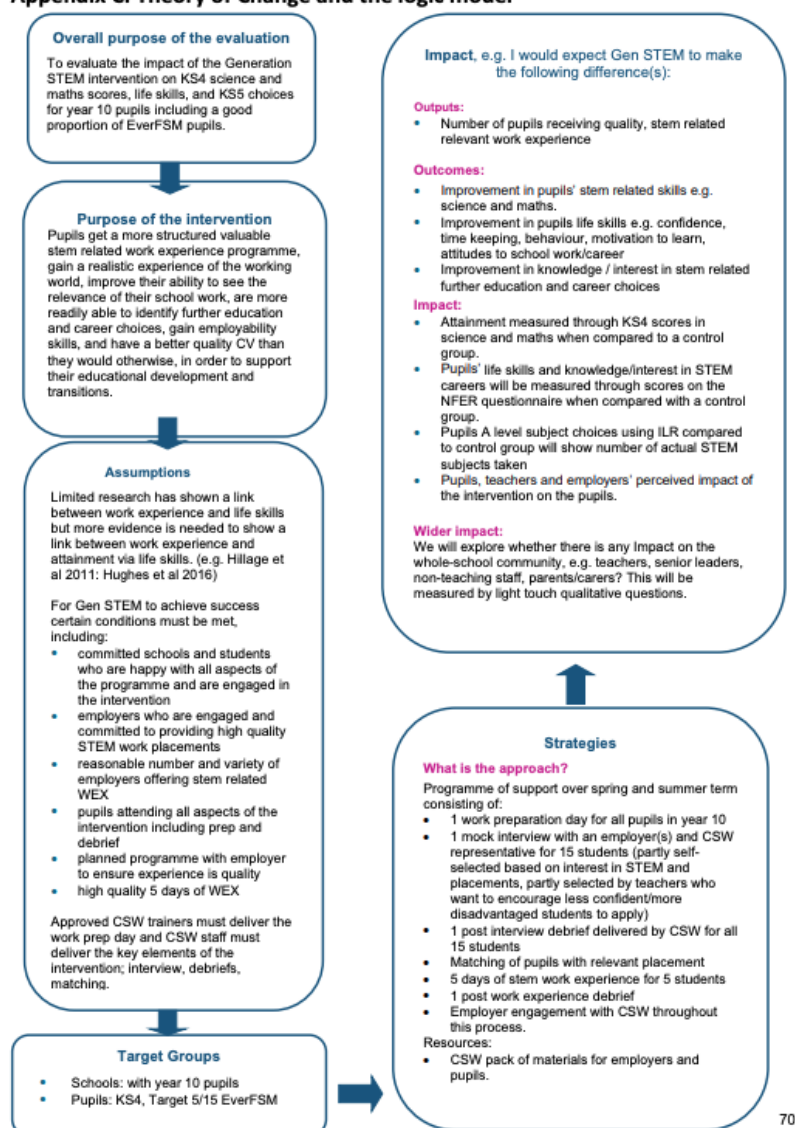
Resource list

- Efficacy trial ([EEF Page](#))
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant: N/A
- Suggested further research (*copied in from Background Evidence*):
 - Hillage et al. 2001
 - Hughes et al. 2016
 - Linnehan 2001
 - Golden et al. 2005
 - Hooley et al. 2014

Programme logic model/ theory of change (see p.70)

Evaluation Report

Appendix C: Theory of Change and the logic model



70

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes

% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **increased pupil confidence**
 - "It **increased young people's confidence**. For example, one employer described how the young person who carried out work experience with them was **very quiet initially** and did not talk much as she thought she did not want to work there. By the **end of the week**, she **had grown in confidence a lot and even did the presentation**...Another employer felt that their young person departed with '**increased confidence and communication skills** as a result of rising to the challenge of the presentation to 100 people" (p.54)*

FSM Pupils

- May **increase FSM pupils'** post-16 attainment through **increased pupil confidence**
 - Same as above

RQ2: Educational Progression

% of pupils progressing to KS5

- May **increase** post-16 progression through **improved pupil engagement with STEM subjects**
 - *"Employers reported that young people appeared to have developed, for example, 'an increased knowledge about the routes into engineering', and an insight into what they might want to do when they had previously been unsure about a future career. The work experience was also reported to challenge young people's thoughts about future jobs" (p.54)*
 - *"They had increased motivation—for example, pupils now see the connection between their GCSEs and work (two staff members)" (p.55)*

FSM Pupils

- May **increase FSM pupils'** post-16 progression through **improved pupil engagement with STEM subjects**
 - Same as above
 - Supported by the finding that effect size was larger (albeit non-significant) for FSM pupils (1 month of progress, $p=0.31$)

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **increased pupil confidence/exposure to working life**
 - *'The majority of employers viewed the quality of the Gen STEM work experience programme to be good and believed it had a positive impact on young people through: increasing young people's confidence; providing an insight into working life; and progressing their career decision-making.'* (p.45)
 - *"All of the young people who undertook the work experience believed that it had confirmed or expanded their interest in STEM and had increased their self-confidence. Additionally, most of the young people felt that the experience confirmed their career interests or intentions."* (p.46)
- May **increase** participation in labour market/education/ training through **improving pupil's employability skills (interviews/ recruitment)**
 - *"They noted that their pupils had gained valuable interview skills and experience"* (p.45)

FSM Pupils

- May **increase FSM pupils'** participation in labour market/education/ training through **increased pupil exposure to STEM careers**
 - Same as above
 - *"It provided the opportunity for a diverse group of pupils with different characteristics to participate in the programme (one staff member)."* (p.53)
 - Not mentioned in the report, but it is conceivable that FSM pupils have lower exposure to such opportunities/ careers and as such, this may benefit FSM pupils even more.
- May **increase FSM pupils'** participation in labour market/education/ training through **improving pupil's employability skills (interviews/ recruitment)**

- Same as above

Any Additional Evidence [OPTIONAL]

N/A

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - "Of the analysed sample, **21% of intervention pupils fully complied according to the above criteria**. This was higher than the cut-off point specified in the SAP (10%). Therefore, both compliance measures were used in two separate CACE (complier average causal effect) analyses." (p.22)
 - "The process of **selecting young people for the trial varied between intervention schools** (for example, in terms of FSM pupils)." (p.9)
 - "**Low implementation fidelity** makes it harder to accurately estimate the size of the impact on pupils in the trial." (p. 5)
 - "**Inconsistent application of elements of the logic model** calls into question whether the lack of impact seen was due to the mechanisms of the programme per se or its implementation." (p. 62)

3. Risk of Spillover (high/low): Low

Maths in Context

Background info

- **Population:** KS4 (Year 10)
- **Intervention:** Lessons focused on a specific area of mathematics in the context of financial capability. Seeks to improve children's financial capability, and specifically their financial knowledge and understanding, applied numeracy and problem-solving skills. (p.6-7)
 - **Deliverer:** Teacher
 - **Mode of delivery:** Full Classroom
 - **Dosage:** 10-12 lessons over 1 year

Impact and Outcomes

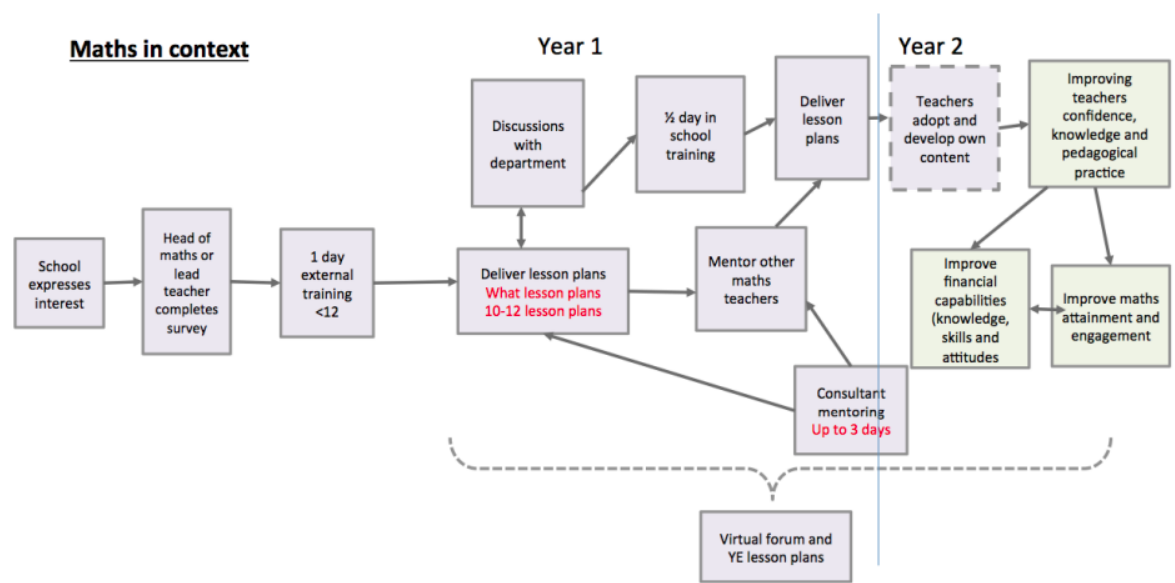
Outcome Measure / Finding	Direction	Significance
GCSE Maths scores: 0.02 [-0.10, 0.14]	+	<i>Not significant</i>
GCSE Math (FSM): 0.03 [-0.10, 0.17]	+	<i>Not significant</i>

- **Outcomes:** copied from Key Conclusions (p.4 & p.46)
 - “Pupils in the MiC schools made **no additional progress in Mathematics attainment compared to pupils in the other schools.** This result has a high security rating.”
 - “Pupils in the MiC schools made **no additional progress on questions in GCSE Mathematics involving problem-solving or in contexts involving money, or in financial literacy, compared to pupils in the other schools.**”
 - “Teachers showed a high level of support for the MiC intervention. Eighty-one percent reported that they believed the intervention improved pupils’ understanding and learning of financial literacy. Close to half of lead teachers felt that lessons had improved pupils’ attitudes towards Mathematics.”
 - “The MiC intervention was carried out with **high levels of compliance in terms of delivery and high fidelity in terms of the lesson aims and design.**”
- **Any mention of longer-term outcomes:**
 - “There were similarly high levels of agreement or strong agreement reported by respondents of their awareness of how financial capability can affect young people’s future well-being (lead teachers 95%, teachers 95%, senior leaders: 98%)” (p.35)

Resource list

- Efficacy trial ([EEF Page](#)):
- [Final Report](#)
- Effectiveness trial (EEF page):
- Any cited studies that might be relevant:
 - “The intervention is based on an adapted form of an earlier project funded by the London Schools Excellence Fund (PFEF, 2015), which was developed by the charity Young Enterprise” (p.6)
- Suggested further research:
 - N/A

Programme logic model/ theory of change (see p.7 of Evaluation Protocol)



Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **more positive pupil attitudes towards Mathematics**
 - Some evidence for this, although it is mixed
 - *“Compared to teachers’ strong support for the programme in terms of financial education, they were notably less convinced that the intervention had a positive impact on Mathematics learning. Although there **was some evidence that lead teachers (47%) and teachers (21%) agreed with the statement that pupils’ attitudes towards Mathematics had improved (see Figure 10),** only 15% of lead teachers and 9% of teachers agreed or strongly agreed that their perception was that Mathematics attainment had improved as a result of the programme. Most, 78% of lead teachers and 60% of teachers, neither agreed nor disagreed that this was the case” (p.42)*

FSM Pupils

- May **increase FSM pupils’** post-16 attainment through **more positive pupil attitudes towards Mathematics**
 - Same as above

RQ2: Educational Progression
% of pupils progressing to KS5

- May **increase** post-16 progression through **improved pupil engagement**
 - *“Because of the focus of the lessons on meaningful money contexts, teachers found ...**considerable interest and often enthusiasm in their pupils.** Some teachers reported this as a challenge, which they also reported they enjoyed,...lead teacher pointed to a particular way in which the **lessons helped them motivate their pupils in a particular lesson.**” (p.39)*
 - *“Although small in number, these observations suggested a relatively high level of pupil engagement and interest in financial literacy issues. In response to this interest, teachers were both observed and self-reported that they spent time in lessons explaining financial processes to pupils, often providing insights from their own personal experiences which pupils were keen to hear” (p.39)*

FSM Pupils

- May **increase FSM pupils’** post-16 progression through **improved pupil engagement**
 - Same as above

RQ3: Labour Market Activities
of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **increasing pupil financial literacy**
 - *“Improvement in pupils’ knowledge of financial literacy (agreement or strong agreement from 93% of lead teachers and 77% of teachers).” (p.41)*
 - *“an increase in pupils’ motivation to learn about financial aspects of life with 75% of lead teachers agreeing or strongly agreeing that this was the case*

