REVIEW OF EEF PROJECTS

Technical Annex

August 2021

Sean Demack, Bronwen Maxwell, Mike Coldwell, Anna Stevens, Claire Wolstenholme, Sarah Reaney-Wood, Bernadette Stiell (Sheffield Institute of Education, Sheffield Hallam University)

Hugues Lortie-Forgues (University of York)

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

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

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

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

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

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

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

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Overview

This document provides supplementary statistical tables to support our review of EEF evaluations. This includes:

1. Descriptive (univariate) tables for all explanatory variables
2. Tables for the meta-analyses of primary ITT effect sizes
3. Tables for the meta-analyses of secondary ITT effect sizes
4. Tables for the meta-analyses of FSM subsample primary / secondary effect sizes
5. Tables for the analyses of cost effectiveness
6. Tables for the analyses of pupil-level attrition.

The analyses of pupil-level attrition identified a clear (and expected) association with the type of primary ITT outcome that was used in a trial (i.e., commercial tests tended to have higher attrition compared with official / NPD outcomes). For this reason, a limited follow-on elaboration analysis was undertaken for the attrition analyses. Specifically, analyses were undertaken separately for trials that used a commercial and trials that used an official / NPD outcome. This resulted in additional tables:

7. Tables for the elaboration analyses of pupil-level attrition.

Contents

Tables

Presenting the explanatory variables

This section provides supplementary tables drawn on in the discussion / interpretation for the main report. The tables present explanatory variables under the five overarching themes and 22 subthemes.

1 The intervention

- Focus of intervention
- Intensity of intervention (minutes per week)
- Who implements with direct targets?
- Perceived quality of support resources
- Cost of intervention
- EEF intervention school theme areas
- EEF rating as promising project.

2 Theory & evidence

- Empirical evidence and theoretical detail
- Causal processes and mechanisms

3 Context

- External context
- Characteristics of participating organisations (barriers and enablers)
- Characteristics of participating individuals (barriers and enablers)

4 Implementation & fidelity

- Developer characteristics
- Focus, planning, time and SLT support
- Professional development (CPD)
- Support and monitoring
- Fidelity

5 Evaluation design

- Trial description
- Length and size of trial
- Statistical sensitivity, attrition and trial quality
- Evaluation burden
- Primary outcome

The intervention

Table 1: Distributions of explanatory variables for the intervention theme

	No. of trials (%)	Headline, ITT primary outcome effect sizes (%)	Secondary attainment outcome ITT effect sizes (%)	FSM attainment outcome effect sizes (%)	Psychological outcome effect sizes (%)
All trials	82 (100%)	133 (100%)	78 (100%)	149 (100%)	88 (100%)
School phase					
Primary (including Early Years)	51 (62%)	88 (66%)	65 (83%)	102 (69%)	70 (80%)
Primary–secondary transition	6 (7%)	7 (5%)	5 (6%)	7 (5%)	11 (13%)
Secondary	25 (30%)	38 (29%)	8 (10%)	40 (27%)	7 (8%)
Key Stage of pupils					
Early Years	2 (2%)	2 (2%)	1 (1%)	1 (< 1%)	0 (–)
Primary KS1	13 (16%)	23 (17%)	16 (21%)	29 (20%)	21 (24%)
Primary KS2	33 (40%)	57 (43%)	47 (60%)	66 (44%)	49 (56%)
Primary (multiple Key Stages)	3 (4%)	6 (5%)	1 (1%)	6 (4%)	0 (–)
Primary–secondary transition	6 (7%)	7 (5%)	5 (6%)	7 (5%)	11 (13%)
Secondary KS3	20 (24%)	26 (20%)	6 (8%)	27 (18%)	7 (8%)
Secondary KS4	4 (5%)	9 (7%)	2 (3%)	10 (7%)	0 (–)
Secondary (multiple Key Stages)	1 (1%)	3 (2%)	0 (–)	3 (2%)	0 (–)
Cross-curriculum	29 (35%)	67 (50%)	38 (49%)	70 (47%)	46 (52%)
English	36 (44%)	48 (36%)	27 (35%)	61 (41%)	25 (28%)
Maths	14 (17%)	15 (11%)	11 (14%)	15 (10%)	13 (15%)
Science	3 (4%)	3 (2%)	2 (3%)	3 (2%)	4 (5%)
Intensity (minutes per week) – measured at the trial-level					
Mean (SD)	94 (74.1)	–	–	–	–
Median	70	–	–	–	–
Min : max	10 : 300	–	–	–	–
n =	51	–	–	–	–
Intensity (categorised)					
30 minutes or less per week	12 (15%)	16 (12%)	4 (5%)	18 (12%)	20 (23%)
31–60 minutes per week	13 (16%)	21 (16%)	12 (15%)	29 (20%)	26 (30%)
61–120 minutes per week	15 (18%)	27 (20%)	11 (14%)	32 (22%)	10 (11%)
Over 120 minutes per week	11 (13%)	16 (12%)	11 (14%)	13 (9%)	6 (7%)

No intensity detail	31 (38%)	53 (40%)	40 (51%)	57 (38%)	26 (30%)
Teacher-led	37 (45%)	57 (43%)	40 (51%)	63 (42%)	32 (36%)
Externally-led (e.g., delivery partner)	18 (22%)	30 (23%)	10 (13%)	32 (22%)	18 (21%)
TA-led	12 (15%)	15 (11%)	5 (6%)	17 (11%)	10 (11%)
Parent-led	2 (2%)	7 (5%)	3 (4%)	10 (7%)	11 (13%)
Resource-led	2 (2%)	2 (2%)	1 (1%)	3 (2%)	0 (–)
Other school staff-led	1 (1%)	2 (2%)	6 (8%)	2 (1%)	6 (7%)
Other	10 (12%)	20 (15%)	13 (17%)	22 (15%)	11 (13%)
High	20 (24%)	27 (20%)	25 (32%)	30 (20%)	13 (15%)
Variation	27 (33%)	40 (30%)	19 (24%)	59 (40%)	25 (28%)
Low	5 (6%)	6 (5%)	4 (5%)	5 (3%)	0 (–)
Not mentioned	30 (37%)	60 (45%)	30 (39%)	55 (37%)	50 (57%)
Total cost of delivery (£)					
Mean (SD)	493,655 (292,416.3)	–	–	–	–
Median	469,467	–	–	–	–
Min : max	70,575: 1,410,000	–	–	–	–
n =	82	–	–	–	–
Total cost of delivery (categorised)					
< £100k	4 (5%)	7 (5%)	0 (–)	6 (4%)	0 (–)
£100k–<£250k	14 (17%)	16 (12%)	12 (15%)	16 (11%)	9 (10%)
£250k–<£500k	28 (34%)	44 (33%)	27 (35%)	53 (36%)	28 (32%)
£500k–<£750k	21 (26%)	33 (25%)	23 (30%)	30 (20%)	17 (19%)
£750k–<£1 million	9 (11%)	15 (11%)	10 (13%)	22 (15%)	21 (24%)
£1 million +	6 (7%)	18 (14%)	6 (8%)	22 (15%)	13 (15%)
Cost per pupil (£)					
Mean (SD)	174 (322.4)	–	–	–	–
Median	54	–	–	–	–
Min : max	1: 1,750	–	–	–	–
n =	82	–	–	–	–
Cost per pupil (categorised)					
<£10	12 (15%)	17 (13%)	7 (9%)	21 (14%)	10 (11%)

£10–<£25	14 (17%)	28 (21%)	15 (19%)	25 (17%)	16 (18%)
£25–<£50	10 (12%)	12 (9%)	14 (18%)	20 (13%)	16 (18%)
£50–<£100	15 (18%)	24 (18%)	20 (26%)	28 (19%)	10 (11%)
£100–<£250	18 (22%)	27 (20%)	11 (14%)	31 (21%)	12 (14%)
£250–<£1000	10 (12%)	20 (15%)	10 (13%)	21 (14%)	24 (27%)
£1,000+	3 (4%)	5 (4%)	1 (1%)	3 (2%)	0 (–)
Language and literacy	38 (46%)	53 (40%)	31 (40%)	68 (46%)	25 (28%)
Staff deployment and development	36 (44%)	46 (35%)	32 (41%)	52 (35%)	25 (28%)
Organising your school	18 (22%)	33 (25%)	21 (27%)	32 (22%)	16 (18%)
Developing effective learners	17 (21%)	23 (17%)	12 (15%)	21 (14%)	19 (22%)
Mathematics	16 (20%)	18 (14%)	14 (18%)	18 (12%)	15 (17%)
Feedback and monitoring pupil progress	10 (12%)	16 (12%)	18 (23%)	26 (17%)	4 (5%)
Behaviour	8 (10%)	16 (12%)	8 (10%)	17 (11%)	16 (18%)
Character and essential life skills	7 (9%)	15 (11%)	2 (3%)	16 (11%)	23 (26%)
Parental engagement	6 (7%)	14 (11%)	6 (8%)	16 (11%)	11 (13%)
Enrichment	4 (5%)	7 (5%)	2 (3%)	10 (7%)	8 (9%)
Science	3 (4%)	3 (2%)	2 (3%)	3 (2%)	4 (5%)
Early years	3 (4%)	4 (3%)	1 (1%)	5 (3%)	0 (–)
Special educational needs and disabilities	2 (2%)	3 (2%)	1 (1%)	6 (4%)	0 (–)
EEF promising intervention					
Classed as promising	17 (21%)	30 (23%)	16 (21%)	35 (24%)	13 (15%)
Not classed as promising	65 (79%)	103 (77%)	62 (80%)	114 (77%)	75 (85%)

* Under this intervention theme, the overall (curriculum) focus of the intervention was extracted from the EEF trial websites. A separate (effect-size-level) measure is also shown under the evaluation design theme below. It is possible that an intervention has a cross-curriculum focus overall with impact measured using separate (distinct) outcomes (e.g., KS2 maths, reading etc) or a composite (cross-curriculum) outcome measure by combining the separate outcomes.
** The EEF intervention themes are taken from the evaluation website on EEF. These categories are not mutually exclusive – a trial might be included in two or more of the themes.

Theory & evidence

Table 2: Distributions of explanatory variables for the theory & evidence theme

	No. of trials (%)	Headline, ITT primary outcome effect sizes (%)	Secondary attainment outcome ITT effect sizes (%)	FSM attainment outcome effect sizes (%)	Psychological outcome effect sizes (%)
All trials	82 (100%)	133 (100%)	78 (100%)	149 (100%)	88 (100%)
Empirical evidence and theoretical detail					
Strength of empirical evidence					
Strong evidence	17 (21%)	31 (23%)	22 (28%)	45 (30%)	10 (11%)
Some evidence	56 (68%)	87 (65%)	49 (63%)	91 (61%)	74 (84%)
Minimal or no evidence	9 (11%)	15 (11%)	7 (9%)	13 (9%)	4 (5%)
Detail on theory behind causal impact					
Highly detailed	17 (21%)	27 (20%)	18 (23%)	26 (17%)	22 (25%)
Some detail	28 (34%)	44 (33%)	29 (37%)	51 (34%)	39 (44%)
Minimal or no detail	37 (45%)	62 (47%)	31 (40%)	72 (48%)	27 (31%)
Empirical* theory intersection					
Strong evidence, detailed theory	5 (6%)	10 (8%)	5 (6%)	12 (8%)	2 (2%)
Strong evidence, limited / no theory	12 (15%)	21 (16%)	17 (22%)	33 (22%)	8 (9%)
Detailed theory, limited / no evidence	12 (15%)	17 (13%)	13 (17%)	14 (9%)	20 (23%)
Some evidence and theory	20 (24%)	29 (22%)	14 (18%)	29 (20%)	31 (35%)
Some evidence, minimal / no theory	25 (30%)	42 (32%)	22 (28%)	49 (33%)	23 (26%)
Some theory, minimal / no evidence	1 (1%)	2 (2%)	0 (–)	2 (1%)	0 (–)
Minimal / no evidence or theory	7 (9%)	12 (9%)	7 (9%)	10 (7%)	4 (5%)
Causal processes and mechanisms					
Direct or training-based					
Training-based	64 (78%)	97 (73%)	63 (81%)	114 (77%)	65 (74%)
Direct	16 (20%)	27 (20%)	8 (10%)	29 (5%)	16 (18%)
Other	2 (2%)	9 (7%)	7 (9%)	6 (4%)	7 (8%)
Focus of change					
Learning focus	69 (84%)	106 (80%)	74 (95%)	119 (77%)	61 (69%)
Teacher change focus	3 (4%)	4 (3%)	2 (3%)	8 (5%)	0 (–)
Wider pupil outcomes focus	9 (11%)	21 (16%)	2 (3%)	20 (13%)	23 (26%)

Other	1 (1%)	2 (2%)	0 (–)	2 (1%)	4 (5%)

Context

Table 3: Distributions of explanatory variables for the context theme

	No. of trials (%)	Headline, ITT primary outcome effect sizes (%)	Secondary attainment outcome ITT effect sizes (%)	FSM attainment outcome effect sizes (%)	Psychological outcome effect sizes (%)
All trials	82 (100%)	133 (100%)	78 (100%)	149 (100%)	88 (100%)
External environment					
Geography					
National	25 (30%)	45 (34%)	22 (28%)	46 (31%)	26 (30%)
One geographical location	19 (23%)	31 (23%)	22 (28%)	36 (24%)	27 (31%)
Two or three geographical areas	22 (27%)	35 (26%)	25 (32%)	42 (28%)	20 (23%)
Other	16 (20%)	22 (17%)	9 (12%)	25 (17%)	15 (17%)
OFSTED					
Mentioned as barrier	16 (20%)	26 (20%)	16 (21%)	30 (20%)	16 (18%)
Not mentioned as barrier	66 (80%)	107 (80%)	62 (80%)	119 (80%)	72 (82%)
Characteristics of participating organisations					
… perceived barriers					
Specialist facilities and space					
Mentioned as barrier	35 (43%)	51 (38%)	28 (36%)	53 (36%)	32 (36%)
Not mentioned as barrier	47 (57%)	82 (62%)	50 (64%)	96 (64%)	56 (64%)
Staff time and availability					
Mentioned as barrier	54 (66%)	88 (66%)	50 (64%)	105 (71%)	57 (65%)
Not mentioned as barrier	28 (34%)	45 (34%)	28 (36%)	44 (30%)	31 (35%)
Workforce capacity					
Mentioned as barrier	31 (38%)	47 (35%)	35 (45%)	51 (34%)	37 (42%)
Not mentioned as barrier	51 (62%)	86 (65%)	43 (55%)	98 (66%)	51 (58%)
… perceived enablers					
Alignment of intervention & existing practice					
Mentioned as enabler	19 (23%)	30 (23%)	15 (19%)	35 (24%)	24 (27%)

Not mentioned as enabler	63 (77%)	103 (77%)	63 (81%)	114 (77%)	64 (73%)
Staff teamwork					
Mentioned as enabler	22 (27%)	31 (23%)	24 (31%)	35 (24%)	16 (18%)
Not mentioned as enabler	60 (73%)	102 (77%)	54 (69%)	114 (77%)	72 (82%)
Characteristics of participating individuals					
… perceived barriers					
Pupil behaviour					
Mentioned as barrier	26 (32%)	34 (26%)	18 (23%)	42 (28%)	43 (49%)
Not mentioned as barrier	56 (68%)	99 (74%)	60 (77%)	107 (72%)	45 (51%)
… perceived barriers & enablers					
SLT buy-in					
Barrier	8 (10%)	8 (6%)	8 (10%)	10 (7%)	7 (8%)
Both barrier & enabler	11 (13%)	17 (13%)	6 (8%)	23 (15%)	23 (26%)
Enabler	18 (22%)	30 (23%)	20 (26%)	34 (23%)	3 (3%)
Not mentioned or unclear	45 (55%)	78 (59%)	44 (56%)	82 (55%)	55 (63%)
Staff expectations and motivations					
Barrier	12 (15%)	18 (14%)	3 (4%)	18 (12%)	12 (14%)
Both barrier & enabler	15 (18%)	20 (15%)	11 (14%)	23 (15%)	15 (17%)
Enabler	18 (22%)	27 (20%)	26 (33%)	30 (20%)	14 (16%)
Not mentioned or unclear	37 (45%)	68 (51%)	38 (49%)	78 (52%)	47 (53%)

Implementation & fidelity

Table 4: Distributions of explanatory variables for the implementation & fidelity theme

	No. of trials (%)	Headline, ITT primary outcome effect sizes (%)	Secondary attainment outcome ITT effect sizes (%)	FSM attainment outcome effect sizes (%)	Psychological outcome effect sizes (%)
All trials	82 (100%)	133 (100%)	78 (100%)	149 (100%)	88 (100%)
Developer characteristics					
Type of developer					
Not for profit / charity	32 (39%)	48 (36%)	30 (39%)	47 (32%)	51 (58%)
University	19 (23%)	42 (32%)	19 (24%)	41 (28%)	14 (16%)

Private company	9 (11%)	13 (10%)	13 (17%)	24 (16%)	8 (9%)
School, academy or MAT	9 (11%)	9 (7%)	6 (8%)	10 (7%)	4 (5%)
Council / LA	8 (10%)	12 (9%)	7 (9%)	15 (10%)	11 (13%)
Mixed	5 (6%)	9 (7%)	3 (4%)	12 (8%)	0 (–)
Planning, time & support					
Clarity of implementation plan					
Clearly understood	33 (40%)	49 (37%)	32 (41%)	54 (36%)	32 (36%)
Variation in understanding	23 (28%)	37 (28%)	30 (39%)	55 (37%)	23 (26%)
Unclear or not mentioned	26 (32%)	47 (35%)	16 (21%)	40 (27%)	33 (38%)
Lead-in time for preparation					
Sufficient time	5 (6%)	8 (6%)	2 (3%)	10 (7%)	4 (5%)
Variation in perception	14 (17%)	22 (17%)	13 (17%)	20 (13%)	22 (25%)
Insufficient time	24 (29%)	34 (26%)	19 (24%)	39 (26%)	5 (6%)
Not mentioned	39 (48%)	69 (52%)	44 (56%)	80 (54%)	45 (51%)
Senior leader support					
Strong	11 (13%)	17 (13%)	12 (15%)	20 (13%)	2 (2%)
Some	22 (27%)	37 (28%)	15 (19%)	43 (29%)	31 (35%)
Limited or minimal	5 (6%)	5 (4%)	5 (6%)	7 (5%)	5 (6%)
Not mentioned	44 (54%)	74 (56%)	46 (59%)	79 (53%)	50 (57%)
Professional development (CPD)					
Is CPD provided to support implementation?					
YES, only to direct implementers	46 (56%)	71 (53%)	47 (60%)	78 (52%)	45 (51%)
YES, only to direct implementers and other stakeholders	30 (37%)	47 (35%)	21 (27%)	59 (40%)	26 (30%)
YES, only to stakeholders who are not direct implementers	1 (1%)	1 (<1%)	0 (–)	1 (<1%)	0 (–)
NO CPD or unclear	5 (6%)	14 (11%)	10 (13%)	11 (7%)	17 (19%)
Is CPD subject / curriculum specific or general?					
Predominantly subject- or curriculum-specific	49 (60%)	63 (47%)	39 (50%)	74 (50%)	47 (53%)
Predominantly generic	22 (27%)	46 (35%)	25 (32%)	53 (36%)	25 (28%)

Mixed generic and subject-specific	7 (9%)	15 (11%)	4 (5%)	16 (11%)	1 (1%)
Not mentioned	4 (5%)	9 (7%)	10 (13%)	6 (4%)	15 (17%)
Sequencing of CPD					
Pre-intervention only	18 (22%)	25 (19%)	15 (19%)	30 (20%)	10 (11%)
During the intervention only	10 (12%)	19 (14%)	15 (19%)	24 (16%)	8 (9%)
Pre and during the intervention	47 (57%)	76 (57%)	36 (46%)	83 (56%)	53 (60%)
Not mentioned	7 (9%)	13 (10%)	12 (15%)	12 (8%)	17 (19%)
Who delivers CPD?					
Delivery partner	53 (65%)	82 (62%)	44 (56%)	91 (61%)	57 (65%)
Another external organisation	8 (10%)	8 (6%)	6 (8%)	10 (7%)	6 (7%)
Leaders / teachers from schools in the trial	1 (1%)	2 (2%)	6 (8%)	2 (1%)	6 (7%)
Mixed	12 (15%)	21 (16%)	10 (13%)	27 (18%)	2 (2%)
Not mentioned	8 (10%)	20 (15%)	12 (15%)	19 (13%)	17 (19%)
Types of CPD (* not mutually exclusive categories, see below)					
Face to face	74 (90%)	119 (89%)	68 (87%)	139 (93%)	69 (78%)
Online	11 (13%)	15 (11%)	5 (6%)	17 (11%)	5 (6%)
Coaching or mentoring	13 (16%)	22 (17%)	9 (12%)	27 (18%)	7 (8%)
Cascade 'train the trainer' model	16 (20%)	24 (18%)	13 (17%)	37 (25%)	14 (16%)
Support & monitoring					
Does delivery partner provide support (other than CPD)?					
Before the intervention only	1 (1%)	2 (2%)	6 (8%)	2 (1%)	6 (7%)
Before and during the intervention	12 (15%)	22 (17%)	11 (14%)	20 (14%)	14 (16%)
During the intervention only	47 (57%)	70 (53%)	38 (49%)	83 (56%)	43 (49%)
Other or not mentioned	22 (27%)	39 (29%)	23 (30%)	44 (30%)	25 (28%)
Monitoring of implementation					
Robust monitoring	14 (17%)	24 (18%)	16 (21%)	38 (26%)	17 (19%)
Some monitoring	28 (34%)	47 (35%)	24 (31%)	51 (34%)	29 (33%)
No monitoring	8 (10%)	10 (8%)	2 (3%)	13 (9%)	5 (6%)
Not mentioned	32 (39%)	52 (39%)	36 (46%)	47 (32%)	37 (42%)
Fidelity					

Intended fidelity					
Faithful adoption	37 (45%)	52 (39%)	36 (46%)	60 (40%)	31 (35%)
Adaptation to context	31 (38%)	57 (43%)	35 (45%)	66 (44%)	44 (50%)
Not mentioned	14 (17%)	24 (18%)	7 (9%)	23 (15%)	13 (15%)
Fidelity related to CPD					
High	12 (15%)	18 (14%)	8 (10%)	22 (15%)	3 (3%)
Varied or moderate	26 (32%)	40 (30%)	24 (31%)	53 (36%)	25 (28%)
Limited	6 (7%)	10 (8%)	12 (15%)	11 (7%)	8 (9%)
Not mentioned	38 (46%)	65 (49%)	34 (44%)	63 (42%)	52 (59%)
Actual fidelity of implementation					
High	13 (16%)	20 (15%)	18 (23%)	25 (17%)	16 (18%)
Varied or moderate	46 (56%)	72 (54%)	34 (44%)	89 (60%)	41 (47%)
Limited	14 (17%)	28 (21%)	16 (21%)	25 (17%)	19 (22%)
Not mentioned	9 (11%)	13 (10%)	10 (13%)	10 (7%)	12 (14%)

Notes: *The categorised types of CPD are not mutually exclusive and so one trial may appear in 2+ categories.