(although the teachers involved in the intervention were less convinced, with only 45% agreeing or strongly agreeing and 42% neither agreeing or disagreeing)" (p.41-42)

FSM Pupils

- May **increase FSM pupils'** participation in labour market/education/ training through **increasing pupil financial literacy**
 - Same as above

Any Additional Evidence [OPTIONAL]

- There are no FSM specific details found in IPE but some notes on the differences between high attaining & low attaining pupils:
 - "For example, one lead teacher discussed how they had provided more structure to the materials to better cater to lower attaining pupils." (p.40)
 - "Through an interview one lead teacher provided insight into two ways (one mathematical and one in relation to providing structured support) they were using to support pupils in their lower attaining sets" (p.40)

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)

Compliance / Overall fidelity were high

3. Risk of Spillover (high/low): Low

Increasing Pupil Motivation

Background info

- **Population:** KS4
- **Intervention:** Financial and non-financial (event/trip) incentives for Year 11 pupils to complete effort tasks related to attendance, behaviour, classwork and homework & feedback letters detailing performance (p. 7)
 - **Deliverer:** Teacher/ School
 - **Mode of delivery:** Full classroom
 - **Dosage:** Each half-term

Impact and Outcomes

Finding	Direction	Significance
Financial Incentives		

GCSE Maths: 0.037 [-0.06, 0.13]	+	<i>Not significant</i>
GCSE English: 0.022 [-0.08, 0.12]	+	<i>Not significant</i>
GCSE Science: -0.058 [-0.27, 0.16]	-	<i>Not significant</i>
GCSE Maths (FSM): 0.062	+	<i>Not significant</i>
GCSE English (FSM): 0.030	+	<i>Not significant</i>
GCSE Science (FSM): -0.014	-	<i>Not significant</i>
Classwork Effort (2nd term out of 4): 0.176	+	<i>p<0.05 (exact value not reported); see Table A10</i>
Classwork Effort (FSM): Range 0.025 - 0.122 (4 terms)	+	<i>Not significant</i>
Event Incentives		
GCSE Maths: 0.084 [-0.01, 0.17]	+	<i>Not significant</i>
GCSE English: 0.042 [-0.08, 0.16]	+	<i>Not significant</i>
GCSE Science: -0.063 [-0.25, 0.13]	-	<i>Not significant</i>
GCSE Maths (FSM): 0.091	+	<i>Not significant</i>
GCSE English (FSM): 0.076	+	<i>Not significant</i>
GCSE Science(FSM): -0.094	+	<i>Not significant</i>
Classwork Effort: 0.028 - 0.122	+	<i>Not significant</i>
Classwork Effort (FSM): 0.004 - 0.122	+	<i>Not significant</i>

- **Outcomes:** copied from Key Conclusions (p.4 & p.46)
 - “Event Incentives – There is no evidence of a significant positive impact of event incentives on GCSE attainment in Maths, English or Science.”
 - “Financial Incentives – There is no evidence of a significant positive impact of financial incentives on GCSE attainment in Maths, English or Science.”
 - “There is a **statistically significant improvement in classwork effort** across English, Maths and Science for the **financial incentive treatment**. There is no evidence of impact on behaviour, attendance or homework effort. **This may suggest that even when there is a marked improvement in effort in classwork, this does not translate into higher GCSE attainment.**”
 - “There was a **positive impact of both the event and financial incentives on GCSE Maths for pupils with low levels of prior attainment** – equivalent to about one quarter of a GCSE grade in Maths, although this is not statistically significant for the financial incentive treatment.”
 - “Schools found it difficult to organise and pay for events before they knew how many pupils were likely to meet their targets. Schools should also consider the cost of monitoring and providing feedback about pupil effort.”

- **Any mention of longer-term outcomes:**
 - *“Further research should explore ...the long-term impact of such schemes. ...future studies **should explore why incentives appear to change classwork effort but do not necessarily translate into higher attainment.**” (p.4 and p.46)*
 - *“Previous literature demonstrates some positive impact ... It is **unclear whether these short-term impacts affect long-term outcomes, however, such as progression to further education or employment.** Future research should therefore include the scope for assessing the long-term implications of providing incentives to pupils, which could be positive if increased attainment leads to higher outcomes or **negative if the provision of incentives permanently displaced some intrinsic motivation for effort.**” (p.37)*

Resource list

- [Efficacy trial \(EEF Page\)](#)
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant: N/A
- Suggested further research:
 - The report covers the literature on financial incentives for students quite well but does not mention anything on long-term outcomes. Useful for if we need references to motivations, particularly differences between intrinsic and extrinsic (p.7).

Programme logic model/ theory of change (diagram or equivalent in-text)

Not found in text or diagram but the literature review suggests the mechanism is via **increasing extrinsic motivation** (p.7).

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **increased pupils' classwork effort**
 - While there was no significant effect on attainment in this study, there was evidence of increased effort towards classwork (p.52). In the long run, students' increased effort may lead to more positive post-16 outcomes
 - *Worth noting that the report does mention that this doesn't translate to attainment increase in the short-term. Although coefficients are positive, it may be worth exploring whether this effect is delayed through classwork effort*

FSM Pupils

- May **increase FSM pupils' post-16 attainment** through increasing **pupil's classwork effort**
 - There was a significant positive impact observed for all pupils' classwork effort (p.52, Table A10). Subgroup analysis showed that the coefficient (not effect size) was approximately 20% larger for pupils eligible for FSM meals. However, this effect was only for one term.

RQ2: Educational Progression
% of pupils progressing to KS5

- May **increase** post-16 progression through increased **pupil motivation**
 - Motivation is not explicitly measured in this study but it is the hypothesised mechanism for increased attainment. Specifically, extrinsic rewards are suggested to increase extrinsic motivation. If this continued, it could increase longer-term outcomes: *“Despite the lack of funding, they report that 5 treatment schools were planning to continue with the incentives.”* (p.35)

FSM Pupils

- May **increase FSM pupils’** post-16 progression through increased **pupil motivation**
- May **increase FSM pupils’** post-16 progression through increased **pupil effort**
 - The suggested mechanisms are the same as above for all students. Given some differential impacts on effort (p.52) for FSM pupils, the effect may be stronger in the longer-term for FSM pupils too.

RQ3: Labour Market Activities

of months not in employment, education or training from August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase participation** in labour market through **increasing pupils motivation**
 - Given the experience of receiving financial rewards as a form of extrinsic motivation may not be familiar to kids, such an intervention could serve to motivate students to participate in the labour market to reap similar rewards.
 - *“Previous literature demonstrates some positive impact ... It is unclear whether these short-term impacts affect long-term outcomes, however, such as progression to further education or employment.”* (p.37)

FSM Pupils

- May **increase FSM pupils’** participation in labour market through **increasing pupils motivation**
 - Same as above.

Any Additional Evidence [OPTIONAL]

“However, for both sets of incentives, there is a more positive effect of incentives on Maths scores for pupils with low levels of prior attainment – equivalent to about one quarter of a grade in GCSE Maths, although this is only significant at the 10% level for the financial incentive treatment. Across English and Science, there is no evidence of an effect of incentives on pupil attainment and we can again rule out medium-to-large effects.” (p.36)

Quality of Implementation

1. Is the answer to any of the following a ‘yes’ in the Trial classifier (‘Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?

2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)

- 18 out of 33 control schools did not provide data on secondary outcomes i.e. pupil effort
 - This could have biased the results as the primary effect was on pupil effort but robustness checks were conducted to check this: *'To mitigate this potential problem, a robustness check was conducted to estimate the effect of both the financial and event incentive treatment relative to the subset of control schools that monitored pupil effort (those that submitted effort data to the project team)'* (p.6)
- No explicit mention of spillover due to design but schools offered incentives already. While this was accounted for by the team, it could have affected fidelity but an IPE was not conducted. Specifically:
 - *'Before the start of the intervention, a number of schools already offered existing incentives to pupils. The project team conducted a small-scale survey and found that 18 out of 33 schools who responded to the survey offered some kind of incentive (mostly trophies or trips). Across the treatment and control groups, this means we know that at least 27% of each treatment group offered some kind of incentive scheme and at least 30% of control schools did so Data on whether schools offered existing incentives was also used in the randomisation procedure (see later in this section) and there is therefore no difference across treatment or control groups in terms of whether schools offered existing incentives. The only potential concern is if control schools offered new incentives in light of the intervention or if treatment schools scaled down existing schemes in light of the new ones. Schools were explicitly told not to do this as part of the initial training and information event in July 2012, but the extent to which schools followed this guidance isn't known.'* (p.11)

3. Risk of Spillover (high/low): Low

Texting Parents

Background info

- **Population:** KS3, KS4
- **Intervention:** School-based, low-cost and low-effort process, involving text messages being sent directly by schools to parents of students in years 7, 9, and 11 at specific time-points and intervals during the school year (p.6)
 - **Deliverer:** Teacher/ School
 - **Mode of delivery:** Full classroom
 - **Dosage:** Approximately 30 texts/year

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
GCSE English Scores: 0.033 [0.004, 0.064]	+	$p=0.024$
GCSE English (FSM): -0.021 [-0.106, -0.065]	-	Not significant
GCSE Maths Scores: 0.067 [0.041, 0.092]	+	$p<0.001$

GCSE Maths Scores (FSM): -0.004 [-0.081, 0.073]	-	<i>Not significant</i>
GCSE Science Scores: -0.013 [-0.048, 0.022]	-	<i>Not significant</i>
GCSE Science Scores (FSM): -0.077 [-0.173, 0.18]	-	<i>Not significant</i>
Absenteeism: -0.054 [-0.079,-0.029]	+	<i>p<0.001</i>
Absenteeism (FSM): -0.064 [-0.149, 0.021]	+	<i>Not significant (p=0.052)</i>

- **Outcomes:** Copied from Key Conclusions (p.4 & p.44)
 - “Children who had the intervention experienced about **one month of additional progress in maths compared to other children**. This positive result is unlikely to have occurred by chance”
 - “Children who had the intervention had **reduced absenteeism compared to other children**. This positive result is unlikely to have occurred by chance.”
 - “Children who had the intervention appeared to experience about **one month of additional progress in English** compared to other children. However, analysis suggests that this finding **might have been affected by bias introduced by missing data**, so we cannot reliably draw this conclusion.”
 - “**Schools embraced the programme** and liked its immediacy and low cost.”
 - “The vast majority of parents were accepting of the programme, including the content, frequency, and timing of texts.”
- **Any mention of longer-term outcomes:**
 - “While there is evidence to suggest that school-based parental engagement can result in raising achievement, **the evidence relating to the long-term impact of such programmes on academic outcomes is limited and typically studies have not employed rigorous methodological designs...it is still necessary to understand and identify what parents want and need in order to help them to better engage and participate in their children’s learning experiences**” (p.8)

Resource list

- Efficacy trial ([EEF Page](#))
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant: N/A
 - Page 8, second last paragraph discusses long-term outcomes / previous studies / lack of evidence on this
- Suggested further research: N/A

Programme logic model/ theory of change (diagram or equivalent in-text)

Not Found

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **increased pupil attainment**
 - See outcomes table above and p.26-7 for full results in the report. Significant improvements in attainment is supported by an overall positive effect in the literature:

- *“The results of this evaluation are commensurate with other studies which have found that providing parents with regular, concise updates regarding their child’s progress, missing homework, and exam performance can improve academic attainment (Bergman, 2014; Kraft and Rogers, 2015; York and Loeb, 2014). However, the current trial provides evidence of only a very small positive effect and was unable to confirm the larger effect sizes” (p.45)*
- May **increase** post-16 attainment through **lowered pupil absenteeism**
 - See outcomes in table above; absenteeism was reduced by half a day.
 - *“Given that no absence-related texts were in fact sent, this is a particularly curious finding which is perhaps related to the increased monitoring of children’s school-related activities overall, creating an environment in which pupils felt less able or willing to truant. In particular, the intervention was more effective in reducing absenteeism in the KS4 cohort (ES = -0.11) compared to KS3 (ES = -0.01), and this **result is of particular note given that absenteeism increases as pupils progress through school, is more likely if parents do not value education, and is closely linked to lower academic attainment** (Taylor, 2012)” (p.46)*
- May **increase** post-16 attainment through **increased parent engagement**
 - *“Teacher quote: “On the whole, I’d say that it has had a positive influence because, certainly, my Year 9 students have come back and said, oh yeah, I had a conversation about Romeo when he tells Juliet he loves her for the first time. You know, I talked to my mum and dad about that. And I think that’s really pleasing to know that what you’re doing is reaching beyond the classroom and their parents are being made aware of it, but actually, not just aware of it, being involved in it.” (p.40)*
 - *“Positive consequences of the programme, which were widely mentioned, include the programme’s ability to develop and promote a three-way relationship between school/teacher, parents, and pupils. This was achieved by breaking down barriers between school/teacher and parents while stimulating communication between parents and pupils in the home environment” (p.40)*
 - *‘Student quote: “I think it just means that they partially understand what I’ve been doing so I don’t have to explain it all from like...it makes it easier to talk about what you’ve been doing at school.” (p.41)*