Evaluation design

Table 5: Distributions of explanatory variables for the evaluation design theme

	No. of trials (%)	Headline, ITT primary outcome effect sizes (%)	Secondary attainment outcome ITT effect sizes (%)	FSM attainment outcome effect sizes (%)	Psychological outcome effect sizes (%)
All trials	82 (100%)	133 (100%)	78 (100%)	149 (100%)	88 (100%)
Trial description					
Trial design					
RCT	27 (33%)	41 (31%)	13 (17%)	50 (34%)	23 (26%)
Clustered RCT	55 (67%)	92 (69%)	65 (83%)	99 (66%)	65 (74%)
Level of randomisation					
School	49 (60%)	82 (62%)	56 (72%)	85 (57%)	51 (58%)
Pupil	25 (30%)	34 (26%)	10 (13%)	40 (27%)	12 (14%)
Class or teacher	3 (4%)	4 (3%)	3 (4%)	6 (4%)	8 (9%)
Key Stage or year group	2 (2%)	5 (4%)	6 (8%)	5 (3%)	6 (7%)
Parent	2 (2%)	7 (5%)	3 (4%)	10 (7%)	11 (13%)
Other / complex	1 (1%)	1 (< 1%)	0 (–)	3 (2%)	0 (–)
Type of trial (EEF-defined)					

Efficacy	41 (50%)	69 (52%)	30 (39%)	73 (49%)	39 (44%)
Effectiveness	41 (50%)	64 (48%)	48 (62%)	76 (51%)	49 (56%)
Type of evaluator					
Non-university	30 (37%)	54 (41%)	19 (24%)	55 (37%)	29 (33%)
University	52 (63%)	79 (59%)	59 (76%)	94 (63%)	59 (67%)
Trial length and size					
Length of trial (weeks)					
Mean (SD)	38 (30.0)	–	–	–	–
Median	30	–	–	–	–
Min : max	4 : 97	–	–	–	–
***n* =**	82	–	–	–	–
Length of trial (categorised)					
15 weeks or less (1 term)	23 (28%)	37 (28%)	17 (22%)	45 (30%)	29 (33%)
Above 15–30 weeks (2 terms)	21 (26%)	31 (23%)	16 (21%)	31 (21%)	22 (25%)
31–45 weeks (3 terms / one year)	21 (26%)	39 (29%)	18 (23%)	34 (23%)	24 (27%)
46+ weeks–more than one year	17 (21%)	26 (20%)	27 (35%)	39 (26%)	13 (15%)
Number of schools					
Mean (SD)	64 (47.5)	–	–	–	–
Median	51	–	–	–	–
Min : max	3 : 205	–	–	–	–
***n* =**	82	–	–	–	–
Number of schools (categorised)					
20 or less	15 (18%)	21 (16%)	4 (5%)	27 (18%)	4 (5%)
21–40	16 (20%)	26 (20%)	18 (23%)	25 (17%)	13 (15%)
41–60	16 (20%)	30 (23%)	11 (14%)	31 (21%)	13 (15%)
61–80	8 (10%)	18 (14%)	3 (4%)	19 (13%)	14 (16%)
81–100	10 (12%)	15 (11%)	12 (15%)	24 (16%)	30 (34%)
101 or more	17 (21%)	23 (17%)	30 (39%)	23 (15%)	14 (16%)
Number of pupils					
Mean (SD)	3,696 (4,969.1)	–	–	–	–
Median	2,006	–	–	–	–
Min : max	36 : 25,000	–	–	–	–
***n* =**	80	–	–	–	–
Number of pupils (categorised)					

500 or less	19 (23%)	25 (19%)	11 (14%)	26 (18%)	6 (7%)
501–1,000	12 (15%)	16 (12%)	6 (8%)	23 (16%)	18 (21%)
1,001–2,500	15 (18%)	27 (20%)	18 (23%)	25 (17%)	13 (16%)
2,501–5,000	14 (17%)	23 (17%)	10 (13%)	28 (19%)	26 (31%)
5,001 or more	20 (24%)	39 (29%)	32 (42%)	43 (30%)	21 (25%)
Statistical sensitivity, attrition and trial quality					
Reported MDES					
Mean (SD)	0.22 (0.081)				
Median	0.20				
Min : max	0.07 : 0.45				
***n* =**	78				
Categorised MDES					
Lower than 0.15 SD	11 (13%)	21 (16%)	17 (22%)	20 (30%)	23 (29%)
0.15 to lower than 0.25 SD	33 (40%)	56 (42%)	30 (40%)	68 (49%)	32 (36%)
0.25 to lower than 035 SD	28 (34%)	40 (30%)	22 (29%)	44 (31%)	20 (25%)
0.35 SD or higher	6 (7%)	9 (7%)	7 (9%)	8 (6%)	5 (6%)
Pupil-level % attrition					
Mean (SD)	19.4 (16.54)	–	–	–	–
Median	15.2	–	–	–	–
Min : max	0 : 75	–	–	–	–
n	79	–	–	–	–
Pupil-level % attrition categorised					
Zero	2 (2%)	8 (6%)	0 (–)	6 (4%)	4 (5%)
<10%	25 (30%)	37 (28%)	36 (46%)	55 (37%)	35 (40%)
10% to < 20%	24 (29%)	37 (28%)	13 (17%)	43 (29%)	34 (39%)
20% to < 30%	14 (17%)	22 (17%)	16 (21%)	21 (14%)	10 (11%)
30%+	17 (21%)	29 (22%)	13 (17%)	24 (16%)	5 (6%)
Trial quality (EEF padlock rating) treated as scale					
Mean (SD)	3.1 (1.25)	–	–	–	–
Median	3.0	–	–	–	–
Min : Max	0 : 5	–	–	–	–
***n* =**	82	–	–	–	–
Trial quality (EEF padlock rating)					
0	3 (4%)	4 (3%)	3 (4%)	3 (2%)	0 (–)

1	7 (9%)	9 (7%)	3 (4%)	8 (5%)	1 (1%)
2	13 (16%)	30 (23%)	15 (19%)	22 (15%)	22 (25%)
3	27 (33%)	44 (33%)	26 (33%)	49 (33%)	29 (33%)
4	23 (28%)	35 (26%)	14 (18%)	50 (34%)	36 (41%)
5	9 (11%)	11 (8%)	17 (22%)	17 (11%)	0 (–)
Evaluation burden					
Testing burden					
Low (just NPD)	9 (11%)	19 (14%)	15 (19%)	16 (11%)	0 (–)
Medium (one external test)	24 (29%)	35 (26%)	19 (24%)	43 (29%)	15 (17%)
High (two or more external tests)	49 (60%)	79 (59%)	44 (56%)	90 (60%)	73 (83%)
IPE data collection burden					
Low (no surveys or interviews)	12 (15%)	16 (12%)	6 (8%)	16 (11%)	6 (7%)
Medium (just surveys or just interviews)	27 (33%)	57 (43%)	22 (28%)	71 (48%)	42 (48%)
High (interviews and surveys)	43 (52%)	60 (45%)	50 (64%)	62 (42%)	40 (46%)
Overall (IPE and testing) burden					
Low / medium IPE and testing activity / burden	12 (15%)	24 (18%)	9 (12%)	28 (19%)	10 (11%)
Low / medium IPE but high testing	27 (33%)	49 (37%)	19 (24%)	59 (40%)	38 (43%)
High IPE but low / medium testing	21 (26%)	30 (23%)	25 (32%)	31 (21%)	5 (6%)
High IPE and testing activity / burden	22 (27%)	30 (23%)	25 (32%)	31 (21%)	35 (40%)

Table 6: Distributions of explanatory variables for the evaluation design theme

- **– Primary outcome subtheme**
- **– Variables measured at both trial and effect size levels**

Trial or ES level?		Number of trials (%)	Number of effect sizes (%)
Trial or ES level?	All trials	N_T = 82 (100%)	N_{ES} = 133 (100%)
	Number of primary outcomes		
Trial level	One	50 (61%)	50 (38%)
Trial level	Two	22 (27%)	44 (33%)
Trial level	Three or more	10 (12%)	39 (29%)
	Alignment between intervention focus and primary outcome		
Trial level	Direct match	47 (57%)	60 (45%)
Trial level	Associated match	25 (30%)	43 (32%)
Trial level	Limited match	10 (12%)	30 (23%)
	Types of primary outcome (simple)		
Effect size level	Commercial	51 (62%)	79 (59%)
Effect size level	Official / SATs	22 (27%)	45 (34%)
Effect size level	Other / mixed	9 (11%)	9 (7%)
	Types of primary outcome (detailed)		
	… Commercial:		
Effect size level	GL Assessment	33 (40%)	46 (35%)
Effect size level	CEM	11 (13%)	20 (15%)
Effect size level	Hodder	7 (9%)	8 (6%)
Effect size level	Pearson	2 (2%)	5 (4%)
	… Official / SATs:		
Effect size level	KS1	1 (1%)	1 (1%)
Effect size level	KS2	15 (18%)	30 (23%)
Effect size level	KS3	2 (2%)	3 (2%)
Effect size level	KS4	5 (6%)	11 (8%)
	Types of primary outcome (very detailed)		
	Commercial:		
Effect size level	**GL Assessment:**		
Effect size level	NGRT	–	23 (17%)
Effect size level	PiE / PTE	–	13 (10%)

	PiM / PTM	–	7 (5%)
	CEM:		
	InCAS maths	–	7 (5%)
	InCAS reading	–	5 (4%)
	InCAS combined reading and maths	–	4 (3%)
	Hodder:		
	HGRT	–	4 (3%)
	Other commercial	–	16 (12%)
	Official / NPD:		
	KS2 maths	–	9 (7%)
	KS2 reading	–	5 (4%)
	KS2 writing	–	5 (4%)
	GCSE maths	–	3 (2%)
	GCSE English	–	3 (2%)
	GCSE overall	–	3 (2%)
	Other official / NPD	–	17 (13%)
	Other / mixed	–	9 (7%)
	Primary outcome curriculum area*		
	Cross-curriculum	–	11 (8%)
Effect size level	English / literacy	–	77 (58%)
	Maths / numeracy	–	38 (29%)
	Science	–	7 (5%)
	Trial / ES levels reconciled		
	Cross- curriculum trial and outcome(s)	8 (10%)	11 (8%)
Trial-level	Cross- curriculum trial, multiple subject outcomes	16 (20%)	50 (38%)
	English / literacy trial and outcome(s)	40 (49%)	53 (40%)
	Maths / numeracy trial and outcome(s)	15 (18%)	16 (12%)

	Science trial and outcome(s)	3 (4%)	3 (2%)

Tables for meta-analyses of primary ITT effect sizes

Reported effect size for headline ITT analyses of primary outcome

There are a total of 133 effect sizes for headline ITT analyses of primary outcome(s) across the 82 trials. 50 trials reported a single primary outcome effect size, 22 reported two effect sizes and 10 report three or more effect sizes.

This section presents the analyses of these 133 effect sizes across explanatory variables in each of the five themes of the review's theoretical framework.

The analyses are summarised statistically using five tables (one for each of the themes). To highlight a finding of interest, the weighted mean effect size is shown in red text where it differs notably from the other categories within that variable.

- Table 7 presents the average effect sizes across categories of explanatory variables included in the intervention theme using unweighted mean and median statistics. Alongside these descriptive unweighted statistics, meta-analysis means and standard errors are shown, along with 95% confidence intervals.
- Table 8 uses the same approach to present average effect sizes across categories of explanatory variables included in the theory & evidence theme.
- Table 9 uses the same approach to present average effect sizes across categories of explanatory variables included in the context theme.
- Table 10 uses the same approach to present average effect sizes across categories of explanatory variables included in the implementation & fidelity theme.
- Table 11 uses the same approach to present average effect sizes across categories of explanatory variables included in the evaluation design theme.

Table 7: Primary ITT effect size and the intervention

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SD)	95% CI
All trials	133	+0.03	+0.06 (0.128)	+0.04 (0.01)	+0.03; +0.06
Focus of intervention					
School phase (#)		$p > 0.10$	$eta^2 = 0.05$**	$p > 0.10$	
Primary (including Early Years)	88	+0.03	+0.05 (0.09)	+0.04 (0.01)	+0.03 : +0.06
Primary–secondary transition	7	+0.13	+0.19 (0.27)	+0.12 (0.07)	−0.01 : +0.25
Secondary	38	+0.03	+0.06 (0.15)	+0.04 (0.02)	0.00 : +0.08
Key Stage (#)		$p > 0.10$	$eta^2 = 0.08$	$p > 0.10$	
Early Years	2	–	–	–	–
Primary KS1	23	+0.08	+0.09 (0.09)	+0.08 (0.02)	+0.04 : +0.11
Primary KS2	57	+0.02	+0.04 (0.09)	+0.03 (0.01)	+0.01 : +0.05
Primary (multiple Key Stages)	6	+0.01	+0.03 (0.09)	+0.03 (0.04)	−0.05 : +0.12
Primary–secondary transition	7	+0.13	+0.19 (0.27)	+0.12 (0.07)	−0.01 : +0.25
Secondary KS3	26	0.00	+0.08 (0.18)	+0.06 (0.03)	−0.01 : +0.12

Secondary KS4	9	+0.04	+0.02 (0.06)	+0.04 (0.02)	+0.01 : +0.07
Secondary (multiple Key Stages)	3	–	–	–	–
Curriculum focus of intervention (*)		p = 0.09*	eta^2 = 0.08**	p = 0.09*	
Cross-curriculum	67	+0.02	+0.03 (0.08)	+0.03 (0.01)	+0.01 : +0.04
English	48	+0.08	+0.11 (0.17)	+0.08 (0.02)	+0.04 : +0.12
Maths	15	+0.08	+0.07 (0.08)	+0.05 (0.02)	+0.01 : +0.08
Science	3	–	–	–	–
Intensity of intervention					
Intensity of intervention (#)		p > 0.10	eta^2 = 0.00	p > 0.10	
30 minutes or less per week	16	+0.02	+0.07 (0.12)	+0.04 (0.02)	0.00 : +0.08
31–60 minutes per week	21	+0.03	+0.05 (0.09)	+0.05 (0.02)	+0.01 : +0.08
61–120 minutes per week	27	+0.01	+0.06 (0.17)	+0.04 (0.03)	−0.02 : +0.10
Over 120 minutes per week	16	+0.04	+0.06 (0.13)	+0.04 (0.03)	−0.02 : +0.09
No intensity details	53	+0.05	+0.07 (0.13)	+0.05 (0.01)	+0.03; +0.07
Who implements with direct target?					
Direct implementers (***)		p = 0.02**	eta^2 = 0.11**	p < 0.01***	
Teacher-led	57	+0.03	+0.05 (0.13)	+0.03 (0.01)	+0.01 : +0.04
Externally-led (e.g., delivery partner)	30	+0.01	+0.04 (0.13)	+0.02 (0.02)	−0.02 : +0.05
TA-led	15	+0.18	+0.18 (0.14)	+0.17 (0.04)	+0.10 : +0.25
Parent-led	7	+0.02	+0.02 (0.04)	+0.03 (0.03)	−0.04 : +0.09
Resource-led	2	–	–	–	–
Other school staff-led	2	–	–	–	–
Other	20	+0.07	+0.05 (0.08)	+0.06 (0.02)	+0.03 : +0.09
Supporting resources					
Perceived quality of supporting resources (#)		p > 0.10	eta^2 = 0.00	p > 0.10	
High	27	+0.07	+0.07 (0.10)	+0.06 (0.02)	+0.02 : +0.09
Variation	40	+0.01	+0.07 (0.15)	+0.05 (0.02)	+0.01 : +0.09
Low	6	+0.04	+0.06 (0.09)	+0.03 (0.03)	−0.03 : +0.08
Not mentioned	60	+0.03	+0.06 (0.13)	+0.04 (0.01)	+0.02 : +0.06
Cost					
Total cost of delivery (***)		p < 0.01***	eta^2 = 0.11**	p < 0.01***	
<£100k	7	+0.03	+0.05 (0.13)	+0.01 (0.03)	−0.06 : +0.08
£100k–<£250k	16	+0.04	+0.07 (0.10)	+0.02 (0.01)	−0.01 : +0.05
£250k–<£500k	44	+0.10	+0.11 (0.15)	+0.09 (0.01)	+0.06 : +0.11

£500k–<£750k	33	+0.01	+0.06 (0.13)	+0.05 (0.02)	+0.01 : +0.09
£750k–<£1 million	15	+0.01	0.00 (0.06)	0.00 (0.02)	−0.03 : +0.04
£1 million+	18	+0.01	+0.01 (0.05)	+0.01 (0.01)	−0.01 : +0.04
Cost per pupil (**)		$p > 0.10$	$eta^2 = 0.05$	$p = 0.02$**	
<£10	17	+0.03	+0.06 (0.08)	+0.04 (0.01)	+0.01 : +0.06
£10–<£25	28	+0.02	+0.04 (0.10)	+0.02 (0.02)	−0.01 : +0.05
£25–<£50	12	+0.02	+0.06 (0.10)	+0.07 (0.03)	+0.01 : +0.12
£50–<£100	24	+0.04	+0.08 (0.17)	+0.04 (0.01)	+0.02 : +0.06
£100–<£200	27	+0.02	+0.05 (0.13)	+0.02 (0.02)	−0.01 : +0.05
£200–<£1,000	20	+0.04	+0.08 (0.15)	+0.08 (0.03)	+0.01 : +0.14
£1,000+	5	+0.21	+0.20 (0.13)	+0.20 (0.07)	+0.06 : +0.34
EEF promising interventions					
Whether classed as promising (***)		$p < 0.01$***	$eta^2 = 0.13$***	$p < 0.01$***	
Classed as promising	30	+0.14	+0.15 (0.08)	+0.12 (0.02)	+0.09; +0.15
Not classed as promising	103	+0.01	+0.04 (0.13)	+0.01 (0.01)	0.00; +0.03

EEF school themes (*see below)	*n* / *n*'	Median ES / ES' [diff]	Unweighted mean ES / ES' [diff]	Weighted mean difference (SE)	95% CI for weighted mean difference
Language and literacy	53 / 80	0.09 / 0.02 [+0.07]	0.11 / 0.03 [+0.07]	+0.08 (0.03)	+0.05 : +0.12
Staff deployment and development	46 / 87	0.05 / 0.03 [+0.02]	0.08 / 0.05 [+0.03]	+0.06 (0.01)	+0.03 : +0.09
Organising your school	33 / 100	0.03 / 0.03 [0.00]	0.06 / 0.06 [0.00]	+0.05 (0.02)	+0.02 : +0.09
Developing effective learners	23 / 110	0.03 / 0.03 [0.00]	0.08 / 0.06 [+0.02]	+0.05 (0.02)	+0.01 : +0.08
Mathematics	18 / 115	0.09 / 0.03 [+0.06]	0.09 / 0.06 [+0.03]	+0.05 (0.02)	+0.02 : +0.06
Feedback and monitoring pupil progress	16 / 117	0.03 / 0.03 [0.00]	0.09 / 0.06 [+0.03]	+0.02 (0.02)	−0.01 : +0.06
Behaviour	16 / 117	0.03 / 0.03 [0.00]	0.02 / 0.07 [−0.05]	+0.02 (0.02)	−0.01 : +0.06
Character and essential life skills	15 / 118	0.01 / 0.03 [−0.02]	0.01 / 0.07 [−0.06]	0.00 (0.02)	−0.04 : +0.04
Parental engagement	14 / 119	0.01 / 0.03 [−0.02]	0.01 / 0.07 [−0.06]	+0.03 (0.01)	0.00 : +0.05
Enrichment	7 / 126	0.01 / 0.03 [−0.02]	0.02 / 0.07 [−0.05]	+0.01 (0.05)	−0.06 : +0.08
Science	3 / 130	–	–	–	–
Early Years	4 / 129	0.13 / 0.03 [+0.10]	0.13 / 0.06 [+0.07]	+0.11 (0.06)	−0.01 : +0.23
Special educational	3 / 130	–	–	–	–

needs and disabilities

* **Key:** The EEF school themes are not mutually exclusive: effect sizes for a particular trial can be included in more than one of the school themes. Table 7 takes each school theme to illustrate the number of effect sizes included (and not included) in each along. The table also shows unweighted averages (median and mean) for effect sizes included and not included. Finally, the table shows the weighted mean difference in effect sizes (included–not included) from the meta-analyses.

Label	Details
n / *n*':	• *n* is the number of effect sizes attached to a specific EEF school theme • *n*' is the number of effect sizes not included
med / med' [diff]	• med is the unweighted median effect size for trials included in a specific EEF school theme • med' is the unweighted median effect size for trials not included • [diff] is the difference between the two medians (i.e., med – med')
mean / mean' [diff]	• mean is the unweighted mean effect size for trials included in a specific EEF school theme • mean' is the unweighted mean effect size for trials not included • [diff] is the difference between the two mean (i.e., mean – mean')
Weighted mean difference (SE)	• This is the weighted mean difference between effect sizes included in a specific EEF school theme and effect sizes not included obtained from the meta-analyses. • (SE) is the standard error of the weighted mean difference.

Table 8: Primary ITT effect size and theory & evidence

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	133	+0.03	+0.06 (0.128)	0.04 (0.01)	+0.03; +0.06
Empirical evidence and theoretical detail					
Strength of empirical evidence (*)		$p > 0.10$	$eta^2 = 0.01$	$p = 0.06$*	
Strong evidence	31	+0.01	+0.07 (0.14)	+0.06 (0.02)	+0.01 : +0.11
Some evidence	87	+0.04	+0.07 (0.13)	+0.04 (0.01)	+0.03 : +0.06
Minimal or no evidence	15	0.00	+0.03 (0.11)	−0.02 (0.02)	−0.06 : +0.03
Theoretical detail		$p > 0.10$	$eta^2 = 0.01$	$p > 0.10$	
Highly detailed	27	+0.03	+0.04 (0.09)	+0.03 (0.01)	0.00 : +0.06
Some detail	44	+0.02	+0.07 (0.16)	+0.05 (0.02)	+0.02 : +0.08
Minimal or no detail	62	+0.03	+0.06 (0.11)	+0.05 (0.01)	+0.02 : +0.07
Causal processes					
Direct or training-based		$p > 0.10$	$eta^2 = 0.01$	$p > 0.10$	
Training-based	97	+0.03	+0.07 (0.13)	+0.04 (0.01)	+0.03 : +0.06
Direct	27	+0.01	+0.04 (0.12)	+0.02 (0.02)	−0.02 : +0.06
Other	9	+0.10	+0.08 (0.09)	+0.10 (0.02)	+0.06 : +0.14
Focus of change (***)		$p = 0.01$**	$eta^2 = 0.07$**	$p < 0.01$***	
Learning focus	106	+0.05	+0.08 (0.14)	+0.06 (0.01)	+0.04 : +0.08
Teacher change focus	4	−0.02	−0.01 (0.03)	−0.01 (0.02)	−0.06 : +0.04

Wider pupil outcomes focus	21	+0.02	+0.01 (0.07)	+0.02 (0.01)	0.00 : +0.04
Other	2	–	–	–	–

Table 9: Primary ITT effect sizes and evaluation context

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	133	+0.03	+0.06 (0.128)	+0.04 (0.01)	+0.03; +0.06
External context					
Geography (*)		*p* = 0.05*	eta^2 = 0.09***	*p* = 0.06*	
National	45	+0.02	+0.02 (0.07)	+0.03 (0.01)	+0.01 : +0.05
One geographical location	31	+0.05	+0.12 (0.19)	+0.09 (0.03)	+0.03 : +0.14
Two or three geographical areas	35	+0.09	+0.08 (0.10)	+0.06 (0.02)	+0.03 : +0.09
Other	22	+0.01	+0.04 (0.12)	+0.01 (0.02)	−0.03 : +0.05
OFSTED (#)		*p* = 0.05**	eta^2 = 0.03**	*p* > 0.10	
Mentioned as barrier	26	+0.01	+0.02 (0.08)	+0.03 (0.02)	0.00 : +0.06
Not mentioned as barrier	107	+0.04	+0.07 (0.14)	+0.05 (0.01)	+0.03 : +0.07
Characteristics of participating organisations					
… perceived barriers					
Specialist facilities and space (***)		*p* = 0.08*	eta^2 = 0.03**	*p* < 0.01***	
Mentioned as barrier	51	+0.07	+0.09 (0.14)	+0.08 (0.02)	+0.05 : +0.12
Not mentioned as barrier	82	+0.03	+0.05 (0.12)	+0.03 (0.01)	+0.01 : +0.04
Staff time and availability		*p* > 0.10	eta^2 = 0.02	p > 0.10	
Mentioned as barrier	88	+0.03	+0.05 (0.11)	+0.04 (0.01)	+0.02 : +0.06
Not mentioned as barrier	45	+0.03	+0.09 (0.16)	+0.06 (0.02)	+0.02 : +0.09
Workforce capacity		*p* > 0.10	eta^2 = 0.00	*p* > 0.10	
Mentioned as barrier	47	+0.03	+0.06 (0.13)	+0.05 (0.02)	+0.02 : +0.09
Not mentioned as barrier	86	+0.03	+0.07 (0.13)	+0.04 (0.01)	+0.02 : +0.05
… perceived enablers					
Alignment of intervention and existing practice (*)		*p* = 0.04**	eta^2 = 0.04**	*p* = 0.10*	
Mentioned as enabler	30	+0.01	+0.02 (0.08)	+0.02 (0.01)	−0.01 : +0.04
Not mentioned as enabler	103	+0.04	+0.08 (0.14)	+0.05 (0.01)	+0.03 : +0.07
Staff teamwork (*)		*p* > 0.10	eta^2 = 0.03**	*p* = 0.08*	
Mentioned as enabler	31	+0.06	+0.10 (0.18)	+0.08 (0.02)	+0.03 : +0.13
Not mentioned as enabler	102	+0.03	+0.05 (0.10)	+0.03 (0.01)	+0.02 : +0.05
Characteristics of participating individuals					
… perceived barriers					

Pupil behaviour		*p* > 0.10	eta^2 = 0.00	*p* > 0.10	
Mentioned as barrier	34	+0.03	+0.06 (0.13)	+0.04 (0.02)	0.00 : +0.07
Not mentioned as barrier	99	+0.03	+0.06 (0.13)	+0.05 (0.01)	+0.03 : +0.06
… perceived barriers and enablers					
SLT buy-in		*p* > 0.10	eta^2 = 0.01	*p* > 0.10	
Mentioned as barrier	8	+0.01	+0.08 (0.17)	+0.01 (0.02)	−0.03 : +0.06
Mentioned as both barrier and enabler	17	0.00	+0.02 (0.09)	0.00 (0.02)	−0.03 : +0.04
Mentioned as enabler	30	+0.01	+0.07 (0.14)	+0.06 (0.02)	+0.01 : +0.11
Not mentioned	78	+0.04	+0.07 (0.13)	+0.05 (0.01)	+0.03 : +0.07
Staff expectations and motivations		*p* > 0.10	eta^2 = 0.01	*p* > 0.10	
Mentioned as barrier	18	+0.02	+0.04 (0.12)	+0.04 (0.02)	0.00 : +0.08
Mentioned as both barrier and enabler	20	+0.02	+0.04 (0.09)	+0.01 (0.02)	−0.02 : +0.05
Mentioned as enabler	27	+0.09	+0.08 (0.09)	+0.06 (0.02)	+0.03 : +0.09
Not mentioned	68	+0.03	+0.07 (0.15)	+0.05 (0.01)	+0.02 : +0.07

Table 10: Primary ITT effect size and implementation & fidelity

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SD)	95% CI
All trials	133	+0.03	+0.06 (0.128)	+0.04 (0.01)	+0.03; +0.06
Developer characteristics					
Type of developer (**)		*p* = 0.05**	eta^2 = 0.07	*p* = 0.05**	
Not for profit / charity	48	+0.03	+0.05 (0.11)	+0.03 (0.01)	+0.01; +0.06
University	42	+0.02	+0.05 (0.12)	+0.03 (0.01)	+0.01; +0.06
Private company	13	+0.01	+0.10 (0.22)	+0.05 (0.03)	−0.01; +0.11
School, academy or MAT	9	+0.12	+0.17 (0.11)	+0.15 (0.04)	+0.07; +0.23
Council / LA	12	+0.03	+0.04 (0.08)	+0.02 (0.02)	−0.02; +0.06
Mixed	9	+0.10	+0.09 (0.11)	+0.09 (0.03)	+0.03; +0.15
Planning, time and support					
Clarity of implementation plan		*p* > 0.10	eta^2 = 0.00	*p* > 0.10	
Clearly understood	49	+0.04	+0.06 (0.10)	+0.04 (0.01)	+0.02 : +0.06
Variation in understanding	37	+0.01	+0.06 (0.14)	+0.04 (0.02)	0.00 : +0.08
Unclear or not mentioned	47	+0.06	+0.07 (0.14)	+0.05 (0.01)	+0.02 : +0.07
Lead in time for preparation (#)		*p* > 0.10	eta^2 = 0.01	*p* > 0.10	
Sufficient time	8	−0.01	+0.06 (0.13)	+0.05 (0.04)	−0.03 : +0.13

Variation in perceptions	22	+0.01	+0.05 (0.14)	+0.04 (0.02)	−0.01 : +0.08
Insufficient time	34	+0.04	+0.05 (0.10)	+0.03 (0.01)	+0.01 : +0.05
Not mentioned	69	+0.04	+0.07 (0.14)	+0.05 (0.01)	+0.03 : +0.07
Senior leadership support		p = 0.08*	eta^2 = 0.03	p > 0.10	
Strong	17	+0.02	+0.06 (0.12)	+0.06 (0.03)	0.00 : +0.11
Some	37	0.00	+0.03 (0.12)	+0.02 (0.02)	−0.01 : +0.05
Limited or minimal	5	+0.12	+0.13 (0.20)	+0.08 (0.06)	−0.04 : +0.20
Not mentioned	74	+0.05	+0.08 (0.13)	+0.05 (0.01)	+0.03 : +0.07
Professional development (CPD)					
Is CPD provided to support implementation?		p > 0.10	eta^2 = 0.01	p > 0.10	
YES, only to direct implementers	71	+0.03	+0.07 (0.14)	+0.05 (0.01)	+0.03 : +0.06
YES, to implementers and other stakeholders	47	+0.01	+0.05 (0.11)	+0.04 (0.01)	+0.02 : +0.07
YES, only to other stakeholders	1	–	–	–	–
NO CPD or unclear	14	+0.05	+0.05 (0.11)	+0.02 (0.03)	−0.03 : +0.08
Is CPD subject / curriculum- specific or general? (**)		p < 0.01***	eta^2 = 0.07**	p = 0.02**	
Predominantly subject-specific or curriculum-specific	63	+0.08	+0.09 (0.13)	+0.07 (0.02)	+0.04 : +0.10
Predominantly generic	46	+0.02	+0.03 (0.13)	+0.02 (0.01)	0.00 : +0.04
Mixed generic and subject-specific	15	+0.01	+0.03 (0.09)	+0.03 (0.02)	0.00 : +0.06
Not mentioned	9	+0.10	+0.11 (0.09)	+0.09 (0.03)	+0.03 : +0.14
Types of CPD (see note below)					
Face-to-face	119	+0.03	+0.06 (0.13)	+0.04 (0.01)	+0.02 : +0.05
Online	15	0.00	+0.03 (0.10)	+0.03 (0.02)	−0.02 : +0.07
Coaching or mentoring	22	+0.03	+0.05 (0.11)	+0.02 (0.02)	−0.01 : +0.06
Cascade 'train the trainer' model	24	+0.02	+0.05 (0.11)	+0.04 (0.02)	+0.01 : +0.08
Sequencing of CPD (#)		p > 0.10	eta^2 = 0.02	p > 0.10	
Pre-intervention only	25	+0.11	+0.09 (0.15)	+0.09 (0.03)	+0.03 : +0.15
During the intervention only	19	+0.01	+0.03 (0.09)	+0.02 (0.02)	−0.02 : +0.05
Pre-intervention and during the intervention	76	+0.03	+0.06 (0.13)	+0.03 (0.01)	+0.02 : +0.05
Not mentioned	13	+0.08	+0.06 (0.11)	+0.05 (0.03)	−0.02 : +0.11
Who delivers CPD?		p > 0.10	eta^2 = 0.04	p > 0.10	
Delivery partner	82	+0.03	+0.06 (0.12)	+0.04 (0.01)	+0.02 : +0.06
Other external organisation	8	+0.06	+0.16 (0.25)	+0.07 (0.03)	+0.01 : +0.12
Leaders / teachers from schools in the trial	2	–	–	–	–
Mixed	21	+0.02	+0.05 (0.11)	+0.05 (0.02)	+0.01 : +0.09

Not mentioned	20	+0.06	+0.06 (0.11)	+0.05 (0.02)	+0.01 : +0.1
Support and monitoring					
Does delivery partner provide support (other than CPD)?		$p > 0.10$	$eta^2 = 0.01$	$p > 0.10$	
Before the intervention only	2	–	–	–	–
During the intervention only	22	+0.08	+0.06 (0.11)	+0.05 (0.02)	0.00 : +0.09
Before and during the intervention	70	+0.03	+0.08 (0.15)	+0.05 (0.01)	+0.03 : +0.07
Other or not mentioned	39	+0.02	+0.05 (0.10)	+0.03 (0.01)	+0.01 : +0.06
Monitoring of implementation (***)		$p < 0.01$***	$eta^2 = 0.05$	$p < 0.01$***	
Robust monitoring	24	+0.03	+0.06 (0.12)	+0.04 (0.02)	0.00 : +0.07
Some monitoring	47	0.00	+0.03 (0.14)	+0.01 (0.01)	−0.01 : +0.03
No monitoring	10	+0.06	+0.08 (0.13)	+0.06 (0.04)	−0.01 : +0.13
Not mentioned	52	+0.08	+0.09 (0.12)	+0.09 (0.01)	+0.06 : +0.11
Fidelity					
Intended fidelity		$p > 0.10$	$eta^2 = 0.03$	$p > 0.10$	
Faithful adoption	52	+0.02	+0.05 (0.12)	+0.02 (0.01)	0.00 : +0.04
Adaptation to context	57	+0.02	+0.05 (0.13)	+0.04 (0.01)	+0.02 : +0.07
Not mentioned	24	+0.09	+0.11 (0.15)	+0.09 (0.03)	+0.04 : +0.15
Fidelity related to CPD (**)		$p = 0.09$*	$eta^2 = 0.03$	$p = 0.04$**	
High	18	+0.11	+0.10 (0.11)	+0.09 (0.03)	+0.04 : +0.14
Varied or moderate	40	+0.02	+0.06 (0.16)	+0.03 (0.01)	0.00 : +0.05
Limited	10	0.00	0.00 (0.07)	0.00 (0.02)	−0.04 : +0.04
Not mentioned	65	+0.03	+0.06 (0.12)	+0.05 (0.01)	+0.02 : +0.07
Actual fidelity of implementation (#)		$p > 0.10$	$eta^2 = 0.04$	$p > 0.10$	
High	20	+0.05	+0.07 (0.10)	+0.05 (0.02)	+0.01 : +0.08
Varied or moderate	72	+0.01	+0.06 (0.14)	+0.03 (0.01)	+0.01 : +0.05
Limited	28	+0.04	+0.05 (0.09)	+0.04 (0.01)	+0.02 : +0.07
Not mentioned	13	+0.10	+0.13 (0.16)	+0.12 (0.05)	+0.03; +0.21

Table 11: Primary ITT effect size and evaluation design

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	133	+0.03	+0.06 (0.128)	+0.04 (0.01)	+0.03 : +0.06
Trial description					
Trial design (***)		$p < 0.01$***	$eta^2 = 0.06$***	$p < 0.01$***	
RCT	41	+0.08	+0.11 (0.15)	+0.10 (0.02)	+0.05; +0.14
Clustered RCT	92	+0.02	+0.04 (0.11)	+0.03 (0.01)	+0.01; +0.04
Level of randomisation (**)		$p = 0.05$**	$eta^2 = 0.10$**	$p = 0.02$**	
School	81	+0.02	+0.04 (0.12)	+0.03 (0.01)	+0.01 : +0.05
Pupil	34	+0.13	+0.13 (0.16)	+0.11 (0.03)	+0.06 : +0.17
Class or teacher	5	0.00	+0.03 (0.11)	+0.03 (0.05)	−0.07 : +0.12
Key Stage or year group	5	+0.03	+0.02 (0.08)	+0.02 (0.02)	−0.02 : +0.07
Parent	7	+0.02	+0.02 (0.04)	+0.03 (0.03)	−0.04 : +0.09
Complex (multiple)	1	–	–	–	–
Type of trial		$p > 0.10$	$eta^2 = 0.03$*	$p > 0.10$	
Efficacy	69	+0.03	+0.04 (0.09)	+0.04 (0.01)	+0.02 : +0.06
Effectiveness	64	+0.03	+0.09 (0.16)	+0.05 (0.01)	+0.03 : +0.08
Type of evaluator		$p > 0.10$	$eta^2 = 0.00$	$p > 0.10$	
Non-University	54	+0.02	+0.06 (0.12)	+0.05 (0.01)	+0.03 : +0.08
University	79	+0.03	+0.07 (0.13)	+0.04 (0.01)	+0.02 : +0.05
Trial length and size					
Length of trial (#)		$p = 0.06$*	$eta^2 = 0.05$*	$p > 0.10$	
Up to 15 weeks (one term)	37	+0.05	+0.07 (0.12)	+0.04 (0.01)	+0.01 : +0.06
16–30 weeks (two terms)	31	+0.07	+0.09 (0.15)	+0.08 (0.02)	+0.03 : +0.13
31–45 weeks (three terms / one year)	39	+0.04	+0.07 (0.14)	+0.05 (0.01)	+0.02 : +0.08
More than 45 weeks / three terms / one year	26	0.00	+0.01 (0.07)	+0.01 (0.01)	−0.01 : +0.03
Number of schools in trial (#)		$p > 0.10$	$eta^2 = 0.09$**	$p > 0.10$	
20 or less	21	+0.03	+0.09 (0.14)	+0.06 (0.03)	+0.01 : +0.12
21–40	26	+0.10	+0.13 (0.20)	+0.10 (0.03)	+0.03 : +0.16
41–60	30	+0.04	+0.05 (0.10)	+0.04 (0.02)	+0.01 : +0.07
61–80	18	+0.03	+0.03 (0.09)	+0.04 (0.02)	−0.01 : +0.08
81–100	15	0.00	+0.01 (0.06)	0.00 (0.01)	−0.02 : +0.03
101 or more	23	+0.02	+0.05 (0.07)	+0.04 (0.01)	+0.01 : +0.06

Number of pupils in trial (***)		$p <$ 0.01***	eta^2 = 0.21***	$p <$ 0.01***	
500 or less	25	+0.17	+0.18 (0.20)	+0.16 (0.04)	+0.09 : +0.23
501–1,000	16	+0.04	+0.05 (0.10)	+0.05 (0.02)	+0.01 : +0.10
1,001–2,500	27	+0.04	+0.06 (0.11)	+0.05 (0.02)	+0.01 : +0.08
2,501–5,000	23	+0.01	+0.02 (0.08)	+0.02 (0.02)	−0.01 : +0.05
5,001 or more	39	+0.02	+0.02 (0.06)	+0.03 (0.01)	+0.01 : +0.05
Statistical sensitivity, attrition and trial quality					
MDES estimate for design		$p <$ 0.01***	eta^2 = 0.18***	$p <$ 0.01***	
Lower than 0.15 SD	21	+0.02	+0.03 (0.06)	+0.03 (0.01)	+0.01; +0.05
0.15 to lower than 0.25 SD	56	+0.01	+0.01 (0.07)	0.00 (0.01)	−0.01; +0.02
0.25 to lower than 035 SD	40	+0.13	+0.13 (0.13)	+0.12 (0.02)	+0.08; +0.16
0.35 SD or higher	9	+0.09	+0.12 (0.27)	+0.08 (0.05)	−0.01; +0.18
Pupil-level attrition		$p >$ 0.10	eta^2 = 0.03	$p >$ 0.10	
Zero	8	−0.02	−0.01 (0.07)	−0.01 (0.03)	−0.07 : +0.04
<10%	37	+0.08	+0.07 (0.10)	+0.06 (0.01)	+0.03 : +0.09
10% to <20%	37	+0.03	+0.06 (0.11)	+0.04 (0.01)	+0.02 : +0.07
20% to <30%	22	+0.06	+0.09 (0.16)	+0.07 (0.03)	+0.01 : +0.13
30%+	29	+0.01	+0.06 (0.17)	+0.02 (0.01)	0.00 : +0.05
EEF padlock rating (#)		$p =$ 0.02**	eta^2 = 0.10*	$p >$ 0.10	
0	4	+0.23	+0.25 (0.13)	+0.20 (0.07)	+0.06; +0.35
1	9	−0.04	−0.02 (0.11)	0.00 (0.03)	−0.06; +0.05
2	25	+0.02	+0.08 (0.19)	+0.05 (0.03)	−0.01; +0.10
3	44	+0.06	+0.07 (0.11)	+0.06 (0.01)	+0.03; +0.08
4	40	+0.03	+0.06 (0.10)	+0.04 (0.01)	+0.02; +0.07
5	11	+0.01	+0.02 (0.07)	+0.02 (0.01)	−0.01; +0.04
Evaluation burden					
Testing burden (#)		$p >$ 0.10	eta^2 = 0.03	$p >$ 0.10	
Low (just NPD)	19	+0.01	+0.01 (0.06)	+0.03 (0.01)	0.00 : +0.05
Medium (one external test)	35	+0.07	+0.09 (0.15)	+0.04 (0.01)	+0.02 : +0.06
High (two or more external tests)	79	+0.03	+0.06 (0.13)	+0.05 (0.01)	+0.03 : +0.08
IPE data collection burden (#)		$p =$ 0.02**	eta^2 = 0.07***	$p >$ 0.10	
Lowest (no surveys / interviews)	16	+0.19	+0.14 (0.15)	+0.10 (0.04)	+0.02 : +0.18
Medium (just interviews or surveys but not both)	57	+0.02	+0.04 (0.10)	+0.04 (0.01)	+0.01 : +0.06