FSM Pupils

- May **increase FSM pupils’** post-16 attainment through **lowered pupil absenteeism**
 - Same as above (see effect on absenteeism holds for FSM pupils)
- May **increase FSM pupils’** post-16 attainment through **increased parent engagement**
 - Same as above

RQ2: Educational Progression **% of pupils progressing to KS5**

- May **increase** post-16 progression through **increased pupil attainment**
 - See outcomes table above and p.26-7 for full results in the report. Significant improvements in attainment is supported by an overall positive effect in the literature (quote is above for RQ1)
- May **increase** post-16 progression through **lowered pupil absenteeism**
 - See outcomes in table above; absenteeism was reduced by half a day.
 - *“Given that no absence-related texts were in fact sent, this is a particularly curious finding which is perhaps related to the increased monitoring of children’s school-related activities overall, creating an environment in which*

*pupils felt less able or willing to truant. In particular, the intervention was more effective in reducing absenteeism in the KS4 cohort (ES = -0.11) compared to KS3 (ES = -0.01), and this **result is of particular note given that absenteeism increases as pupils progress through school, is more likely if parents do not value education, and is closely linked to lower academic attainment** (Taylor, 2012)” (p.46)*

FSM Pupils

- May **increase FSM pupils’** post-16 progression through **lowered pupil absenteeism**
 - Same as above

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **higher pupil attainment**
 - See outcomes table above and p.26-7 for full results in the report. Significant improvements in attainment is supported by an overall positive effect in the literature (quote is above for RQ1)

FSM Pupils

Any Additional Evidence [OPTIONAL]

“It represents an opportunity to conduct a rigorous experimental evaluation that will enable us to better understand the causal mechanism between parental engagement (as defined by PEP) and pupil attainment. Furthermore, as the data from this trial will be linked to the National Pupil Database (NPD), and potentially to other relevant statutory datasets in the future, there is the potential to follow up participants to better understand the long term effects, if any.” (p.9)

Quality of Implementation

1. Is the answer to any of the following a ‘yes’ in the Trial classifier (‘Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were ‘yes’, please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - There were some issues in implementation of the programme but the effect on ToC is expected to be minor. The issues include:
 - sometimes a miscommunication/ inconsistency of text messages being sent to parents for e.g. 1 quote notes that one week 4 texts were sent while the next there were 0 (p.41-42)
 - and it sometimes encouraged parents to push students to revise when this was not relevant for them (p.41)

3. Risk of Spillover (high/low): Low

Texting Students and Study Supporters

Background info

- **Population:** KS5 (further education colleges)
- **Intervention:** Text messages sent to students re-sitting GCSE English or Maths, or sent to named study supporters, over the course of an academic year. (p.7)
 - **Deliverer:** Tutors
 - **Mode of delivery:** 1:1
 - **Dosage:** Approximately 36 weekly text messages / year

Impact and Outcomes

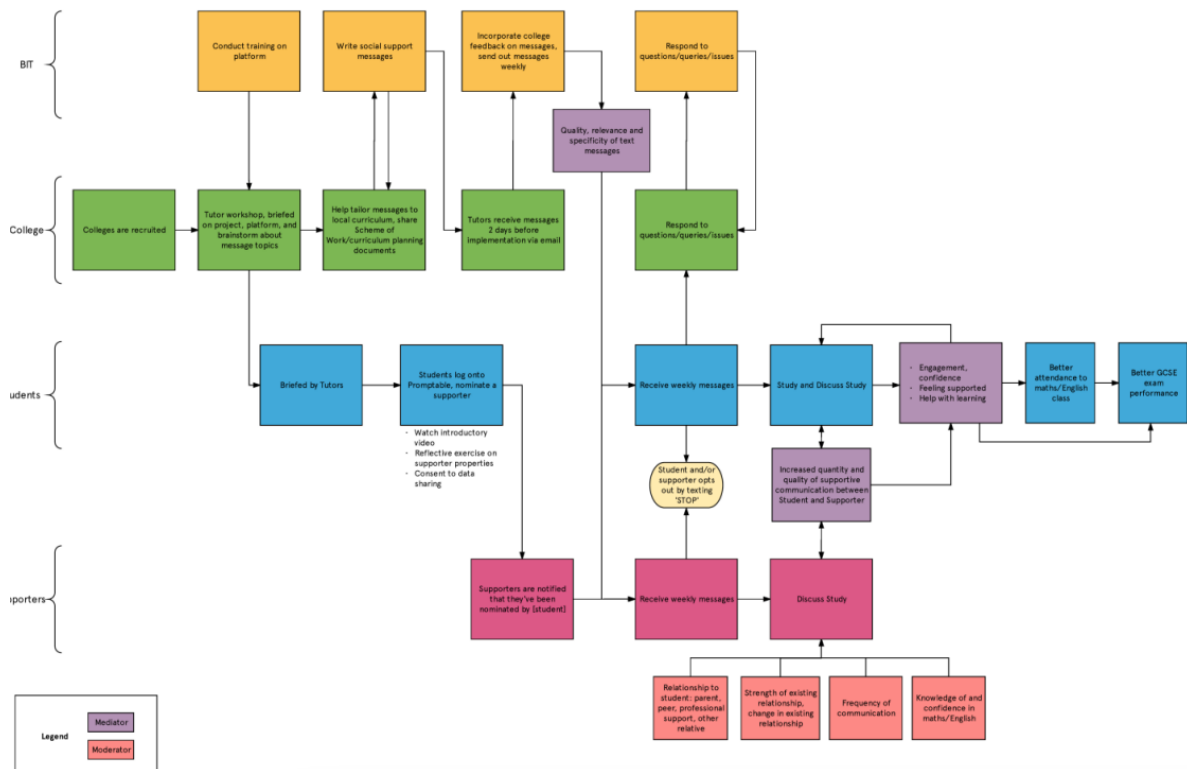
Outcome Measure / Finding	Direction	Significance
Treatment = Texts to Students Proportion achieving GCSE pass rate: 0.01 [-0.05;0.07]	+	<i>Not significant</i>
Treatment = Texts to Study supporters Proportion achieving GCSE pass rate: 0.01 [-0.05, 0.08]	+	<i>Not significant</i>
Treatment = Texts to both Proportion achieving GCSE pass rate: -0.03 [-0.09, 0.04]	-	<i>Not significant</i>
Treatment = All treatment arms (FSM) Proportion achieving GCSE pass rate: -0.03 [-0.12,0.08]	-	<i>Not significant</i>
Primary analysis indicates no effect on attendance (secondary outcome) for all intervention groups	N/A	<i>Not significant</i>

- Outcomes (see Key conclusions p.55):
 - “There is **no evidence that the Project Success intervention had any impact on the GCSE English or maths re-sit pass rate for further education college students.**”
 - “There is **no evidence that the Project Success intervention had any impact on the attendance of further education college students re-sitting GCSE English or maths.**”
 - “The **intervention did not have a differential impact on the GCSE re-sit pass rate by gender or by eligibility for free school meals** (at the end of KS4). The subject being re-examined or the number of re-sits being taken also did not lead to differential effects from the intervention.”
 - “There were significant limitations to the programme’s ability to engage those who may need it the most as it was the **highly motivated students** that were more engaged with their studies and with college generally who were **more likely to sign up to the intervention**”
- **Any mention of longer-term outcomes:**
 - Nothing found on this

Resource list

- Efficacy trial ([EEF Page](#)):
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant: N/A
- Suggested further research:
 - N/A

Programme logic model/ theory of change (see project protocol, p.17)



Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
 % of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **higher pupil attainment**
 - “It is possible that improvements in student attendance or overall academic performance might have been facilitated by the intervention; **project leads did report improvements in mock results where raw marks on a continuous scale were available as opposed to categorical grade boundaries used in the impact evaluation.** Project leads also reported changes in drop-out rates in some colleges, though they were reluctant to attribute this to the intervention without formal evaluation.” (p.56)
- May **increase** post-16 attainment through **increased pupil effort**
 - “According to students, **text messages sent either to them or their study supporters encouraged revision by functioning as reminders and planning tools and by providing useful web resources.** Motivational texts were perceived to be less useful. Interviews with students and college tutors, however, suggest that this increased engagement in revision outside of college did not translate into increased engagement in maths and English classes.” (p.55)

- "...**some study supporters reported that they helped students with their revision after receiving the texts.** 'Well, if he knew I had an exam coming up, he'd tell me. He'd come and help me. We'd sit at the table and do like a paper together. And we'd just do it together' (student, re-sitting GCSE maths)." (p.47)
- May **increase** post-16 attainment through **improving attendance**
 - "Importantly, the impact evaluation suggests that the intervention had no impact on college attendance in any of the intervention arms. **This contrasts with findings from two recent studies of similar texting interventions (Miller et al., 2016 and Groot et al., 2017) which both found evidence that such interventions may improve college attendance,** which is hypothesised to improve attainment. Kraft and Rogers (2015) argue that teacher-to-parent communication, such as texts, may improve college attendance"

FSM Pupils

- May **increase FSM pupils'** post-16 attainment through **higher pupil attainment**
 - Same as above; subgroup analysis indicated no differences (see table 17, p.37). Might **expect smaller effects for FSM students**:
 - "The pass rate for GCSE re-sits in English or maths was lower for pupils who had been eligible for free school meals by the end of KS4 relative to their peers, in both intervention and control, though this was not statistically significant. This reflects the known attainment gap between disadvantaged students and their peers. The subgroup analysis indicates that the intervention had no effect on pupils eligible for free school meals the end of KS4, or the preceding six years. Furthermore, the intervention effect did not differ for these pupils relative to their peers" (p.37)
- May **increase FSM pupils'** post-16 attainment through **improving attendance**
 - Same as above
- May **increase FSM pupils'** post-16 attainment through **increased pupil effort**
 - Same as above

RQ2: Educational Progression
 % of pupils progressing to KS5

RQ3: Labour Market Activities
 # of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

Any Additional Evidence [OPTIONAL]

- Given the trial recruited students on an opt-in basis; survey results at sign up suggest the sample was positively skewed by motivated / engaged students. Based on self reported measures, "students who signed up to receive BIT text messages were significantly more likely to:
 - believe if they put in enough effort, they could succeed in their English GCSEs (78% agreeing or strongly agreeing compared to 69%) and maths GCSEs (75% compared to 68%);
 - state that if they had different teachers, they would try harder in their English GCSEs (29% agreeing or strongly agreeing compared to 25%) and maths GCSEs (28% compared to 24%);

- *think if they wanted to, they could do well in their English GCSEs (74% agreeing or strongly agreeing compared to 64%) and maths GCSEs (70% compared to 62%);*
- *report they work hard at college (80% agreeing or strongly agreeing compared to 72%) and complete their college work regularly (81% compared to 75%)” (p.40)*

Quality of Implementation

1. Is the answer to any of the following a ‘yes’ in the Trial classifier (‘*Robustness & Relevance sheet, columns S to V*’)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?

2. If any of the answers above were ‘yes’, please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)

There were issues of adherence and fidelity which may have affected the trial’s ability to detect an impact:

- Only 37% of students/supporters received all intended texts, with particularly low adherence in the "students & supporters" arm (p.34-35)
- 3 issues of fidelity identified, discussed on p.45:
 - *“was intended that all students and/or their study supporters would receive texts for the duration of the intervention*
 - *period. However, as a few students had changed their mobile number and did not inform BIT, they stopped receiving texts”*
 - *“ The majority of project leads reported planning the texts according to the scheduled plan of work for their college. However, project and subject leads commented that it was difficult to schedule texts for the whole term, especially because the schedule of work may change. They expressed a preference to prepare messages for each half-term.”*
 - Contamination (details below)

3. Risk of Spillover (high/low): High

“Therefore, it is possible a student who was assigned to the control group was receiving text messages as a study supporter, which would have affected their validity as a member of the control group. There was also a risk of students passing on content from the text messages they received themselves or from interaction with their study supporter to students in the control group. This may have affected outcomes for the control group and therefore the evaluation’s ability to detect an impact.” (p.36)

The RISE Project

Background info

- **Population:** KS4
- **Intervention:** Equip a senior teacher in the role of research advocate (Research Leads), to help teachers at their school to use educational research to develop their pedagogy It had a package of training to the Research Leads and included on-going

support via a newsletter, website, a network of peers, and school visits by RISE team. (p.6)

- **Deliverer:** External Provider
- **Mode of delivery:** Full classroom
- **Dosage:** 8 CPD sessions & termly follow-up meetings over 30 months

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
Mathematics Year 1: 0.09 [-0.03, 0.20] Mathematics Year 2: 0.04 [-0.09, 0.16]	+	<i>Not significant</i>
English Year 1: 0.05 [-0.07, 0.17] English Year 2: 0.03 [-0.11, 0.18]	+	<i>Not significant</i>
Maths (FSM) Year 1: 0.05 [-0.10, 0.21]	+	<i>Not significant</i>
English (FSM) Year 1: -0.02[-0.20,0.16]	-	<i>Not significant</i>

- **Outcomes:** copied from Key conclusions (p.3 & p.49):
 - “For both the one-year and two-year cohorts, children in RISE schools made **a small amount of additional progress in mathematics and English compared to children in the comparison schools**. This result has a high security rating. The differences were small and not statistically significant. This means that the statistical evidence does not meet the threshold set by the evaluator to conclude that the true impact was not zero.”
 - “There was **no evidence that RISE had an impact on the outcomes of pupils eligible for free school meals**.”
 - “Uptake by schools and attendance at the intervention training was high and **sustained over the 30 months of the intervention**.”
- **Any mention of longer-term outcomes:**
 - “Finally, taking a longer view of the time that is required for changing culture and teaching practice should be taken into account. This intervention was **conducted over a relatively short time, but had great ambitions**. A longer time scale would provide greater opportunity for embedding the aims of the programme at school level.” (p.47)
 - “Other studies of training for research use have had longer intervention timescales just for changing the practice in the public sector, for example the four year B-CURE programme [Approaches to Build Capacity to Use Research Evidence] which worked with government stakeholders to make effective use of research in policy making decisions. The B-CURE evaluation authors concluded that even after four years of intervention, it was too soon to begin to measure an impact” (p.50)

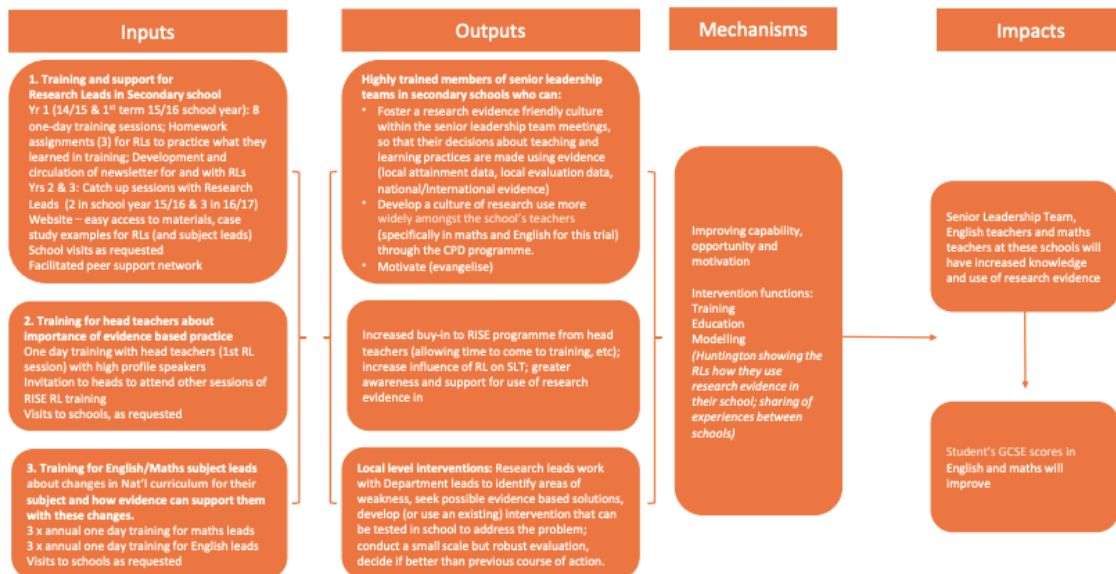
Resource list

- Efficacy trial ([EEF Page](#)):
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant: N/A
- Suggested further research:
 - Introduction discusses some other related research around evidence use in schools and cites 3 other trials funded by the EEF on this topic (p.12)

- Langer et al. is frequently cited; their framework was used in the assessment of this trial

Programme logic model/ theory of change (see p.11)

Figure 2. RISE Logic Model, based on developers' views



Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes

% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **improved teaching quality**
 - While there was no significant impact on attainment, there is evidence that teachers felt positively about the programme:
 - *"Without doubt the best CPD I've ever had. It has transformed teaching and learning in our school."* (Research Lead) (p.41)
 - *"I have found this year a great experience, in terms of my own knowledge and personal development. It has been extremely well organised, and it has been fantastic meeting new people."* (Research Lead) (p.41)
 - *"The English sessions run by Alex were highly informative and we were given useful resources. The research was English specific and easy to share with colleagues."* (Department Lead) (p. 42)

FSM Pupils

- May **increase FSM pupils'** post-16 attainment through **improved teaching quality**
 - Same as above
 - No indication that the programme focused more / less on attainment of FSM pupils & subgroup analysis finds no differences (p.49). Report also outlines that data on FSM pupils may not be as robust / has higher missingness (p.37)

RQ2: Educational Progression

% of pupils progressing to KS5

- May **increase** post-16 progression through **improved school practices**

- *“These included raising their awareness and positive attitude towards evidence-informed decision making in their school; understanding around the types of questions that needed to be answered using evidence; access (in different forms) to quality research evidence; the skills to both access and interpret the evidence they needed. Some Research Leads found being involved in RISE also gave them the structure and process to use the evidence to enact change within their school – the final mechanism in the Langer model.” (p.42)*

FSM Pupils

- **May increase FSM pupils’ post-16 progression through improved school practices**
 - Same as above

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- **May increase participation in labour market/education/ training through improved school practices**
 - This would likely be highly variable though as schools noted it would be difficult to continue without the support of the RISE programme:
 - *“By the end of the initial training, this group also felt that they had learned a great deal and had been given the appropriate skills for understanding the evidence and identifying areas for school improvement.” (p.41)*
 - *“For this subgroup, they found the RISE intervention was transformative at school level, as they were able to use the knowledge and skills from training and the leverage of being part of the project to springboard change in their schools.” (p.42)*

FSM Pupils

- **May increase FSM pupils’ participation in labour market/education/ training through improved school practices**
 - Same as above

Any Additional Evidence [OPTIONAL]

N/A

Quality of Implementation

1. Is the answer to any of the following a ‘yes’ in the Trial classifier (*‘Robustness & Relevance sheet, columns S to V’*)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were ‘yes’, please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)

- No major issues but evidence that there is highly variable implementation across schools. Analysis may need to take this into account.
 - *“Schools’ adoption of the research-informed school improvement model was highly variable and influenced by schools’ context and relationships, and the stability of the Research Lead role. The teacher in the role of Research Lead changed in 40% of the schools during the project.”* (p.3 & p.49)

3. Risk of Spillover (high/low): Low

Tutor Trust

Background info

- **Population:** KS3, KS4
- **Intervention:** University students/recent graduates were recruited and trained by Tutor Trust to deliver one-on-one or small group mathematics/English tutoring sessions to years 9,10,11 (p.6-7)
 - **Deliverer:** Tutor
 - **Mode of delivery:** Varied (Table 4, p.18 shows this)
 - **Dosage:** Variable but approx 8-10 hours in each phase (see p.17)

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
Mathematics GCSE: 0.051 [0.001,0.101]	+	$p<0.05$
Mathematics GCSE (FSM): 0.090 [-0.008, 0.188]	+	<i>Not significant</i>
English GCSE: -0.161 [-0.291, -0.03]	-	$p<0.05$
English GCSE (FSM): -0.026 [-0.219,0.168]	-	<i>Not significant</i>

- **Outcomes:** copied from Key Conclusions (p.3 & p.35)
 - *“Participating pupils achieved slightly **higher mathematics GCSE** scores than pupils in the comparison group, and **lower English GCSE** scores than pupils in the comparison group. However, it is **not possible to attribute either change to the tuition provided.**”*
 - The lack of confidence in the impact evaluation is due to lower security (see *Quality of Implementation*)
 - *“Schools involved in the qualitative interviews were **positive about the tuition**, keen to work with the Tutor Trust again, and **largely confident that the tuition was beneficial for their pupils.**”*
 - *“It is **recommended that the Tutor Trust continues to monitor tutor performance and identifies mechanisms to increase the consistency of tuition.**”*
- **Any mention of longer-term outcomes:**
 - *“**Pupil aspirations:** Some pupils and tutors reported that the tutoring had helped to raise pupils’ academic aspirations by contributing to improvements in their own knowledge and skills and through opportunities to discuss ‘university life’ with a current or recently graduated university student.”* (p.31)
 - *“...one teacher observed that while his pupils were well behaved with their tutor, they continued to ‘play up’ in class. He remarked that it would be*

interesting to see if, in the longer term, this improved behaviour with the tutor would be carried over into the classroom.” (p.30)

Resource list

- Efficacy trial (EEF Page): N/A
- [Final Report](#)
- Effectiveness trial ([EEF page](#))
- Any cited studies that might be relevant:
 - “...a small number of US studies have found that programmes led by college students can lead to improvements in student outcomes (Elbaum et al., 2000), suggesting that this is an area worthy of further evaluation in the UK” (p.5)
- Suggested further research: N/A

Programme logic model/ theory of change

Not Found in report

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **higher pupil attainment in Mathematics**
 - The intervention increased mathematics attainment for students (see table above).
 - “Most interviewees reported that the tuition had helped to develop pupils’ understanding of different topics within the relevant GCSE syllabi and had contributed to developing their skills and examination technique.” (p.30)
- May **decrease** post-16 attainment through **lower pupil attainment in English**
 - Not much evidence in the process evaluation but significant negative effect observed
- May **increase** post-16 attainment through **higher pupil confidence**
 - “There was general agreement that the programme had had a positive effect on pupils’ confidence in both mathematics and English. Pupils reported gaining confidence in their abilities: ‘It’s made me more confident. I’ve got a better overall understanding of the different parts of English and mathematics.’” (p.30)
- May **increase post-16** attainment through **higher pupil enjoyment**
 - “Some pupils in earlier rounds of tuition reported they were enjoying their mathematics and English classes more: ‘I now love mathematics!’ However, for most pupils, there was little evidence to suggest the tuition had raised their interest in or made them more enthusiastic about mathematics and English beyond their tuition session” (p.30)

FSM Pupils

- May **increase FSM pupils’** post-16 attainment through **higher pupil attainment in Mathematics**
 - The intervention had a significant positive effect on mathematics attainment and this held for FSM pupils but the finding was not significant
- May **decrease FSM pupils’** post-16 attainment through **lower pupil attainment in English**

- Negative effect held for FSM pupils but not significant
- May **increase FSM pupils'** post-16 attainment through **higher pupil confidence**
 - Same as above
- May **increase FSM pupils' post-16** attainment through **higher pupil enjoyment**
 - Same as above

RQ2: Educational Progression
 % of pupils progressing to KS5

- May **increase** post-16 progression through **raising pupil's aspirations**
 - *"The findings from the process evaluation suggest the Tutor Trust tuition was perceived to impact most successfully on 'softer' outcomes such as pupils' confidence in mathematics and English, and on their aspirations to continue on to further study." (p.36)*

FSM Pupils

- May **increase FSM pupils'** post-16 progression through **raising pupil's aspirations**
 - Same as above

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **raising pupil's aspirations**
 - *"Pupil aspirations: Some pupils and tutors reported that the tutoring had helped to raise pupils' academic aspirations by contributing to improvements in their own knowledge and skills and through opportunities to discuss 'university life' with a current or recently graduated university student." (p.31)*

FSM Pupils

- May **increase FSM pupils'** participation in labour market/education/ training through **exposure to university students / discussions of future with them**
 - Same as above

Any Additional Evidence [OPTIONAL]

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - No mention of low fidelity, high flexibility to the programme delivery overall hence it was a bit difficult to measure. Details on p.34, see section:
"The Tutor Trust model of tutoring is flexible ...rather more fluid than was anticipated"