High (interviews and surveys)	60	+0.03	+0.07 (0.14)	+0.04 (0.01)	+0.02 : +0.06

Primary outcome(s)					
Number of primary outcomes		$p > 0.10$	$eta^2 = 0.03$	$p > 0.10$	
One	50	+0.04	+0.09 (0.15)	+0.05 (0.01)	+0.02 : +0.07
Two	44	+0.02	+0.06 (0.14)	+0.05 (0.02)	+0.01 : +0.09
Three or more	39	+0.03	+0.04 (0.07)	+0.04 (0.01)	+0.02 : +0.06
Alignment between intervention focus and primary outcome(s) (***)		$p < 0.01$***	$eta^2 = 0.13$***	$p < 0.01$***	
Direct match	60	+0.10	+0.11 (0.16)	+0.09 (0.02)	+0.06 : +0.12
Associated match	43	+0.02	+0.03 (0.09)	+0.02 (0.01)	0.00 : +0.04
Limited match	30	+0.01	+0.01 (0.07)	+0.03 (0.01)	0.00 : +0.05
Types of primary outcome (simple) (#)		$p > 0.10$	$eta^2 = 0.03$	$p > 0.10$	
Commercial	79	+0.03	+0.07 (0.15)	+0.05 (0.01)	+0.03; +0.08
Official / SATs	45	+0.02	+0.03 (0.08)	+0.02 (0.01)	0.00; +0.05
Other / mixed	9	+0.10	+0.11 (0.10)	+0.08 (0.03)	+0.03; +0.14
Types of primary outcome (detailed) (#)		$p > 0.10$	$eta^2 = 0.10$	$p = 0.06$*	
… Commercial:					
GL Assessment	46	+0.07	+0.10 (0.18)	+0.08 (0.02)	+0.03; +0.12
CEM	20	+0.01	+0.02 (0.09)	+0.01 (0.01)	−0.01; +0.03
Hodder	8	+0.06	+0.09 (0.10)	+0.07 (0.03)	+0.01; +0.14
Pearson	5	+0.05	+0.05 (0.02)	+0.05 (0.03)	0.00; +0.11
… Official / SATs:					
KS1	1	–	–	–	–
KS2	30	+0.03	+0.05 (0.08)	+0.04 (0.01)	+0.01; +0.07
KS3	3	–	–		
KS4	11	+0.02	+0.01 (0.06)	+0.03 (0.01)	0.00; +0.06
Types of primary outcome (very detailed) (#)		$p > 0.10$	$eta^2 = 0.10$	$p > 0.10$	
Commercial:					
GL Assessment					
GL NGRT	23	+0.06	+0.10 (0.17)	+0.09 (0.04)	+0.02; +0.16
GL PiE or PTE	14	+0.08	+0.11 (0.24)	+0.06 (0.05)	−0.03; +0.16
GL PiM or PTM	8	+0.03	+0.04 (0.09)	+0.03 (0.03)	−0.04; +0.09
CEM					
CEM InCAS maths	7	+0.02	+0.06 (0.11)	+0.02 (0.02)	−0.02; +0.07
CEM InCAS reading	5	0.00	0.00 (0.07)	−0.02 (0.03)	−0.09; +0.05
CEM InCAS reading and maths combined	4	0.00	−0.03 (0.08)	0.00 (0.04)	−0.07; +0.07

Hodder GRT	4	+0.01	+0.01 (0.02)	−0.01 (0.03)	−0.06; +0.04
Other commercial	16	+0.08	+0.08 (0.07)	+0.07 (0.02)	+0.03; +0.10
Official / NPD					
KS2					
KS2 maths	9	0.00	+0.03 (0.08)	−0.01 (0.02)	−0.05; +0.02
KS2 reading	5	0.00	0.00 (0.08)	−0.01 (0.03)	−0.08; +0.05
KS2 writing	5	+0.10	+0.07 (0.13)	−0.01 (0.03)	−0.06; +0.05
KS4 / GCSE					
GCSE English	3	–	–		
GCSE maths	3	–	–		
GCSE overall	3	–	–		
Other official / NPD	15	+0.03	+0.06 (0.07)	+0.06 (0.02)	+0.03; +0.10
Other / mixed	9	+0.10	+0.11 (0.10)	+0.08 (0.03)	+0.03; +0.14
Primary outcome curriculum area		$p > 0.10$	$eta^2 = 0.03$	$p > 0.10$	
Cross-curriculum	11	+0.01	0.01 (0.06)	+0.02 (0.01)	0.00 : +0.05
English / literacy	77	+0.03	0.08 (0.15)	+0.05 (0.01)	+0.03 : +0.08
Maths / numeracy	38	+0.05	0.06 (0.09)	+0.04 (0.01)	+0.02 : +0.07
Science	7	−0.01	0.03 (0.10)	+0.04 (0.04)	−0.04 : +0.11
Trial / ES-level reconciled		$p > 0.10$	$eta^2 = 0.07$**	$p > 0.10$	
Cross-curriculum trial and outcome(s)	11	+0.01	+0.01 (0.06)	+0.02 (0.01)	0.00; +0.05
Cross-curriculum trial, multiple subject outcomes	50	+0.02	+0.03 (0.08)	+0.03 (0.01)	+0.01; +0.05
English / literacy trial and outcome(s)	53	+0.06	+0.10 (0.17)	+0.07 (0.02)	+0.04; +0.11
Maths / numeracy trial and outcome(s)	16	+0.09	+0.09 (0.10)	+0.05 (0.02)	+0.02; +0.08
Science trial and outcome(s)	3	–	–	–	–

Tables for meta-analyses of secondary attainment ITT effect sizes

Reported effect size for headline ITT analyses of secondary attainment outcome

There are a total of 133 effect sizes for headline ITT analyses of primary outcome(s) across the 82 trials. 50 trials reported a single primary outcome effect size, 22 reported two effect sizes and 10 report three or more effect sizes.

This section presents the analyses of these 133 effect sizes across explanatory variables in each of the five themes of the review's theoretical framework.

The analyses are summarised statistically using five tables (one for each of the themes). To highlight a finding of interest, the weighted mean effect size is shown in red text where it differs notably from the other categories within that variable.

- Table 12 presents the average effect sizes across categories of explanatory variables included in the intervention theme using unweighted mean and median statistics. Alongside these descriptive unweighted statistics, meta-analyses, means and standard errors are shown, along with 95% confidence intervals.
- Table 13 uses the same approach to present average effect sizes across categories of explanatory variables included in the theory & evidence theme.
- Table 14 uses the same approach to present average effect sizes across categories of explanatory variables included in the context theme.
- Table 15 uses the same approach to present average effect sizes across categories of explanatory variables included in the implementation & fidelity theme.
- Table 16 uses the same approach to present average effect sizes across categories of explanatory variables included in the evaluation design theme.

Table 12: Secondary ITT effect size and the intervention

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	78	+0.01	+0.01 (0.13)	+0.01 (0.01)	−0.01 : +0.03
Focus of intervention					
School phase				$p > 0.10$	
Primary (including Early Years)	65	+0.01	+0.01 (0.13)	+0.01 (0.01)	−0.01 : +0.03
Primary–secondary transition	5	+0.01	+0.02 (0.14)	+0.02 (0.10)	−0.17 : +0.21
Secondary	8	−0.01	+0.04 (0.18)	−0.03 (0.04)	−0.10 : +0.05
Key Stage				$p < 0.01$***	
Early Years	1	–	–	–	–
Primary KS1	16	+0.08	+0.09 (0.10)	+0.09 (0.02)	+0.05 : +0.12
Primary KS2	47	−0.01	−0.02 (0.12)	−0.01 (0.01)	−0.02 : +0.01
Primary (multiple Key Stages)	1	–	–	–	–
Primary–secondary transition	5	+0.01	+0.02 (0.14)	+0.02 (0.10)	−0.17 : +0.21

Secondary KS3	6	−0.06	+0.04 (0.21)	−0.04 (0.06)	−0.16 : +0.07
Secondary KS4	2	–	–	–	–
Secondary (multiple Key Stages)	0	–	–	–	–
Curriculum focus of intervention (**)			eta^2 = 0.10**	p < 0.01***	
Cross-curriculum	38	0.00	0.00 (0.08)	+0.02 (0.01)	−0.01 : +0.04
English	27	+0.01	0.00 (0.18)	−0.02 (0.03)	−0.07 : +0.03
Maths	11	+0.14	+0.10 (0.11)	+0.06 (0.03)	0.00 : +0.13
Science	2	–	–	–	–
Intensity of intervention					
Intensity of intervention				p > 0.10	
30 minutes or less per week	4	–	–	+0.06 (0.03)	−0.01 : +0.12
31–60 minutes per week	12	0.00	+0.05 (0.13)	+0.01 (0.02)	−0.03 : +0.05
61–120 minutes per week	11	+0.01	+0.06 (0.14)	+0.02 (0.03)	−0.05 : +0.08
Over 120 minutes per week	11	−0.03	−0.06 (0.08)	−0.03 (0.02)	−0.07 : +0.01
No intensity details	40	+0.01	+0.01 (0.13)	+0.01 (0.02)	−0.02 : +0.04
Who implements with direct target?					
Direct implementers (***)			eta^2 = 0.26***	p < 0.01***	
Teacher-led	40	−0.01	−0.02 (0.13)	−0.01 (0.01)	−0.04 : +0.01
Externally-led (e.g., delivery partner)	10	+0.04	+0.08 (0.14)	+0.01 (0.02)	−0.03 : +0.05
TA-led	5	+0.04	+0.01 (0.04)	+0.03 (0.03)	−0.04 : +0.09
Parent-led	3	–	–	–	–
Resource-led	1	–	–	–	–
Other school staff-led	6	−0.09	−0.10 (0.10)	−0.10 (0.05)	−0.19 : 0.00
Other	13	+0.10	+0.08 (0.06)	+0.10 (0.02)	+0.06 : +0.13
Supporting resources					
Perceived quality of supporting resources (*)			eta^2 = 0.09*	p < 0.01***	
High	25	+0.04	+0.05 (0.09)	+0.01 (0.01)	−0.01 : 0.03
Variation	19	−0.01	−0.03 (0.18)	−0.04 (0.03)	−0.10 : +0.01
Low	4	–	–	−0.12 (0.03)	−0.18 : −0.05
Not mentioned	30	0.00	+0.02 (0.12)	+0.03 (0.02)	−0.01 : +0.07
Cost					
Total cost of delivery (**)			eta^2 = 0.15**	p < 0.01***	
<£100k	0	–	–	–	–

£100k–<£250k	12	+0.01	+0.03 (0.07)	0.00 (0.02)	−0.03 : +0.04
£250k–<£500k	27	+0.04	+0.05 (0.11)	+0.06 (0.02)	+0.03 : +0.09
£500k–<£750k	23	−0.01	−0.03 (0.09)	−0.02 (0.02)	−0.05 : +0.01
£750k–<£1 million	10	−0.20	−0.06 (0.23)	−0.14 (0.06)	−0.25 : −0.02
£1 million+	6	+0.01	+0.01 (0.15)	+0.05 (0.04)	−0.02 : +0.12
Cost per pupil (**)			eta^2 = 0.01	p < 0.01***	
< £10	7	+0.02	+0.02 (0.11)	0.00 (0.04)	−0.08 : +0.08
£10–<£25	15	+0.07	+0.04 (0.09)	+0.05 (0.02)	+0.01 : +0.09
£25–<£50	14	−0.05	−0.10 (0.13)	−0.07 (0.03)	−0.14 : −0.01
£50–<£100	20	0.00	0.00 (0.05)	+0.01 (0.01)	−0.02 : +0.03
£100–<£250	11	+0.17	+0.14 (0.14)	+0.12 (0.05)	+0.03 : +0.21
£250–<£1,000	10	−0.05	0.00 (0.18)	−0.02 (0.05)	−0.12 : +0.08
£1,000+	1	–	–	–	–
EEF promising interventions					
Whether classed as promising				p < 0.01***	
Classed as promising	16	+0.07	+0.06 (0.08)	+0.08 (0.02)	+0.05 : +0.11
Not classed as promising	62	−0.01	0.00 (0.14)	−0.01 (0.01)	−0.03 : +0.01

EEF school themes (* see below)	*n* / *n*'	Median ES / ES' [diff]	Unweighted mean ES / ES' [diff]	Weighted mean difference (SE)	95% CI for weighted mean difference
Language and literacy	31 / 47	0.01 / 0.01 [0.00]	0.00 / 0.02 [−0.02]	−0.01 (0.02)	−0.05 : +0.03
Staff deployment and development	32 / 46	0.02 / −0.01 [+0.03]	0.04 / 0.00 [+0.04]	+0.01 (0.01)	−0.01 : +0.03
Organising your school	21 / 57	0.10 / 0.00 [+0.10]	0.07 / −0.01 [+0.08]	+0.06 (0.03)	0.00 : +0.11
Developing effective learners	12 / 66	−0.06 / 0.01 [−0.07]	0.00 / 0.02 [−0.02]	−0.02 (0.05)	−0.11 : +0.07
Mathematics	14 / 64	0.06 / 0.01 [+0.05]	0.06 / 0.00 [+0.06]	+0.04 (0.05)	−0.11 : +0.07
Feedback and monitoring pupil progress	18 / 60	−0.07 / 0.03 [−0.10]	−0.10 / 0.05 [−0.15]	−0.07 (0.03)	−0.12 : −0.01
Behaviour	8 / 70	−0.07 / 0.02 [−0.09]	−0.08 / 0.02 [−0.10]	−0.07 (0.04)	−0.14 : 0.00
Character and essential life skills	2 / 76	–	–	–	–
Parental engagement	6 / 72	0.03 / 0.01 [+0.02]	0.09 / 0.01 [+0.08]	+0.07 (0.05)	−0.03 : +0.18
Enrichment	2 / 76	–	–	–	–
Science	2 / 76	–	–	–	–
Early Years	1 / 77	–	–	–	–
Special educational	1 / 77	–	–	–	–

needs and disabilities

*** Key:** The EEF school themes are not mutually exclusive: effect sizes for a particular trial can be included in more than one of the school themes. Table 12 takes each school theme to illustrate the number of effect sizes included (and not included) in each along. The table also shows unweighted averages (median and mean) for effect sizes included and not included. Finally, the table shows the weighted mean difference in effect sizes (included – not included) from the meta-analyses.

Label	Details
n / *n*':	• *n* is the number of effect sizes attached to a specific EEF school theme • *n*' is the number of effect sizes not included
med / med' [diff]	• med is the unweighted median effect size for trials included in a specific EEF school theme • med' is the unweighted median effect size for trials not included • [diff] is the difference between the two medians (i.e., med – med')
mean / mean' [diff]	• mean is the unweighted mean effect size for trials included in a specific EEF school theme • mean' is the unweighted mean effect size for trials not included • [diff] is the difference between the two means (i.e., mean – mean')
Weighted mean difference (SE)	• This is the weighted mean difference between effect sizes included in a specific EEF school theme and effect sizes not included obtained from the meta-analyses • (SE) is the standard error of the weighted mean difference

Table 13: Secondary ITT effect size and theory & evidence

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	78	+0.01	+0.01 (0.13)	+0.01 (0.01)	−0.01 : +0.03
Empirical evidence and theoretical detail					
Strength of empirical evidence (***)			eta^2 = 0.20***	p < 0.01***	
Strong evidence	22	−0.02	−0.07 (0.12)	−0.04 (0.02)	−0.08 : −0.01
Some evidence	49	+0.01	+0.03 (0.12)	+0.03 (0.01)	0.00 : +0.05
Minimal or no evidence	7	+0.15	+0.13 (0.11)	+0.04 (0.05)	−0.06 : +0.15
Theoretical detail (***)			eta^2 = 0.30***	p < 0.01***	
Highly detailed	18	+0.05	+0.06 (0.07)	+0.05 (0.02)	+0.02 : +0.08
Some detail	29	−0.06	−0.08 (0.12)	−0.06 (0.02)	−0.10 : −0.02
Minimal or no detail	31	+0.02	+0.07 (0.13)	+0.02 (0.01)	−0.01 : +0.04
Focus of change				n/a	
Learning focus	74	+0.01	+0.01 (0.13)	0.00 (0.01)	−0.02 : +0.03
Teacher change focus	2	–	–	–	–
Wider pupil outcomes focus	2	–	–	–	–
Other	0	–	–	–	–

Table 14: Secondary ITT effect sizes and evaluation context

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	78	+0.01	+0.01 (0.13)	+0.01 (0.01)	−0.01 : +0.03
External context					
Geography				$p > 0.10$	
National	22	+0.04	+0.04 (0.09)	+0.03 (0.02)	0.00 : +0.07
One geographical location	22	0.00	+0.03 (0.16)	−0.01 (0.03)	−0.06 : +0.05
Two or three geographical areas	25	+0.01	−0.01 (0.15)	−0.01 (0.01)	−0.04 : +0.02
Other	9	+0.01	−0.01 (0.09)	+0.01 (0.02)	−0.03 : +0.05
OFSTED (#)				$p > 0.10$	
Mentioned as barrier	16	+0.04	+0.06 (0.10)	+0.03 (0.02)	−0.01 : +0.06
Not mentioned as barrier	62	+0.01	0.00 (0.14)	0.00 (0.01)	−0.02 : +0.03
Characteristics of participating organisations					
… perceived barriers					
Specialist facilities and space (**)			$eta^2 = 0.07$**	$p < 0.01$***	
Mentioned as barrier	28	+0.05	+0.06 (0.11)	+0.06 (0.01)	+0.03 : +0.09
Not mentioned as barrier	50	−0.01	−0.01 (0.14)	−0.02 (0.01)	−0.04 : 0.00
Staff time and availability				$p > 0.10$	
Mentioned as barrier	50	+0.02	+0.02 (0.14)	+0.01 (0.01)	−0.01 : +0.04
Not mentioned as barrier	28	−0.01	0.00 (0.12)	−0.01 (0.01)	−0.03 : 0.00
Workforce capacity				$p < 0.01$***	
Mentioned as barrier	35	+0.04	+0.04 (0.07)	+0.03 (0.01)	+0.01 : +0.06
Not mentioned as barrier	43	−0.02	−0.01 (0.16)	−0.03 (0.02)	−0.07 : 0.00
… perceived enablers					
Alignment of intervention and existing practice (*)				$p > 0.10$	
Mentioned as enabler	15	+0.03	+0.03 (0.06)	+0.01 (0.01)	−0.01 : +0.04
Not mentioned as enabler	63	0.00	+0.01 (0.14)	0.00 (0.01)	−0.03 : +0.03
Staff teamwork (*)				$p < 0.01$***	
Mentioned as enabler	24	+0.03	+0.04 (0.07)	+0.04 (0.01)	+0.01 : +0.06
Not mentioned as enabler	54	−0.01	0.00 (0.15)	−0.02 (0.01)	−0.05 : +0.01
Characteristics of participating individuals					
… perceived barriers					

Pupil behaviour				$p < 0.01$***	
Mentioned as barrier	18	−0.03	−0.02 (0.14)	−0.04 (0.02)	−0.08 : +0.01
Not mentioned as barrier	60	+0.02	+0.02 (0.13)	+0.02 (0.01)	0.00 : +0.04
… perceived barriers and enablers					
SLT buy-in				$p > 0.10$	
Mentioned as barrier	8	−0.04	+0.03 (0.18)	−0.05 (0.03)	−0.11 : +0.02
Mentioned as both barrier and enabler	6	+0.01	+0.03 (0.07)	+0.01 (0.02)	−0.03 : +0.05
Mentioned as enabler	20	+0.03	+0.03 (0.05)	+0.01 (0.01)	−0.01 : +0.03
Not mentioned	44	0.00	0.00 (0.15)	0.00 (0.02)	−0.04 : +0.05
Staff expectations and motivations		$p > 0.10$	$eta^2 = 0.01$	$p < 0.01$***	
Mentioned as barrier	3	–	–	–	–
Mentioned as both barrier and enabler	11	+0.01	+0.03 (0.07)	+0.01 (0.01)	−0.02 : +0.04
Mentioned as enabler	26	+0.07	+0.08 (0.08)	+0.05 (0.01)	+0.03 : +0.08
Not mentioned	38	−0.03	−0.03 (0.16)	−0.05 (0.02)	−0.09 : −0.01