- “Due to the limitations of the study design and the **absence of a high-quality comparison group**, this evaluation has not provided a secure estimate of the impact of the project on academic outcomes.” (p.3 and p.35)
- “...in a deviation from the original planned analysis, we were **unable to control a number of potentially influential variables ... pupil-level special educational needs (SEN); English as an additional language (EAL); ethnicity; and looked after children (LAC) status**.... excluded from the final analysis due to a change in the level of consent required by the NPD in order to release this data, after we had completed our data collection in schools. Failing to account for important variables (known and/or unobserved) has been shown to result in spurious effects, certainly those of a magnitude we see here (Coe, 2009).” (p.35)

3. Risk of Spillover (high/low): Low

Best Practice in Setting

Background info

- **Population:** KS3
- **Intervention:** Trained schools in a best practice approach to grouping Year 7 and 8 students in English and mathematics by attainment (p.6)
 - **Deliverer:** Teachers
 - **Mode of delivery:** Full classroom (all of year 7 and 8)
 - **Dosage:** 4 teacher workshops across 2 years

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
GL Assessment in Maths: -0.01 [-0.11, 0.10]	-	Not significant
GL Assessment in English: -0.08 [-0.21, 0.05]	-	Not significant
GL Assessment in Maths (FSM): 0.02 [-0.14, 0.18]	+	Not significant
GL Assessment in English (FSM): -0.01 [-0.25, 0.22]	-	Not significant
Maths Self-Confidence: 0.02 [-0.08, 0.13]	+	Not significant
English Self-Confidence: 0.09 [-0.08, 0.25]	+	Not significant

- **Outcomes:** copied from Key Conclusions (p.4 & p.52)
 - “The project found **no evidence that the Best Practice in Setting intervention improves maths or English attainment for children in Years 7 and 8**. For English, pupils in the intervention schools made slightly less progress than the control pupils, but this finding has no meaning given its level of uncertainty and very low security rating.”
 - “There was **no evidence of impact on pupils’ self-confidence in maths**. For English there was a small positive difference in pupil self-confidence, but this result has no meaning due to its level of uncertainty and the large amount of measurement attrition.”
 - “The process evaluation **revealed mixed views from participants**. While some were largely positive about the intervention, **many interviewees**

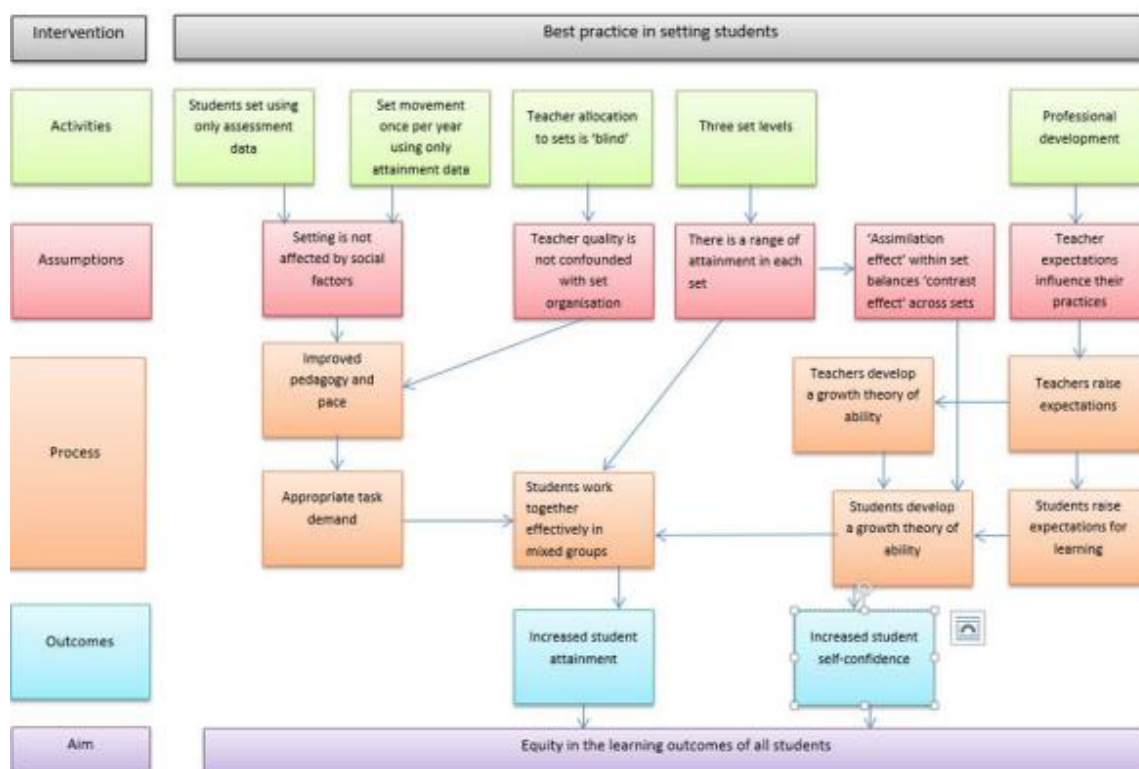
thought that what they were being asked to do represented little change from what they already do. Some schools reported that the intervention was onerous and difficult to deliver.”

- Any mention of longer-term outcomes: N/A

Resource list

- Efficacy trial ([EEF Page](#))
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant: N/A
- Suggested further research: N/A

Programme logic model/ theory of change (copied in from p.83, Appendix E)



Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **increased pupil confidence**
 - Table 12 on p.36 shows that there is a slight increase in pupil confidence in both subjects, although this result is not statistically significant. There are further notes of how students (particularly the top and bottom performers) benefitted from having higher aspirations
 - *“In one school, the Heads of English and mathematics both reported that that their ‘low’ and ‘top’ set Year 7 classes had particularly benefited from the intervention, but that this appeared to have been at the expense of the middle set, who ‘lacked the aspiration to do better’” (p.50)*

- May **increase** post-16 attainment through **better teaching quality**
 - *“Most interviewees appeared to find the opportunities to interact with teachers from other schools, to exchange information, and to develop professional or social contacts to be the most valuable part of the training events. This was particularly the case where staff were teaching the same subject but in different schools” (p.44)*

FSM Pupils

- May **increase FSM pupils’** post-16 attainment through **increased pupil confidence**
 - Same as above.
- May **increase FSM pupils’** post-16 attainment through **better teaching quality**
 - Same as above

RQ2: Educational Progression % of pupils progressing to KS5

- May **increase** post-16 progression through **better teaching quality**
 - *“These impacts included: staff being more proactive in checking that their students were in the most appropriate groups; students feeling ‘more settled’ in their groups; the pace of lessons being more appropriate to the academic attainment of the group; and staff applying more effective differentiation strategies.” (p.50)*

FSM Pupils

- May **increase FSM pupils’** post-16 progression through **better teaching quality**
 - Same as above

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

Any Additional Evidence [OPTIONAL]

N/A

Quality of Implementation

1. Is the answer to any of the following a ‘yes’ in the Trial classifier (‘Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were ‘yes’, please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - There were significant issues in the implementation fidelity - *“Due to the low fidelity programme implementation by schools we cannot conclude that the underlying idea – optimising practices associated with setting to ameliorate poorer outcomes for those in lower sets – does not work, ... findings of this evaluation indicate that more attention needs to be paid to school requirements to inform future research” (p.5)*
 - Low security rating of 2

- “Thirty (out of 61) maths intervention schools and 27 (out of 43) English intervention schools dropped out of delivering the programme. Attendance at workshops reduced dramatically during the course of intervention delivery.” (p.52)
 - Attrition resulted in insufficient power to detect an effect
- “School and teacher buy-in was low. Half of the schools in the maths trial and more than half of the schools in the English trial ceased intervention delivery before follow-up. ... treatment attrition, being able to engage schools with the programme was demonstrated to be at least as important as the programme itself.” (p.4 & p.51)
- “Some schools struggled to implement the intervention in full (for example, there was poor fidelity to teachers being randomly allocated to sets), this seemed to be due to schools finding it difficult to make some changes and therefore adapting principles to make them more implementable.” (p.4 & p.51)

3. Risk of Spillover (high/low): Low

Catch up Literacy (2014)

Background info

- **Population:** KS2, KS3
- **Intervention:** Structured literacy intervention for pupils between 6-14 years old who are struggling to learn to read. The intervention matches books to pupils according to their reading ability, which pupils then read to a teaching assistant (TA). (p.4)
 - **Deliverer:** TA
 - **Mode of delivery:** 1:1
 - **Dosage:** Two 15-minute sessions / week over 30 weeks

Impact and Outcomes

Finding	Direction	Significance
KS2 Reading: 0.21 [-0.02, 0.25]	+	<i>Not significant</i>
KS2 Reading (FSM): 0.00 [-0.3, 0.3] <i>Analysis of FSM subsample</i>	null	<i>Not significant</i>
FSM / Interaction Term: -0.45 (<i>No confidence intervals</i>)	-	<i>Not significant (in-text, p.23)</i>
Reading Confidence & Ability: 0.32 [0.15 - 0.5]	+	<i>p=0.00</i>
Positive Attitude towards school: 0.25 [0.08, 0.42]	+	<i>p=0.00</i>
Writing confidence /enjoyment: 0.23 [0.06, 0.40]	+	<i>p=0.01</i>

- **Outcomes:** copied from Key Conclusions (p.4 & p.)
 - “The **pupils that received Catch Up Literacy made more progress than pupils that did not**. However, this difference was not statistically significant so we cannot be confident that it was not due to chance.”
 - “Catch Up Literacy did have a **statistically significant impact on pupils’ attitudes to school, self-assessed ability in reading, and their confidence in and enjoyment of writing.**”
 - “Teaching assistants reported a number of benefits for their own **professional development**. These include increases in confidence, knowledge of literacy support and overall job satisfaction.”
- **Any mention of longer-term outcomes:** N/A

Resource list

- Efficacy trial ([EEF Page](#))
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant: N/A
- Suggested further research: N/A

Programme logic model/ theory of change

Not found

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **better teaching assistant quality**
 - *“Teaching assistants were also very positive about the impact that Catch Up Literacy had on them as professionals. The ability to monitor and track pupil progress through the programme also enabled TAs to measure the impact of their own work. They, and a number of programme coordinators, described this as fundamental to focusing, tracking, and improving their own performance as professionals. In turn, this is said to give TAs confidence and contribute to their enjoyment of the role” (p.27)*
- May **increase** post-16 attainment through **improved pupil reading ability and confidence**
 - See table 8 on p.23 noting the significant improvement in pupil reading ability and confidence. The effect size was **0.32** [0.15-0.50], $p = 0.00$
- May **increase** post-16 attainment through **more positive pupil attitudes towards school**
 - See table 8 on p.23 noting the significant improvement in positive attitudes towards school. The effect size was **0.25** [0.08-0.42], $p = 0.00$
- May **increase** post-16 attainment through **improved writing confidence and enjoyment**
 - See table 8 on p.23 noting the significant improvement in writing confidence and enjoyment. The effect size was **0.25** [0.08-0.42], $p = 0.00$
 - *“A one-to-one teaching approach...TAs felt that the pupils could ask questions and make mistakes without the risk of embarrassment. It also offered opportunities for praise and to listen to the pupil’s own individual concerns and opinions which contributed to raising self-esteem and a sense of self-worth. According to staff, this resulted in enjoyment of the sessions and a more positive attitude to attending.” (p.27)*

FSM Pupils

- May **increase FSM pupils’** post-16 attainment through **better teaching assistant quality**
 - Same as above
- May **increase FSM pupils’** post-16 attainment through **improved pupil reading ability and confidence**
 - Same as above
- May **increase FSM pupils’** post-16 attainment through **more positive pupil attitudes towards school**
 - Same as above

- May **increase FSM pupils'** post-16 attainment through **improved writing confidence and enjoyment**
 - Same as above

RQ2: Educational Progression
% of pupils progressing to KS5

- May **increase** post-16 progression through **increased pupil attainment**
 - There was a positive effect of the treatment - although non-significant - on pupil's outcomes but several notes on the impact of the treatment in the process evaluation
 - *"Half of interviewees reported progress in pupils' general reading skills. A third specified an increase in pupils' reading age (by as much as 24 months for some pupils). Around a third detailed improvements in pupils' spelling skills. Around a quarter felt that comprehension and understanding of texts was getting better. Survey results show that 19 of the 27 TAs (70%) believed that Catch Up Literacy has improved pupils' attainment in literacy to some or a great extent" (p.26)*

FSM Pupils

- May **increase FSM pupils'** post-16 progression through **increased pupil attainment**
 - Same as above, however, there is a caveat as the interaction term between the intervention and FSM was negative; but this was not significant (see outcomes table above)

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **increased pupil attainment**
 - There was a positive effect of the treatment - although non-significant - on pupil's outcomes but several notes on the impact of the treatment in the process evaluation (see RQ2)
 - Additional notes on embeddedness which may promote a longer-term impact:
 - *"Teaching assistants explained that they were now more analytical about their approach to teaching, considered new methods, and had also gained a different perspective on how pupils learn." (p.37)*

FSM Pupils

- May **increase FSM pupils'** participation in labour market/education/ training through **increased pupil attainment**
 - Same as above, however, there is a caveat as the interaction term between the intervention and FSM was negative; but this was not significant (see outcomes table above)

Any Additional Evidence [OPTIONAL]

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('Robustness & Relevance sheet, columns S to V)?