Table 15: Secondary ITT effect size and implementation & fidelity

	$n_{ES} =$	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SD)	95% CI
All trials	78	+0.01	+0.01 (0.13)	+0.01 (0.01)	−0.01 : +0.03
Developer characteristics					
Type of developer (***)			$eta^2 = 0.28$***	$p < 0.05$**	
Not for profit / charity	30	+0.04	+0.03 (0.13)	+0.02 (0.02)	−0.02 : +0.06
University	19	+0.01	+0.01 (0.05)	0.00 (0.01)	−0.02 : +0.03
Private company	13	−0.13	−0.13 (0.12)	−0.10 (0.04)	−0.17 : −0.02
School, academy or MAT	6	+0.09	+0.07 (0.17)	+0.04 (0.06)	−0.08 : +0.17
Council / LA	7	+0.02	+0.11 (0.15)	+0.08 (0.06)	−0.03 : +0.19
Mixed	3	–	–	–	–
Planning, time and support					
Clarity of implementation plan			$eta^2 = 0.06$*	$p < 0.01$***	
Clearly understood	32	+0.01	+0.01 (0.10)	0.00 (0.01)	−0.02 : +0.02
Variation in understanding	30	−0.01	−0.02 (0.16)	−0.03 (0.02)	−0.07 : +0.01
Unclear or not mentioned	16	+0.10	+0.07 (0.09)	+0.10 (0.02)	+0.07 : +0.14
Lead in time for preparation				$p > 0.10$	
Sufficient time	2	–	–	–	–

Variation in perceptions	13	−0.01	+0.05 (0.11)	+0.01 (0.02)	−0.03 : +0.05
Insufficient time	19	−0.02	−0.01 (0.11)	−0.03 (0.03)	−0.09 : +0.02
Not mentioned	44	+0.02	+0.02 (0.14)	+0.02 (0.01)	0.00 : +0.05
Senior leadership support				$p > 0.10$	
Strong	12	+0.04	+0.03 (0.05)	+0.01 (0.01)	−0.01 : +0.04
Some	15	+0.02	+0.07 (0.10)	+0.02 (0.02)	−0.01 : +0.05
Limited or minimal	5	−0.07	+0.06 (0.24)	−0.03 (0.08)	−0.17 : +0.12
Not mentioned	46	−0.01	−0.01 (0.14)	0.00 (0.02)	−0.04 : +0.03
Professional development (CPD)					
Is CPD provided to support implementation?			$eta^2 = 0.16$***	$p < 0.01$***	
YES, only to direct implementers	47	+0.01	+0.03 (0.13)	0.00 (0.01)	−0.02 : +0.02
YES, to implementers and other stakeholders	21	−0.03	−0.07 (0.13)	−0.03 (0.02)	−0.08 : +0.01
YES, only to other stakeholders	0	–	–	–	–
NO CPD or unclear	10	+0.10	+0.09 (0.06)	+0.08 (0.02)	+0.03 : +0.13
Is CPD subject-specific / curriculum-specific or general? (**)			$eta^2 = 0.16$**	$p < 0.01$***	
Predominantly subject-specific or curriculum-specific	39	+0.01	+0.04 (0.14)	0.00 (0.02)	−0.04 : +0.03
Predominantly generic	25	−0.01	−0.06 (0.12)	−0.01 (0.01)	−0.03 : +0.02
Mixed generic and subject-specific	4	–	–	+0.01 (0.03)	−0.05 : +0.07
Not mentioned	10	+0.10	+0.09 (0.06)	+0.08 (0.02)	+0.03 : +0.13
Types of CPD (see note below)					
Face-to-face	68	0.00	0.00 (0.14)	−0.01 (0.00)	−0.03 : +0.01
Online	5	+0.06	+0.07 (0.10)	+0.06 (0.03)	0.00 : +0.13
Coaching or mentoring	9	+0.11	+0.09 (0.14)	+0.03 (0.03)	−0.08 : +0.13
Cascade 'train the trainer' model	13	−0.04	−0.07 (0.17)	−0.06 (0.04)	−0.14 : +0.03
Sequencing of CPD (#)		$p > 0.10$	$eta^2 = 0.31$***	$p < 0.01$***	
Pre-intervention only	15	−0.06	−0.10 (0.12)	−0.07 (0.03)	−0.11 : −0.02
During the intervention only	15	−0.04	−0.04 (0.10)	−0.05 (0.03)	−0.11 : +0.01
Pre and during the intervention	36	+0.02	+0.06 (0.13)	+0.02 (0.01)	0.00 : +0.04
Not mentioned	12	+0.10	+0.09 (0.06)	+0.08 (0.03)	+0.04 : +0.12
Who delivers CPD?			$eta^2 = 0.12$*	$p < 0.01$***	
Delivery partner	44	+0.01	+0.01 (0.16)	0.00 (0.01)	−0.03 : +0.03
Other external organisation	6	−0.03	−0.02 (0.09)	−0.01 (0.04)	−0.09 : +0.08
Leaders / teachers from schools in the trial	6	−0.09	−0.10 (0.10)	−0.10 (0.05)	−0.19 : 0.00
Mixed	10	+0.02	0.00 (0.04)	0.00 (0.01)	−0.03 : +0.03

Not mentioned	12	+0.10	+0.09 (0.13)	+0.08 (0.02)	+0.04 : +0.12
Support and monitoring					
Does delivery partner provide support (other than CPD)?			eta^2 = 0.15**	*p* < 0.01***	
Before the intervention only	6	−0.08	−0.10 (0.10)	−0.09 (0.05)	−0.19 : 0.00
During the intervention only	11	+0.11	+0.10 (0.06)	+0.10 (0.02)	+0.06 : +0.14
Before and during the intervention	38	−0.01	−0.01 (0.15)	−0.02 (0.02)	−0.05 : +0.02
Other or not mentioned	23	+0.01	+0.05 (0.10)	0.00 (0.01)	−0.01 : +0.02
Monitoring of implementation (**)			eta^2 = 0.12**	*p* < 0.01***	
Robust monitoring	16	+0.05	−0.02 (0.20)	−0.04 (0.04)	−0.12 : +0.05
Some monitoring	24	−0.03	−0.04 (0.09)	−0.03 (0.02)	−0.06 : 0.00
No monitoring	2	–	–	–	–
Not mentioned	36	+0.02	+0.06 (0.11)	+0.04 (0.01)	+0.01 : +0.06
Fidelity					
Intended fidelity			eta^2 = 0.14**	*p* > 0.10	
Faithful adoption	36	+0.02	+0.04 (0.12)	+0.01 (0.01)	−0.02 : +0.03
Adaptation to context	35	−0.01	−0.03 (0.13)	−0.01 (0.02)	−0.05 : +0.03
Not mentioned	7	+0.15	+0.12 (0.11)	+0.04 (0.03)	−0.03 : +0.10
Fidelity related to CPD (**)			eta^2 = 0.20***	*p* < 0.01***	
High	8	+0.05	+0.11 (0.09)	+0.05 (0.03)	0.00 : +0.11
Varied or moderate	24	−0.04	−0.04 (0.17)	−0.04 (0.03)	−0.10 : +0.01
Limited	12	−0.05	−0.05 (0.08)	−0.05 (0.03)	−0.10 : +0.01
Not mentioned	34	+0.03	+0.05 (0.10)	+0.03 (0.01)	+0.01 : +0.06
Actual fidelity of implementation (#)			eta^2 = 0.17**	*p* < 0.01***	
High	18	0.00	−0.06 (0.14)	−0.02 (0.01)	−0.04 : +0.01
Varied or moderate	34	+0.03	+0.06 (0.13)	+0.04 (0.01)	+0.01 : +0.06
Limited	16	+0.04	+0.03 (0.09)	+0.04 (0.02)	−0.01 : +0.08
Not mentioned	10	−0.04	−0.05 (0.11)	−0.04 (0.03)	−0.10 : +0.01

Table 16: Secondary ITT effect size and evaluation design

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	78	+0.01	+0.01 (0.13)	+0.01 (0.01)	−0.01 : +0.03
Trial description					

Trial design (***)			eta^2 = 0.08***	p > 0.10	
RCT	13	+0.02	+0.10 (0.17)	+0.07 (0.04)	−0.01 : +0.15
Clustered RCT	65	0.00	0.00 (0.12)	0.00 (0.01)	−0.02 : +0.02
Level of randomisation (**)			eta^2 = 0.16**	p < 0.05**	
School	56	+0.02	+0.01 (0.12)	+0.01 (0.01)	−0.01 : +0.03
Pupil	10	+0.01	+0.06 (0.16)	+0.02 (0.04)	−0.07 : +0.11
Class or teacher	3	–	–	–	–
Key Stage or year group	6	−0.08	−0.10 (0.10)	−0.10 (0.05)	−0.19 : 0.00
Parent	3	–	–	–	–
Complex (multiple)	0	–	–	–	–
Type of trial			eta^2 = 0.13***	p < 0.05**	
Efficacy	30	+0.03	+0.07 (0.14)	+0.04 (0.02)	+0.01 : +0.07
Effectiveness	48	0.00	−0.02 (0.12)	−0.01 (0.01)	−0.03 : +0.02
Type of evaluator			eta^2 = 0.06**	p < 0.01***	
Non-university	19	+0.07	+0.07 (0.10)	+0.07 (0.02)	+0.04 : +0.10
University	59	0.00	0.00 (0.14)	−0.01 (0.01)	−0.03 : +0.01
Trial length and size					
Length of trial (#)			eta^2 = 0.28***	p < 0.01***	
Up to 15 weeks (one term)	17	+0.04	+0.10 (0.14)	+0.03 (0.02)	−0.01 : +0.07
16–30 weeks (two terms)	16	−0.05	−0.06 (0.08)	−0.04 (0.02)	−0.08 : 0.00
31–45 weeks (three terms / one year)	18	+0.09	+0.08 (0.10)	+0.06 (0.02)	+0.02 : +0.10
More than 45 weeks / three terms / one year	27	−0.01	−0.04 (0.12)	−0.02 (0.02)	−0.05 : +0.01
Number of schools in trial (#)			eta^2 = 0.28***	p < 0.01***	
20 or less	4	–	–	+0.08 (0.07)	−0.07 : +0.22
21–40	18	−0.03	0.00 (0.14)	−0.04 (0.03)	−0.10 : +0.02
41–60	11	−0.03	+0.02 (0.15)	−0.01 (0.04)	−0.08 : +0.07
61–80	3	–	–	–	–
81–100	12	−0.12	−0.13 (0.12)	−0.09 (0.03)	−0.15 : −0.03
101 or more	30	+0.04	+0.06 (0.06)	+0.04 (0.03)	+0.02 : +0.06
Number of pupils in trial				p > 0.10	
500 or less	11	+0.01	+0.06 (0.20)	+0.02 (0.04)	−0.06 : +0.10
501–1,000	6	+0.11	+0.13 (0.15)	+0.11 (0.06)	−0.01 : +0.22
1,001–2,500	18	−0.02	+0.01 (0.14)	−0.01 (0.03)	−0.06 : +0.05
2,501–5,000	10	0.00	+0.01 (0.05)	0.00 (0.02)	−0.03 : +0.04

5,001 or more	32	+0.01	−0.02 (0.13)	0.00 (0.02)	−0.03 : +0.04
Statistical sensitivity, attrition and trial quality					
EEF padlock rating (#)			$eta^2 = 0.27$***	$p < 0.01$***	
0	3	–	–	–	–
1	3	–	–	–	–
2	15	+0.07	+0.05 (0.09)	+0.08 (0.02)	+0.05 : +0.12
3	26	+0.01	+0.03 (0.14)	−0.01 (0.02)	−0.05 : +0.03
4	14	−0.10	−0.09 (0.15)	−0.09 (0.04)	−0.17 : −0.01
5	17	+0.01	+0.01 (0.06)	+0.01 (0.01)	−0.01 : +0.04
Evaluation burden					
Testing burden (#)				$p > 0.10$	
Low (just NPD)	15	+0.01	+0.01 (0.05)	0.00 (0.01)	−0.02 : +0.30
Medium (one external test)	19	−0.04	−0.04 (0.18)	−0.05 (0.04)	−0.13 : +0.03
High (two or more external tests)	44	+0.02	+0.04 (0.12)	+0.02 (0.02)	−0.01 : +0.05
IPE data collection burden (#)			$eta^2 = 0.07$*	$p > 0.10$	
Lowest (no surveys / interviews)	6	+0.16	+0.13 (0.18)	+0.14 (0.08)	−0.02 : +0.30
Medium (just interviews or surveys but not both)	22	0.00	−0.01 (0.17)	0.00 (0.03)	−0.07 : +0.07
High (interviews and surveys)	50	+0.01	+0.01 (0.10)	0.00 (0.01)	−0.02 : +0.01

Primary outcome(s)					
Types of primary outcome (simple) (#)				$p < 0.05$**	
Commercial	26	−0.03	0.00 (0.14)	−0.04 (0.02)	−0.08 : +0.01
Official / SATs	51	+0.02	+0.02 (0.08)	+0.02 (0.01)	0.00 : +0.04
Other / mixed	1	–	–	–	–

Tables for meta-analyses of FSM attainment effect sizes

Reported effect size for of primary / secondary attainment outcomes for FSM subsamples.

There are a total of 133 effect sizes for headline ITT analyses of primary outcome(s) across the 82 trials. 50 trials reported a single primary outcome effect size, 22 reported two effect sizes and 10 report three or more effect sizes.

This section presents the analyses of these 133 effect sizes across explanatory variables in each of the five themes of the review's theoretical framework.

The analyses are summarised statistically using five tables (one for each of the themes). To highlight a finding of interest, the weighted mean effect size is shown in red text where it differs notably from the other categories within that variable.

- Table 17 presents the average effect sizes across categories of explanatory variables included in the intervention theme using unweighted mean and median statistics. Alongside these descriptive unweighted statistics, meta-analyses, means and standard errors are shown, along with 95% confidence intervals.
- Table 18 uses the same approach to present average effect sizes across categories of explanatory variables included in the theory & evidence theme.
- Table 19 uses the same approach to present average effect sizes across categories of explanatory variables included in the context theme.
- Table 20 uses the same approach to present average effect sizes across categories of explanatory variables included in the implementation & fidelity theme.
- Table 21 uses the same approach to present average effect sizes across categories of explanatory variables included in the evaluation design theme.

Table 17: FSM effect size and the intervention

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	149	+0.02	+0.04 (0.23)	+0.03 (0.010)	+0.01 : +0.05
Focus of intervention					
School phase			eta^2 = 0.09***	p < 0.05**	
Primary (including Early Years)	102	+0.02	+0.02 (0.21)	+0.04 (0.01)	+0.02 : +0.06
Primary–secondary transition	7	+0.20	+0.35 (0.57)	+0.13 (0.07)	−0.02 : +0.28
Secondary	40	+0.01	+0.03 (0.18)	0.00 (0.02)	−0.04 : +0.03
Key Stage			eta^2 = 0.10**	p < 0.05**	
Early Years	1	–	–	–	–
Primary KS1	29	+0.02	−0.02 (0.31)	+0.07 (0.03)	+0.02 : +0.12
Primary KS2	66	+0.04	+0.04 (0.15)	+0.04 (0.01)	+0.01 : +0.06
Primary (multiple Key Stages)	6	−0.02	−0.06 (0.09)	−0.01 (0.05)	−0.11 : +0.09
Primary–secondary transition	7	+0.20	+0.35 (0.57)	+0.13 (0.07)	−0.02 : +0.28

Secondary KS3	27	−0.01	+0.04 (0.21)	−0.01 (0.03)	−0.08 : +0.06
Secondary KS4	10	+0.04	+0.04 (0.03)	+0.04 (0.02)	0.00 : +0.08
Secondary (multiple Key Stages)	3	–	–	–	–
Curriculum focus of intervention (**)				$p > 0.10$	
Cross-curriculum	70	+0.02	+0.02 (0.10)	+0.03 (0.01)	0.00 : +0.05
English	61	+0.03	+0.05 (0.34)	+0.04 (0.025)	−0.01 : +0.09
Maths	15	+0.01	+0.01 (0.14)	+0.04 (0.03)	−0.02 : +0.10
Science	3	–	–	–	–
Intensity of intervention					
Intensity of intervention				$p > 0.10$	
30 minutes or less per week	18	+0.02	−0.03 (0.40)	+0.03 (0.03)	−0.03 : +0.10
31–60 minutes per week	29	+0.03	+0.06 (0.15)	+0.07 (0.03)	+0.02 : +0.12
61–120 minutes per week	32	0.00	+0.03 (0.18)	0.00 (0.03)	−0.06 : +0.06
Over 120 minutes per week	13	+0.01	+0.03 (0.14)	0.00 (0.02)	−0.03 : +0.04
No intensity details	57	+0.03	+0.05 (0.25)	+0.03 (0.01)	0.01 : +0.05
Who implements with direct target?					
Direct Implementers (***)				$p > 0.10$	
Teacher-led	63	+0.03	+0.07 (0.24)	+0.03 (0.01)	+0.01 : +0.06
Externally-led (e.g., delivery partner)	32	+0.02	+0.03 (0.17)	+0.02 (0.02)	−0.03 : +0.07
TA-led	17	0.00	−0.05 (0.42)	+0.07 (0.06)	−0.05 : +0.19
Parent-led	10	0.00	+0.02 (0.08)	+0.01 (0.04)	−0.07 : +0.09
Resource-led	3	–	–	–	–
Other school staff-led	2	–	–	–	–
Other	22	+0.03	+0.03 (0.11)	+0.03 (0.02)	−0.01 : +0.07
Supporting resources					
Perceived quality of supporting resources (*)			$eta^2 = 0.06$**	$p < 0.01$***	
High	30	+0.11	+0.13 (0.17)	+0.10 (0.03)	+0.05 : +0.15
Variation	59	−0.01	−0.03 (0.24)	0.00 (0.01)	−0.02 : +0.02
Low	5	+0.10	+0.08 (0.07)	+0.05 (0.03)	−0.02 : +0.12
Not mentioned	55	+0.02	+0.05 (0.25)	+0.02 (0.02)	−0.01 : +0.05
Cost					
Total cost of delivery (**)			$eta^2 = 0.11$**	$p < 0.01$***	
<£100k	6	+0.06	+0.09 (0.12)	+0.05 (0.04)	−0.02 : +0.12
£100k–<£250k	16	+0.06	+0.10 (0.15)	+0.04 (0.02)	0.00 : +0.08

£250k–<£500k	53	+0.05	+0.11 (0.26)	+0.07 (0.02)	+0.04 : +0.11
£500k–<£750k	30	−0.02	−0.07 (0.31)	−0.01 (0.03)	−0.06 : +0.05
£750k–<£1 million	22	−0.04	−0.07 (0.11)	−0.04 (0.02)	−0.08 : +0.01
£1 million+	22	+0.03	+0.04 (0.08)	+0.03 (0.02)	0.00 : +0.07
Cost per pupil (**)			–	$p > 0.10$	
<£10	21	+0.01	+0.06 (0.14)	+0.02 (0.01)	−0.01 : +0.05
£10–<£25	25	+0.03	+0.05 (0.13)	+0.05 (0.02)	+0.01 : +0.10
£25–<£50	20	+0.02	+0.01 (0.17)	+0.01 (0.03)	−0.06 : +0.07
£50–<£100	28	+0.01	−0.01 (0.44)	+0.02 (0.02)	−0.02 : +0.06
£100–<£250	31	+0.03	+0.08 (0.19)	+0.03 (0.03)	−0.03 : +0.09
£250–<£1,000	21	0.00	+0.01 (0.12)	−0.01 (0.03)	−0.07 : +0.05
£1,000+	3	–	–	–	–
EEF promising interventions					
Whether classed as promising				$p < 0.01$***	
Classed as promising	35	+0.11	+0.07 (0.32)	+0.11 (0.02)	+0.06 : +0.15
Not classed as promising	114	+0.01	+0.03 (0.21)	+0.01 (0.01)	−0.01 : +0.03

EEF school themes (*see below)	*n* / *n*'	Median ES / ES' [diff]	Unweighted mean ES / ES' [diff]	Weighted mean difference (SE)	95% CI for weighted mean difference
Language and literacy	68 / 81	0.05 / 0.01 [+0.04]	0.06 / 0.02 [+0.04]	+0.05 (0.02)	0.00 : +0.09
Staff deployment and development	52 / 97	0.03 / 0.02 [+0.01]	0.01 / 0.05 [−0.04]	+0.04 (0.01)	+0.01 : +0.06
Organising your school	32 / 117	0.02 / 0.02 [0.00]	0.07 / 0.03 [+0.04]	+0.04 (0.02)	−0.01 : +0.09
Developing effective learners	21 / 128	0.09 / 0.02 [+0.07]	0.14 / 0.02 [+0.12]	+0.07 (0.03)	+0.01 : +0.14
Mathematics	18 / 131	0.03 / 0.02 [+0.01]	0.03 / 0.04 [−0.01]	+0.05 (0.03)	0.00 : +0.10
Feedback and monitoring pupil progress	26 / 123	−0.02 / 0.03 [−0.05]	0.03 / 0.04 [−0.01]	−0.02 (0.02)	−0.06 : +0.02
Behaviour	17 / 132	0.01 / 0.02 [−0.01]	0.02 / 0.04 [−0.02]	+0.04 (0.02)	−0.01 : +0.08
Character and essential life skills	16 / 133	0.03 / 0.02 (+0.01)	0.01 / 0.04 (−0.03)	0.00 (0.03)	−0.07 : +0.06
Parental engagement	16 / 133	0.00 / 0.03 [−0.03]	−0.01 / 0.04 [−0.05]	−0.01 (0.02)	−0.04 : +0.02
Enrichment	10 / 139	0.01 / 0.02 (−0.01)	−0.04 / 0.04 (−0.08)	−0.06 (0.05)	−0.15 : +0.03
Science	3 / 146	–	–		
Early Years	5 / 144	−0.26 / 0.02 (−0.28)	−0.38 / 0.05 (−0.43)	+0.01 (0.02)	−0.04 : +0.05

Special educational needs and disabilities	6 / 143	−0.24 / 0.02 (−0.26)	−0.37 / 0.05 (−0.42)	−0.15 (0.20)	−0.55 : +0.25

* **Key**: The EEF school themes are not mutually exclusive: effect sizes for a particular trial can be included in more than one of the school themes. Table 17 takes each school theme to illustrate the number of effect sizes included (and not included) in each along. The table also shows unweighted averages (median and mean) for effect sizes included and not included. Finally, the table shows the weighted mean difference in effect sizes (included – not included) from the meta-analyses.

Label	Details
n / *n*':	• *n* is the number of effect sizes attached to a specific EEF school theme • *n*' is the number of effect sizes not included
med / med' [diff]	• med is the unweighted median effect size for trials included in a specific EEF school theme • med' is the unweighted median effect size for trials not included • [diff] is the difference between the two medians (i.e., med – med')
mean / mean' [diff]	• mean is the unweighted mean effect size for trials included in a specific EEF school theme • mean' is the unweighted mean effect size for trials not included • [diff] is the difference between the two mean (i.e., mean – mean')
Weighted mean difference (SE)	• This is the weighted mean difference between effect sizes included in a specific EEF school theme and effect sizes not included obtained from the meta-analyses. • (se) is the standard error of the weighted mean difference

Table 18: FSM effect size and theory & evidence

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	149	+0.02	+0.04 (0.23)	+0.03 (0.01)	+0.01 : +0.05
Empirical evidence and theoretical detail					
Strength of empirical evidence (*)			eta^2 = 0.04*	*p* < 0.01***	
Strong evidence	45	0.00	+0.02 (0.14)	+0.02 (0.02)	−0.01 : +0.05
Some evidence	91	+0.04	+0.06 (0.27)	+0.04 (0.01)	+0.02 : +0.07
Minimal or no evidence	13	−0.09	−0.09 (0.15)	−0.08 (0.04)	−0.17 : +0.01
Theoretical detail			–	*p* > 0.10	
Highly detailed	26	+0.02	+0.02 (0.11)	+0.04 (0.02)	+0.01 : +0.08
Some detail	51	+0.03	+0.03 (0.35)	+0.02 (0.01)	−0.01 : +0.04
Minimal or no detail	72	+0.02	+0.05 (0.16)	+0.04 (0.02)	0.00 : +0.07
Focus of change				*p* < 0.01***	
Learning focus	119	+0.02	+0.05 (0.26)	+0.04 (0.01)	+0.02 : +0.06
Teacher change focus	8	+0.02	+0.02 (0.04)	+0.02 (0.02)	−0.02 : +0.06

Wider pupil outcomes focus	20	+0.03	+0.01 (0.09)	+0.02 (0.01)	−0.01 : +0.05
Other	2	–	–	–	–

Table 19: FSM effect sizes and evaluation context

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	149	+0.02	+0.04 (0.23)	+0.03 (0.01)	+0.01 : +0.05
External context					
Geography			eta^2 = 0.09**	*p* > 0.10	
National	46	+0.02	+0.05 (0.13)	+0.03 (0.01)	0.00 : +0.05
One geographical location	36	+0.07	+0.14 (0.29)	+0.06 (0.02)	+0.02 : +0.09
Two or three geographical areas	42	−0.01	−0.03 (0.28)	+0.02 (0.02)	−0.02 : +0.07
Other	25	+0.01	−0.02 (0.16)	−0.01 (0.03)	−0.07 : +0.05
OFSTED (#)				*p* > 0.10	
Mentioned as barrier	30	+0.03	+0.01 (0.13)	+0.03 (0.02)	−0.01 : +0.07
Not mentioned as barrier	119	+0.02	+0.04 (0.25)	+0.03 (0.01)	+0.01 : +0.05
Characteristics of participating organisations					
… perceived barriers					
Specialist facilities and space				*p* < 0.05**	
Mentioned as barrier	53	+0.06	+0.05 (0.29)	+0.07 (0.02)	+0.03 : +0.12
Not mentioned as barrier	96	+0.01	+0.03 (0.20)	+0.01 (0.01)	−0.01 : +0.03
Staff time and availability			eta^2 = 0.04**	*p* > 0.10	
Mentioned as barrier	105	+0.02	+0.01 (0.21)	+0.02 (0.01)	0.00 : +0.04
Not mentioned as barrier	44	+0.05	+0.11 (0.27)	+0.04 (0.02)	+0.01 : +0.08
Workforce capacity				*p* > 0.10	
Mentioned as barrier	51	+0.02	+0.02 (0.16)	+0.03 (0.02)	−0.01 : +0.06
Not mentioned as barrier	98	+0.02	+0.05 (0.27)	+0.02 (0.01)	0.00 : +0.04
… perceived enablers					
Alignment of intervention and existing practice (*)				*p* > 0.10	
Mentioned as enabler	35	+0.04	+0.02 (0.10)	+0.03 (0.01)	0.00 : +0.06
Not mentioned as enabler	114	+0.02	+0.04 (0.26)	+0.03 (0.01)	0.00 : +0.05
Staff teamwork (*)				*p* > 0.10	
Mentioned as enabler	35	+0.02	+0.01 (0.39)	+0.03 (0.01)	0.00 : +0.05
Not mentioned as enabler	114	+0.02	+0.04 (0.16)	+0.03 (0.01)	+0.01 : +0.05
Characteristics of participating individuals					
… perceived barriers					
Pupil behaviour			eta^2 = 0.03**	*p* < 0.10*	

Mentioned as barrier	42	+0.01	−0.02 (0.28)	0.00 (0.02)	−0.05 : +0.04
Not mentioned as barrier	107	+0.03	+0.06 (0.21)	+0.04 (0.01)	+0.02 : +0.06
… perceived barriers and enablers					
SLT buy-in				$p > 0.10$	
Mentioned as barrier	10	+0.01	+0.04 (0.19)	−0.03 (0.04)	−0.10 : +0.04
Mentioned as both barrier and enabler	23	+0.02	−0.01 (0.10)	0.00 (0.02)	−0.04 : +0.04
Mentioned as enabler	34	+0.01	0.00 (0.31)	+0.04 (0.02)	0.00 : +0.08
Not mentioned	82	+0.03	+0.06 (0.23)	+0.04 (0.01)	+0.01 : +0.07
Staff expectations and motivations		$p > 0.10$	$eta^2 = 0.01$	$p > 0.10$	
Mentioned as barrier	18	+0.02	+0.02 (0.17)	0.00 (0.01)	−0.02 : +0.03
Mentioned as both barrier and enabler	23	+0.01	−0.10 (0.32)	−0.02 (0.03)	−0.08 : +0.04
Mentioned as enabler	30	+0.04	+0.04 (0.13)	+0.04 (0.02)	+0.01 : +0.07
Not mentioned	78	+0.04	+0.08 (0.24)	+0.04 (0.01)	+0.01 : +0.07

Table 20: FSM effect size and implementation & fidelity

	$n_{ES} =$	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SD)	95% CI
All trials	149	+0.02	+0.04 (0.23)	+0.03 (0.010)	+0.01 : +0.05
Developer characteristics					
Type of developers (***)			$eta^2 = 0.07$*	$p < 0.01$***	$eta^2 = 0.06$
Not for profit / charity	47	+0.02	+0.03 (0.16)	+0.03 (0.02)	−0.01 : +0.07
University	41	+0.02	+0.04 (0.11)	+0.03 (0.01)	0.00 : +0.05
Private company	24	−0.01	+0.06 (0.38)	−0.02 (0.03)	−0.07 : +0.03
School, academy or MAT	10	+0.13	+0.15 (0.17)	+0.14 (0.05)	+0.04 : +0.24
Council / LA	15	+0.03	+0.06 (0.13)	−0.01 (0.03)	−0.06 : +0.04
Mixed	12	+0.03	+0.04 (0.44)	+0.06 (0.04)	−0.01 : +0.13
Planning, time and support					
Clarity of implementation plan			$eta^2 = 0.04$*	$p > 0.10$	
Clearly understood	54	+0.03	0.00 (0.25)	+0.04 (0.02)	+0.01 : +0.07
Variation in understanding	55	0.00	+0.02 (0.14)	0.00 (0.01)	−0.03 : +0.02
Unclear or not mentioned	40	+0.05	+0.11 (0.30)	+0.05 (0.02)	+0.01 : +0.09
Lead in time for preparation			$eta^2 = 0.06$**	$p > 0.10$	
Sufficient time	10	−0.02	−0.17 (0.47)	+0.02 (0.03)	−0.05 : +0.09

Variation in perceptions	20	+0.02	+0.04 (0.13)	+0.03 (0.03)	−0.03 : +0.08
Insufficient time	39	0.00	+0.04 (0.15)	+0.01 (0.01)	−0.01 : +0.04
Not mentioned	80	+0.04	+0.06 (0.24)	+0.03 (0.01)	+0.01 : +0.06
Senior leadership support				$p > 0.10$	
Strong	20	−0.02	−0.02 (0.40)	+0.07 (0.04)	−0.01 : +0.15
Some	43	+0.01	+0.01 (0.10)	+0.01 (0.01)	−0.02 : +0.04
Limited or minimal	7	0.00	+0.05 (0.22)	−0.02 (0.04)	−0.10 : +0.07
Not mentioned	79	+0.04	+0.07 (0.23)	+0.04 (0.01)	+0.01 : +0.07
Professional development (CPD)					
Is CPD provided to support implementation?			$eta^2 = 0.06$**	$p > 0.10$	
YES, only to direct implementers	78	+0.04	+0.09 (0.23)	+0.05 (0.01)	+0.02 : +0.08
YES, to implementers and other stakeholders	59	0.00	−0.02 (0.24)	+0.01 (0.01)	−0.02 : +0.03
YES, only to other stakeholders	1	–	–	–	–
NO CPD or unclear	11	+0.02	−0.01 (0.014)	−0.01 (0.04)	−0.08 : +0.07
Is CPD subject-specific / curriculum-specific or general? (**)				$p > 0.10$	
Predominantly subject-specific or curriculum-specific	74	+0.02	+0.04 (0.25)	+0.05 (0.02)	+0.01 : +0.08
Predominantly generic	53	+0.02	+0.04 (0.25)	+0.02 (0.01)	−0.01 : +0.04
Mixed generic and subject-specific	16	0.00	+0.02 (0.11)	+0.01 (0.02)	−0.02 : +0.04
Not mentioned	6	+0.05	+0.08 (0.08)	+0.06 (0.03)	+0.01 : +0.12
Type of CPD (see note below)				$p > 0.10$	
Face-to-face	139	+0.04	+0.02 (0.20)	+0.03 (0.01)	+0.01 : +0.04
Online	17	+0.02	+0.04 (0.15)	+0.04 (0.03)	−0.02 : +0.09
Coaching or mentoring	27	+0.02	−0.03 (0.33)	+0.03 (0.02)	−0.01 : +0.06
Cascade 'train the trainer' model	37	+0.02	+0.03 (0.16)	+0.02 (0.01)	0.00 : +0.05
Sequencing of CPD (#)			–	$p > 0.10$	
Pre-intervention only	30	−0.01	−0.04 (0.34)	0.00 (0.04)	−0.07 : +0.08
During the intervention only	24	+0.03	+0.04 (0.12)	+0.03 (0.02)	−0.01 : +0.07
Pre-intervention and during the intervention	83	+0.02	+0.07 (0.22)	+0.03 (0.01)	−0.01 : +0.05
Not mentioned	12	+0.04	+0.02 (0.17)	0.00 (0.05)	−0.10 : +0.10
Who delivers CPD?			$eta^2 = 0.08$**	$p > 0.10$	
Delivery partner	91	+0.02	+0.04 (0.15)	+0.02 (0.01)	−0.01 : +0.04
Other external organisation	10	+0.07	+0.23 (0.51)	+0.07 (0.04)	0.00 : +0.14
Leaders / teachers from schools in the trial	2	–	–	–	–
Mixed	27	+0.01	+0.05 (0.15)	+0.04 (0.01)	+0.01 : +0.07

Not mentioned	19	+0.04	−0.08 (0.37)	+0.01 (0.04)	−0.06 : +0.08
Support and monitoring					
Does delivery partner provide support (other than CPD)?				$p > 0.10$	
Before the intervention only	2	–	–	–	–
During the intervention only	83	+0.01	+0.05 (0.23)	+0.03 (0.04)	−0.06 : +0.11
Before and during the intervention	20	0.00	−0.07 (0.37)	+0.01 (0.01)	−0.01 : +0.03
Other or not mentioned	44	+0.05	+0.06 (0.15)	+0.05 (0.01)	+0.02 : +0.07
Monitoring of implementation				$p < 0.05$**	
Robust monitoring	38	+0.03	+0.03 (0.16)	+0.05 (0.02)	+0.01 : +0.08
Some monitoring	51	+0.01	+0.04 (0.26)	−0.01 (0.01)	−0.03 : +0.02
No monitoring	13	+0.01	−0.08 (0.46)	+0.04 (0.06)	−0.07 : +0.16
Not mentioned	47	+0.04	+0.07 (0.15)	+0.06 (0.02)	+0.03 : +0.10
Fidelity					
Intended fidelity				$p > 0.10$	
Faithful adoption	60	+0.02	+0.05 (0.17)	+0.02 (0.02)	−0.01 : +0.06
Adaptation to context	66	+0.01	+0.04 (0.23)	+0.03 (0.01)	0.00 : +0.05
Not mentioned	23	+0.07	−0.01 (0.37)	+0.06 (0.04)	−0.02 : +0.13
Fidelity related to CPD (**)				$p > 0.10$	
High	22	+0.10	+0.02 (0.39)	+0.09 (0.04)	+0.02 : +0.17
Varied or moderate	53	+0.03	+0.06 (0.26)	+0.02 (0.01)	−0.01 : +0.04
Limited	11	+0.01	+0.01 (0.13)	+0.02 (0.03)	−0.05 : +0.08
Not mentioned	63	+0.02	+0.03 (0.14)	+0.02 (0.01)	−0.01 : +0.05
Actual fidelity of implementation (#)			$eta^2 = 0.17$**	$p > 0.10$	
High	25	−0.01	0.00 (0.17)	+0.01 (0.03)	−0.05 : +0.07
Varied or moderate	89	+0.02	+0.03 (0.27)	+0.02 (0.01)	−0.01 : +0.05
Limited	25	+0.06	+0.09 (0.16)	+0.05 (0.02)	+0.02 : +0.09
Not mentioned	10	+0.01	+0.03 (0.13)	+0.04 (0.03)	−0.01 : +0.10

Table 21: FSM effect size and evaluation design

	n_{ES} =	Unweighted descriptive analyses		Weighted meta-analyses	
		Median	Mean (SD)	Mean (SE)	95% CI
All trials	149	+0.02	+0.04 (0.23)	+0.03 (0.01)	+0.01 : +0.05
Trial description					
Trial design				*p* > 0.10	
RCT	50	+0.02	+0.03 (0.27)	+0.04 (0.02)	0.00 : +0.09
Clustered RCT	99	+0.02	+0.04 (0.22)	+0.02 (0.01)	0.00 : +0.05
Level of randomisation (**)				*p* > 0.10	
School	85	+0.03	+0.05 (0.23)	+0.03 (0.01)	+0.01 : +0.06
Pupil	40	+0.05	+0.03 (0.30)	+0.06 (0.03)	0.00 : +0.12
Class or teacher	6	−0.03	−0.02 (0.06)	−0.02 (0.03)	−0.07 : +0.03
Key Stage or year group	5	−0.02	−0.04 (0.06)	−0.03 (0.02)	−0.08 : +0.01
Parent	10	0.00	+0.02 (0.08)	+0.01 (0.04)	−0.07 : +0.09
Complex (multiple)	3	–	–	–	–
Type of trial			eta^2 = 0.08***	*p* > 0.10	
Efficacy	73	+0.04	+0.10 (0.23)	+0.04 (0.01)	+0.01 : +0.06
Effectiveness	76	0.00	−0.03 (0.22)	+0.02 (0.01)	−0.01 : +0.04
Type of evaluator			eta^2 = 0.02*	*p* > 0.10	
Non-university	55	+0.01	−0.01 (0.23)	+0.03 (0.01)	+0.01 : +0.05
University	94	+0.02	+0.06 (0.23)	+0.03 (0.01)	0.00 : +0.06
Trial length and size					
Length of trial (#)			eta^2 = 0.09**	*p* < 0.01***	
Up to 15 weeks (one term)	45	+0.05	+0.08 (0.16)	+0.03 (0.01)	0.00 : +0.06
16–30 weeks (two terms)	31	+0.01	−0.05 (0.31)	−0.01 (0.03)	−0.06 : +0.05
31–45 weeks (three terms / one year)	34	+0.05	+0.13 (0.30)	+0.07 (0.02)	+0.03 : +0.11
More than 45 weeks / three terms / one year	39	−0.01	−0.02 (0.12)	0.00 (0.01)	−0.03 : +0.03
Number of schools in trial (#)			–	*p* < 0.10*	
20 or less	27	+0.01	+0.09 (0.17)	+0.04 (0.03)	−0.02 : +0.10
21–40	25	0.00	+0.01 (0.47)	0.00 (0.03)	−0.06 : +0.06
41–60	31	+0.07	+0.06 (0.15)	+0.07 (0.03)	+0.02 : +0.12
61–80	19	+0.05	+0.04 (0.19)	+0.04 (0.03)	−0.03 : +0.10
81–100	24	−0.01	−0.03 (0.11)	−0.01 (0.02)	−0.04 : +0.02
101 or more	23	+0.03	+0.03 (0.11)	+0.04 (0.01)	+0.01 : +0.07

Number of pupils in trial			–	*p* > 0.10	
500 or less	26	+0.14	+0.09 (0.48)	+0.10 (0.04)	+0.01 : +0.18
501–1,000	23	+0.01	+0.05 (0.18)	+0.03 (0.03)	−0.04 : +0.10
1,001–2,500	25	+0.10	+0.08 (0.15)	+0.08 (0.03)	+0.02 : +0.13
2,501–5,000	28	+0.03	+0.02 (0.12)	+0.03 (0.02)	−0.01 : +0.07
5,001 or more	43	+0.02	0.00 (0.10)	+0.02 (0.01)	0.00 : +0.04
Statistical sensitivity, attrition and trial quality					
EEF padlock rating (#)			eta^2 = 0.07*	*p* < 0.01***	
0	3	–	–	–	–
1	8	−0.13	−0.02 (0.30)	−0.12 (0.04)	−0.19 : −0.05
2	22	+0.03	+0.09 (0.36)	+0.02 (0.03)	−0.04 : +0.08
3	49	+0.05	+0.08 (0.15)	+0.05 (0.02)	+0.02 : +0.09
4	50	0.00	−0.03 (0.25)	+0.01 (0.02)	−0.02 : +0.04
5	17	+0.02	+0.05 (0.07)	+0.03 (0.01)	0.00 : +0.05
Evaluation burden					
Testing burden				*p* > 0.10	
Low (just NPD)	16	−0.01	+0.02 (0.08)	+0.02 (0.02)	−0.02 : +0.05
Medium (one external test)	43	+0.01	+0.05 (0.28)	+0.01 (0.02)	−0.02 : +0.04
High (two or more external tests)	90	+0.03	+0.03 (0.23)	+0.04 (0.02)	+0.01 : +0.07
IPE data collection burden (#)			eta^2 = 0.04*	*p* > 0.10	
Lowest (no surveys / interviews)	16	+0.16	+0.12 (0.21)	+0.07 (0.06)	−0.05 : +0.19
Medium (just interviews or surveys but not both)	71	+0.01	−0.01 (0.23)	+0.02 (0.01)	−0.01 : +0.05
High (interviews and surveys)	62	+0.03	+0.07 (0.24)	+0.04 (0.01)	+0.01 : +0.06

Primary outcome(s)					
Types of primary outcome (simple) (#)				*p* > 0.10	
Commercial	84	+0.02	+0.08 (0.23)	+0.04 (0.01)	+0.01 : +0.07
Official / SATs	54	+0.02	0.00 (0.12)	+0.02 (0.01)	−0.01 : +0.04
Other / mixed	11	−0.02	−0.12 (0.48)	+0.05 (0.05)	−0.06 : +0.16

Tables for cost effectiveness analyses

Cost effectiveness of interventions given that a positive impact was reported

Please see the *Presenting the outcome variables* section in the main report for detail on how the cost effectiveness outcome variable was derived. To summarise, two criteria were applied to the trials for their inclusion in the cost effectiveness outcome variable: First, at the effect size level, only effect sizes above +0.05 SD were included; second, at the trial-level, only trials where at least half of the effect sizes were above the +0.05 SD threshold were included. These two criteria ensured that only trials where there was reasonable evidence of positive impact are included in this outcome. A total of 40 trials fulfilled these two criteria and the cost (per pupil) for an effect size of 0.10 SD was calculated for these trials as the cost effectiveness outcome.