- a. Were there any protocol violations?
- b. Is there any evidence of serious non-compliance?
- c. Did the trial have low fidelity?
- d. Did the trial have low adherence?

2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)

3. Risk of Spillover (high/low): Low

British Science Association CREST Silver Award

Background info

- **Population:** KS3
- **Intervention:** A nationwide award scheme set up to help students build their science, technology, engineering and maths skills. It involves students developing their own projects and conducting desk-based and practical research with support from a teacher. (p.4)
 - **Deliverer:** Teachers
 - **Mode of delivery:** full classroom
 - **Dosage:** 30 hours of project work (minimum) over one year (p.4-5)

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
GL Progress in Science: -0.01 [-0.22, 0.19]	-	<i>Not significant</i>
GL Progress in Science (FSM): 0.00 [-0.20, 0.20]	null	<i>Not significant</i>
Student Confidence: 0.12 [-0.01, 0.24]	+	<i>Not significant</i>
Student Attitudes: 0.10 [-0.05, 0.24]	+	<i>Not significant</i>
Student Self-Efficacy: 0.01 [-0.12, 0.14]	+	<i>Not significant</i>
Student Aspiration to pursue STEM: 1.03 [0.96, 1.09] (measured as RRR - relative risk reduction; binary)	+	<i>Not significant</i>
GL Progress in Science (in class): -0.23 [-0.54, 0.0]	-	<i>p=0.035</i>
GL Progress in Science (club): 0.16 [-0.08, 0.40]	+	<i>Not significant</i>

- **Outcomes:** copied from Key Conclusions (p.4 & p.54)
 - "Pupils that were offered participation in the CREST Silver Award made, on average, **no additional progress in science attainment in comparison to those in the control group**, equivalent to zero months of additional progress. This is our best estimate, which has a moderate to low-security rating."
 - "There is **no evidence to suggest that the CREST Silver Award improved self-efficacy in science or increased the proportion of students aspiring**

to a STEM career. Small positive impacts were found for pupil confidence and attitudes to school. This is our best estimate.”

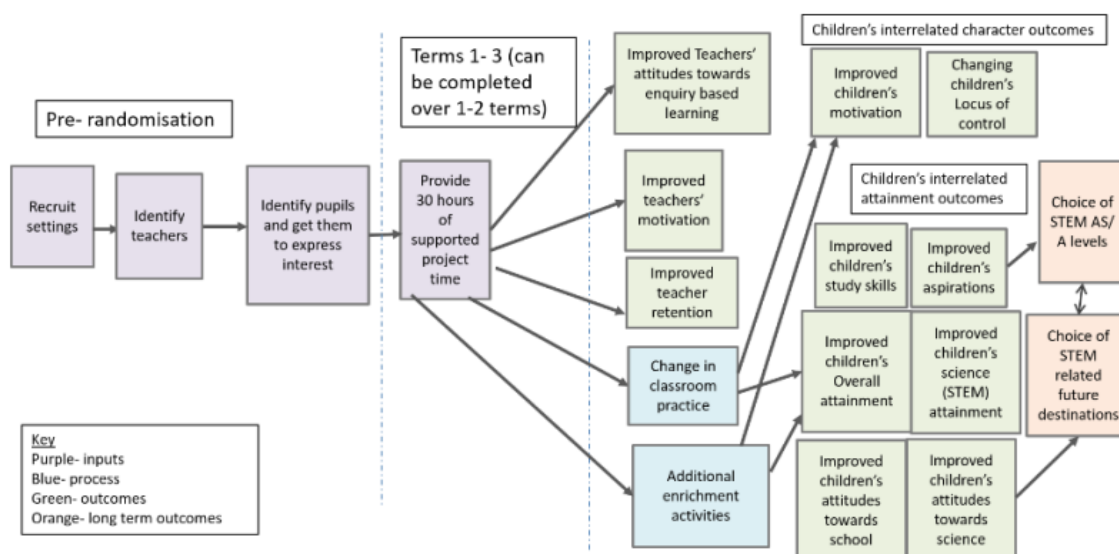
- “There is some indication that the **delivery model of CREST might influence the programme’s impact on science attainment.** Exploratory analysis found that pupils participating in CREST outside of class time made **two months’ additional progress while those that participated during class time made three months less progress than similar pupils in control schools.** These results should be treated with caution as the trial was not designed to measure the difference between the groups and the statistical uncertainty around the results is consistent with larger or null impacts in both cases.”
- **Any mention of longer-term outcomes:**
 - A follow-up qualitative study investigating the drivers of STEM choices found students were more likely to choose STEM subjects if they had parent and teacher support for continuing in STEM, an interest in learning science, the belief that science is useful for obtaining a good job, and confidence in their ability to do well in STEM (AT Kearney, 2016). This research suggests that effort must be made to engage student’s interest in STEM courses and careers throughout secondary school.

Resource list

- [Efficacy trial \(EEF Page\)](#)
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant:
 - Yes, lots of previous studies on p.6-7 including some on longer-term outcomes:
 - Fisher, Titford and Gammon, 2011
 - Moote, 2014
 - Stock Jones et al., 2016
 - Also on p.76:
 - ‘Previous research by Pro Bono economics (2016) suggested that students who do a CREST Silver Award are more likely to improve their educational outcomes and choose STEM A-levels.’
 - Suggested further research: N/A

Programme logic model/ theory of change (see p.10)

Figure 1: CREST Silver Award logic model



Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes

% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment (in STEM) through **increased confidence**
 - While non-significant, confidence was higher in the treatment group compared to the control group and student testimonials support this:
 - Student Testimonial 1 (p.47): *'It made you, like, more sure of yourself, so instead of doubting yourself and, like, being like, "Oh, I should have done what they've done, 'cause they've done it a lot better than me", you could have been like, "Oh, yeah, I've, like, I can do the work the way I want, and what I like to do with it", and all that' (Student 1151 CSFG).*
 - Student Testimonial 2 (p.47): *"They spend their whole school life feeling that they can't achieve. Because they're struggling all the time. And they're not the high fliers, and, and I think it's very undermining for students, to their self-confidence. And this has made them feel that, actually, yeah, I can do it. I have achieved. I've got this' (CREST Educator 1113 CSCO)."*
- May **increase** post-16 attainment (in STEM) through **better pupil-teacher relationships**
 - *"CREST participation was felt to benefit teacher-student interactions. The majority of CREST educators (75%, n = 49) reported an improvement in interpersonal relationships. This was thought to result from a teacher promoting a more independent style of learning and students taking on more responsibility. They felt that participating students seemed to have more respect for them because of the support given to students to complete their CREST projects" (p.47)*
- May **decrease** post-16 attainment through **disruption to regular teaching**
 - *"A concern raised by CREST educators was that CREST participation was disruptive to regular lessons, particularly in schools that had opted to deliver the intervention as a whole class activity. CREST educators indicated that for students with additional learning needs—such as those with English as an*

Additional Language (EAL) or special educational needs (SEN)—CREST may have been detrimental to their progress.” (p.48)

FSM Pupils

- May **increase FSM pupils’** post-16 attainment through **increased confidence**
 - Same as above
- May **increase FSM pupils’** post-16 attainment through **better pupil-teacher relationships**
 - The study was designed and randomised with FSM pupils in mind and the sample in the treatment and control group contained 34% and 45% FSM pupils respectively (p.26) suggesting all relationships noted above should also apply to FSM pupils, especially given no differential impacts were observed qualitatively or quantitatively (p.34)
- May **increase FSM pupils’** post-16 attainment through **higher attainment**
 - Based on an older cited study testing the same intervention:
 - *‘Additionally, the CREST students eligible for free school meals (FSM) saw a larger increase in their best GCSE science score compared to a matched comparison group of other students also eligible for FSM (Stock Jones et al., 2016).’ (p.7)*

RQ2: Educational Progression

% of pupils progressing to KS5

- May **increase** post-16 progression through pupils **increased engagement with science subjects**
 - *“Students also enjoyed the inquiry-based learning approach required by CREST which was a change from the typical didactic style of classroom teaching and learning. In some cases, it was reported that this appeared to aid learning and engagement in science.” (p.45)*
- May **increase** post-16 progression through pupils **increased motivation and positive attitude to school**
 - Teacher testimonial 1 (p.45):
 - *“They’re very motivated, so they’re much more motivated than they were at the start. They feel like they’ve been rewarded for all their hard work’ (CREST Educator 1170 CSCO).”*

FSM Pupils

- May **increase FSM pupils’** post-16 progression through **pupils science attainment**
 - Based on an older cited study testing the same intervention:
 - *‘ Additionally, the CREST students eligible for free school meals (FSM) saw a larger increase in their best GCSE science score compared to a matched comparison group of other students also eligible for FSM (Stock Jones et al., 2016).’ (p.7)*

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **increased pupil aspirations** in STEM careers
 - Although this secondary outcome was not significant between groups, the process evaluation provides some evidence of students’ being more interested in STEM careers after the programme:

- Student Testimonial (p.46): *“I think I know much more about science than I did before. I think I’m more clearer based on my future pathways of careers and work, which college I want to go to, actually ... CREST and everything. I think it’s really helped out” (Student 1110 CSFG).”*
- “Student survey responses did not indicate that a high level of interest in pursuing a STEM careers was as a result of CREST participation. Although the level of interest in a STEM career fell slightly at follow-up, both intervention and control groups indicated a high level of interest in a STEM career” (p.46)

FSM Pupils

- May **increase FSM pupils’** participation in labour market/education/ training through **increased aspiration** to pursue a science career
 - Same as above

Any Additional Evidence [OPTIONAL]

Quality of Implementation

1. Is the answer to any of the following a ‘yes’ in the Trial classifier (*‘Robustness & Relevance sheet, columns S to V’*)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were ‘yes’, please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - There were issues with compliance and the flexible model of implementation:
 - *“There are a number of conflicting reports of compliance. Only 10% of the pupils were compliant according to the definition agreed at set-up of submitting a project that passed (being awarded a Silver award). But then other records say 22-29% submission rate and pupil reports suggest 51% submission rates. Only 76% of pupils who submitted a project to the BSA could be matched to trial data. This means at best there is low quality of data but moderate compliance, or at worst, very low compliance.” (p.58)*
3. **Risk of Spillover** (high/low): Low (but control schools may have taken up the intervention after the CREST trial; no measure of this)

Ark Mathematics Mastery

Background info

- **Population:** KS3
- **Intervention:** A ‘mastery curriculum’ where pupils progress at the same pace, with differentiation through depth of exploration rather than accelerated content coverage. This approach reduces the need to repeatedly revisit topics, and promotes depth of understanding over memorised procedures. Appears to replace entire maths curriculum (p.9)
 - **Deliverer:** Teachers (with training from External provider)

- **Mode of delivery:** full classroom
- **Dosage:** Over 1 year in all maths lessons

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
PiM Post-test: 0.055 [-0.037, 0.147]	+	<i>Not significant</i>
PiM Post-test (FSM): 0.066 [-0.037, 0.169]	+	<i>Not significant</i>
PiM Post-Test (material covered in MM): 0.10 - <i>No 95% CIs stated (see p.39)</i>	+	<i>p=0.04</i>
PiM Post-Test (material covered in MM & FSM): 0.104 [0, 0.21]	+	<i>p<0.05</i>

- **Outcomes:** copied from Key Conclusions (p.5 & p.52)
 - “On average, Year 7 pupils in schools adopting Mathematics Mastery made a **small amount more progress than pupils in schools that did not**. However, the effect detected was **not statistically significant**, meaning that it is not possible to rule out chance as an explanation.”
 - “There is **no strong evidence that the approach had a greater impact on lower-attaining pupils than on higher-attaining pupils**.”
 - “Combining the findings from this study and a second randomised controlled trial of Mathematics Mastery involving Year 1 pupils may strengthen the overall evidence for the approach.”
- **Any mention of longer-term outcomes:**
 - ‘It would be worthwhile to track the medium- and long-term impact of the approach, to assess whether there is a cumulative effect to the approach and whether it has an impact on performance in **high-stakes tests**.’ (p.5)
 - ‘Low-stakes tests: The post-test used in this trial was a ‘low-stakes’ assessment, meaning that neither participating children nor their schools had anything riding on the results. It has been suggested that this may lead to a lack of effort among children taking the test (Wise and DeMars, 2005). Moreover, this factor could potentially differ between intervention and control groups. For instance, if intervention schools are more engaged in the trial than control schools, they may have put more effort into completing the post-test. This would potentially lead to an upwards bias in the estimated intervention effect.’ (p.52-53)