In constructing the cost effectiveness outcome variable, 42 trials were dropped because they did not meet the two criteria outlined above. This process led to the derivation of a supplementary outcome variable measuring the probability for inclusion in the cost effectiveness outcome. Overall this probability was (40/82 =) 0.49 across the 82 trials. This can be interpreted as the probability of an EEF trial reporting a positive impact (i.e., an effect size above +0.05 SD). Unlike the meta-analyses, this measure of a positive impact does not take account of uncertainty in effect size estimates but does provide a second (trial-level) perspective to supplement the effect size meta-analyses and provide context for the analyses of the cost effectiveness of interventions.

The analyses are summarised statistically using five tables (one for each of the themes). To highlight a finding of interest, the weighted mean effect size is shown in red text where it differs notably from the other categories within that variable.

- Table 22 and 23 presents the cost effectiveness variables across categories of explanatory variables included in the intervention theme using (unweighted) mean and median statistics.
- Table 24 uses the same approach to present the cost effectiveness variables across categories of explanatory variables included in the theory & evidence theme.
- Table 25 uses the same approach to present the cost effectiveness variables across categories of explanatory variables included in the context theme.
- Table 26 uses the same approach to present the cost effectiveness variables across categories of explanatory variables included in the implementation & fidelity theme.
- Table 27 uses the same approach to present the cost effectiveness variables across categories of explanatory variables included in the evaluation design theme.

Cost effectiveness and the intervention

Table 22: Cost effectiveness and the intervention

	Total number of trials	Number of trials included in CE outcome	Probability of inclusion in CE outcome.	Cost effectiveness (£ per pupil for an effect size above +0.05)	
				Median	Mean (SD)
All trials	82	40	0.49	£54	£150 (£229)
Focus of intervention					
School phase (*)			$eta^2 = 0.01$	$p = 0.09$*	$eta^2 = 0.12$
Primary (including Early Years)	51	25	0.49	£43	£130 (£223)
Primary–secondary transition	6	4	0.67	£376	£385 (£383)
Secondary	25	11	0.44	£69	£109 (£126)
School Key Stage (*)			$eta^2 = 0.03$	$p = 0.10$*	$eta^2 = 0.20$

Early Years	2	1	0.50	–	–
Primary (KS1)	13	7	0.54	£11	£25 (£23)
Primary (KS2)	33	16	0.48	£60	£188 (£264)
Primary (multiple Key Stages)	3	1	0.33	–	–
Transition KS2–KS3	6	4	0.67	£376	£385 (£383)
Secondary KS3	20	10	0.50	£79	£119 (£128)
Secondary KS4	4	1	0.25	–	–
Secondary (multiple Key Stages)	1	0	0.00	–	–
Intervention curriculum area (#)			$eta^2 = 0.06$	$p > 0.10$	$eta^2 = 0.05$
Cross-curriculum	29	10	0.34	£24	£196 (£315)
English	36	20	0.56	£65	£171 (£232)
Maths	14	9	0.64	£62	£67 (£48)
Science	3	1	0.33	–	–
Intensity of delivery					
Minutes per week (*)			$eta^2 = 0.02$	$p = 0.08$*	$eta^2 = 0.11$
30 mins or less	12	5	0.42	£62	£162 (£269)
31–60 mins	13	5	0.38	£10	£19 (£23)
61–120 mins	15	7	0.47	£69	£119 (£118)
Over 120 mins per week	11	7	0.64	£183	£285 (£273)
No Intensity data	31	16	0.52	£40	£141 (£259)
Direct implementers					
Direct implementers (*)			$eta^2 = 0.07$	$p = 0.08$*	$eta^2 = 0.27$***
Teacher-led	37	19	0.51	£33	£50 (£53)
TA-led	12	9	0.75	£62	£139 (£203)
Externally-led	18	6	0.33	£257	£364 (£358)
Other	15	6	0.40	£171	£269 (£307)
Perceived quality of support resources (#)			$eta^2 = 0.01$	$p > 0.10$	$eta^2 = 0.16$*
High	20	11	0.55	£43	£48 (£50)
Variation	27	13	0.48	£62	£106 (£115)
Low	5	3	0.60	–	–
Total cost			$eta^2 = 0.21$***	$p > 0.10$	$eta^2 = 0.10$
<£100k	4	2	0.50	–	–
£100–<£250k	14	6	0.43	£101	£256 (£322)
£250–<£500k	28	22	0.79	£34	£96 (£160)
£500–<£750k	21	7	0.33	£62	£249 (£355)

£750–<£1 million	9	2	0.22	–	–
£1 million+	6	1	0.17	–	–
EEF promising intervention					
EEF promising intervention (***)			eta^2 = 0.22***	$p < 0.01$***	eta^2 = 0.10**
Yes	17	16	0.94	£17	£61 (£111)
No	65	24	0.37	£89	£209 (£268)

Table 23: Cost effectiveness and EEF school themes

EEF intervention themes	*N / N'*	*n / n'*	*p / p'*	Med / med'	Mean / mean'
Language and literacy	38 / 44	22 / 18	0.58 / 0.41	£55 / £51	£157 / £140
Staff deployment and development	36 / 46	18 / 22	0.50 / 0.48	£55 / £51	£89 / £200
Organising your school	18 / 64	9 / 31	0.50 / 0.48	£163 / £43*	£335 / £96***
Developing effective learners	17 / 65	10 / 30	0.59 / 0.46	£13 / £60	£46 / £184*
Mathematics	16 / 66	11 / 29	0.69 / 0.44	£59 / £48	£56 / £185
Feedback and monitoring pupil progress	10 / 72	7 / 33	0.70 / 0.46	£33 / £62	£41 / £173
Behaviour	8 / 74	3 / 37	–	–	–
Character and essential life skills	7 / 75	2 / 38	–	–	–
Parental engagement	6 / 76	0 / 40	–	–	–
Enrichment	4 / 78	1 / 39	–	–	–
Science	3 / 79	1 / 39	–	–	–
Early years	3 / 79	2 / 38	–	–	–
Special educational needs and disabilities	2 / 80	1 / 39	–	–	–

Key: The EEF school themes are not mutually exclusive: trials can be included in more than one of the school themes. Table 23 takes each school theme to illustrate the number of trials included (and not included) in each.

Label	Details
N / *N*'	Of the 82 trials included in the review: • *N* is the number of trials included in a specific EEF school theme • *N*' is the number of trials not included
n / *n*'	Of the 40 trials included in the cost effectiveness outcome: • *n* is the number of trials included in a specific EEF school theme that were included in the cost effectiveness outcome • *n*' is the number of trials not included in the cost effectiveness outcome
p / *p*':	• *p* is the probability of trials included in a specific EEF school theme being included in the cost effectiveness outcome • *p*' is the probability of trials not included in a specific EEF school theme being included in the cost effectiveness outcome
med / med'	• med is the median cost effectiveness for trials included in a specific EEF school theme • med' is the median cost effectiveness for trials not included
mean / mean'	• mean is the mean cost effectiveness for trials included in a specific EEF school theme

- mean' is the mean cost effectiveness for trials not included

Cost effectiveness and theory & evidence

Table 24: Cost effectiveness and theory & evidence

	Total number of trials	Number of trials included in CE outcome	Probability of inclusion in CE outcome	Cost effectiveness (£ per pupil for an effect size above +0.05)	
				Median	Mean (SD)
All trials	82	40	0.49	£54	£150 (£229)
Empirical evidence and theoretical detail					
Empirical evidence (**)			$eta^2 = 0.00$	$p = 0.02$**	$eta^2 = 0.25$***
Strong evidence	17	9	0.53	£43	£74 (£126)
Some evidence	56	27	0.48	£48	£126 (£198)
Minimal / none	9	4	0.44	£482	£483 (£358)
Theoretical detail			$eta^2 = 0.02$	$p > 0.10$	$eta^2 = 0.04$
Highly detailed	17	6	0.35	£54	£68 (£60)
Some detail	28	15	0.54	£40	£125 (£230)
Minimal / none	37	19	0.51	£89	£195 (£258)
Causal processes and mechanisms					
Direct or training-based			$eta^2 = 0.05$	$p > 0.10$	$eta^2 = 0.21$**
Training	64	33	0.52	£45	£108 (£179)
Direct	16	5	0.31	£403	£421 (£370)
Other	2	2	1.00	–	–
Focus of change			$eta^2 = 0.13$**	$p > 0.10$	$eta^2 = 0.01$
Learning focus	69	39	0.57	£59	£153 (£231)
Teacher change focus	3	0	0.00	–	–
Wider pupil outcome focus	9	1	0.11	–	–
Other focus	1	0	0.00	–	–

Cost effectiveness and context

Table 25: Cost effectiveness and context

	Total number of trials	Number of trials included in CE outcome	Probability of inclusion in CE outcome	Cost effectiveness (£ per pupil for an effect size above +0.05)	
				Median	Mean (SD)
All trials	82	40	0.49	£54	£150 (£229)
External context					
Geography			$eta^2 = 0.09$	$p < 0.10$	$eta^2 = 0.10$
National	25	9	0.36	£15	£152 (£249)
One geographical location	19	11	0.58	£183	£254 (£282)
Two or three geographical areas	22	15	0.68	£59	£111 (£196)
Other	16	5	0.31	£14	£33 (£43)

Cost effectiveness and implementation & fidelity

Table 26: Cost effectiveness and implementation & fidelity

	Total number of trials	Number of trials included in CE outcome	Probability of inclusion in CE outcome	Cost effectiveness (£ per pupil for an effect size above +0.05)	
				Median	Mean (SD)
All trials	82	40	0.49	£54	£150 (£229)
Developer characteristics					
Type of developer (#)			$eta^2 = 0.15$**	$p > 0.10$	$eta^2 = 0.10$
Not-for-profit / charity	32	13	0.41	£62	£238 (£306)
University	19	6	0.32	£28	£74 (£118)
Private company	9	6	0.67	£41	£159 (£318)
School, academy chain or MAT	9	8	0.89	£76	£107 (£132)
Council / local authority	8	3	0.38	/	/
Mixed	5	4	0.80	£25	£26 (£19)
Professional development (CPD)					
Generic or subject-specific (#)			$eta^2 = 0.07$	$p > 0.10$	$eta^2 = 0.09$
Predominantly subject-specific or curriculum-specific	49	29	0.59	£62	£159 (£231)
Predominantly generic	22	7	0.32	£33	£70 (£109)
Mixed generic and subject-specific	7	2	0.29	/	/
Not mentioned	4	2	0.50	/	/

Was CPD provided?			$eta^2 = 0.02$	$p > 0.10$	$eta^2 = 0.07$
YES, only to direct implementers	46	21	0.46	£59	£154 (£237)
YES, only to direct implementers and other stakeholders	30	17	0.57	£45	£115 (£174)
YES, only to stakeholders who are not direct implementers	1	0	0.00	/	/
No CPD or unclear	5	2	0.40	/	/
Type of CPD					
Face-to-face training (***)	74	36	0.49	£47	£119 (£199)***
Online training	11	5	0.45	£45	£42 (£37)
Coaching or mentoring	13	6	0.46	£43	£56 (£56)
Cascade 'train the trainer'	16	9	0.56	£34	£107 (£147)
Fidelity					
Intended fidelity			$eta^2 = 0.02$	$p > 0.10$	$eta^2 = 0.00$
Faithful adoption	37	16	0.43	£60	£163 (£253)
Adaptation to context	31	15	0.48	£45	£135 (£202)
Not mentioned	14	9	0.64	£40	£150 (£252)
CPD fidelity			$eta^2 = 0.03$	$p > 0.10$	$eta^2 = 0.12$
High	12	8	0.67	£41	£47 (£51)
Varied or moderate	26	13	0.50	£33	£133 (£197)
Limited	6	2	0.33	£408	£408 (£451)
Not mentioned or unclear	38	17	0.45	£62	£180 (£265)
Actual fidelity of implementation			$eta^2 = 0.02$	$p > 0.10$	$eta^2 = 0.08$
High	13	7	0.54	£43	£79 (£105)
Varied or moderate	46	20	0.43	£43	£146 (£224)
Limited	14	7	0.50	£62	£110 (£187)
Not mentioned or unclear	9	6	0.67	£79	£290 (£361)

Cost effectiveness and evaluation design

Table 27: Cost effectiveness and evaluation design

	Total number of trials ($N_T = 82$)	Number of trials included in CE outcome	Probability of inclusion in CE outcome	Cost effectiveness (£ per pupil for an effect size above +0.05)	
				Median	Mean (SD)
All trials	82	40	0.49	£54	£150 (£229)
Trial description					
Trial design (**)			eta^2 = 0.04*	p = 0.03**	eta^2 = 0.07*
RCT	27	17	0.63	£107	£221 (£263)
Clustered RCT	55	23	0.42	£34	£97 (£190)
Level of randomisation (**)			eta^2 = 0.10	p = 0.04**	eta^2 = 0.28**
School	48	21	0.44	£34	£71 (£137)
Pupil	25	17	0.68	£107	£221 (£263)
Class or teacher	4	1	0.25	–	–
Key Stage or year group	2	1	0.50	–	–
Parent	2	0	0.00	–	–
Complex or multiple	1	0	0.00	–	–
Type of trial (EEF-defined)			eta^2 = 0.00	p > 0.10	eta^2 = 0.02
Efficacy	41	19	0.46	£43	£112 (£205)
Effectiveness	41	21	0.51	£62	£184 (£249)
Length and size of trial					
Length of trial (categorised) (#)			eta^2 = 0.08	p > 0.10	eta^2 = 0.09
Within one term (up to 15 weeks)	23	14	0.61	£84	£223 (£283)
Within two terms (16–30 weeks)	21	12	0.57	£55	£161 (£249)
Within 3 terms (1 year, 31–45 weeks)	21	10	0.48	£13	£78 (£125)
46+ weeks	17	4	0.24	£33	£39 (£36)
Number of schools in trial (#)			eta^2 = 0.08	p > 0.10	eta^2 = 0.11
20 or less	15	9	0.60	£112	£176 (£187)
21–40	16	10	0.63	£88	£222 (£295)
41–60	16	9	0.56	£62	£131 (£256)
61–80	8	3	–	–	–
81–100	10	3	–	–	–
101 or more	17	6	0.35	£11	£22 (£23)
Number of pupils in trial (#)			eta^2 = 0.14**	p > 0.10	eta^2 = 0.11

500 or less	19	14	0.74	£88	£232 (£284)
501–1,000	12	6	0.50	£52	£158 (£245)
1,001–2,500	15	9	0.60	£59	£153 (£238)
2,501–5,000	14	3	0.21	–	–
5,001 or more	20	7	0.35	£33	£42 (£38)
Statistical sensitivity, attrition and trial quality					
EEF padlock rating (#)			$eta^2 = 0.09$	$p > 0.10$	$eta^2 = 0.19$
0	3	3	1.00	–	–
1	7	2	0.29	–	–
2	12	6	0.50	£88	£218 (£309)
3	27	15	0.56	£62	£137 (£196)
4	24	12	0.50	£36	£95 (£179)
5	9	2	0.22	–	–
Types of primary outcome (simple) (*)			$eta^2 = 0.05$	$p = 0.09$*	$eta^2 = 0.12$*
Commercial	51	27	0.53	£89	£204 (£262)
Official / SATs	22	7	0.32	£15	£42 (£57)
Other / mixed	9	6	0.67	£36	£33 (£21)
Trial / outcome curriculum area (#)			$eta^2 = 0.08$	$p > 0.10$	$eta^2 = 0.108$
Cross-curriculum trial and outcome	8	1	0.13	–	–
Cross-curriculum trial, multiple subject and outcomes	16	7	0.44	£34	£175 (£298)
English trial and outcome	40	21	0.53	£69	£197 (£257)
Maths trial and outcome	15	10	0.67	£60	£61 (£49)
Science trial and outcome	3	1	0.33	–	–

Tables for attrition analyses

Note

On undertaking these analyses of the attrition outcome, it became apparent that this outcome variable was not aligned with some of the explanatory variables included in the review. Specifically, variables under the intervention and / or implementation & fidelity themes capture descriptive aspects of an intervention and how it was implemented within an evaluation. These variables will therefore be focused on intervention group samples and will have little / no relevance for control group samples. However, the pupil-level attrition outcome is a measure of overall attrition (i.e., in both intervention and control schools). Whilst analyses presented in this section do examine how explanatory variables are associated with overall pupil attrition, future reviews may want to collect attrition rates for intervention and control group samples separately. In doing this, the analyses on how explanatory variables relating to the intervention and / or implementation are associated with intervention group attrition would be more meaningful because the variables would be more closely aligned.

The review identified the pupil-level attrition rates for 79 of the 82 (96%).

The analyses are summarised statistically using five tables (one for each of the themes). To highlight a finding of interest, the weighted mean effect size is shown in red text where it differs notably from the other categories within that variable.

- Table 28 presents the average attrition across categories of explanatory variables included in the intervention theme using (unweighted) mean and median statistics.
- Table 29 uses the same approach to present average attrition across categories of explanatory variables included in the theory & evidence theme.
- Table 30 uses the same approach to present average attrition across categories of explanatory variables included in the context theme.
- Table 31 uses the same approach to present average attrition across categories of explanatory variables included in the implementation & fidelity theme.
- Table 32 uses the same approach to present average attrition across categories of explanatory variables included in the evaluation design theme.

Attrition and the intervention

Table 28: Attrition and the intervention

	Total number of trials	Number of trials included in attrition outcome	Overall attrition (% pupil-level attrition)	
			Median	Median
All trials	82	79	15.2	19.4 (16.54)
Focus of intervention				
School phase (#)			$p > 0.10$	$eta^2 = 0.06$
Primary (including Early Years)	51	49	15.2	18.8 (15.07)
Primary–secondary transition	6	6	16.7	32.8 (30.91)
Secondary	25	24	15.6	17.4 (13.93)
School Key Stage (#)			$p > 0.10$	$eta^2 = 0.14$
Early Years	2	2	–	–

Primary (KS1)	13	13	11.0	15 (10.58)
Primary (KS2)	33	31	16.0	20.6 (17.06)
Primary (multiple Key Stages)	3	3	–	–
Transition KS2–KS3	6	6	16.7	32.8 (30.91)
Secondary KS3	20	19	21.0	18.7 (13.37)
Secondary KS4	4	4	5.7	5.4 (5.11)
Secondary (multiple Key Stages)	1	1	–	–
Intervention curriculum area (#)			$p > 0.10$	$eta^2 = 0.06$
Cross-curriculum	29	27	15.2	18.8 (15)
English	36	35	16.2	22.9 (19.59)
Maths	14	14	7.4	12.2 (9.3)
Science	3	3	–	–
Intensity				
Intensity of delivery (#)			$p > 0.10$	$eta^2 = 0.04$
30 mins or less	12	12	13.0	13.9 (8.03)
31–60 mins	13	13	10.7	15.9 (14.51)
61–120 mins	15	15	16.2	21.4 (18.09)
Over 120 mins per week	11	11	21.3	23.7 (19.33)
N/A or no detail on intensity	31	28	17.0	20.7 (18.2)
EEF promising intervention				
EEF promising intervention (**)			$p = 0.05$**	$eta^2 = 0.04$*
Not classed as promising	65	63	16.2	21.2 (17.45)
Classed as promising	17	16	10.8	12.5 (9.95)

Table 29: Attrition and EEF school themes

EEF intervention school themes (see key)	*N* / *N*'	Med / med'	Mean / mean'
Language and literacy	37 / 42	18.3 / 12.9*	23.1 / 16.1*
Staff deployment and development	35 / 44	16.2 / 13.6	18.1 / 20.5
Organising your school	17 / 62	22.0 / 14.9	22.6 / 18.5
Developing effective learners	17 / 62	16.0 / 14.0	22.6 / 18.5
Mathematics	16 / 63	10.9 / 15.2	13.7 / 20.9
Feedback and monitoring pupil progress	9 / 70	8.1 / 15.2	20.5 / 19.3
Behaviour	8 / 71	10.5 / 15.2	11.1 / 20.4
Character and essential life skills	6 / 73	12.6 / 15.2	15.5 / 19.7
Parental engagement	6 / 73	27.9 / 15.2	29.8 / 18.6

Enrichment	3 / 76	–	–
Science	3 / 76	–	–
Early years	3 / 76	–	–
Special educational needs and disabilities	2 / 77	–	–

Key: *N* / *N*' – of the 79 trials in the review with a pupil-level attrition rate, *N* = number placed in the theme; *N** = number not placed in the theme.
Med / med' – med : median attrition rate for trials that are placed in a theme; med*: median attrition rate for trials that are not placed in a theme.
Mean / mean' – mean : mean attrition rate for trials placed in a theme; mean* : mean attrition rate for trials that are not placed in a theme.