Resource list

- Efficacy trial (EEF Page): N/A
- [Final Report](#)
- Effectiveness trial ([EEF page](#))
- Any cited studies that might be relevant: N/A
- Suggested further research:
 - <https://ideas.repec.org/s/qss/dqsswp.html>

Programme logic model/ theory of change (diagram or equivalent in-text)

None found

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes

% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **higher pupil attainment**
 - In the PiM, sub-analysis on questions addressed by the Maths Mastery programme showed that pupils performed significantly better suggesting the programme had a differential effect based on the topics covered rather than improving overall Maths attainment. This positive effect may carry on in the longer-term (results on p.38)
 - *“Another school [A] observed in a telephone interview that the test results did not show as much progress as expected from the class work. The suggestion was that the learning witnessed in class time was different in nature. Whereas previously the pupils could answer questions without understanding, now they really had a proper grasp of the topics”* (p.48)
- May **increase** post-16 attainment through **better maths teaching quality / newer techniques**
 - *‘More than half of the 27 teachers from intervention schools who completed both questionnaires reported increases in the regularity of introducing concepts or skills through problem-solving and pupils working on unfamiliar problems, and less regularity in practice exercises’* (p.48)

FSM Pupils

- May **increase FSM pupils’** post-16 attainment through **higher pupil attainment**
 - While the results were no significant, the improvement in FSM pupil’s PiM scores were higher than the overall effect suggesting that it could have positive knock-on effects for longer-term outcomes (p.32)
- May **increase FSM pupils’** post-16 attainment through **better maths teaching quality / newer techniques**
 - *‘More than half of the 27 teachers from intervention schools who completed both questionnaires reported increases in the regularity of introducing concepts or skills through problem-solving and pupils working on unfamiliar problems, and less regularity in practice exercises’* (p.48)

RQ2: Educational Progression

% of pupils progressing to KS5

- May **increase** post-16 progression through **higher pupil attainment in higher-stakes tests**
- May **increase** post-16 progression through **increased pupil effort**
 - *‘The post-test used in this trial was a ‘low-stakes’ assessment...’* (p.52), full quote above which suggests that a higher-stakes test is better for capturing learning as pupils will maximise their effort here
 - *‘... children are likely to follow the Mathematics Mastery programme for a number of years (e.g. throughout secondary school), whereas this evaluation has only considered the impact of a one-year ‘dose’. Long-term effects after sustained exposure to the programme could be significantly higher’* (p.53)

FSM Pupils

- May **increase FSM pupils’** post-16 progression through **higher pupil attainment**
- May **increase FSM pupils’** post-16 progression through **increased pupil effort**
 - Same as above

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training **better overall quantitative skills**
 - Across pages 52-53, there are discussions of potential longer-term effects of the intervention. While there is no explicit mention of labour market outcomes, this maths-specific programme could improve and encourage students' quantitative skills

FSM Pupils

- May **increase FSM pupils'** participation in labour market/education/ training through **better overall quantitative skills**
 - Across pages 52-53, there are discussions of potential longer-term effects of the intervention. While there is no explicit mention of labour market outcomes, this maths-specific programme could improve and encourage students' quantitative skills

Any Additional Evidence [OPTIONAL]

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
3. Risk of Spillover (high/low): Low

Peer Tutoring in Secondary Schools

Background info

- **Population:** KS3 (Year 7 students)
- **Intervention:** Year 9 students were paired with Year 7 students to tutor them in reading skills
 - **Deliverer:** Students (External provider conducted the training & teachers oversaw)
 - **Mode of delivery:** 1:1
 - **Dosage:** One 20 minute session / week for 16 weeks

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
Overall reading score (Year 7): -0.02 [-0.15, 0.11]	-	Not significant

Overall reading score (Year 7, FSM): -0.04 [-0.21, 0.13]	-	<i>Not significant</i>
Overall Reading Score (Year 9): -0.06 [-0.14, 0.02]	-	<i>Not significant</i>
Overall Reading Score (Year 9, FSM): -0.06 [-0.14, 0.02]	-	<i>Not significant</i>

- **Outcomes:** copied from Key Conclusions (p. & p.52)
 - “This evaluation does **not provide any evidence that the Paired Reading programme had an impact on overall reading ability, sentence completion and passage comprehension of participating pupils.**”
 - “There was **no evidence** of the Paired Reading programme having an effect on overall reading ability, sentence completion and passage comprehension of **FSM** pupils.”
 - “There was **some variation in the intervention group schools in terms of the programme setup and delivery.** There was also a varying level of support provided to pupils within the intervention by the teachers involved, based mainly on the reading ability of the pupils. However, these appear to be natural variations between the settings of the schools involved and are unlikely to have affected the dosage of the intervention for the pupils involved.”
- **Any mention of longer-term outcomes:** N/A

Resource list

- Efficacy trial ([EEF Page](#))
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant:
 - Yes; despite the lack of effects found here; previous research outlined on p.6-7 has found this kind of intervention to be successful
- Suggested further research: N/A

Programme logic model/ theory of change (diagram or equivalent in-text)

Not found

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
% of pupils with level 2 as highest qualification

- May **decrease** post-16 attainment [of Year 7s] through **pupils falling behind.**⁴³
 - “Teachers felt that participating pupils fell behind their peers in subjects that the Paired reading sessions replaced. In total, they were taken out for at least 30 minutes a week over 16 weeks and there was a worry that this may result in them doing worse in their assessments than the control group pupils. Impact of existing programmes. Some senior leaders felt that their school’s

⁴³ As the quotes illustrate, this hypothesis presents a lower certainty as it is backed by relatively limited evidence and is not supported by statistically significant results from the original impact evaluation.

own existing literacy programmes were better than the Paired Reading programme.” (p.50)

- *“Schools were concerned about the implementation of the programme including timetabling the lessons. There were concerns that missing other subject lessons was having an adverse effect on attainment.” (p.50)*
- *“When lower ability Year 9s were paired with lower ability Year 7s teachers felt neither improved their reading ability nor their confidence. It was often the case that the lower ability Year 7 had a higher reading age than the Year 9 and this resulted in a considerable loss of confidence and morale for the Year 9 pupil” (p.49)*
- May **decrease** post-16 attainment [of Year 9s] through **lower pupil confidence**
 - *“Loss of confidence for lower ability Year 9s. While some pupils in Year 9 gained in confidence, teachers felt that especially for the lower ability Year 9 pupils, the programme had the opposite effect. It was felt they were embarrassed about having a reading ability the same as or lower than a Year 7 pupil and so disengaged from learning.” (p.50)*

FSM Pupils

- May **decrease FSM pupils’** post-16 attainment [of Year 7s] through **pupils falling behind**
 - Same as above
- May **decrease FSM pupils’** post-16 attainment [of Year 9s] through **lower pupil confidence**
 - Same as above

RQ2: Educational Progression % of pupils progressing to KS5

- May **decrease** post-16 progression [of Year 9s] through **lower pupil confidence**
 - Same as above

FSM Pupils

- May **decrease FSM pupils’** post-16 progression [of Year 9s] through **lower pupil confidence**
 - Same as above

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **improved pupil relationships**
 - While the process evaluation was generally negative about the impact on academic attainment, there were some notes on how relationships improved through tutoring:
 - *“Improved relationships. The programme was felt to improve relationships between pupils with the pairs having a positive social experience from peer tutoring. In some cases, teachers felt that pupils who had never spoken to someone two years above them had gained in confidence as a result of their interactions on the programme.” (p.50)*

FSM Pupils

- May **increase FSM pupils'** participation in labour market/education/ training through **improved pupil relationships**
 - Same as above

Any Additional Evidence [OPTIONAL]

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)

No major issues here

3. Risk of Spillover (high/low): Low

Quest

Background info

- **Population:** KS3
- **Intervention:** A whole year group approach for teaching English designed to help improve reading comprehension. Key components include ability grouping of students, emphasis on cooperative learning, consistent use of a cycle of instruction, and regular assessments with regrouping. (p.7-8)
 - **Deliverer:** Teachers (p.7)
 - **Mode of delivery:** full classroom (p.7)
 - **Dosage:** Every English lesson in the full school year (p.7)

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
NGRT Reading Comprehension: -0.04 [-0.14, 0.07]	-	Not significant
NGRT Reading Comprehension (FSM): -0.11 [-0.24, 0.02]	-	Not significant

- **Outcomes:** copied from Key Conclusions (p.3 & p.30)
 - "The evaluation was **unable to provide a secure estimate of the impact of Quest on reading comprehension outcomes among Year 7 pupils**, primarily due to a high level of drop-out from the project."
 - "Some of the **main barriers to successful implementation included the difficulty in covering the expected material in a single school lesson; a perception of an insufficient focus on writing activities; and its adoption as a whole-year group intervention.**"

- *“Many schools were positive about the range of resources provided by the programme, and about the co-operative learning aspects of the programme.”*
- *“Though it is **not possible to draw a conclusive statement** about the impact of the programme, on average, pupils who received the programme made less progress than those who did not.”*
- Any mention of longer-term outcomes:

Resource list

- Efficacy trial ([EEF Page](#))
- [Final Report](#)
- Effectiveness trial (EEF page): N/A
- Any cited studies that might be relevant:
 - Reading Edge program in the US (cited on p.5) - *Quest was based on this*
 - Slavin, R.E., Chamberlain, A., Daniels, C., and Madden, N.A. (2009) ‘The Reading Edge: A randomized evaluation of a middle school co-operative reading program’. Effective Education, 1 (1), 13–26.
- Suggested further research: N/A

Programme logic model/ theory of change (diagram or equivalent in-text)

None found

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
% of pupils with level 2 as highest qualification

- May **increase** post-16 attainment through **increased pupil cooperative learning**
 - *“Facilitators and teachers were, on the whole, positive about the co-operative learning structure provided by Quest. Two of the schools already used co-operative learning techniques: one facilitator felt Quest had built on these, but the other thought it had diminished what they did” (p.22)*
 - *“Some schools had even adopted the techniques with other year groups. They especially welcomed the group working, the competitive element and the opportunities for talk and discussion.” (p.23)*

FSM Pupils

- May **increase FSM pupils’** post-16 attainment through **increased pupil cooperative learning**
 - Same as above

RQ2: Educational Progression
% of pupils progressing to KS5

- May **decrease** post-16 progression through **lowered pupil reading comprehension**
 - *“On average, pupils who received the programme scored lower in terms of reading comprehension compared to those who were taught English using standard practice.” (p.30)*
 - Lower comprehension scored observed for the treatment group (see outcomes table above). Note these were non-significant.

FSM Pupils

- May **decrease FSM pupils'** post-16 progression through **lowered pupil reading comprehension**
 - Same as above & the magnitude of the (non-significant) negative effect was larger for FSM pupils

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **improving pupil's social skills**
 - *"One positive consequence of the programme, which was mentioned by three of the teachers interviewed, was that students' social skills developed, potentially as a result of both the ability grouping and the teamwork aspects of the programme" (p.26)*
 - *'The majority of pupils liked working with partners or in small groups/teams 'very much' or 'quite a lot', regardless of which arm of the trial they were in.'* (p.27)

FSM Pupils

- May **increase FSM pupils'** participation in labour market/education/ training through **improving pupil's social skills**
 - Same as above

Any Additional Evidence [OPTIONAL]

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - There were issues of dropout (46% in the intervention group, p.30), data attrition and some imbalance in the randomisation but the mechanisms noted above i.e. "better cooperation" and "better social skills" should not be as affected by these.
 - **"Few, if any, schools implemented the programme as designed, suggesting that substantial adaptation may be required if the approach is to gain wider adoption in English schools."** (p.3 & p.30)

3. Risk of Spillover (high/low): Low

Sci-napse

Background info

- **Population: KS3**

- **Intervention:** Sci-napse is a programme which aims to improve the science knowledge of Year 8 pupils through using quizzes in science lessons. An online quizzing tool, called Q-Fire, provides teachers with banks of questions, which can be added to or adapted to fit with lesson plans. (p.3)
 - **Deliverer:** Teachers
 - **Mode of delivery:** Full classroom (broke out into small groups)
 - **Dosage:** 6 quiz questions / science lesson over 1 year
 - There were 2 variations: *game-based or test-based*

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
Game Based		
GL Post-test: -0.05 [-0.23, 0.12]	-	<i>Not significant</i>
GL Post-test (FSM*GameBased) 0.64 [-0.31,1.59] - <i>Interaction effect, not hedges' g</i>	+	<i>Not significant</i>
GL Post-test (minimally compliant): 0.05 [-0.17, .28]	+	<i>Not significant</i>
Test Based		
GL Post-test: -0.08 [-0.25, 0.09]	-	<i>Not significant</i>
GL Post-test (FSM*TestBased): 0.65 [-0.41, 1.70]	+	<i>Not significant</i>
GL Post-test (minimally compliant): 0.19 [-0.17, 0.54]	+	<i>Not significant</i>
Test Vs Game Based (higher score indicates that test-based is better)		
GL Post-test: 0.02 [-0.14, 0.19]	+	<i>Not significant</i>
GL Post-test (minimally compliant): 0.17 [-0.4, 0.7]	+	<i>Not significant</i>

- **Outcomes:** *copied in from Key conclusions (p.3)*
 - *“Children in both test- and game-based Sci-napse classes made a **small amount less progress than pupils in the comparison classes**. The difference in scores is not “statistically significant”...This finding has moderate to high security.”*
 - *“**Exploratory analysis found that the test-based Sci-napse classes that met the minimum requirements of the intervention made an average of three additional months progress**. A small positive impact was found in the game-based classes that met the minimum requirements. The small number of pupils included in these analyses means that they should be treated with caution.”*
 - *“Teachers and pupils **reported that Sci-napse was valuable as a revision tool** but indicated that it was difficult to fit six questions into each lesson and that there were challenges with the technology – particularly around adding their own questions.”*
- **Any mention of longer-term outcomes:** N/A

Resource list

- Efficacy trial ([EEF Page](#))
- [Final Report](#)
- Effectiveness trial (EEF page)
- Any cited studies that might be relevant:
 - The introduction cites some background research about reward uncertainty and learning games (p.6)
- Suggested further research: N/A

Programme logic model/ theory of change

None found

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes
 % of pupils with level 2 as highest qualification

- May **decrease** post-16 attainment through **improved pupil's attainment**
 - *"a number of teachers (12.9%) in game and 14.3% in test responded that they disagreed or strongly disagreed with the statement and 48.4% of teachers in game and 49.4% in test neither agreed or disagreed" (p.56)*
 - Relevant statement: *"using the Sci-napse quiz questions had a positive impact on the attainment of their students"*
 - *"Scepticism from teachers and pupils with regards to the potential benefits (or not) of the notion of points for a correct answer that can be risked; A sometimes difficult to reconcile trade-off between striking a balance between fitting six Q-Fire questions into a lesson, maintaining the pace of the lesson, managing group/team work and covering the required content in a lesson." (p.56)*
- May **increase** post-16 attainment through **improved pupil's attainment**
 - Particularly for the compliant schools (see outcomes table above), the intervention had a positive effect (non-significant) on pupil's outcomes. For the test-based intervention, this is equivalent to 3 months progress.
 - *"Contrary to the results of the impact evaluation, 38.7% of teachers who taught the game version and 36.4% who taught the test version responding either agreed or strongly agreed that they thought using the Sci-napse quiz questions had a positive impact on the attainment of their students" (p.56)*
 - *"Many teachers commented in the end of evaluation survey that it would be a useful revision tool, but it seems that this is how they saw it, rather than an intervention with the specific purpose of raising attainment." (p.57)*

FSM Pupils

- May **decrease** post-16 attainment through **improved pupil's attainment**
 - Same as above; while the interaction term for FSM*Intervention for both arms were positive, these were not significant and not supported when broken down into mean scores, there is a negligible difference. Hence homogenous results are reasonable to expect.
 - *"At post-test, the raw mean (SD) for the FSM subgroup was 29.0 (9.3): test-based group 29.0 (9.1); game-based group 28.8 (9.1); and control group 29.2 (9.9). Within the non-FSM subgroup, the raw mean (SD) post-test score was 32.3 (9.4): test-based group 31.7 (9.5); game-based group 32.1 (8.9); and control group 33.3 (9.9). Therefore, regardless of group allocation, pupils eligible for FSM tended to perform worse on the pre and post-tests than non-FSM pupils." (p.43)*

- May **increase FSM pupils'** post-16 attainment through **improved pupil's attainment**
 - Same as above; all effects held for FSM

RQ2: Educational Progression
% of pupils progressing to KS5

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

Any Additional Evidence [OPTIONAL]

N/A

Quality of Implementation

1. Is the answer to any of the following a 'yes' in the Trial classifier ('*Robustness & Relevance sheet, columns S to V*)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were 'yes', please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - Low adherence and low fidelity:
 - *"Only 54% of teachers in the test-based Sci-napse classes and 29% of in the game-based classes met the minimum requirements of the interventions. This meant using questions in some, most or all of their lessons – and, for the game-based intervention, making use of escalating points and the wheel of fortune."* (p.3)
 - *"The data gathered in the IPE (as discussed in previous sections) explains how **various barriers to the implementation of both the game and testing conditions** (e.g. technology problems, inconsistent implementation) are **likely to have influenced any potential for impact on student achievement**"* (p.56)
3. **Risk of Spillover** (high/low): Low
 - Minor crossover noted though: *"One respondent who had taught in game- and test-based and control conditions did report some crossover when asked in an open question)"* (p.55)

Youth United

Background info

- **Population:** KS3
- **Intervention:** A programme of activities provided by four uniformed youth organisations; where possible, schools were offered a choice of uniformed groups (members include Army Cadet Force, Fire Cadets, Girlguiding, JLGB, Royal Air Force Air Cadets, Sea Cadets, St John Ambulance, The Boys' Brigade, The Girls'

Brigade England & Wales, The Scout Association, and Volunteer Police Cadets). (p.7)

- **Deliverer:** External provider
- **Mode of delivery:** Varied between full classroom or small group (p.15)
- **Dosage:** 24 sessions / year

Impact and Outcomes

Outcome Measure / Finding	Direction	Significance
KS3 Test Score gains in English: -0.09	-	<i>No confidence intervals reported</i>
KS3 Test Score gains in Maths: -0.09	-	<i>No confidence intervals reported</i>
Progress in Self-Confidence: 0.10	+	<i>No confidence intervals reported</i>
Progress in Teamwork: 0.04	+	<i>No confidence intervals reported</i>
KS3 Test Score gains in English (FSM): -0.24	-	<i>No confidence intervals reported</i>
KS3 Test Score gains in Maths (FSM): -0.29	-	<i>No confidence intervals reported</i>
Progress in Self-Confidence (FSM): 0.05	+	<i>No confidence intervals reported</i>
Progress in Teamwork (FSM): -0.11	+	<i>No confidence intervals reported</i>

- **Outcomes:** copied from Key Conclusions (p.4 & p.43)
 - “There is **no evidence that the intervention had any benefit in terms of pupils’ academic performance**. Although the attainment data suggests a small negative impact, the quality of this data is too low to draw this conclusion with confidence. The **data quality was compromised due to changes in national testing**.”
 - “Participation in the intervention is associated with a small improvement in self-reported nonattainment outcomes including self-confidence and teamwork. It is possible that these small effects are an underestimate due to technical issues regarding the groups of children that were compared in the analysis.”
 - “For pupils eligible for free schools meals, there is **no evidence that the intervention had a positive impact on academic attainment or self-reported character attributes**. Again, the attainment data suggests a negative impact, but the quality of this data is too low to draw this conclusion with confidence.”
 - “**Study participants were extremely positive** about the intervention and many felt it had a positive effect on the behaviour and skills of participating pupils.”
- **Any mention of longer-term outcomes:**
 - “Some longer-term outcomes such as enhanced opportunities for subsequent employment were also considered, but would not be measured in the time-scale of this trial.” (p.12)

Resource list

- Efficacy trial (EEF Page):
- [Final Report](#)

- Effectiveness trial ([EEF page](#)⁴⁴)
- Any cited studies that might be relevant:
 - *“A large-scale experimental study by the Cabinet Office in the U.K. evaluated the impact of Youth Social Action (YSA) on young people aged 10 to 20 in 73 schools (Kirkman et al. 2016). The evaluation consisted of three RCTs and one matched pair trial of YSA interventions provided by four providers. All of the programmes included of an element of citizenship activities” (p.8)*
- Suggested further research: N/A

Programme logic model/ theory of change (diagram or equivalent in-text)

Hypotheses based on documents review

RQ1: Post-16 and HE attainment Outcomes

% of pupils with level 2 as highest qualification

RQ2: Educational Progression

% of pupils progressing to KS5

- May **increase** post-16 progression through **improved pupil behaviour**
 - *“Teachers and delivery staff observed positive changes in pupils’ behaviour as a result of the experience. There were many anecdotal accounts of improvements in pupil behaviour and skills. We could only report a summary of these” (p.36)*
- May **increase** post-16 progression through **improved pupil confidence and self-esteem**
 - *“Delivery staff also reported obvious improvements in pupils’ self-confidence and leadership skills. One pupil had gained so much in the short time that he is now helping to run some of the sessions. As pupils were awarded for their achievements and attendance, delivery staff and teachers reported that this boosted their self-esteem” (p.39)*

FSM Pupils

- May **increase FSM pupils’** post-16 progression through **improved pupil behaviour**
 - Same as above
- May **increase FSM pupils’** post-16 progression through **improved pupil confidence and self-esteem**
 - Same as above

RQ3: Labour Market Activities

of months not in employment, education or training from the August 31 after they turned 16 or participated in the trial (whichever is later), to the latest point for which we have data

- May **increase** participation in labour market/education/ training through **increasing exposure and experience with ‘real-life’ programmes**
 - Student testimonial
 - *“I am wanting to join the Royal Marines Cadets at City of Liverpool Sea Cadets. I have already been to look around. I enjoy army cadets and sea cadets so Royal Marines cadets is a mix of the two. I want a career in the army so think this will help. The Sea Cadets after school club has been great!” (p.39)*

⁴⁴ This document and the trial classifier sheet have recorded this is an effectiveness trial, based on the EEF webpage, the archive [metadata](#) states it is an efficacy trial.

- *“They liked the fact that they were learning something useful in real life, skills that they could actually put into practice. Pupils talked enthusiastically about the first aid skills they had learnt.” (p.37)*
- *“Pupils also liked the fact that they had a certificate at the end that they could put on their CV. The pupils said they would strongly recommend the course to other pupils” (p.37)*
- Existing evidence on such programmes suggest:
 - *“Many studies have been conducted suggesting positive effects of such programmes on a range of young people’s outcomes, such as employability, self-esteem, confidence, and other useful skills.” (p.7)*
- May **increase** participation in labour market/education/ training through **increasing exposure to non-academic dependent pathways**
 - *‘I think I may fail my GCSE’s so I need something else for the future on my CV and thought the Sea Cadets would help me gain qualifications.’ (p.39)*

FSM Pupils

- May **increase FSM pupils’** participation in labour market/education/ training through **increasing exposure and experience with ‘real-life’ programmes**
 - Same as above
 - *“...achieved the aims in **offering opportunities to disadvantaged children to learn important life skills** and to participate in youth social action and volunteering activities. Feedback from stakeholders (parents, teachers, delivery staff, and pupils) suggests that pupils have benefited from these experiences and their lives enriched.” (p.36)*

Any Additional Evidence [OPTIONAL]

Quality of Implementation

1. Is the answer to any of the following a ‘yes’ in the Trial classifier (‘Robustness & Relevance sheet, columns S to V)?
 - a. Were there any protocol violations?
 - b. Is there any evidence of serious non-compliance?
 - c. Did the trial have low fidelity?
 - d. Did the trial have low adherence?
2. If any of the answers above were ‘yes’, please provide details below (**prioritise elaborating on issues/ violations that have implications for the ToC and/or the specified hypotheses above. If not, can be very brief & copy/paste**)
 - **Inconsistent reporting by schools:**
 - *“As discussed in the outcomes section, these findings must be treated with great caution because the KS3 scores were provided by schools in different formats, and many of these ‘survey volunteer’ pupils did not actually participate in the intervention itself.” (p.34)*
 - *“Almost a quarter of schools did not deliver the intervention due to issues such as lack of teacher volunteers, and other schools did not deliver a full programme of activity. Support from senior leaders, dedicated space, school staff time, and a dedicated slot in the school day or after school were all identified as necessary conditions for successful implementation.” (p.4 & p.43)*

3. Risk of Spillover (high/low): Low