Theory & evidence

Table 30: Attrition and theory & evidence

	Total number of trials	Number of trials included in attrition outcome	Overall attrition (% pupil-level attrition)	
			Median	Mean (SD)
All trials	82	79	15.2	19.4 (16.54)
Causal processes and mechanisms				
Direct or training-based			$p > 0.10$	$eta^2 = 0.00$
Training	64	61	15.5	19.2 (14.75)
Direct	16	16	10.5	20.1 (22.48)
Other	2	2	–	–
Focus of change			$p > 0.10$	$eta^2 = 0.01$
Learning focus	69	67	15.2	20.2 (17.22)
Teacher change focus	3	3	–	–
Wider pupil outcome focus	9	8	17.0	16.7 (12.09)
Other focus	1	1	–	–

Attrition and evaluation context

Table 31: Attrition and context

	Total number of trials	Number of trials included in attrition outcome	Overall attrition (% pupil-level attrition)	
			Median	Mean (SD)
All trials	82	79	15.2	19.4 (16.54)
External context				
Geography			$p > 0.10$	$eta^2 = 0.03$
National	25	23	15.2	21.9 (18.28)
One geographical location	19	19	16.0	22.5 (19.93)
Two or three geographical areas	22	22	14.3	16.5 (11.91)
Other	16	15	14.8	16.1 (15.05)

Implementation & fidelity

Table 32: Attrition and implementation & fidelity

	Total number of trials $(N_T = 82)$	Number of trials included in attrition outcome	Overall attrition (% pupil-level attrition)	
			Median	Median
All trials	82	79	15.2	19.4 (16.54)
Developer characteristics				
Type of developer			$p > 0.10$	$eta^2 = 0.07$
Not-for-profit / charity	32	30	17.2	22.9 (20.12)
University	19	18	12.9	16.9 (12.43)
Private company	9	9	25.7	24.6 (22.34)
School, academy chain or MAT	9	9	15.0	12.9 (7.58)
Council / local authority	8	8	10.2	12.8 (9.86)
Mixed	5	5	21.0	20.7 (9.66)
Focus, planning, time and support				
Clarity of implementation plan			$p > 0.10$	$eta^2 = 0.02$
Clearly understood	33	31	15.2	18.7 (12.7)
Variation in understanding	23	23	9.4	16.8 (15.2)
Unclear or not mentioned	26	25	18.0	22.7 (21.4)
Lead-in time			$p > 0.10$	$eta^2 = 0.01$
Sufficient time	5	5	11.0	16.7 (11.9)

Variation in perceptions	14	14	13.1	19.5 (16.3)
Insufficient time	24	22	16.5	21.3 (17.2)
Not mentioned	39	38	15.4	18.7 (17.2)
Professional development (CPD)				
Is CPD provided			$p > 0.10$	$eta^2 = 0.01$
YES, only to implementers	46	45	15.2	20.7 (16.8)
YES, implementers and others	30	28	14.8	17.2 (13.3)
YES, to non-direct implementers	1	1	–	–
No CPD or unclear	5	5	8.0	21.3 (31)
Generic or subject-specific			$p > 0.10$	$eta^2 = 0.03$
Predominantly subject-specific	49	49	15.5	18.6 (14.0)
Predominantly generic	22	19	15.2	22.1 (20.2)
Mixed generic / subject-specific	7	7	12.5	13.7 (10.4)
Not mentioned	4	4	15.0	27.3 (32.9)
Sequencing of CPD			$p > 0.10$	$eta^2 = 0.02$
Pre-intervention only	18	17	15.2	22.1 (17.6)
During the intervention only	10	9	22.0	23.7 (12.9)
Pre-intervention and during the intervention	47	46	15.4	18.0 (15.3)
Not mentioned	7	7	8.0	16.8 (26.3)
Support and monitoring				
Non-CPD support			$p > 0.10$	$eta^2 = 0.01$
Before the intervention only	1	1	–	–
During the intervention only	12	12	12.5	20.5 (21.3)
Before and during intervention	47	44	15.4	20.3 (17.7)
Other or not mentioned	22	22	14.2	17.0 (11.3)
Monitoring of implementation			$p > 0.10$	$eta^2 = 0.04$
Robust monitoring	14	13	15.2	13.8 (7.8)
Some monitoring	28	27	21.3	21.0 (16.1)
No monitoring	8	8	10.8	14.4 (10.1)
Not mentioned	32	31	14.0	21.7 (20.3)
Fidelity				
Intended fidelity			$p > 0.10$	$eta^2 = 0.02$
Faithful adoption	37	35	18.2	21.0 (16.3)
Adaptation to context	31	30	16.0	19.8 (15.7)
Not mentioned	14	14	10.8	14.7 (19.1)

Fidelity relating to CPD			$p > 0.10$	$eta^2 = 0.02$
High	12	12	14.6	15.2 (7.7)
Varied or moderate	26	26	13.0	18.5 (16.1)
Limited	6	4	27.5	26.1 (17.8)
Not mentioned	38	37	15.5	20.7 (18.9)
Implementation fidelity			$p > 0.10$	$eta^2 = 0.04$
High	13	13	13.3	14.7 (9.8)
Varied or moderate	46	45	12.5	18.7 (16.2)
Limited	14	12	19.0	22.0 (21.0)
Not mentioned	9	9	21.0	26.5 (19.2)

Evaluation design

Table 33: Attrition and evaluation design

	Total number of trials	Number of trials included in attrition outcome	Overall attrition (% pupil-level attrition)	
			Median	Mean (SD)
All trials	82	79	15.2	19.4 (16.54)
Trial description				
Trial design			$p > 0.10$	$eta^2 = 0.00$
RCT	27	27	12.0	19.5 (18.91)
Clustered RCT	55	52	16.0	19.4 (15.36)
Level of randomisation			$p > 0.10$	$eta^2 = 0.02$
School	48	45	16.0	18.9 (14.88)
Pupil	25	25	12.0	20.2 (19.49)
Class or teacher	4	4	12.0	21.7 (23.77)
Year or Key Stage	2	2	–	–
Parent	2	2	–	–
Other / complex	1	1	–	–
Type of trial (EEF-defined) (#)			$p > 0.10$	$eta^2 = 0.04$*
Efficacy	41	40	15.8	22.9 (19.1)
Effectiveness	41	39	13.3	15.9 (12.71)
Type of evaluator			$p > 0.10$	$eta^2 = 0.02$
University	30	29	12.0	16.2 (13.37)
Non-university	52	50	17.1	21.3 (17.98)

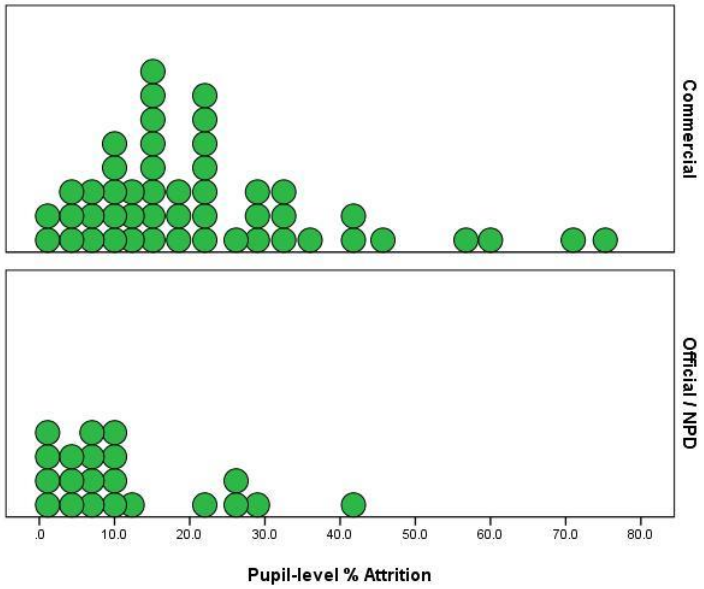
Length and size of trial				
Length of trial			$p > 0.10$	$eta^2 = 0.01$
Within one term (up to 15 weeks)	23	23	14.8	21.1 (19.08)
Within two terms (15–30 weeks)	21	21	12.0	17.6 (16.10)
Within 3 terms (1 year, 30–14 weeks)	21	21	16.0	19.4 (15.41)
More than one academic year	17	14	15.8	19.3 (15.81)
Number of schools in trial (**)			$p = 0.01$**	$eta^2 = 0.18$**
20 or less	15	15	11.0	14.8 (16.72)
21–40	16	16	25.5	28.8 (21.15)
41–60	16	15	23.0	26.8 (16.27)
61–80	8	7	18.3	13.1 (11.83)
81–100	10	10	8.8	11.1 (6.44)
101 or more	17	16	12.0	15.5 (11.62)
Number of pupils in trial			$p > 0.10$	$eta^2 = 0.05$
500 or less	19	19	15.0	23.2 (23.56)
501–1,000	12	12	11.0	17.3 (13.29)
1,001–2,500	15	14	16.9	21.7 (14.59)
2,501–5,000	14	13	18.8	22.6 (16.09)
5,001 or more	20	19	13.3	14.7 (10.76)
Statistical sensitivity, attrition and trial quality				
EEF padlock rating (***)			$p = < 0.01$***	$eta^2 = 0.45$***
0	3	3	/	/
1	7	7	42.8	41.2 (14.35)
2	12	11	29.0	25.8 (20.42)
3	27	25	16.0	16.3 (11.22)
4	24	24	10.8	12.1 (4.97)
5	9	9	9.1	11.2 (8.94)
Evaluation burden				
Testing burden (**)			$p = < 0.01$***	$eta^2 = 0.10$**
Low (just NPD)	9	9	3.7	7.5 (8.03)
Medium (one external test)	24	24	18.3	25 (21.58)
High (two or more external tests)	49	46	16.0	18.9 (13.39)
IPE data collection burden			$p > 0.10$	$eta^2 = 0.01$
Lowest (no surveys or interviews)	12	12	15.3	17.5 (20.34)
Medium (surveys or interviews)	27	25	12.0	18.6 (16.11)

High (interviews and teacher surveys)	43	42	16.7	20.5 (15.96)
Primary ITT outcome				
Types of primary outcome (***)			$p < 0.01$***	$eta^2 = 0.09$**
Commercial	51	49	18.2	22.3 (16.97)
Official / NPD	22	21	8.0	11.4 (10.95)
Other / mixed	9	9	16.0	22.3 (20.24)

Elaborating the attrition analyses to account for type of primary outcome

A clear and expected link between type of primary outcome (commercial or official / NPD) and overall pupil-level attrition rate was observed under the evaluation design theme. Figure 1.1 illustrates the distributions of attrition rates for the 49 trials that used a commercial test as a primary outcome and the 21 trials that used an official / NPD outcome. On average, evaluations that used commercial tests reported higher attrition (mean = 22.3%; median = 18.2%) compared evaluations that used NPD or other official data (mean = 11.4%; median = 8.0%).

Figure 1.1: Dot-plot of pupil-level % attrition by type of primary outcome



The association between type of primary outcome and attrition is likely to confound the interpretation of how other explanatory variables are associated with attrition. For example, differences in attrition rates across school phases and key stages may be explained by greater use of commercial tests for some phases / years compared with others. To examine this, elaboration analyses were undertaken for the following selection of explanatory variables under the five thematic groupings:

- The intervention
  - School phase and key stage
  - Curriculum focus of intervention
  - Intensity of intervention
  - EEF intervention themes
  - EEF promising intervention identifier
- Theory & evidence
  - – (none)
- Context

    - Publication year

- Implementation & fidelity
    - Type of developer

- Evaluation design
    - Type trial
    - Type of evaluator
    - Testing burden

The elaboration analyses compare the association between pupil-level attrition and the above variables for trials that used a commercial or official / NPD primary outcome using median attrition rates.

Elaborating intervention

Commercial and official / NPD outcomes

In terms of school phase, overall attrition rates for primary to secondary school transition interventions were observed to be higher than those seen with transitions located in secondary or primary schools. However, when the type of primary outcome is accounted for, a different pattern emerges. First, five of the six primary to secondary school transition interventions used a commercial test for the primary outcome. Second, for evaluations using commercial tests, attrition rates for primary to secondary school transition interventions were lower (median = 11.4%) compared with those seen with interventions in secondary (18.6%) or primary (18.3%) schools. The overall attrition rates for interventions in secondary or primary schools are smaller because of the use of an NPD primary outcome in seven secondary school interventions (median attrition = 9.4%) and 14 primary school interventions (median = 7.5%).

In terms of school key stage, median attrition rates ranged between 5.7% for the four KS4 interventions (all of which used an NPD outcome) to 26.6% for the 15 KS2 interventions that used a commercial test outcome. The vast majority of KS3 interventions used a commercial test (16 out of 20) as did the majority of KS2 interventions (15 out of 33). When comparisons are possible, the use of commercial tests results in higher attrition rates compared with the use of NPD outcomes.

Whilst the overall median attrition rate for interventions that focused on maths (7.4%) was notably lower than English (16.2%) or cross-curriculum (16.0%) interventions, this seems to relate primarily to the type of primary outcome used (i.e., commercial or NPD). The use of an NPD primary outcome was more common in maths (five out of 14, 36%) and cross-curriculum (13 out of 27, 48%) compared with English (two out of 35, 6%) interventions. Amongst evaluations that used a commercial test, attrition rates for maths (16.0%) were comparable to English (15.7%) but a higher rate was observed for cross-curriculum interventions (21.5%). Amongst evaluations that used an NPD outcome, attrition rates for maths (6.9%) were slightly lower than cross-curriculum interventions (9.1%).[1]

The association between attrition and the intensity of an intervention remains unclear when type of primary outcome is accounted for. A suggestion of a weak positive correlation between the two remains. However, as discussed earlier, the problem of alignment between the attrition outcome (for intervention and control school pupil samples) and the focus of this explanatory variable (intensity of the intervention) serve to obscure interpretation. The use of separate rates of attrition for intervention and control samples would be one way of addressing this lack of alignment in future reviews.

The median attrition rate for evaluations of interventions classed as 'promising' by EEF was observed to be consistently lower than rates for evaluations of interventions not classed as promising. This pattern is seen overall (11.0% promising; 16.1% other) whether the primary outcome used was a commercial test (14.6% promising; 18.8% other) or taken from official / NPD data (5.4% promising; 9.1% other).

[1] The number of English interventions that used an NPD primary outcome was too few (*n* = 3) to analyse.

Table 34: Attrition and the intervention (commercial and official / NPD outcomes)

Type of primary outcome	Commercial test		Official / NPD	
	n =	Median	*n* =	Median
All trials with attrition rate	49	18.3	21	8.0
School phase		$p > 0.10$		$p > 0.10$
Primary (including Early Years)	28	18.3	14	7.5
Primary–secondary transition	5	11.4	0	–
Secondary	16	18.6	7	9.4
School Key Stage		$p > 0.10$		$p > 0.10$
Early Years	2	–	0	–
Primary (KS1)	8	12.0	2	–
Primary (KS2)	15	26.6	12	7.5
Primary (multiple Key Stages)	3	–	0	–
Transition KS2–KS3	5	11.4	0	–
Secondary KS3	16	18.6	3	–
Secondary KS4	0	–	4	5.7
Secondary (multiple Key Stages)	0	–	0	–
Intensity of intervention		$p > 0.10$		$p > 0.10$
30 mins or less	10	15.3	1	–
31–60 mins	8	15.2	5	6.9
61–120 mins	11	16.2	2	–
Over 120 mins per week	8	21.5	3	–
No intensity detail	12	22.0	10	9.3
Intervention curriculum area		$p > 0.10$		$p > 0.10$
Cross-curriculum	12	21.5	13	9.1
English	30	15.7	2	–
Maths	7	16.0	5	6.9
Science	0	–	1	–
EEF promising intervention		p > 0.10		p > 0.10
Not classed as promising	43	18.8	17	9.1
Classed as promising	8	14.6	5	5.4

Table 35 shows that once type of primary outcome is accounted for, much higher average attrition rates are consistently seen for interventions with a commercial test primary outcome; across the EEF themes, rates ranged between 13.6% (parental engagement) and 26.0% (organising your school). For evaluations that used an NPD outcome, attrition rates ranged between 6.9% (mathematics) and 17.3% (staff deployment and development).

Table 35: Attrition and the intervention (commercial and official / NPD outcomes)

Type of primary outcome	Commercial test		Official / NPD	
	N	Median	*N*	Median
Trials with attrition rate	49	18.3	21	8.0
EEF intervention themes	N_c / *Nc*'	Med_c / Med_c'	N_o / N_o'	Med_o / Med_o
Language and literacy	32 / 17	17.2 / 18.2	2 / 19	–
Staff deployment and development	21 / 28	18.2 / 18.7	10 / 11	17.3 / 3.9
Organising your school	10 / 39	26.0 / 15.5	6 / 15	6.9 / 10
Developing effective learners	9 / 40	21.0 / 17.2	4 / 17	8.8 / 8
Mathematics	9 / 40	16.0 / 18.5	5 / 16	6.9 / 9.7
Feedback and monitoring pupil progress	4 / 45	24.3 / 18.2	2 / 19	–
Behaviour	6 / 43	14.9 / 18.2	2 / 19	–
Character and essential life skills	2 / 47	–	3 / 18	–
Parental engagement	4 / 45	13.6 / 18.3	1 / 20	–
Enrichment	1 / 48	–	2 / 19	–
Science	0 / 49	–	1 / 20	–
Early years	2 / 47	–	0 / 21	–
Special educational needs and disabilities	1 / 48	–	0 / 21	–

Key: N_c / N_c' – of the 49 trials in the review that used a commercial test as the primary outcome, N_c = number placed in the theme; N_c' = number not placed in the theme.
Med_c / Med_c' – Med_c: The median cost effectiveness for trials with a commercial test that are placed in a theme; Medc': the median cost effectiveness for trials with a commercial test that are not placed in a theme.
N_o / N_o' - of the 21 trials in the review that used official / NPD data as the primary outcome, N_o = number placed in the theme; N_o' = number not placed in the theme.
Med_o / Med_o' – Med_o: The median cost effectiveness for trials with official/NPD outcome(s) that are placed in a theme; Med_o': the median cost effectiveness for trials with official / NPD outcome(s) that are not placed in a theme.

Elaborating context

Commercial and official / NPD outcomes

In terms of publication year, earlier trials were much more likely to use a commercial test as a primary outcome, but the use of an official / NPD outcome is seen to increase over time.

Median attrition rates for evaluations using commercial tests are seen to reduce from 21.3% in 2014 to 14.8% in 2018. However, for evaluations using NPD / official data as primary outcome(s), attrition rates are smaller and fluctuate from a median of 3.8% in 2016 up to 9.1% in 2017 and back down to 8.0% in 2018. This suggests that observed drop in attrition rates is at least in part accounted for by the reduced use of commercial tests in trials along with declining attrition rates for commercial tests that are used.

Table 36: Attrition and context (commercial and official / NPD outcomes)

Type of primary outcome	Commercial test		Official / NPD	
	n =	Median	*n* =	Median
All trials	51	18.3	22	8.0
Publication year		$p > 0.10$		$p > 0.10$
2014	13	21.3	2	–
2015	16	16.1	1	–
2016	8	16.9	6	3.8
2017	3	–	5	9.1
2018	7	14.8	6	8.0
2019	2	–	1	–

Elaborating implementation & fidelity

Commercial and official / NPD outcomes

Across types of developers, attrition rates ranged between 5.4% (charity developers for eight evaluations used an NPD outcome) to 30% (four private company developers that used a commercial test outcome). Across most developers, commercial tests were more likely to be used as the primary outcome compared with an NPD outcome. The only exception was developers from universities where 50% (nine evaluations) used an NPD outcome and 39% (seven evaluations) used a commercial test.

Table 37: Attrition and implementation & fidelity (commercial and official / NPD outcomes)

Type of primary outcome	Commercial test		Official / NPD	
	n =	Median	*n* =	Median
All trials	51	18.3	22	8.0
Type of developer		$p > 0.10$		$p > 0.10$
Not-for-profit / charity	19	18.8	8	5.4
University	7	16.0	9	9.4
Private company	4	30.0	3	8.0
School, academy chain or MAT	7	15.0	1	–
Council / local authority	8	10.2	0	–
Mixed	4	21.2	0	–

Elaborating evaluation design

Commercial and official / NPD outcomes

Efficacy trials were observed to be associated with higher attrition compared with effectiveness trials. However, this seems to relate to the higher use of commercial tests in efficacy trials (31 out of 41 trials, 76%) compared with their use in effectiveness trials (18 out of 41, 44%). The five efficacy and 16 effectiveness trials that used an official / NPD outcome had very similar rates of attrition (8%). The 31 efficacy trials that used a commercial test had a slightly lower average attrition (16%) compared with the 18 effectiveness trials that used a commercial test (18%).

A higher proportion of evaluations undertaken by a non-university used commercial tests (21 evaluations, 72%) compared with university evaluators (29 evaluations, 58%). However, on average, universities had higher attrition for evaluations using either commercial (median of 21.5% compared with 14.8%) or NPD / official outcomes (median of 9.1% compared with 5.5%).

Once type of primary outcome is controlled for, there is scant evidence for an association between testing burden and attrition. The type of primary outcome (commercial or NPD) seems to be the key determinant. On average, evaluations that used an NPD / official primary outcome but also collected data for a single external test had a comparable rate of attrition (median = 9.4%) compared with evaluations with an NPD outcome that had two or more external tests (9.1%). Evaluations that collected no external test data had the lowest observed attrition (3.7%). For evaluations that used a commercial test for the primary outcome, attrition rates for the use of a single test (median = 21.0%) were higher than attrition rates for two or more tests (17.2%).

Table 38: Attrition and evaluation design (commercial and official / NPD outcomes)

Type of primary outcome	Commercial test		Official / NPD	
	n =	Median	*n* =	Median
All trials	51	18.3	22	8.0
Type of trial		$p > 0.10$		$p > 0.10$
Efficacy	31	16.2	5	8.0
Effectiveness	18	18.3	16	8.2
Type of evaluator		$p > 0.10$		$p > 0.10$
Non-university	21	14.8	6	5.5
University	29	21.5	15	9.1
Testing burden		$p > 0.10$		$p > 0.10$
Low (just NPD)	0	–	9	3.7
Medium (1 external test)	15	21.0	6	9.4
High (2+ external tests)	34	17.2	7	9.1

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

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn