EEF Projects Review

Jake Anders, David Godfrey, Rebecca Nelson

Contents

Acknowledgements

Executive Summary

Introduction

The EEF's underlying aim is to ensure that children from all backgrounds can fulfil their potential and make the most of their talents. This review aims to support this goal by systematically considering the outcomes of published trials. It includes both quantitative and qualitative elements. The former sketches out some of the patterns of relationships between context and outcomes within trial designs based on quantitative indicators. The qualitative element of the report is intended to identify learning that will help in the design of interventions and trials that could succeed in schools. Evaluation reports were scrutinised for the impact on pupil attainment but also for ease of implementation and for the enthusiasm with which they were received by leaders, teachers, support staff and by pupils themselves.

Methodology and Scope of the Review

The review includes both quantitative and qualitative elements and sets out to show emerging relationships based on project contexts, mechanisms and outcomes. The quantitative analysis focuses on the patterns of relationships between context and outcomes within trial designs and includes all the trials published to date. The latter brings out further features and details that the quantitative analysis misses, including intervention characteristics, implementation factors and likelihood of sustainability. The qualitative analysis also includes all published trials, however the report draws out some richer illustrative features of a sub-set of projects that had key characteristics in common.

Summary of findings from the quantitative analysis

- The average EEF trial has achieved a security rating of three padlocks. Effectiveness trials have achieved a higher mean security rating than efficacy trials. The most common limiting factor for the overall rating has been attrition, followed by power.

- Trials of projects designed by charities, rather than school groups, universities or local authorities, tend to have lower security ratings, although it seems likely this could be due to the types of interventions they are more likely to propose, which could still be very valuable. Those designed by school groups have had larger average effect sizes.

- The average effect size of interventions that have been evaluated is equivalent to one month's progress. However, few trials have been powered to detect effect sizes of this magnitude.

- Attrition is notably higher in trials that span the transition between these two phases, highlighting the importance of planning for this in future trials of this type.

- Attrition is also higher in trials that are more intensive. This highlights the importance of finding ways of reducing the impact of the higher intensity on burden and, hence, attrition despite the potential need for relatively intensive interventions in order to achieve larger impacts.

Summary of findings from the qualitative analysis

Different characteristics of EEF programmes that have been successful or unsuccessful in raising pupil attainment.

Programmes are more likely to be successful in raising pupil attainment if:
- Senior leaders have a good understanding of the intervention and show commitment and support.
- Timescales of the trial sufficient are sufficient for effective implementation and for the effect on pupil attainment to be demonstrated.

- The timing of the intervention in the school year suits school patterns of activity.
- The target group is accurately identified.
- Delivery of the intervention takes place within relevant, timetabled lessons.
- Materials are of high quality, with essential elements identified and sufficient flexibility to be adapted to school timetables and for appropriate differentiation
- Staff receive high- quality training before the programme and sufficient time for preparation and collaboration during implementation.
- Monitoring and support, whether from external teams or within the school, is effective.

Programmes are less likely to be successful in raising pupil attainment if one or more of these characteristics is not met.

Common characteristics of programmes that have been easy or difficult to implement
- Programmes developed by groups of schools appear to be relatively easy to implement. Recruitment tends to be less problematic than observed in other types of trials and retention of schools is high.
- Where there are problems of implementation these often appear to be linked to a lack of shared understanding among senior leaders and teachers of what is involved. The amount of preparation required for introducing the interventions is a common issue that occurs across all programmes.
- Programmes which require pupils to be withdrawn from subjects across the curriculum are more likely to experience resistance from pupils and these subject staff. Senior leader support is required for ensuring suitable spaces, preparation time for TAs and timetabling arrangements. Programmes for use with a whole class are easier to implement.
- External support from the delivery team facilitates implementation. It is a feature of nearly all efficacy trials and may be more available in effectiveness trials than in a 'real life' situation.

Intensity of the programme and its bearing on implementation and success
- Intensive interventions requiring significant timetable adjustments find it more difficult to recruit and retain schools for the trial.
- Intensive interventions requiring attendance by pupils or parents out- of- school hours experience problems with attendance and attrition, with little evidence for impact on attainment.
- Interventions which require a change in teacher's whole- class pedagogy may benefit from a long term programme.
- The success of one- to- one tuition appears to be related to the quality of tuition, but may also be linked to the length of the programme.

Models of CPD/training associated with more successful programmes
- Efficacy trials that take into account the preparation time required for introducing the new initiative into the school have been more successful in achieving consistent implementation. Lack of time to assimilate training and prepare for implementation may be more significant than the model of training prior to the programme.
- The quality and relevance of programme guidelines, protocols and teaching resources is variable. The amount of expected content is frequently found to be too high for times allotted for programme delivery. Resources provided are rarely ideal for direct use with pupils, and considerable time may be needed to prepare additional material and make adaptations. In some cases, resources did not allow for sufficient differentiation. Some resources, such as pupil worksheets, are not well- matched to current expectations of good teaching practice.
- External or inexperienced tutors need training which enables them to use subject or academic expertise with the pupils in the trial. Those who have not previously worked in schools need training in behaviour management.
- Training that relies on cascading within the school may require additional time and resources.

- The extent to which the contribution of external monitoring and support provided in efficacy trials is essential to the success of the programme might be taken into account when planning further trials.

The success of domain-general vs curriculum-specific programmes
- The measure of pupil academic attainment and the timeframe of trials may not be appropriate for domain- general programmes.
- Domain- general programmes may have to compete for space in the curriculum.

Key differences between programmes delivered by charities, schools and universities
- School group trials experience fewer issues with recruitment and retention of schools and fewer problems with implementation.
- Programmes delivered by charities are more likely than other programmes to involve external tutors working with school pupils and to be general than curriculum specific.

Differences between targeted, whole class and whole school programmes
- The time frames for trials are such that programmes intended to be implemented throughout the school or across a number of year groups are insufficient to determine whether this is a factor in the success of the intervention.
- Targeted programmes which require pupils to be withdrawn from subjects across the curriculum are more likely to experience resistance from pupils and these subject staff. Senior leader support is required for ensuring suitable spaces, preparation time for TAs and timetabling arrangements. Programmes for use with a whole class are easier to implement.

Differences between projects in the early years, primary and secondary phases.
- The success of a programme is more likely to be related to factors other than the phase for delivery. Trials in the transition phase between primary and secondary school are least likely to be successful.

In addition, we have identified a number of cross cutting issues that are relevant to the design and implementation of all trials and that would merit further investigation:
- The importance of securing the commitment and support of senior leaders for the trial and its implementation.
- Recruitment and retention of schools.
- School capacity to introduce an intervention and to fulfil trial requirements.
- Overall timescales for trials to include planning, training, preparation, implementation and testing.
- The need for further evidence on critical factors related to the effectiveness of one- to- one and small- group interventions.
- The potential sustainability of the intervention after the trial.

1. Introduction and purpose

The EEF's underlying aim is to ensure that children from all backgrounds can fulfil their potential and make the most of their talents. It has now published 59 reports on trials it has funded. We must learn as much as possible from these trials, both to inform practice and future research. This report aims to support this goal by systematically considering the outcomes of the trials in terms of substantive foci of interventions but also trial design and context, since these may interact to affect trial effectiveness. In particular, we explore these outcomes in terms of:

- Different characteristics of EEF programmes that have been successful or unsuccessful in raising pupil attainment
- Common characteristics of programmes that have been easy or difficult to implement
- Intensity of the programme and its bearing on implementation and success
- Models of CPD/training associated with more successful programmes
- The success of domain-general vs curriculum-specific programmes
- Key differences between programmes delivered by charities, schools and universities
- Differences between targeted, whole class and whole school programmes
- Differences between projects in the early years, primary and secondary phases

Our approach is informed by that of Pawson and Tilley (1997), in particular the identification of context – mechanism – outcome configurations. In other words, in a particular context (e.g. primary school with high proportion FSM), if particular features of a programme are in place (e.g. intensive CPD, use of professional learning communities), what is the likely outcome (e.g. improvement in pupil motivation, attainment, reported increases in teacher pedagogical skills)? This CMO model allows us to begin to put forward programme theories that make explicit the theory of action behind the most promising interventions. They can also look for negative and unintended outcomes from less successful interventions.

The review includes both quantitative and qualitative elements. The former sketches out some of the patterns of relationships between context and outcomes within trial designs based on quantitative indicators. The latter focuses more on intervention characteristics and implementation factors in a number of key contexts or outcomes and explores the mechanisms that evaluators have identified as relevant to connecting these. A more detailed set of examples of the kinds of context, mechanisms and outcomes that are explored by the two main evaluation approaches is given in Appendix A.

	QUANTITATIVE	QUALITATIVE
CONTEXT	✔	✔
MECHANISM		✔
OUTCOME	✔	✔

We begin by describing the dataset assembled primarily for the quantitative analysis but also used as a sampling frame for the qualitative analysis. This is followed by the quantitative and qualitative analyses themselves, and, finally, our conclusions and recommendations.

2. Data

In order to conduct this review, we read through the EEF Evaluation Reports in order to extract a number of trial quality markers and contextual factors from each. For much of the analysis we excluded those evaluations which were pilots, since these did not tend to have the quality markers necessary for our analyses. In addition, a small number of early trials did not report features that have become consistent quality factors, and a number of unpublished trials (at time of writing) had not yet been awarded security ratings, although these ratings were estimated where possible.

We extracted the following quality characteristics from the trials:

- Security Rating (0 padlocks = weak security; 1 padlock = low security; 2 padlocks = moderate to low security; 3 padlocks = moderate security; 4 padlocks = moderate to high security; 5 padlocks = high security) and sub-ratings:
  – Power Rating
  – Attrition Rating
  – Balance Rating
  – Threats to Validity
- Minimum detectable effect size
- Effect size
- Effect size for FSM
- Rating Limiting Factor (Design, Power, Attrition, Balance, Threats to Validity)
- Statistical Significance of Main Effect
- Attrition at Randomisation Level (%)
- Attrition at Pupil Level (%)

We extract the following trial contextual factors:

- Cost per pupil (average over 3 years)
- Intervention Length (Weeks)
- Intervention Intensity (Minutes per Week)
- EEF Website Topic Tags
- Developer Categorisation (Charity, University, School Group, Local Authority)
- Evaluator Categorisation (University, Non-University)
- Phase (Primary, Secondary or Transition)
- Focus (Maths, Literacy, Numeracy, Science, General)
- Level of Randomisation (Pupil, Class, School)
- In School Training (Teachers, Teaching Assistants, None)
- Targeting (Targeted Pupils, Year Group, Whole School)

These are all reported in the trial database assembled as part of this project. We do not report all relationships in this report, instead exploring particularly interesting narratives that emerge from examining the data.

3. Quantitative analysis

This quantitative analysis largely considers descriptive differences in the quality characteristics of trials. We examine some of these relationships further in the qualitative analysis in this report. It should be stressed that this is very much an exploratory analysis, generating hypotheses that it may be interesting to test further, rather than testing existing ones.

3.1   Security Rating

We begin by considering the types of trials that tend to gain higher security ratings, before delving further into some of the relevant sub-ratings in order to analyse these further. A summary of the security ratings awarded to trials sub-divided by type of trial and phase of education is reported in Table 1.

Among the 49 trials we considered, the most common overall Security Rating awarded (where one was reported) was 3 padlocks, which the official guidance describes as meaning findings are 'reasonably' secure or approximately equivalent to a "well-matched quasi-experiment". 13 trials (representing just over a quarter of the sample) have achieved a rating of 4 padlocks (moderate to high security) or greater. Just two have achieved the top rating of 5 padlocks (high security), both of which were conducted in primary schools.

Some of the lower ratings might partly reflect that a large number of the EEF's early trials were efficacy trials, many of which were only designed to achieve a minimum detectable effect size of approximately 0.3, which would limit their maximum rating to 3 padlocks. However, this was not the case for all efficacy trials and some have been awarded higher security ratings (including one of the two trials to be awarded 5 padlocks). Efficacy trials have tended to be awarded lower Security Ratings, with a mean rating of 2.25 padlocks for efficacy trials and 3.24 for effectiveness trials.

Table 1: Security Ratings by category of trial

Security Rating	Trial	Effectiveness	Efficacy	Primary	Secondary	Transition
0	5	0	5	1	2	2
1	5	2	3	2	3	0
2	6	1	5	3	2	1
3	20	9	11	8	9	3
4	11	8	3	6	4	1
5	2	1	1	2	0	0
Total	49	21	28	22	20	7

Notes: Trials categorised into Effectiveness/Efficacy and Primary/Secondary/Transition

What is the limiting factor for trial quality in each case? We considered this by encoding which of the five sub-ratings was the lowest for each trial (Table 2). Where more than one of the sub-ratings was the same, we followed the hierarchy taken by the Security Rating system, with design being the first factor considered, followed by Power, Attrition, Balance and then Threats to Validity. As noted above, two trials achieve the full five padlocks rating and, so, have no limiting factor.

The most common limiting factor is Attrition (40% of trials) followed by Power, then both Balance and Threats to Validity, with only one trial's rating limited by its Design (this is due to there being only a small number of non-RCT designs funded by the EEF). Attrition (which is calculated at randomisation-level in the calculation of Securing Ratings) appears to have been a particularly common limiting factor in Primary School trials. By contrast, the limiting factor has been more balanced across different factors in trials conducted in Secondary Schools. Perhaps surprisingly, given the different design decisions inherent in efficacy and effectiveness trials, there is no evidence of systematic differences in the limiting factors across these two types of trial.

Table 2: Rating Limiting Factor by category of trial

Rating Limiting Factor	Overall	Effectiveness	Efficacy	Primary	Secondary	Transition
Design	1	0	1	0	0	1
Power	11	4	7	4	3	4
Attrition	17	8	9	10	5	2
Balance	6	1	5	0	6	0
Validity	6	4	2	2	4	0
None	2	1	1	2	0	0
Total	43	18	25	18	18	7

Notes: Trials categorised into Effectiveness/Efficacy and Primary/Secondary/Transition

Table 3 reports differences in the average Security Rating of trials by category of both developer/grantee and evaluator. We stress that there are many factors that could be driving this that aren't captured here, such as some developers proposing/evaluators being more likely to apply for projects that are more challenging to conduct successfully but are, nevertheless, valuable trials.

All that said, the projects developed by Charities are awarded, on average, one fewer padlock than those developed by Local Authorities, School Groups or Universities. This is entirely driven by the presence of 10 trials that have only achieved 0 or 1 padlocks, all of which were of interventions proposed by Charities. The differences between security ratings awarded to different types of evaluators are much smaller (much less than a whole padlock category) with a slightly higher success rate for non-university evaluators.

Table 3: Mean Security Rating by trial category

Category	Mean Security Rating	Frequency
Overall	2.67	49
Charity	2.26	27
Local Authority	3.20	5
School Group	3.14	7
University	3.20	10
Non-University	2.85	20
University	2.55	29

Notes: Trials categorised into Effectiveness/Efficacy and Primary/Secondary/Transition

3.2   Effect sizes

While not necessarily a quality factor in its own right, since it also reflects the underlying effectiveness of the intervention, the size of the effects we see from EEF trials is clearly of some interest; we report this in Table 4. The average effect size of the primary outcomes reported in trials to this point has been 0.09, which is classified by the EEF technical appendices as equivalent to 1 months' progress or a "low" effect. The average effect for individuals who have ever been classified as eligible for Free School Meals is slightly larger, an effect size of 0.10 (2 months' progress) but still described by the EEF's official guidance as a "low" effect. Possible reasons for these small effects are explored in subsequent analyses.

Table 4: Mean effect size

Effect Size	Mean	N
Overall	0.09	49
FSM sub-sample	0.10	40
Overall (trials define FSM sub-sample)	0.08	40

Notes: 9 trials do not report an estimated effect size for an FSM sub-sample.

Are there particular types of trials where we tend to see larger effects? We address this question in Table 5. So far, literacy trials have produced the largest effect size on average, although still classified as 2 months' progress and low effect. Interventions targeted at specific pupils and those in which pupil-level randomisation, rather than class- or school-level, has been used (there is likely to be a significant cross-over between these categories) have seen larger average effect sizes. Finally, thus far, School Groups have developed the trials with the largest average effect sizes (this finding remains even if we drop their most effective trial from the analysis, which seemed rather an outlier and had a low security rating).

Table 5: Mean effect size by trial category

Category	Mean Effect Size	Frequency
Focus: General	0.00	14
Focus: Literacy	0.14	27
Focus: Numeracy	0.10	6
Focus: Science	0.10	2
Targeting: School	-0.01	5
Targeting: Pupils	0.15	24
Targeting: Year	0.05	20
Targeting: Class	0.02	4
Randomisation: None	-0.01	2
Randomisation: Pupil	0.13	24
Randomisation: School	0.07	19
Developer: Charity	0.06	27
Developer: Local Authority	0.04	5
Developer: School Group	0.28	7
Developer: University	0.10	10
Overall	0.09	49

Notes: Trials categorised into General/Literacy/Numeracy/Science focus, School/Targeted/Year/Class-level intervention, None/Pupil/School-level randomisation, and Charity/Local Authority/School Group/University developer.

In which trials do we see larger effects for disadvantaged pupils than those for pupils in general? Unfortunately, Table 6 demonstrates there is little obvious systematic pattern evident from the data predicting trials in which this is the case; there is a fairly average distribution across subject focus, extent of targeting, and which (if any) type of training of in-school staff is used as part of the trial.

Table 6: Whether effect size is larger for FSM sub-sample by category

Focus	FSM ES ≤ Overall ES	FSM ES > Overall ES	Total
General	7	6	13
Literacy	10	11	21
Numeracy	1	3	4
Science	1	1	2
Total	19	21	40
Targeting	**FSM ES ≤ Overall ES**	**FSM ES > Overall ES**	**Total**
School	2	3	5
Targeted	9	11	20
Year	8	7	15
Total	19	21	40
Level of Randomisation	**FSM ES ≤ Overall ES**	**FSM ES > Overall ES**	**Total**
Class	0	4	4
None	1	1	2
Pupil	9	9	18
School	9	7	16
Total	19	21	40
In-School Training	**FSM ES ≤ Overall ES**	**FSM ES > Overall ES**	**Total**
None	8	10	18
Teachers	9	9	18
Teaching Assistants	2	2	4
Total	19	21	40

Notes: Trials categorised into General/Literacy/Numeracy/Science focus, School/Targeted/Year/Class-level intervention, Pupil/School/None-level randomisation, Charity/Local Authority/School Group/University developer, and Teachers/Teaching Assistants/No In-School Training.

3.3   Power

As noted in the previous section, the average EEF trial has found an effect size equivalent to one month of progress (using the EEF's official conversion system). We report the number of trials reporting each level of months' progress in Table 7. Over 70% of trials have found an effect size of no greater than two months of progress in either direction. This is, at most, an effect size of 0.18.

However, according to Table 8, over half of trials have a power rating of 4 padlocks or less, suggesting they are powered to detect a minimum effect size of greater than 0.20. As such, they are only powered to detect the type of effect size seen in under 30% of EEF trials. We report the proportion of all trials that find a statistically significant result in Table 9. Even among trials with a Power Rating of 5 padlocks (Minimum Detectable Effect Size (MDES)<0.2), which are included in Table 10, only 3 (15%) have found a statistically significant effect. In fact, Table 11 demonstrates that only trials with 4 or 5 padlocks have ever detected a statistically significant result as conventionally understood. Of course, it could be that none of these interventions had any effect but it seems suggestive that there is insufficient power to detect the type of effect sizes that are plausible within EEF trials.

Table 7: Months' progress

Months' Progress	Frequency	%
-2	2	4.1
-1	10	20.4
0	9	18.4
1	9	18.4
2	5	10.2
3	9	18.4
4	2	4.1
5	2	4.1
9	1	2.0
Total	49	100

Table 8: Power Rating

Power Rating	Frequency	%
2	2	4.3
3	8	17.0
4	16	34.0
5	21	44.7
Total	47	100

Table 9: Main Effect Statistical Significance

Main Effect Statistically Significant?	Frequency	%
No	33	79
Yes	9	21
Total	42	100

Table 10: Main Effect Statistical Significance in Trials with Power Rating of 5

Main Effect Statistically Significant?	Frequency	%
No	17	85
Yes	3	15
Total	20	100

To explore this further, we consider whether effect sizes for trials are larger or smaller than the minimum detectable effect size that is achieved. The two characteristics are plotted in a scatter plot in Figure 1 with a diagonal line highlighting the point at which the two are equal. Points above/to the left of the line have effect size greater than minimum detectable effect size; points below/to the right of the line have effect size less than minimum detectable effect size. Only four trials are above and to the left of this line, indicating that the effect size estimate is greater than the minimum detectable effect size. This is summarised in Table 11, in which it is also cross-tabulated with the power rating of the trial, while Table 12 cross-tabulates it with whether a statistically significant effect is found.

Figure 1: Scatter plot of Effect Size and MDES
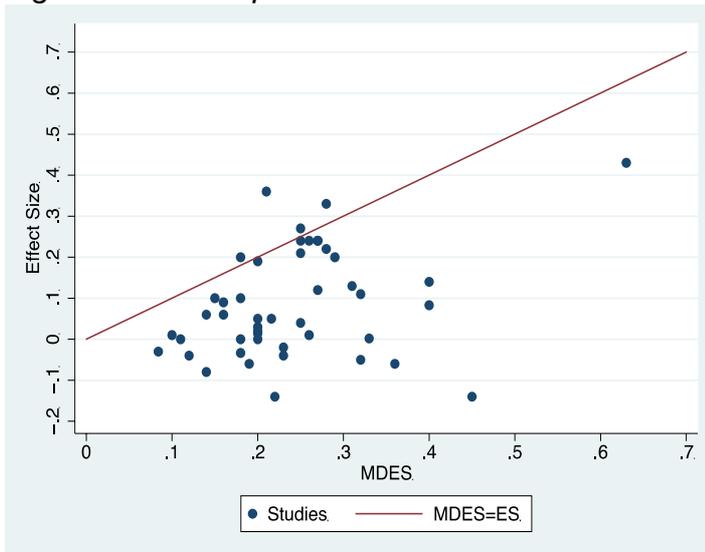
Notes: Red line shows points at which minimum detectable effect size is equal to trial's estimated effect size (NB. not a line of best fit). Points above/to the left of the line have effect size greater than minimum detectable effect size; points below/to the right of the line have effect size less than minimum detectable effect size.

Table 11: Cross-tabulation of whether Effect Size is greater than Minimum Detectable Effect Size (MDES) with Power Rating given to trial

Effect Size > MDES	Power Rating				
	2	3	4	5	Total
No	2	8	13	18	41
Yes	0	0	3	1	4
Total	2	8	16	19	45

Table 12: Cross-tabulation of whether Effect Size is greater than Minimum Detectable Effect Size (MDES) with Main Effect Statistical Significance

Effect Size > MDES	Main Effect Statistically Significant?		
	No	Yes	Total
No	33	4	37
Yes	0	4	4
Total	33	8	41

3.4   Attrition

What limits the overall security rating that projects receive? The most common limiting factor is attrition (see Tables in Security Rating section above), followed by power. Together these make up two thirds of the limiting factor on the rating awarded. Furthermore, since achieved power is also a function of attrition it may well be an important contributor to limiting the security rating even in cases that we attribute to insufficient power. Attrition is clearly a concern in trials, since it reduces the confidence we can have in internal validity, especially if balance on observables is undermined as a result.

Attrition seems to vary significantly by some of the contextual characteristics classified. Table 13 reports that attrition at pupil level has been lowest in primary school trials but lowest at the randomisation level in secondary school trials. (This difference might be related to the differing

types of trial that have been more common in these two settings or because of the larger year group sizes in secondary schools.) Either way, it is clear that attrition is highest among projects that span the transition between the two; this is perhaps unsurprising given that pupils move schools during the period making them more difficult to track.

Table 13: Attrition at Randomisation-Level and Pupil-Level by Category of Trial

Category	Randomisation Level	Pupil Level
Primary	16.63	18.27
Secondary	13.89	20.09
Transition	25.00	27.14
Overall	16.47	20.39

In particular, we might be concerned that attrition is related to the length or intensity of the intervention, if these factors contribute to burden on schools that encourage them to drop out. The remainder of the modelling concentrates on attrition at the pupil level in order to allow comparability across all trials. Figure 2 considers the correlation between these two indicators and finds that there is no indication that longer interventions are associated with higher attrition.

Figure 2: Scatter plot of Intervention Length and Attrition at Pupil-Level
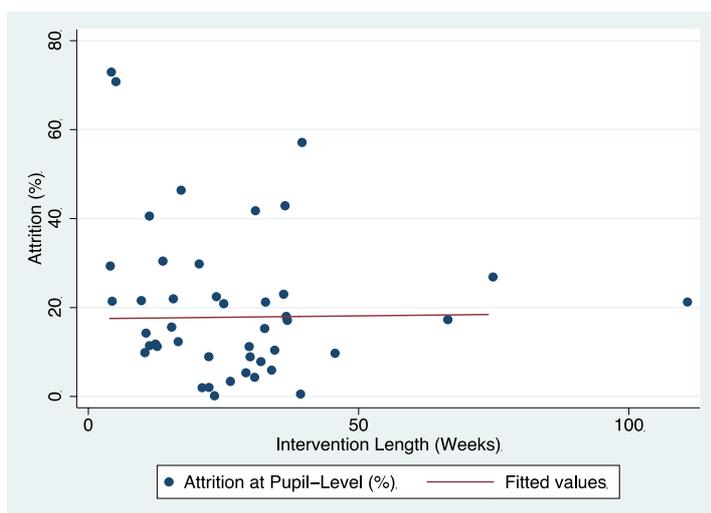
Notes: Red line is line of best fit between intervention length and attrition (excludes a small number of outliers).

By contrast, there is more of an indication that interventions with greater minutes per week are associated with higher attrition (shown using the line of best fit on the scatter plot in Figure 3) although we should note that information on intensity is not present in a significant number of reports either because the intervention is more diffuse or it was simply not reported. Nevertheless, this suggests it is important to consider the burden that high intensity interventions may be placing on schools and ways to mitigate this to ensure that evaluations are not undermined by the threat to interval validity that this may pose.

These factors remain predictive when we take into account the potential interplay between length and intensity by including both in a single regression model (not reported), although perhaps weakening slightly for intensity and strengthening slightly for length. These directions of these associations seem robust to the inclusion or exclusion of outliers.

There is not a very strong association between intervention intensity and effect size, shown using the line of best fit on the scatter plot in Figure 4, although it exists to some extent. This places an importance on finding ways of reducing the impact of the higher intensity on burden and, hence,

attrition despite the potential need for relatively intensive interventions in order to achieve larger impacts.

Figure 3: Scatter plot of Intervention Intensity and Attrition at Pupil-Level
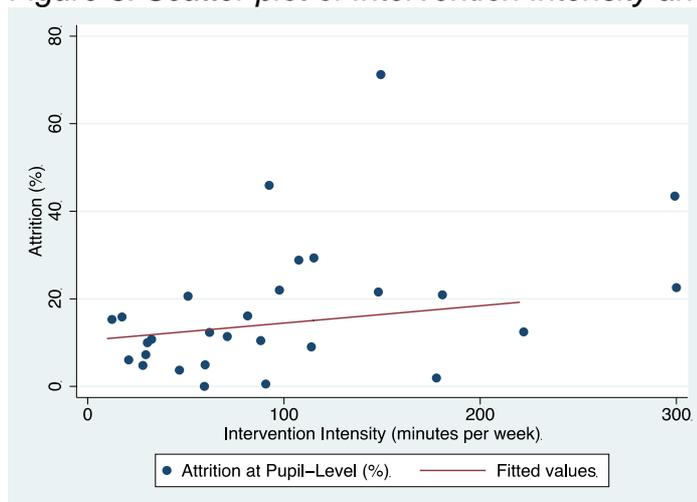
Notes: Red line is line of best fit between intervention intensity and attrition (excludes a small number of outliers).

*Figure 4: Scatter Plot of Intervention Intensity and Effect Size*
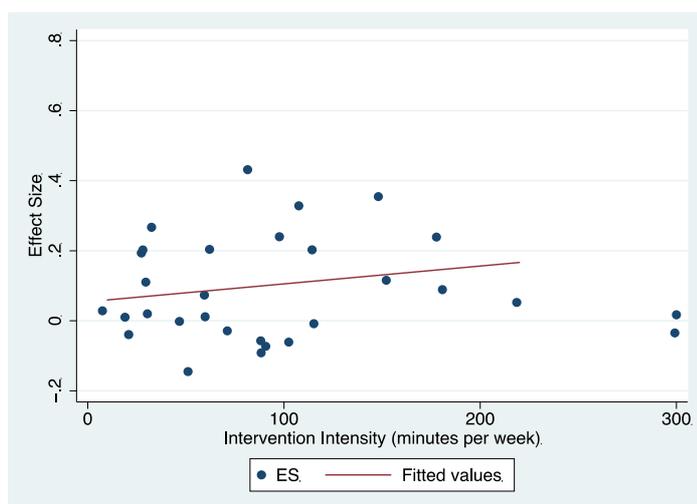
Notes: Red line is line of best fit between intervention intensity and effect size (excludes a small number of outliers).

3.5   Cost per pupil

We explore the relationship between cost per pupil and effect size using a scatter plot (Figure 5). Most points are red, while those in which we have moderate security and a statistically significant effect are reported in blue. While it is not a strong pattern, the general trend is that interventions that cost more per pupil do tend to have larger effect sizes, although there is a lot of variation within this pattern, suggesting that, unsurprisingly, other factors which are not easily quantified (such as the actual content of the intervention) are much more important than just spending more.

Figure 5: Scatter Plot of Cost per pupil and Effect Size
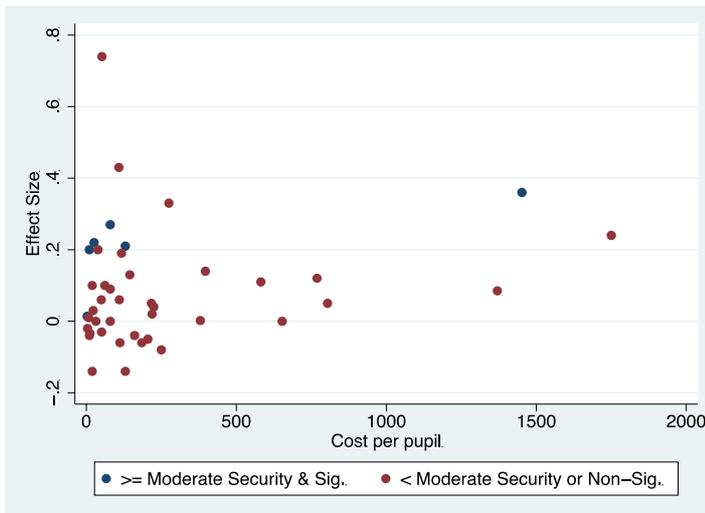
Notes: Blue points are of trials with greater than or equal to moderate security (3 padlocks or more) and with statistically significant estimated effect sizes. Red points are trials that do not meet these two criteria.

Table 14: Cost per Pupil by Category of Trial

Category	Mean	Frequency
General	182.23	13
Literacy	345.85	27
Numeracy	106.67	6
Science	15.00	2
School	160.80	5
Targeted	421.11	23
Year	94.38	20
Class	11.17	3
None	201.00	2
Pupil	426.90	24
School	89.26	19
None	456.83	20
Teachers	50.98	21
Teaching Assistants	310.00	7
Overall	257.85	48

Notes: Trials categorised into General/Literacy/Numeracy/Science focus, School/Targeted/Year/Class-level intervention, Pupil/School/None-level randomisation, and Teachers/Teaching Assistants/No In-School Training.

Cost per pupil by category of trial is reported in Table 14. Excluding those trials that did not report comparable cost per pupil data, thus far the EEF has published 27 trials focussed on literacy, 13 on academic attainment generally, 6 on numeracy and 2 on science. The most expensive (in terms of the estimated cost of the intervention, not the trials themselves) of the EEF's projects have been those focused on literacy, costing an average of £345 per pupil, almost twice as much per pupil as the general purpose interventions that have been trialled. The six published numeracy interventions have cost around £100 per pupil, while the two science trials have cost £15 per pupil, although how much can be learnt from either of these given the particularly small number of trials of this type is a moot point.

It is likely that some of these differences also reflect other differences in the trials. For example, there are a cluster of literacy catch up trials that were focused on primary to secondary transition. These included some very expensive out-of-school trials, driving up the cost per pupil of literacy trials in the analysis sample.

3.6   Summary of quantitative analysis

The average EEF trial has achieved a security rating of three padlocks with the most common limiting factor for this being attrition, especially among trials conducted in primary schools. Trials of projects designed by charities, rather than school groups, universities or local authorities, tend to have lower security ratings, although it seems likely this could be due to the types of interventions they are more likely to propose, which could still be very valuable.

The average effect size of interventions that have been evaluated is equivalent to one month's progress, although it is marginally higher for pupils who have ever been eligible for free school meals. Unfortunately, very few trials have been powered to detect effect sizes of this magnitude. Attrition is notably higher in trials that span the transition between these two phases, highlighting the importance of planning for this in future trials of this type. Attrition is also higher in trials that are more intensive (more minutes per week of delivery), which, combined with the potential need for intensive interventions to achieve larger impacts, demonstrates the importance of a focus on reducing the burden of the evaluation itself to a minimum.

4. Qualitative analysis

4.1 Introduction

The qualitative element of the report is intended to identify learning that will help in the design of interventions and trials that could succeed in schools. Evaluation reports were scrutinised for the impact on pupil attainment but also for ease of implementation and for the enthusiasm with which they were received by leaders, teachers, support staff and by pupils themselves. A further theme that arose in the course of analysis was that of the likelihood that an intervention might be sustained within schools after the end of the trial period. Points that arose from the analysis are presented in this section of our report. Further descriptive detail is provided in Appendix B.

Approach to qualitative analysis

A loose framework of context- mechanism - outcome was used to support the analysis which follows. Trial outcomes may depend on decisions about mechanisms for implementation taken at design stage. These trial mechanisms provide the context in which schools implement the intervention in their schools. Outcomes for schools and for pupils depend both on the mechanism selected for the trial and on the way in which the design is implemented in practice.

A first step in the qualitative review was to narrow the range of projects considered in order to enable deeper insight. We selected some shared contextual factors in a number of projects which were targeted at pupils who do not achieve secure Level 4 in English by the end of Key Stage 2 and which are aimed at helping them catch up in literacy skills. We chose programmes implemented at the end of primary schooling or the beginning of secondary schooling and those which combined the two (transition programmes). Within this group we chose a sample that included programmes designed for whole- class teaching, but where only those in the targeted group were tested for programme effect (whole class interventions). We also included interventions in which the targeted pupils were taught separately from their peers, in small groups or one- to- one (targeted interventions). In most cases, targeted interventions were delivered by teaching assistants (TAs) employed by the school or by other adults without teacher qualifications, as in the *Perry Beeches Graduate Coaching Programme*[1], in which recent graduates were recruited by the school for the intervention but were not otherwise included in the staffing structure of the school.

Points arose in the course of detailed examination of a small number of programmes which were then examined further in relation to other programme factors and to the evaluation reports of a wider range of efficacy and effectiveness trials. As well as whole- class or targeted interventions as described above, this group included both whole- class or targeted interventions taught by volunteers or external tutors and a number of programmes aimed at changing a teaching approach across the whole school, such as *Mathematics Mastery* (whole school interventions). Pilot trials and those trials which were so distinctive that they do not lend themselves to comparison (for example, *Increasing Pupil Motivation, Hampshire Hundreds)* are not included in this analysis.

Trials selected for initial examination were all efficacy trials involving a relatively small number of schools. Initially we considered school- group trials involving three or four schools and then examined a further selection of efficacy trials with between 20 and 70 participating schools. We considered those trials which had a positive effect on attainment for the pupils who completed the intervention and were positively viewed by participating schools, staff and pupils. However, the overall security ratings were modest. In many cases, schools were recruited locally by the developer based on pre- existing relationships and, as such, may have been assumed to represent the 'ideal' conditions expected for an efficacy trial. Programmes tested for efficacy trials are selected on the basis of prior evidence that they can be effective in raising attainment, either

---

- [1] For ease of reading, evaluation reports are referred to by name in the body of the report, with full references included in the list at the end of the document.

through pilot programmes run by EEF or other agencies or on the basis of other research evidence. A key focus in the quantitative section of our report has been those factors in trial design which help ensure that the findings on which school inform their decision are sound. In order to determine if our initial findings applied to programmes where trial findings have been more secure, we considered some high- security efficacy and effectiveness trials in more detail. Points arising from this analysis are summarised in section 4.2.

In section 4.3 we present findings from review of a wider range of efficacy and effectiveness trials. These were selected to provide further insight into themes identified by EEF and to build on findings from quantitative analysis:
- Common characteristics of programmes that have been easy or difficult to implement
- Intensity of the programme and its bearing on implementation and success
- Models of professional development and training associated with more successful programmes
- The success of domain- general versus curriculum- specific programmes
- Key differences between programmes delivered by charities, schools and universities
- Differences between targeted, whole- class and whole- school programmes
- Differences between projects in the early years, primary and secondary phases.

 Section 4.4 provides discussion of an additional theme identified from our review, of the sustainability of a trial intervention. Further descriptive detail to support the analysis that follows is provided in Appendix B.

4.2   Findings from efficacy trials and trials with high security ratings

Small, school- group, efficacy trials

As a sample of school group trials we selected the following: *Accelerated Reader*, *Talk for Literacy* and *Perry Beeches Graduate Coaching Programme.* All of these interventions were among a cluster of projects aimed at improving the literacy level of pupils whose level of attainment is likely to restrict their access to the secondary school curriculum. All three efficacy trials took place for targeted Y7 (year 7) pupils in a small number of secondary schools (3 or 4) and used pupil randomisation within each school to form intervention and control groups. Although overall security ratings were moderate, the effect on pupil attainment was positive. What differs in the trials in the consistency of trial application across different schools, including significant pupil attrition in the *Perry Beeches Graduate Coaching Programme.*

Accelerated Reader was one of a small number of EEF aggregated efficacy trials in which schools had independently applied for funding and were grouped together to conduct and evaluate the project themselves with the support of the external evaluator. It consists of a computer- based management and monitoring programme that aims to foster the habit of independent reading. Pupils are taught, by teachers or teaching assistants (TAs), to use the programme to select books matched to their reading level, which they then read independently. Pupils are also shown how to use the assessment tools within the programme to track their reading progress. Despite some barriers in timetabling and with IT resources, the process evaluation for the project noted that fidelity to the programme was high, with some adjustments to implementation guidelines to help very low attaining pupils access suitable texts.

The *Talk for Literacy* trial was also conducted in a small number of schools, but in this case evaluation was by the external evaluator. Trained TAs worked with small groups of pupils, using specially- designed vocabulary enrichment and narrative materials. The trial required schools to identify TAs to be trained and to deliver the intervention and to find suitable spaces for small- group work during twice- weekly sessions over two terms. One of the barriers reported to implementation was the lack of suitable space for group work, suggesting that despite the small number of schools involved, detailed implications were not fully understood. TAs were enthused by pre- programme training provided by the developer. However, they found it extremely difficult to fit in the expected

number of activities into the small- group sessions and very time consuming to prepare for sessions, resulting in variability in what they chose to include.

Perry Beeches Graduate Coaching Programme was initially piloted at the developer school, who then recruited three other schools with which it had existing links, with all four part of the efficacy trial. It consisted of flexible and individualised one- to- one literacy tuition from school- based coaches over the course of a school year. Coaches were recent university graduates without teaching qualifications. There was considerable inconsistency in all aspects of the trial. Schools made different decisions regarding qualifications and experience of coaches recruited, and there were differing levels of attention to the need for training and support, suitable spaces, provision of resources, timetabling, supervision arrangements and whole school engagement. Many of the coaches reported that they did not feel equipped for their role or have the strategies to meet the diverse learning needs of their tutees. Coaches were insufficiently prepared to manage pupil behaviour effectively. Pupil attrition was high, largely because one school did not correctly test all participating pupils.

Outcomes
As might be expected in efficacy trials, results for all pupils and those eligible for free school meals in these trials are limited by small sample sizes. The evaluator for *Accelerated Reader* found that "there is strong evidence overall that AR improves reading for new Year 7 pupils with weak reading skills and habits, and that it does so especially for those eligible for FSM "(p19). Staff, pupils and their parents responded positively to the programme and TAs reported gaining new skills for supporting reading. With regard to *Talk for Literacy*, TAs reported positive impact on pupils' verbal skills and confidence and, although there was no significant impact on overall reading ability, there was statistically significant evidence of improvement in the secondary measure of passage comprehension. The *Perry Beeches Graduate Coaching Programme* evaluation report notes that "The programme had a positive impact on pupils' attainment in reading, spelling and grammar" (p4). The relationship between coaches and pupils appeared to be important to the views of pupils.

Efficacy trials in larger numbers of schools

The following efficacy trials, *Switch on Reading; Improving Writing Quality* and *Response to Intervention*, were also targeted on improving literacy in low- attaining pupils at the beginning of secondary education and involved larger numbers of schools. They have been selected to illustrate further the issues which appear to be related to a school's capacity to introduce and support a new initiative and of the challenges associated with ensuring full understanding and commitment in relation to the demands of the trial. The section also highlights issues of implementation associated with the mechanism through which the programme is delivered and the implications for outcomes.

Secondary schools in the *Switch On Reading* trial were recruited from within the local authority in Nottinghamshire. Pupils were randomised within each school to a control and an intervention group. The intervention is based on the *Reading Recovery* programme (Clay, 1985) for younger pupils and consists of daily, one- to- one support sessions for weak Y7 readers with teaching assistants who have been trained with the materials and processes particular to the programme. Despite a recruitment strategy which built on a local network and relationships, far fewer pupils were identified as eligible for the trial than expected. Pupils selected for the intervention were usually withdrawn from normal lessons for the daily 20 minute reading sessions. Timetabling difficulties within the timeframe for the trial meant that few schools managed to run the programme for the intended 10 weeks. Attendance at sessions was irregular in some cases, with pupils forgetting to excuse themselves from class or teachers unwilling to release pupils. Variation both within and between schools was noted during implementation. However, overall, the judgement of the evaluators was that fidelity to the programme was high, largely because of the extensive monitoring and support provided by the local authority team. The evaluators suggest that differences in quality and adherence to the programme protocol are related to factors such as experience, confidence and enthusiasm of the TAs. Some sessions were affected by resistance and/or poor behaviour of pupils. There were concerns in some schools about the impact of children being removed from other lessons important for their overall progress and about the project methodology, which was

thought by some teachers to be too rigid. There was some confusion with other literacy interventions already in place in schools and doubts about the accuracy of the pre- test. The evaluation report comments that whole school commitment and motivation are important for success.

Improving Writing Quality, was designed for use at the end of Y6 and the beginning of Y7, and includes a structured, whole- class, approach to helping pupils plan, monitor and evaluate their writing. Recruitment was undertaken within the Calderdale Education Partnership and, of 24 primary schools approached, 23 agreed to participate, as did all 3 secondary schools approached. No schools dropped out of the trial. Allocation to intervention or control group was at school level for primary schools, with secondary schools arranging timetables so that control and intervention pupils could be kept distinct. This appears to have been successful with no evidence of contamination. Because a large proportion (34%) of pupils at participating primary schools did not transfer to the participating secondary schools, eligible pupils from this group were lost to the trial, reducing the sample size. *Improving Writing Quality* offers a structured but flexible approach for teaching writing developed in the USA. It is designed for use for all pupils in whole- class teaching. In the trial, schools received resources and support in the programme and funding for 'Memorable Experiences', such as outings or visitors to the school, to provide stimulus and focus for writing activities. Primary school teachers introduced the approach in literacy lessons in the summer term of Year 6, which was then used by secondary school teachers in English lessons which included target pupils during the first term of Y7. Thus, the intervention was experienced by all pupils in these teaching groups, although only pupils with attainment below a secure level 4 at key stage 2 were included in post- tests. Following training by the developer, a working group of participating teachers was established and produced guidelines for schools using the approach in English schools. The evaluation reported that teachers would meet again after the project to develop a toolkit that could be used by other schools.

Response to Intervention offers an individualised, tiered approach to raising achievement in reading. It uses whole- class teaching, plus targeted small- group remedial teaching and one- to- one tuition where necessary, depending on close needs analysis by the teacher. The trial sought to recruit primary schools and ask them to identify potentially eligible Y6 pupils through teacher assessment and use of the trial pre- test. The intervention was timed for implementation in the second half of the summer term. Schools approached for the trial were those already working with the developer, but recruitment fell short of the 80 schools expected. Several schools dropped out before the trial began when they realised the timing, the expectations of school staff and the IT requirements of the testing approach, with only 61 beginning the trial and a further 11 schools dropping out after the pre- test and randomisation. Errors of communication and misunderstanding of trial requirements prevented data on intervention and control groups from reaching the evaluator in a timely fashion and several schools failed to report full data. The short time scale required by the funders was considered by evaluators to be unsuited to the nature of the intervention. Issues affecting the trial led to the evaluator considering it 'spoilt'.

The *Response to Intervention* trial was introduced to schools through training for school 'champions' and 'achievement coaches' who subsequently shared information with other relevant staff in the school. Use of the intervention is complex, requiring use of a formative assessment tool to determine the literacy areas to focus on for each targeted pupil and the pattern of needs across the whole class. The programme is flexible in length, intensity and frequency. Following the formative assessment, a teacher must then choose from a menu of options to decide which intervention to use to tackle identified weaknesses and whether this is best achieved at whole class, small- group or one- to- one level. Other staff working with small groups or one- to- one must be supported and progress monitored using programme tracking sheets, with adjustments made if required. Unsurprisingly, the evaluator notes that 'schools did not feel that they had enough time to carry out the intervention as thoroughly as they would have liked to. "There were some reports that the intervention was not given sufficient time to develop and to complete the full cycle from needs analysis to monitoring and adjustments" (pp 19-20). There were also concerns about the accuracy of the assessment tool. Administration and the amount of preparation required (as for the introduction of any new programme) were additional challenges.

Outcomes

As with other efficacy trials, exacerbated by lower sample sizes than anticipated, the overall security rating for *Switch on Reading* was modest. Despite reservations, many staff were enthusiastic about the project and reported positive effects on some pupils' reading and confidence. TAs valued the opportunity to improve their own skills. The one- to- one attention provided to pupils was considered to be very important in helping pupils make progress. The trial did not test whether a similar programme of one- to- one support other than that based on *Switch on Reading* was equally or more effective.

The enthusiasm of teachers in both primary and secondary schools for continuing to use the *Improving Writing Quality* approach, to use it over a longer period of time and with other year groups is a significant outcome for this intervention in addition to the promising quantitative outcomes. Although Y6 teachers would have welcomed a longer timeframe for the trial, *Improving Writing Quality* is unusual in not reporting significant dosage limitations which typically affect trials in the summer term of Y6. The guidelines for using the approach allowed for sufficient flexibility within an overall structure and ordering of activities which enabled both primary and secondary teachers to include SRSD within their planning.

It is striking that despite the numerous problems with implementation, teachers, pupils and achievement coaches for *Response to Intervention* were positive about the intervention. Pupils responded positively to individualised attention and to activities used in interventions. The development of teaching resources and innovative approaches new to the school was a positive outcome for schools involved. The diagnostic assessment tool was valued by teachers and school-level pupil data showed positive impacts on pupil progress in reading. Many teachers said that they would continue to use the approach with other classes in the future.

Key Points

- When recruiting schools, full support and commitment to the trial is likely to be needed from senior leaders for successful implementation.
- Efficacy trials based in a small number of schools avoid problems of recruitment and may be more likely to retain schools and pupils for the course of the trial.
- Process evaluations suggest that schools may not be fully prepared for the implications of the trial for timetabling, for rooming, for the preparation time required by delivery teams, for monitoring by senior leaders and for ensuring the support of all school staff, including those who are not directly involved.
- When recruiting schools to efficacy trials, the capacity of the school to introduce a new initiative might be considered. Evaluation reports do not note any links between capacity indicators and recruitment strategy, attrition or issues with implementation and capacity, although these might be relevant. For example, schools in Ofsted categories of 'requires improvement' or 'inadequate' are likely to face pressures which might affect leadership support and monitoring.
- Projects which are scheduled for the summer term of Y6 or those intended to begin in primary school and complete in the first term or year of secondary school face capacity barriers which may hinder implementation. Pupils identified as eligible in participating primary schools and who begin in both intervention and control groups transfer to many different secondary schools, which may or may not be part of the trial. The school summer term is particularly busy, with preparation and taking of SATs tests, staff absence for moderation and other training, transition events with secondary schools and school trips and performances. This results in frequent disruption to the normal timetable.
- Any IT requirements of programme delivery or testing might usefully be taken into account at the recruitment stage of a trial.
- Time frames required by the funder for recruitment and implementation of the trial may restrict sample sizes, be insufficient to allow schools to prepare and for impact to be demonstrated. Where a trial requires the acquisition and use of new teaching skills, additional time is likely to be required for staff to prepare lessons and practise these skills to ensure high quality.

- Intervention programmes devised and tested in one context are not always sufficiently adapted prior to use in a different context for EEF trial. Working groups of teachers, such as that formed for the *Improving Writing Quality* intervention, might be considered both in advance of efficacy trials and to make further modifications following experience of use in an English context.
- Highly- structured, prescriptive activities and resources linked to intervention approaches may be difficult to accommodate within the diverse timetabling, staffing arrangements and individual pupil needs and interests typical among schools even in small- scale trials. Expectations of the amount of content to be delivered in intervention sessions were sometimes too high. Both efficacy and effectiveness trials might usefully consider highlighting aspects and activities of programmes that are essential for fidelity and those where there might be adjustments for school and pupil circumstances.
- Programmes which required external and/or inexperienced tutors to work pupils appear to have faced particular challenges in ensuring consistently high quality and/or the engagement and good behaviour of pupils.

High security trials

In this section a sample of efficacy and effectiveness trials with security ratings of 4 or 5 padlocks was examined. Further evidence was found to support the key points made above, with examples from two of these trials provided below. Further examples, and descriptive detail, are included in Appendix B.

Improving Numeracy and Literacy trains teachers in key stage 1 in strategies targeted at specific areas of improvement in either Literacy or Numeracy, with the focus for the intervention not known until after recruited schools had been randomised. 270 schools were initially approached by the University of Oxford delivery team but recruitment of 55 schools fell short of the target of 60. The evaluation report states: "The main reason given by schools for not participating was that they wanted to choose the intervention that best fitted with their current school development plans, rather than be randomised" (p14). The intervention is implemented during whole- class teaching in place of normal, timetabled Literacy or Numeracy lessons using guidance materials and other resources provided by the developers. Issues regarding the amount of content and the prescriptive nature of *Improving Numeracy and Literacy* resulted in high variability in the ways in which different schools and teachers responded to these. In particular, the independent pupil use of computer games, which was an intrinsic part of the programme, was inconsistent among schools as well as among pupils. Resources provided contained numerous errors, but the main issue with the materials was the amount of preparation required.

Nuffield Early Language Intervention, for the end of nursery and the beginning of the Reception year, is a transition programme which requires withdrawal of low- attaining pupils for one- to- one tuition by trained teaching assistants using materials provided. 302 primary schools with nursery provision were approached by the charity I CAN with 34 recruited. Capacity issues given as reasons for declining to participate including, "finding the 'right person' in the school, schools being unable to commit to the amount of staff time that would be needed without funding.." (p12) as well as reasons related to schools' current priorities. Further evidence of the need for senior leadership support arose during implementation and this was reported as variable. The support of senior leaders was important for ensuring sufficient protected release time for teaching assistants delivering the intervention to prepare, as well as to deliver the intervention, and to arrange for suitable spaces to be available when required. Inappropriate targeting for one- to- one support was noted and some of the material for nursery children was felt to be unsuitable for their levels of development. However the main issue with the materials was the amount of preparation required.

Outcomes
The improvement gains for the three way intervention *Improving Numeracy and Literacy* relate to the numeracy intervention, with no effect found for literacy. Teachers anticipated adapting and adopting at least some aspects of the programme in future. Gains in pupil confidence and verbal skills were noted in the targeted intervention *Nuffield Early Language Intervention*, although

decisions about future use were likely to be mediated by factors such as comparison with other intervention approaches, pupil needs and consideration of the high costs of the programmes.

4.3   Other trials and factors

In this section**,** the points identified above will be considered for relevance to a number of other factors and a wider range of efficacy and effectiveness trials.

Common characteristics of programmes that have been easy or difficult to implement. Differences between targeted, whole- class and whole- school programmes

We found that the mechanism used for the programme, as a targeted or whole- class intervention, was closely associated with ease of implementation. The time frames for trials are such that programmes which are eventually intended to be implemented throughout the school or across a number of year groups (such as *Mathematics Mastery, Promoting Alternative Thinking Strategies*, *Let's Think Secondary Science*) precludes direct comparison of whole- class or targeted interventions with whole school programmes.

School group trials, in which a group of schools acted as the developer, experienced no significant implementation issues. This appeared to be the case for whole- class interventions and for those targeted at small groups or individual pupils.

Process implementation reports often mention 'senior leader support' as significant, although details provided are more usually related to experiences in implementation rather than the causes for lack of senior leader support. Although some reports provide details of recruitment strategies where schools attend information meetings prior to sign- up (as in *Promoting Alternative Thinking Strategies*), or are asked to contribute financially (as in *Thinking, Doing, Talking Science*) the effectiveness of these can only be assessed in terms of number of schools recruited rather than in the extent of commitment and understanding of senior leaders to what is expected.

For example, in the *Catch Up Literacy* report, as well as the teaching assistants responsible for delivering the one- one- sessions, secondary schools were asked to identify coordinators, although their level of seniority is not clear. The positive impact of coordinator support is indicated by two interviewees saying that they had made sure that suitable rooms were available for the intervention (the lack of these was considered a barrier to implementation overall). However a coordinator also indicated that they did not have "buy-in from the senior management team, as they preferred interventions that could be run as a group rather than a one-to-one intervention" (p 25) which raises the question of whether the nature and requirements of the intervention had been fully understood by the leaders of the school when they agreed to participate.

In *Promoting Alternative Thinking Strategies (PATHs)* headteachers were asked to sign a Memorandum of Understanding with the evaluation team before participating in the trial. Senior leader support was noted as a "critical factor" (p 25) in implementation and is described as facilitating implementation by active monitoring and by disseminating the approach so that it is used consistently throughout the school. An example is also provided in this report of a headteacher proving a barrier to implementation by not allowing the flexibility required in the timetable.

However, in other types of trials, programmes which require pupils to be withdrawn from their normal lessons for small- group or individual tuition were more problematic in terms of practical issues arising from implementation. Suitable spaces may not be made available and, particularly when small- group or one- to- one sessions are led by TAs, senior leaders may be unwilling to provide sufficient release time for preparation. In the secondary phase, where many teachers are involved and where pupils are often withdrawn from lessons across the curriculum, pupils may feel stigmatised by being taken away from their peers, resent missing all or part of lessons they enjoy and miss core learning in other subjects. Attempts by the school to timetable intensive interventions and to spread withdrawal sessions across many subjects results in pupils forgetting to attend and/or

needing to be collected for the intervention session. Subject staff may resist pupils being extracted from their lessons. In some cases, schools felt that assessment for suitability for a targeted intervention did not select those who might benefit most.

The availability of external support is a factor which appears to aid implementation, as, for example, in the *Catch Up Literacy* intervention. The *Mathematics Mastery* programme includes an ongoing programme of professional development led by the developer and membership of a support network. Many programmes report support from the developer team during the implementation of the trial, through on- line and telephone support and in school visits to monitor implementation. For example, in the *Vocabulary Enrichment Intervention*, trial schools had supervision and support from the developer team and from an educational psychologist. The cost of this support is not included in per- pupil costings for the programmes.

In the *Catch Up Literacy* programme, teaching assistants and coordinators attended face- to- face training prior to introducing the intervention and at the end, in an evaluation and sharing session. Training costs are provided as part of the overall costing to the programme. Teaching assistants and coordinators also reported that they approached trainers outside of the sessions and that the trainers' visits to schools were seen as particularly valuable. It is not clear from the report the extent to which these are normally provided to schools who choose to invest in the *Catch Up Programme.*

Among findings from *Switch on Reading,* evaluators reported:
> "…close monitoring of teachers is necessary to ensure that teachers conduct the sessions as trained. (The intervention project team had a system to monitor and support schools. Each school was assigned a Switch-on trainer, themselves trained by the project leads, whose role was to monitor and support the schools in implementing the programme: they visited the schools regularly to provide extra support to the members of staff, and had regular online communication for feedback between these expert trainers and the project leaders. The enthusiasm of the project and teaching staff was a noticeable feature of this intervention.)" (p 28) and "It is not clear that such a level of commitment could be maintained if the intervention was used more widely" (p33).

In the case of the school- group trials, *Accelerated Reader* and *Fresh Start* support was provided for conduct of the RCT itself, rather than implementation of the intervention and, in both cases was considered to be essential to the rigour of the trial.

Key points

- Programmes developed by groups of schools appear to be relatively successful (also noted in the quantitative evaluation in relation to quality outcomes). Recruitment tends to be less problematic than observed in other types of trials and retention of schools is high.
- Where there are problems of implementation these often appear to be linked to a lack of shared understanding among senior leaders and teachers of what is involved. The amount of preparation required for introducing the interventions is a common issue that occurs across all programmes.
- In some trials, staff felt that the assessment used to identify pupils for targeted interventions did not identify those who would benefit most.
- Programmes which require pupils to be withdrawn from subjects across the curriculum are more likely to experience resistance from pupils and these subject staff. Senior leader support is required for ensuring suitable spaces, preparation time for TAs and timetabling arrangements. Programmes for use with a whole class are easier to implement.
- External support from the delivery team facilitates implementation. It is a feature of nearly all efficacy trials and may be more available in effectiveness trials than in a 'real life' situation.

Intensity of the programme and its bearing on implementation and success

In line with discussion in the quantitative report, we considered programme intensity measured by the number of minutes per week and also its length, measured by the number of weeks on the programme.

Some in-school programmes required an intensity of 50 or more minutes per week and, as such, required significant timetable modification for schools in the trial. Evaluation reports for these projects indicate problems with recruitment of schools and/or drop out of schools from the programme when full implications for timetabling, staffing and resourcing were understood or experienced. Implementation issues were often linked to small- group or one- to- one interventions or a lack of IT resources. Whole- class interventions with high intensity experienced fewer problems. Recommendations from high intensity pilot trials were generally found to be for adaptations other than that of programme intensity.

Other programmes with high intensity in minutes per week are a number of out- of- school learning projects including summer schools, parent classes and extra- curricular activities. All of these experienced significant implementation problems. Evaluation reports suggest that implementation problems were more closely related to factors other than the intensity of the programme.

We considered a sample of programmes with promising pupil outcomes of 30 or more weeks in duration, but where intensity per week is less than an hour per week. Some whole- class interventions, such as *Thinking, Doing, Talking Science,* are designed for flexible use in normal whole- class science lessons by teachers over the course of the school year. Process evaluation outcomes suggest that such an approach is welcomed by teachers, experiences few problems of implementation and has a positive impact on practice with use over the whole year developing confidence and familiarity. Lengthy projects also include several programmes that were general in focus, for which academic attainment gains in a relatively short period may be considered unlikely, such as *Philosophy for Children* and *Promoting Alternative Thinking Strategies.* For all of the long- term interventions, the length of the programme does not seem to have been a contributing factor to its success or otherwise. However, it would appear that programmes which require teachers to change their pedagogy require a long term approach.

Evaluation reports for one- to- one tuition, as with the *Perry Beeches Graduate Coaching Programme*, suggest that success is dependent on the quality of the tuition and the engagement of the pupils developed through relationships, which perhaps may require time to be developed. However, shorter one- to- one and small- group tuition programmes show similarly promising outcomes, suggesting that further research might be needed in this area.

Key points
- Intensive interventions requiring significant timetable adjustments find it more difficult to recruit and retain schools for the trial.
- Intensive interventions requiring attendance by pupils or parents out- of- school hours experience problems with attendance and attrition, with little evidence for impact on attainment.
- Interventions which require a change in teacher's whole- class pedagogy may benefit from a long term programme.
- The success of one- to- one tuition appears to be related to the quality of tuition, but may also be linked to the length of the programme.

Models of professional development and training associated with more successful programmes
The majority of comments made in process evaluations about the quality of **pre- programme training** are positive, both with regard to the quality of delivery and the response of trainees, whether teachers or teaching assistants. In most cases, training is provided face to face, although a small number of programmes use on- line training modules. The number of these is too low to form conclusions about the relative effectiveness of online versus face- to- face training. On the whole, participants felt that training had prepared them well for implementation, but this was not always the case and was noted in some less successful trials. In most cases, training participants are those who are directly responsible for programme delivery to pupils, although in some cases, a 'cascade' model is used, such that participants must return to their schools and train other members of staff who will be involved in delivery. The lead- in time for the programme may be insufficient for this and

might be considered carefully when planning programmes which rely on this model. *Response to Intervention* suffered from insufficient time to cascade training and prepare as a school for the complex, tiered, intervention. In contrast, the successful *Thinking, Doing, Talking* science programme included a whole- school launch event and two days for the two teachers trained in each school to deliver the programme to plan together on how the methods and materials might be integrated into their normal science lessons. These suggest that time between training and implementation might be valuable.

Following initial training, nearly all efficacy programmes provide **additional support** from the developer, through a programme of school visits and/or support on request. The extent to which additional support is important to the programme varies. For example, in *Switch On Reading* each school was assigned a trainer who visited the school regularly to provide extra support to members of staff and frequent online communication. In *Nuffield Early Language Intervention*, a single visit was found sufficient in most schools. The extent to which support might realistically be provided at an affordable cost in extensions to the programme is commented on in some process evaluations, for example, the *Improving Numeracy and Literacy* report.

There were frequent comments about the **quality of resources** and this was the case in successful as well as less successful programmes. In *Improving Numeracy and Literacy*, teachers complained about basic errors in materials, including in the computer games that were an integral part of the programme. In this programme, teachers also complained about the heavy reliance on worksheets, which do not match their usual teaching practice and that materials did not allow for sufficient differentiation. Materials prepared for use in another context may not be suitable. However, the most common issues reported were, firstly, that guidance on prescribed session content was unrealistic with too much expected for the time allotted. Secondly, the time required for preparation was too high and was expected in addition to staff's normal workload. Both of these were reported by TAs and by teachers and for highly prescriptive and for flexible programmes. The *Catch Up Literacy* programme is an example relating to teaching assistant time, with one the key recommendations, "Schools should ensure … that teaching assistants are given adequate time to prepare before each lesson." (p 4). In the Catch Up Literacy trial, secondary schools were provided with funding for salary costs for two teaching assistants who delivered the one- to- one intervention in their own school and in linked primary schools that had agreed to take part. Unlike teachers, teaching assistants are paid on an hourly rate and would not be expected to work outside their contracted hours. The process implementation report indicates that the TAs were not able to deliver the one- to- one sessions in the time originally allocated and that the time required for them to prepare for each session was more than expected. The issue here would appear to be that the amount of funding provided for teaching assistant hours was insufficient for the time actually needed to prepare for, and deliver, each intervention session. Teaching assistants found the actual materials provided easy to use and adapt to individual pupils' needs.

Teachers are expected to plan, prepare and assess pupils as part of their professional role, and are given some timetables 'Planning, Preparation and Assessment (PPA)' time within their timetables. However, teachers in some trials reported that they needed more time for trial requirements and additional time was not costed into the trial. For example, in the *Vocabulary Enrichment Intervention*, teachers attend pre- implementation training in three different elements and are then given flexibility in how these are used within their lesson planning. In the trial teachers reported spending a long time working through the materials provided and preparing suitable lesson plans. They had insufficient planning and preparation time to make as much use of two of the intervention strands as intended in the intervention. The process implementation report states), "Schools intending to deliver the intervention would also need to consider providing extra non-contact time for teachers to prepare the scheme of work for the lessons; some teachers reported that the time required went beyond their normal planning and preparation time allocation" (p 20). These comments indicate that an additional amount of non- contact time might be required for teachers – however, it might also be interpreted as, in part, an issue with the materials provided and the extent to which they have been prepared and trialled by developers for ease of use in realistic school situations. It is not clear whether the predominant influencing factor in high preparation time is unfamiliarity with materials at first use or inherent in the programme itself.

Process evaluations for programmes which involved external staff coming into school reported variability in the quality of delivery, due to some tutors' inexperience of working with school-age children. In all programmes which relied on staff without school experience, there were reports of behaviour issues and recommendations that this should be considered in the training programme.

Key points
- To achieve ideal conditions in efficacy trials, the preparation time required for introducing any new initiative into a school should be considered and costed into the trial.
- Lack of time to assimilate training and prepare for implementation may be more significant than the model of training prior to the programme.
- The quality and relevance of programme guidelines, protocols and teaching resources is variable. The amount of expected content is frequently found to be too high for times allotted for programme delivery. Resources provided are rarely ideal for direct use with pupils, and considerable time may be needed to prepare additional material and make adaptations. In some cases, resources did not allow for sufficient differentiation. Some resources, such as pupil worksheets, are not well-matched to current expectations of good teaching practice.
- External or inexperienced tutors need training which enables them to use subject or academic expertise with the pupils in the trial. Those who have not previously worked in schools need training in behaviour management.
- Training that relies on cascading within the school may require additional time and resources.
- The extent to which the contribution of external monitoring and support provided in efficacy trials is essential to the success of the programme might be taken into account when planning further trials.

The success of domain-general versus curriculum-specific programmes

Domain-general programmes are intended to increase pupil attainment through an intermediate outcome, for example, increased confidence or increased parental engagement. A direct consequence of this is that any potential attainment gains are less likely to be measurable in the timeframe for the project. The lack of effect of some domain-general programmes would seem to be related to factors other than the length of the programme. *Creative Futures: Act, Sing, Play* and *Chess in Schools* were provided through the use of external tutors and took place over the course of a school year or longer. Despite concerns about variability in quality of external tutors, schools and pupils were positive about the interventions and reported impact on outcomes such as confidence and a broadening of opportunities for disadvantaged pupils. However, teachers were concerned about the amount of time taken from the statutory curriculum for these programmes. *Parent Academy*, *Halle SHINE* and *Youth Social Action* involved out-of-school learning with implementation issues discussed earlier.

The quantitative review found that literacy programmes gave a larger effect size, on average. No reason for this could be found when we compared programmes and examined those which used similar approaches, such as the peer coaching programmes, *Durham Shared Maths Project* and the equivalent literacy programme *Paired Reading*. The two science trials were both intended to lead to long-term, fundamental, change in teachers' pedagogy. It would appear that the higher mean effect seen in literacy trials is unrelated to the curriculum focus.

Key points
- The measure of pupil academic attainment and the timeframe of trials may not be appropriate for domain-general programmes.
- Domain-general programmes may have to compete for space in the curriculum.

Differences between programmes delivered by charities, schools and universities.
Differences between projects in the early years, primary and secondary phases.

We found the deliverer or the phase is less likely to be related to the success of the programme than factors considered earlier. Qualitative analysis supports the finding from the quantitative analysis that school group trials are more likely to be successful, with noticeably fewer issues about recruitment or retention of schools. It may also be more likely that materials for use in school group trials being trialled are school ready, with *Mathematics Mastery* being trialled in the ARK group of schools and the *Improving Writing Quality* trial having a working group of teachers supporting implementation. The quantitative analysis notes than trials delivered by a charity had lower security, on average, than other trials. It is noticeable that programmes delivered by charities are more likely than other programmes to involve external tutors working with school pupils and to be general than curriculum specific.

Only one of the efficacy or effectiveness trials was in an Early Years context, with the efficacy trial for *Nuffield Early Language Intervention* run in primary schools with attached nurseries, and did not include nursery schools or private providers. Quantitative analysis indicated that trials in the transition phase between primary and secondary school are least likely to be successful and reasons for this have been discussed in section 4.2. Qualitative analysis indicates that the success of the programme is more likely to be related to factors other than the phase for delivery. A reason for the higher success rate of secondary- phase trials, identified in quantitative analysis, is likely to be that a higher proportion of primary- phase trials were general, rather than curriculum specific, and that these included the two parenting interventions and one of the summer schools, which experienced significant problems.

4.4   Additional themes identified in qualitative review

Sustainability

Some evaluation reports refer to sustainability and this is an important school outcome indicative of the success of a programme. Process evaluations may report the views of delivery staff about sustainability, but rarely report the views of the senior leaders responsible for making decisions. However, the *Nuffield Early Language Intervention*, process evaluation does include the views of both senior leaders and the teaching assistants responsible for delivery, linking a decision about continuing with the intervention to:

"• the perceived need for the programme in the school—including whether other language interventions were already running;
• the perceived or anticipated impact of the programme;
• experiences of delivering the programme—including the use the resources accompanying the programme;
• the eventual financial costs of the programme; and
• the control schools would have in selecting pupils for the programme after the trial" (p 39).

Time pressures may make schools reluctant to engage or continue with programmes which are not directly related to examination targets, as, for example, in *Philosophy for Children*. However, comments in relation to some programmes indicate that they enhanced, rather than competed, with schools' usual approaches, as in the *Response to Intervention* programme and this seems to have contributed to positive comments about maintaining the approach in future years.

Several evaluation reports note an intention, or make a recommendation, to track intervention pupils to assessment at the end of their current phase of schooling (for example, the *Thinking, Doing, Talking Science*), with the aim of identifying if interventions may show impact in the longer, rather than the short term. What does not seem to be considered is the importance of identifying if individualised, expensive interventions intended to help low prior attainers catch up with their peers are sustained when they return to the whole class, or if they again fall behind and require further intervention. Schools need to know if the high investment made in one- to- one or small- group, short- term interventions will have a lasting effect.

Key points
Schools base decisions on continuing with an intervention on:

- The fit with school improvement priorities, current approaches and the needs of pupils.
- Perceived impact when compared with other interventions in place in the school.
- Ease of implementation.
- Control on the selection of pupils for the intervention.
- Quality of materials.
- Cost
- Curriculum time for programmes that are not directly linked to accountability targets.

5.   Limitations of this review

This review has relied on information provided in published evaluation reports. The EEF has already made changes so that more recent reports follow recommendations on *Implementation and Process Evaluation* (Humphrey et al., n.d.) and provide more comprehensive information in an accessible format. However, some of the themes identified in our review are based on incomplete information in relation to trials and these are indicated in this section.

In view of the number of issues about recruitment, it would be useful if more reports included information about reasons for schools engaging with the project, as well as giving information about reasons for declining. It would be also be useful if any capacity indicators such as current Ofsted grading were linked to implementation issues, including withdrawal from the trial. There are few direct references to school capacity in programme reports (an exception is the point 'lack of school capacity' given as a reason for not participating in the *Nuffield Early Language Intervention*). However, there are indicators that this is a barrier to implementation in several reports. Finding time in the school timetable for an intervention may be difficult – this might be because withdrawal from lessons for a one- to- one or small- group intervention risks pupils' access to core learning in other subjects, as reported for *Catch Up Literacy* or there may be competition for space in a crowded curriculum and with preparation for external tests such as SATs, as in *Promoting Alternative Thinking Strategies (PATHs),* where, on average, schools only delivered about half the expected number of sessions over the trial period. The evaluation reports states "Even in cases where PATHS was viewed as critical, instruction in core academic curriculum areas always won out ('*Well obviously it's massively important, but in terms of literacy and numeracy… it is below that*'). As one teacher noted: '*something has to give*'." (p 23). Other indicators of lack of school capacity are problems resulting from changes in staffing (given as a reason for some schools withdrawing from a trial after recruitment in *Durham Shared Maths Project*; *Let's Think Secondary Science*); in communicating with the developer and evaluation team and accessing recommended on- trial support (the *PATHs* programme) and of issues with resources, such as the IT resources required by *Improving Literacy and Numeracy* or spaces for individual tuition as in *Catch Up Literacy.* Measures of school capacity are difficult and one is that of the most recent Ofsted judgement. Being in the categories of 'Requires Improvement' or 'Inadequate' did not appear to be a barrier to implementation in some programmes and, for example, *Thinking, Doing, Talking Science* was successful despite nearly half of the intervention schools being in these categories. However, the evaluation report for the literacy intervention *Quest* noted that of the four 'inadequate' schools, three failed to provide post- test data. The *Durham Shared Maths Project* recruitment strategy excluded schools that had recently been graded as 'inadequate' or that were facing significant reorganisation, such as taking on academy status. There was a low school drop- out rate for this trial, with only 3 schools withdrawing from 82 recruited for this effectiveness trial.

An Ofsted grade may provide an indication of the overall quality of teaching in a school and the significance of this is noted in the evaluation report for *Promoting Alternative Teaching Strategies,* "… teacher preparedness, enthusiasm, and the manner in which they engage children in PATHS lessons are ultimately more strongly associated with children's attainment than the extent to which they adhered to the lesson scripts" (p21).

The importance of the commitment of school leaders to the programme trial emerges clearly from review of process evaluations. More recent evaluation reports demonstrate how this is taken into account in the information given to senior leaders in schools before signing up. For example, the information included in the memorandum of understanding for *Creative Futures: Act, Sing, Play* appears to be comprehensive, although even here there were problems in some schools about storage of musical instruments, variability in the extent to which the pupils were permitted to take instruments home for practice as expected and evidence that not all staff were on- board. A formative finding in this process evaluation was that there be a pre- meeting between tutors and teachers in order to: "help teachers and other school staff to understand the aims and broad content of the programme, and help to bring all staff on-board" (p41)

In view of the expense and difficulties of implementation of small- group and one- to- one interventions and the reservations of senior leaders about continuing with these beyond the trial, it might be helpful for EEF to secure more evidence in this area than was available for this review. Of particular interest might be adding to the evidence base on relatively short and less expensive small- group or one- to- one interventions and determining if it is the nature of the programme or the more intensive support that has a greater effect. There may also be scope for variants of interventions that were originally designed for one- to- one or small- group use, but adapted for more flexible use of some elements, as suggested in formative findings from the *Nuffield Early Language Intervention* evaluation. The enthusiasm of teachers for the *Response to Intervention* programme suggests that a tiered approach to individualised support might be welcomed by schools. Such programmes might be tested further.

Also welcome in evaluation reports would be further analysis of the potential sustainability of the programme. This is already covered to some extent by the section on costs, but sustainable options for training and support might also be included and the views of headteachers and senior leaders on whether they will continue with the intervention following the trial might be presented in all process evaluations.

6. Conclusions and Recommendations

The quantitative analysis documented a number of potentially interesting associations between the context and outcomes factors relevant to trial design, based on analysis of the reports of trials that the EEF has published during its first five years. The average EEF trial has achieved a security rating of three padlocks with the most common limiting factor for this being attrition, this was particularly the case in the case of trials conducted in primary schools. Trials of projects designed by charities, rather than school groups, universities or local authorities, tend to have lower security ratings, although it seems likely this could be due to the types of interventions they are more likely to propose, which could still be very valuable.

The average effect size of interventions that have been evaluated is equivalent to one month's progress, although it is marginally higher for pupils who have ever been eligible for free school meals. Unfortunately, very few trials have been powered to detect effect sizes of this magnitude. The vast majority of EEF trials have an achieved minimum detectable effect size smaller than the effect size that they find; this remains the case even among projects with the maximum security sub-rating for power.

Surprisingly, given that attrition is more likely to be the security rating limiting factor in primary school trials, average attrition at the pupil level is slightly higher in secondary school trials (although this is not logically inconsistent). However, attrition is notably higher in trials that span the transition between these two phases; this doesn't seem that surprising and possible explanations are obvious but it does highlight it must be planned for in future trials of this type. Attrition is also higher in trials that are more intensive (more minutes per week of delivery) and this highlights importance of finding ways of reducing the impact of the higher intensity on burden and, hence, attrition despite the potential need for relatively intensive interventions in order to achieve larger impacts.

We note that while there is a positive association between cost per pupil and effect size found, there is also a lot of variation around this trend, suggesting that some things matter much more than the cost of the intervention.

The qualitative review examined the characteristics of interventions that were relevant to its achieving positive outcomes for pupils and schools. On the basis of our analysis, we summarise the key findings relevant to trial factors identified by EEF:

Different characteristics of EEF programmes that have been successful or unsuccessful in raising pupil attainment.

Programmes are more likely to be successful in raising pupil attainment if:
- Senior leaders have a good understanding of the intervention and show commitment and support.
- Timescales of the trial sufficient are sufficient for effective implementation and for the effect on pupil attainment to be demonstrated.
- The timing of the intervention in the school year suits school patterns of activity.
- The target group is accurately identified.
- Delivery of the intervention takes place within relevant, timetabled lessons.
- Materials are of high quality, with essential elements identified and sufficient flexibility to be adapted to school timetables and for appropriate differentiation
- Staff receive high- quality training before the programme and sufficient time for preparation and collaboration during implementation.
- Monitoring and support, whether from external teams or within the school, is effective.

Programmes are less likely to be successful in raising pupil attainment if one or more of these characteristics is not met.

Common characteristics of programmes that have been easy or difficult to implement

- Programmes developed by groups of schools appear to be relatively easy to implement. Recruitment tends to be less problematic than observed in other types of trials and retention of schools is high.
- Where there are problems of implementation these often appear to be linked to a lack of shared understanding among senior leaders and teachers of what is involved. The amount of preparation required for introducing the interventions is a common issue that occurs across all programmes.
- Programmes which require pupils to be withdrawn from subjects across the curriculum are more likely to experience resistance from pupils and these subject staff. Senior leader support is required for ensuring suitable spaces, preparation time for TAs and timetabling arrangements. Programmes for use with a whole class are easier to implement.
- External support from the delivery team facilitates implementation. It is a feature of nearly all efficacy trials and may be more available in effectiveness trials than in a 'real life' situation.

Intensity of the programme and its bearing on implementation and success
- Intensive interventions requiring significant timetable adjustments find it more difficult to recruit and retain schools for the trial.
- Intensive interventions requiring attendance by pupils or parents out- of- school hours experience problems with attendance and attrition, with little evidence for impact on attainment.
- Interventions which require a change in teacher's whole- class pedagogy may benefit from a long term programme.
- The success of one- to- one tuition appears to be related to the quality of tuition, but may also be linked to the length of the programme.

Models of CPD/training associated with more successful programmes
- To achieve ideal conditions in efficacy trials, the preparation time required for introducing any new initiative into a school should be considered and costed into the trial.
- Lack of time to assimilate training and prepare for implementation may be more significant than the model of training prior to the programme.
- The quality and relevance of programme guidelines, protocols and teaching resources is variable. The amount of expected content is frequently found to be too high for times allotted for programme delivery. Resources provided are rarely ideal for direct use with pupils, and considerable time may be needed to prepare additional material and make adaptations. In some cases, resources did not allow for sufficient differentiation. Some resources, such as pupil worksheets, are not well- matched to current expectations of good teaching practice.
- External or inexperienced tutors need training which enables them to use subject or academic expertise with the pupils in the trial. Those who have not previously worked in schools need training in behaviour management.
- Training that relies on cascading within the school may require additional time and resources.
- The extent to which the contribution of external monitoring and support provided in efficacy trials is essential to the success of the programme might be taken into account when planning further trials.

The success of domain-general vs curriculum-specific programmes
- The measure of pupil academic attainment and the timeframe of trials may not be appropriate for domain- general programmes.
- Domain- general programmes may have to compete for space in the curriculum.

Key differences between programmes delivered by charities, schools and universities
- School group trials experience fewer issues with recruitment and retention of schools and fewer problems with implementation.
- Programmes delivered by charities are more likely than other programmes to involve external tutors working with school pupils and to be general than curriculum specific.

Differences between targeted, whole class and whole school programmes
- The time frames for trials are such that programmes intended to be implemented throughout the school or across a number of year groups are insufficient to determine whether this is a factor in the success of the intervention.
- Targeted programmes which require pupils to be withdrawn from subjects across the curriculum are more likely to experience resistance from pupils and these subject staff. Senior leader support is required for ensuring suitable spaces, preparation time for TAs and timetabling arrangements. Programmes for use with a whole class are easier to implement.

Differences between projects in the early years, primary and secondary phases.
- The success of a programme is more likely to be related to factors other than the phase for delivery.
- Trials in the transition phase between primary and secondary school are least likely to be successful.

Sustainability

Schools base decisions on continuing with an intervention on:
- The fit with school improvement priorities, current approaches and the needs of pupils.
- Perceived impact when compared with other interventions in place in the school.
- Ease of implementation.
- Control on the selection of pupils for the intervention.
- Quality of materials.
- Cost.
- Curriculum time for programmes that are not directly linked to accountability targets.

In addition, we have identified a number of cross cutting issues that are relevant to the design and implementation of all trials and that would merit further investigation:

- The importance of securing the commitment and support of senior leaders for the trial and its implementation.
- Recruitment and retention of schools.
- School capacity to introduce an intervention and to fulfil trial requirements.
- Overall timescales for trials to include planning, training, preparation, implementation and testing.
- The need for further evidence on critical factors related to the effectiveness of one- to- one and small- group interventions.
- The potential sustainability of the intervention after the trial.

This review has identified a number of factors which relevant to the design of a programme and its testing through Random Controlled Trial (Appendix A). Although the variation in programmes and the impact of interrelated factors makes it difficult to make secure recommendations about future programmes and trials, evaluation reports indicate those factors that are critical to the success of a trial design. Based on our analysis, Table 15 provides a sample of configurations of Context, Mechanism and Outcome that appear to be successful, with examples taken from our analysis.

Table 15 Configurations of trial designs that appear to be more or less likely to lead to a positive effect on pupil progress and on other outcomes, with examples of programmes including many of these factors.

CMO configuration	Context	Mechanism	Outcome	Example trials
CMO1- more likely to be successful	Sample size sufficient for MDES to be reached; schools have good capacity for innovation; senior leaders have good understanding and are committed to trial; planning has taken into account resource, space and intensity requirements; timescales of trial sufficient; timing in school year suits school patterns of activity.	Local recruitment with capacity taken into account; clear communication of expectations; accurate identification of target group; delivery within relevant, timetabled lessons; high quality materials, with essential elements identified with flexibility for fitting within school timetables and for appropriate differentiation; well-prepared staff (external or internal); effective monitoring and support.	High implementation fidelity; Low attrition; High security rating; positive impact on pupil attainment and/or other pupil outcomes; pupils, delivery staff and senior leaders enthusiastic; sustainable; clear formative findings for scale up.	Trials with whole class delivery: *Mathematics Mastery: Secondary*; *Accelerated Reader*; *Improving Numeracy and Literacy; Chess in Schools.* Trials with 1-1 or small- group delivery: *Nuffield Early Language Intervention; Talk for Literacy.*
CMO 2 – less likely to be successful.	Trial continues although recruited sample size too small for MDES to be reached; uneven understanding and commitment of senior leaders; timescales too tight; timing of intervention conflicts with normal patterns of activity.	expectations not clearly understood; identified target group not accurate; intervention disrupts normal timetables; school resource and space requirements not available; programme inflexible with unrealistic demands for teaching time and preparation; staff (external or internal) insufficiently prepared; insufficient monitoring and support.	Low fidelity, high attrition, low security, little evidence of impact on attainment; pupils, staff and senior leaders have negative perceptions of intervention; sustainability unlikely.	Trials at Transition and/or out of school programmes. Trials with whole class delivery: *Response to Intervention.* Trials with 1-1 or small- group delivery: *Perry Beeches Graduate Coaching Programme.*

Appendices

Appendix A: Context, Mechanism and Outcome factors in Quantitative and Qualitative analyses

Context	Context variants	QUANTITATIVE	QUALITATIVE
Type of trial	- Pilot - Efficacy - Effectiveness	✔	✔
Developer	- Charity - School group - University - Local authority	✔	✔
Evaluator	- University - Non- university	✔	
Impact model	- Direct impact on pupil attainment - Impact via intermediate outcomes		✔
Sample Size	- Number of schools - Number of pupils	✔	✔
Randomisation	- Intervention and control pupils in same school - Intervention and control pupils in different schools	✔	✔
Targeting	- Whole class - Small- group - One- to- one - Mixed/tiered interventions - Parents/carers - Teachers	✔	✔
School phase	- Early years - Primary - Secondary - Transition	✔	✔
School suitability indicators	- Number of pupils in targeted group - Capacity indicators (Ofsted grading, school capacity, school quality of teaching, previous relationships with developer) - Convenience for developer support and for process evaluator visits		✔
Curriculum focus	- Numeracy/mathematics - Literacy/English - Science - Other curriculum - Domain- general	✔	✔
Logistical factors	- IT resources - Library or equipment resources - Space requirements		✔
Timescale- related factors	Recruitment period	✔	✔

	• Preparation period (memoranda of understanding signed, pre-testing) • Randomisation period • Professional development period: delivery and in- school preparation • Implementation period: timing (e.g. summer, spring autumn term), intensity as number of weeks of implementation, intensity as number of minutes per week, • Post- testing: immediate, medium, long- term • Evaluation period: immediate, long- term.		
Cost per pupil		✔	

Mechanism	Mechanism variants	QUANTITATIVE	QUALITATIVE
Implementation: Recruitment and trial preparation	• Recruitment strategy • Communication of expectations • Memoranda of understanding • Planning and communication • Identification of target group		✔
Implementation: Identification of target group and measuring effect	• Pre- test • Post- test • Use of other data (e.g. teacher assessment, key stage 2 or GCSE results, other outcomes)		✔
Implementation: Intervention delivery mechanism	• Whole class by teacher • Whole group by external/unqualified • Small- group by teaching assistant or other school staff • Small- group by external/unqualified: within related curriculum subject time/withdrawal from other lessons/out of school hours • One- to- one by teaching assistant or other school staff		✔

		QUANTITATIVE	QUALITATIVE
	• One- to- one by external/unqualified • Direct to pupil/target (online, text) • Other		
Implementation: Timetabling	• within related curriculum subject time/withdrawal from other lessons/out of school hours		✔
Implementation: Flexibility	• Highly structured – no deviations or omissions from those prescribed • Essential and optional features identified • Menu of possible options • No prescription • Adaptation of materials to context		✔
Professional development	• Pre- intervention training/face to face, online, cascade • Materials and handbooks • School-based preparation and collaboration/working groups • Intervention related support/developer, senior leaders • Evaluation or research methods support		✔

Outcome	Outcome variants	QUANTITATIVE	QUALITATIVE
Trial quality	• Security rating • Effect size	✔	
Pupil Outcome	• Effect size - attainment directly related to intervention focus • Effect size - attainment indirectly related to intervention focus • Non- academic outcomes	✔	✔
School outcome	• Delivery staff skills		✔

	• Delivery staff enthusiasm and motivation • Senior leader response • Sustainability		
Planning outcome	• Formative findings to inform further trials and scale up.		

Appendix B

Findings from efficacy trials and trials with high security ratings.
Key points are provided in section 4.2.

Small, school- group, efficacy trials

Table 16: Sample of small school- group efficacy trials

Project Name	Type	Security Rating	Months Progress	Months Progress for FSM	Attrition rating
Accelerated Reader	Efficacy	3	3	5	5
Talk for Literacy	Efficacy	4	3	4	5
Perry Beeches Graduate Coaching Programme	Efficacy	3	5	5	3

Accelerated Reader is a computer- based management and monitoring programme that aims to foster the habit of independent reading among primary and early secondary age pupils. Pupils are taught to use the programme to select books matched to their reading level, which they then read independently. Pupils are also shown how to use the assessment tools within the programme to track their reading progress. *Talk for Literacy* is a small- group intervention delivered by trained TAs (teaching assistants) using vocabulary enrichment and narrative materials. *Perry Beeches Graduate Coaching Programme* provides flexible and individualised one- to- one literacy tuition from school- based coaches recruited specifically for the project. All of these interventions were among a cluster of projects aimed at improving the literacy level of pupils whose level of attainment is likely to restrict their access to the secondary school curriculum. All three efficacy trials took place for targeted Y7 (year 7) pupils in a small number (3 or 4) of secondary schools and used pupil randomisation within each school to form intervention and control groups. Although overall security ratings were moderate, outcomes were positive. What differs in the trials in the consistency of trial application across different schools, including significant pupil attrition in the *Perry Beeches Graduate Coaching Programme.*

Accelerated Reader was one of a small number of EEF aggregated efficacy trials in which schools had independently applied for funding and were grouped together to conduct and evaluate the project themselves with the support of the external evaluator. Unlike most trials, there was no need to undergo the process of recruitment and communication of expectations to the schools involved. By bidding for funding, the four schools in this trial had demonstrated senior leadership support and were likely to have already considered the practical implications of running the trial in their schools. Teachers or TAs were responsible for training and supporting pupils in using the computer programme to identify and select suitable books for independent reading and to use the online testing and monitoring elements. Staff involved were trained for their role by the programme developer. Internet access and a supply of sufficient books for pupils to read were required and some participants would have welcomed a greater choice of books to meet pupils' interests. Schools made varying arrangements for ensuring targeted pupils had sufficient access to the programme over the 22 weeks of the trial, with group withdrawal from normal classes and after-school access both noted. Despite some barriers in timetabling and with IT resources, the process evaluation for the project noted that fidelity to the programme was high, with some adjustments to implementation guideline to help very low attaining pupils access suitable texts. All intervention pupils received the recommended amount of time. Staff and most pupils were positive about using the programme and, with help from the external evaluators were able to conduct the trial competently. The evaluators of *Accelerated Reader* suggest that the trial "… represents a more 'real world' approach to evaluation, akin to schools buying an intervention and implementing it themselves with no external support" (p 5). However, they also suggest that it is necessary for schools to be trained and supported by independent experts for the evaluation.

The *Talk for Literacy* trial was also conducted in a small number of schools, but in this case evaluation was by the external evaluator. Of the three schools involved, one was the developer's own school with the other two recruited through her local contacts. Two schools had an Ofsted rating of 'outstanding' with the other rated as 'good'. This would indicate that schools were more likely to have the capacity to implement a trial. The trial required schools to identify TAs to be trained and to deliver the intervention and to find suitable spaces for small- group work during twice-weekly sessions over two terms. Groups of 5 to 8 students were withdrawn from lessons in two schools and attended timetabled sessions in one school. One of the barriers reported to implementation was the lack of suitable space for group work, suggesting that detailed implications for schools were not fully understood. As with *Accelerated Reader*, pupils experiencing *Talk for Literacy* were identified as eligible for the intervention and randomised into intervention and control groups within each school. TAs were enthused by the training provided for running the small- group sessions. However, in the delivery phase of trial, TAs found it extremely difficult to fit in the expected number of activities and very time consuming to prepare for sessions, resulting in variability in what they chose to include. As with most programmes in EEF trials, this was the first time staff involved had delivered the programme and it is not known if with greater experience of delivery, adjustments to content and grouping demands and better forward planning by the school, the programme would have been delivered with greater fidelity.

Perry Beeches Graduate Coaching Programme was initially piloted at the developer school, who then recruited three other schools with which it had existing links, with all four part of the efficacy trial. One of the participating schools had been graded by Ofsted as 'requires improvement', with one 'good' and two 'outstanding'. Any impact of the overall capacity of the schools involved to undertake the trial is not reported. There was considerable inconsistency in all aspects of the trial. In the pilot, coaches were recent university graduates without teaching experience, but for this trial schools made different decisions regarding qualifications and experience of coaches recruited, although few had qualifications or experience related to their teaching role. There were differing levels of attention to the need for training and support, suitable spaces, provision of resources, timetabling, supervision arrangements and whole school engagement. Pupil attrition was high, largely because one school did not correctly test all participating pupils. Many of the coaches reported that they did not feel equipped for their role or have the strategies to meet the diverse learning needs of their tutees. Coaches were insufficiently prepared to manage pupil behaviour effectively. The programme does not prescribe content for one- to- one sessions and there was no consistency in quality of provision. A mixture of one- to- one and small- group support was used and the frequency and length of sessions ranged widely between schools and students.

Outcomes
As might be expected in efficacy trials, results for all pupils and those eligible for free school meals in these trials are limited by small sample sizes. The evaluator for *Accelerated Reader* found that "there is strong evidence overall that AR improves reading for new Year 7 pupils with weak reading skills and habits, and that it does so especially for those eligible for FSM "(p19). Staff, pupils and their parents responded positively to the programme and TAs reported gaining new skills for supporting reading. With regard to *Talk for Literacy*, TAs reported positive impact on pupils' verbal skills and confidence and, although there was no significant impact on overall reading ability, there was statistically significant evidence of improvement in the secondary measure of passage comprehension. According to the evaluation report, "Overall, TAs reported that the intervention is conceptually sound and enjoyable to deliver, but several practical issues need to be overcome to make it more beneficial for individual pupils and not prohibitively time-consuming for TAs" (p26). The *Perry Beeches Graduate Coaching Programme* evaluation report notes that "The programme had a positive impact on pupils' attainment in reading, spelling and grammar, equivalent to approximately five additional months' progress. The evaluation did not seek to prove that the approach would work in all schools, but did identify strong evidence of promise" (p4). The relationship between coaches and pupils appeared to be important to the views of pupils.

Efficacy trials in larger numbers of schools

The following subset of efficacy trials, also targeted on improving literacy in low- attaining pupils at the beginning of secondary education, involved larger numbers of schools. They have been selected to illustrate further the issues which appear to be related to a school's capacity to introduce and support a new initiative and of the challenges associated with ensuring full understanding and commitment in relation to the demands of the trial. The section also highlights issues of implementation associated with the mechanism through which the programme is delivered and the implications for outcomes.

Table 17: Additional sample of efficacy trials

Project Name	Type	Security Rating	Months Progress	Months Progress for FSM	Attrition rating
Switch-on Reading	Efficacy	3	3	4	5
Improving Writing Quality	Efficacy	2	9	18	5
Response to Intervention	Efficacy	1	3	6	2

Secondary schools in the *Switch On Reading* trial were recruited from within the local authority in Nottinghamshire. Pupils were randomised within each school to a control and an intervention group. The intervention consists of daily, one- to- one support sessions for weak Y7 readers with teaching assistants who have been trained with the materials and processes particular to the programme. *Improving Writing Quality,* was designed for use at the end of Y6 and the beginning of Y7, and includes a structured, whole- class, approach to helping pupils plan, monitor and evaluate their writing. *Response to Intervention* offers an individualised, tiered approach to raising achievement in reading. It uses whole- class teaching, plus targeted small- group remedial teaching and one- to- one tuition where necessary, depending on close needs analysis by the teacher. It was trialled in the summer term of Y6.

37 secondary schools were originally approached for the *Switch on Reading* trial by members of Nottinghamshire education department's school improvement team, who are likely to have had both understanding of the proposed intervention and knowledge of the schools' capacity and improvement priorities. Senior leaders in those schools which expressed an interest discussed the project in more detail with members of the LA project team. Eligible pupils were identified by participating schools on the basis of KS2 results or other indications of poor reading attainment. However, despite a recruitment strategy which built on a local network and relationships, far fewer pupils were identified as eligible for the trial than expected (314 rather than 480) in 21 schools. Teaching assistants were trained in the content and protocol expected for each of the sessions by staff from the local authority, who were experienced with the *Reading Recovery* intervention for younger pupils on which *Switch On Reading* is based. Although the recruitment approach might have been expected to support the 'ideal' conditions assumed for an efficacy trial, there is evidence in the evaluation report that full understanding of the implications of introducing *Switch on Reading* to a school was not achieved. Two of the schools who initially signed up withdrew before the trial started because of timetabling restrictions. There is evidence in the process evaluation that requirements of both timetabling and suitable space were barriers to implementation. Pupils selected for *Switch on Reading* were usually withdrawn from normal lessons for the daily one- to- one, 20 minute reading sessions which took place over 10 weeks. Timetabling difficulties within the timeframe for the trial meant that few schools managed to run the programme for the full 10 weeks and attendance at sessions was irregular in some cases, with pupils forgetting to excuse themselves from class or teachers unwilling to release pupils. Variation both within and between schools was noted during implementation. However, overall, the judgement of the evaluators was that fidelity to the programme was high, largely because of the extensive monitoring and support provided by the local authority team. The evaluators suggest that differences in quality and adherence to the programme protocol are related to factors such as experience, confidence and enthusiasm of the TAs. Some sessions were affected by resistance and/or poor behaviour of pupils. There were also concerns in some schools about the impact of children being removed from other lessons important for their overall progress and about the project methodology, which was thought by some teachers to be too rigid. There was some confusion with other literacy

interventions already in place in schools and doubts about the accuracy of the pre- test. The evaluation report comments that whole school commitment and motivation are important for success.

Recruitment for *Improving Writing Quality* was undertaken within the Calderdale Education Partnership and, of 24 primary schools approached, 23 agreed to participate, as did all 3 secondary schools approached. Allocation to intervention or control group was at school level for primary schools, with secondary schools arranging timetables so that control and intervention pupils could be kept distinct. This appears to have been successful with no evidence of contamination. Because a large proportion (34%) of pupils at participating primary schools did not transfer to the participating secondary schools, eligible pupils from this group were lost to the trial, reducing the sample size. All schools were retained in the trial and data was received for nearly all pupils who did transfer. Ofsted gradings or other capacity indicators are not indicated for the trial schools. *Improving Writing Quality* offers a structured but flexible approach for teaching writing developed in the USA, called Self- regulated Strategy Development (SRSD). It is designed for use for all pupils in whole- class teaching. In the trial, schools received resources and support in SRSD and funding for 'Memorable Experiences', such as outings or visitors to the school, to provide stimulus and focus for writing activities. Primary school teachers implemented the whole- class element of SRSD in literacy lessons in the summer term of Year 6, as did secondary school teachers in English lessons which included target pupils during the first term of Y7. Thus, the intervention was experienced by all pupils in these teaching groups, although only pupils with attainment below a secure level 4 at key stage 2 were included in post- tests. Following training by the SRSD developer, a working group of participating teachers was established and produced guidelines for schools using the approach in English schools. The evaluation reported that teachers would meet again after the project to develop a toolkit that could be used by other schools.

The *Response to Intervention* trial sought to recruit schools and ask them to identify potentially eligible pupils, that is pupils who were expected to fail to achieve a secure level 4 in English in KS2 SATs tests through teacher assessment and use of the trial pre- test. Randomisation was by school, so that control schools were expected to pre- and post- test pupils and provide pupil data without being part of the actual intervention (although they were offered the opportunity to participate in the following year). Schools approached for the trial were those already working with the developer on other initiatives, but recruitment fell short of the 80 schools expected. Several schools dropped out before the trial began when they realised the timing, the expectations of school staff and the IT requirements of the testing approach, with only 61 beginning the trial and a further 11 schools dropping out after the pre- test and randomisation. The *Response to Intervention* trial took place with eligible Y6 pupils in primary schools in the second half of the summer term. It was introduced to schools through training for school 'champions' and 'achievement coaches' who subsequently shared information with other relevant staff in the school. Use of the intervention is complex, requiring use of a formative assessment tool to determine the literacy areas to focus on for each targeted pupil and the pattern of needs across the whole class. The programme is flexible in length, intensity and frequency. Following the formative assessment, a teacher must then choose from a menu of options to decide which intervention to use to tackle identified weaknesses and whether this is best achieved at whole class, small group or one- to- one level. Other staff working with small groups or one- to- one must be supported and progress monitored using programme tracking sheets, with adjustments made if required. Unsurprisingly, the evaluator notes that 'schools did not feel that they had enough time to carry out the intervention as thoroughly as they would have liked to. There were some reports that the intervention was not given sufficient time to develop and to complete the full cycle from needs analysis to monitoring and adjustments" (pp 19,20). There were also concerns about the accuracy of the assessment tool. Administration and the amount of preparation required (as for the introduction of any new programme) were additional challenges. Errors of communication and misunderstanding of trial requirements prevented data on intervention and control groups from reaching the evaluator in a timely fashion and several schools failed to report full data. The short time scale required by the funders was considered by evaluators to be unsuited to the nature of the intervention. Issues affecting the trial led to the evaluator considering it 'spoilt'.

Outcomes

As with other efficacy trials, exacerbated by lower sample sizes than anticipated, the overall security rating for *Switch on Reading* was modest. Despite reservations, many staff were enthusiastic about the project and reported positive effects on some pupils' reading and confidence. TAs valued the opportunity to improve their own skills. The one- to- one attention provided to pupils was considered to be very important in helping pupils make progress. The trial did not test whether a similar programme of one- to- one support other than that based on *Switch on Reading* was equally or more effective.

The enthusiasm of teachers in both primary and secondary schools for continuing to use the SRSD approach, to use it over a longer period of time and with other year groups is a significant outcome for the *Improving Writing Quality* intervention in addition to the promising quantitative outcomes. Although Y6 teachers would have welcomed a longer timeframe for the trial, *Improving Writing Quality* is unusual in not reporting significant dosage limitations which typically affect trials in the summer term of Y6. The guidelines for using the approach allowed for sufficient flexibility within an overall structure and ordering of activities which enabled both primary and secondary teachers to include SRSD within their planning.

It is striking that despite the numerous problems with implementation, teachers, pupils and achievement coaches for *Response to Intervention* were so positive about the intervention. Pupils responded positively to individualised attention and to activities used in interventions. The development of teaching resources and innovative approaches new to the school was a positive outcome for schools involved. The diagnostic assessment tool was valued by teachers and school-level pupil data showed positive impacts on pupil progress in reading. Many teachers said that they would continue to use the approach with other classes in the future.

Key Points
- When recruiting schools, full support and commitment to the trial is likely to be needed from senior leaders and school project coordinators for successful implementation.
- Efficacy trials based in a small number of schools avoid problems of recruitment and may be more likely to retain schools and pupils for the course of the trial.
- Even in small- scale trials, process evaluations suggest that schools may not be fully prepared for the implications of the trial for timetabling, for rooming, for the preparation time required by delivery teams, for monitoring by senior leaders and for ensuring the support of all school staff, including those who are not directly involved.
- When recruiting schools to efficacy trials, the capacity of the school to introduce a new initiative might be considered. Evaluation reports do not note any links between recruitment strategy, attrition or issues with implementation and capacity indicators, although these might be relevant. For example, schools in Ofsted categories of 'requires improvement' or 'inadequate' are likely to face pressures which might affect leadership support and monitoring.
- Projects which are scheduled for the summer term of Y6 or those intended to begin in primary school and complete in the first term or year of secondary school face capacity barriers which may hinder implementation. Pupils identified as eligible in participating primary schools and who begin in both intervention and control groups transfer to many different secondary schools, which may or may not be part of the trial. The school summer term is particularly busy, with preparation and taking of SATs tests, staff absence for moderation and other training, transition events with secondary schools and school trips and performances. This results in frequent disruption to the normal timetable.
- Any IT requirements of programme delivery or testing might usefully be taken into account at the recruitment stage of a trial.
- Time frames required by the funder for recruitment and implementation of the trial may restrict sample sizes, be insufficient to allow schools to prepare and for impact to be demonstrated. Where a trial requires the acquisition and use of new teaching skills, I time is likely to be required for staff to prepare lessons and practise these skills to ensure high quality.
- Intervention programmes devised and tested in one context are not always sufficiently adapted prior to use in a different context for EEF trial. Working groups of teachers, such as

that formed for the *Improving Writing Quality* intervention*,* might be considered both in advance of efficacy trials and to make further modifications following experience of use in an English context.

- Highly- structured, prescriptive activities and resources linked to intervention approaches may be difficult to accommodate within the diverse timetabling, staffing arrangements and individual pupil needs and interests typical among schools even in small- scale trials. Expectations of the amount of content to be delivered in intervention sessions were sometimes too high. Both efficacy and effectiveness trials might usefully consider highlighting aspects and activities of programmes that are essential for fidelity and those where there might be adjustments for school and pupil circumstances.
- Programmes which required external and/or inexperienced tutors to work pupils appear to have faced particular challenges in ensuring consistently high quality and/or the engagement and good behaviour of pupils.

High security trials

In this section a sample of efficacy and effectiveness trials is considered in relation to the key points identified above. The following trials were found to have high security ratings, high attrition ratings and a positive effect size. No examples occur here of programmes intended for delivery by external staff or by inexperienced school- based tutors (*Parent Academy*, *Chess in Schools* and *Creative Futures: Act, Sing, Play* are examples of high security, nil or negative effect trials which do use these).

Table 18: Trials with high ratings for security and attrition and positive effect.

Project Name	Type	Security Rating	Months Progress	Months Progress for FSM	Attrition Rating
Improving Numeracy and Literacy	Efficacy	5	3	2	5
Nuffield Early Language Intervention	Efficacy	4	4	.	4
Talk for Literacy[2]	Efficacy	4	3	4	5
Catch Up Literacy	Effectiveness	4	2	0	5
Mathematics Mastery: Secondary	Effectiveness	4	1	1	4
Vocabulary Enrichment Intervention	Efficacy	4	1	0	4

Improving Numeracy and Literacy, Mathematics Mastery and the *Vocabulary Enrichment Intervention* all train teachers in strategies targeted at specific areas of improvement. These are then implemented during whole- class teaching in place of normal, timetabled lessons using guidance materials and other resources provided by the developers. *Catch Up Literacy*, for Y6 and Y7, and *Nuffield Early Language Intervention*, for the end of nursery and the beginning of the Reception year, are transition programmes which require withdrawal of low- attaining pupils for small group or one- to- one tuition by trained teaching assistants using materials provided.

The local authority *Vocabulary Enrichment Intervention* programme recruited 12 schools from the 17 approached and these were known to the developer team, although the capacity indicator provided in the evaluation report indicates variability, "Seven of the schools had a *Good* Ofsted rating, four were rated as *Requires improvement* and one as *Outstanding*" (p 15). In contrast, 270 schools were initially approached by the University of Oxford delivery team for *Improving Numeracy and Literacy* but recruitment of 55 schools fell short of the target of 60. Schools did not know until after randomisation if they would trial the numeracy intervention, the literacy intervention or be in the no- intervention group. The evaluation report states: "The main reason given by schools for not participating was that they wanted to choose the intervention that best fitted with their current school development plans, rather than be randomised" (p14). Also significant for recruitment is that the IT

- [2] *Talk for Literacy* has been discussed earlier in this section.

requirements for programme implementation were not readily available at appropriate times in some schools and this was a cause of variation in the fidelity of implementation. 302 schools were approached by the charity I CAN for *Nuffield Early Language Intervention* with 34 recruited and, although details of number approached for the *Catch Up Literacy* trial are not reported, recruitment of 15 schools was fewer than the 17 planned. Capacity of the participating *Catch Up Literacy* schools is indicated by Ofsted ratings, with 11 'good' or 'outstanding', 1 'requires improvement' and 3 'inadequate'. Capacity issues related to the nature of the intervention were cited in the evaluation report for *Nuffield Early Language Intervention* as reasons for declining to participate, "finding the 'right person' in the school, schools being unable to commit to the amount of staff time that would be needed without funding.." (p12) as well as reasons related to schools' current priorities. The Nuffield Early Language Intervention evaluation report is unusual in including the reasons given by schools which <u>did</u> sign up for the trial. These included: match with the school's aspirations and goals; the needs of children; school capacity; and the reputation of the developer (a university). Recruitment for *Mathematics Mastery* was by the developer, ARK. Details were not reported.

As transition trials*, Catch Up Literacy* and *Nuffield Early Language Intervention* were vulnerable to pupil attrition at school transfer. Both trials minimised this possibility at recruitment stage. *Nuffield Early Language Intervention* only recruited primary schools with nurseries attached as part of their provision. *Catch Up Literacy* recruited secondary schools with screening for eligibility requiring that the pupil was scheduled to attend one of these. Secondary school teaching assistants delivering the intervention travelled between primary schools to minimise demands. However, the process evaluation reported comments that a few primary schools were unwilling to be involved and issues with funding of expenses for travel.

Further evidence of the need for senior leadership support arose in evaluation reports of both *Catch Up Literacy* and *Nuffield Early Language Intervention*. As interventions delivered one- to- one by teaching assistants, the support of senior leaders was important for ensuring sufficient protected release time for teaching assistants to prepare, as well as to deliver the intervention, and to arrange for suitable spaces to be available when required. In secondary schools, senior leader support might have reduced the resistance from subject staff when pupils were withdrawn from class.

Issues regarding the amount of content and the prescriptive nature of *Improving Numeracy and Literacy* resulted in high variability in the ways in which different schools and teachers responded to these. In particular, the independent pupil use of computer games, which was an intrinsic part of the programme, was inconsistent among schools as well as among pupils. The high degree of flexibility allowed in the *Vocabulary Enrichment Intervention* led to considerable variation in the way the teachers combined different elements in their lessons, with some reporting that the programme was not dissimilar to their usual approach when teaching similar pupils. Among issues with the materials for this programme, there were comments that the phonics element, originally designed for a younger age group was not suitable for Y7 pupils. Inappropriate targeting for one- to- one or small group support was noted in *Catch Up Literacy* and *Nuffield Early Language Intervention* and in the latter some of the material for nursery children was felt to be unsuitable for their levels of development. However the main issue with the materials in the latter two programmes, as for *Vocabulary Enrichment Intervention*, was the amount of preparation required.

Outcomes
The importance of considering the overall timescales for trials is suggested in the key conclusion of the *Mathematics Mastery: Secondary* evaluation report: "It would be worthwhile to track the medium- and long-term impact of the approach, to assess whether there is a cumulative effect to the approach and whether it has an impact on performance in high-stakes tests" (p 5). The process evaluation goes on to say that "success of the intervention seemed to depend largely on the prior beliefs of the teachers. Just one year into the intervention, we are unlikely to see a significant impact on deep-seated beliefs." (p 47). Although only small gains were noted in the EEF trial, it has been argued that as an approach intended for use throughout primary and secondary education, the length of the trial was too short for impact to be realised. The *Mathematics Mastery* approach has since been recommended by the National Centre for Excellence in the Teaching of Mathematics (NCETM) and has been adopted in many schools.

The improvement gains in table 18 for the three- way intervention *Improving Numeracy and Literacy* relate to the numeracy intervention, with no effect found for literacy. Teachers involved in the *Vocabulary Enrichment Intervention* noted improvements in pupils' vocabulary and felt that the test used did not allow them to demonstrate this improvement.

Most of these programmes were, on the whole, positively received by teachers and pupils. Reaction was more mixed in the *Mathematics Mastery* trial, particularly when a significant change in pedagogy was needed from individual teachers. Teachers involved in the whole- class, *Improving Literacy and Numeracy in Key Stage 1* and *Vocabulary Enrichment Intervention* anticipated adapting and adopting at least some aspects of the programme in future. Gains in pupil confidence and verbal skills were noted in the targeted interventions, although decisions about future use were likely to be mediated by factors such as comparison with other intervention approaches, pupil needs and consideration of the high costs of the programmes.

Other trials and factors. Key points are presented in section 4.3.

Common characteristics of programmes that have been easy or difficult to implement. Differences between targeted, whole- class and whole- school programmes

We found that the mechanism used for the programme, as a targeted or whole- class intervention, was closely associated with ease of implementation and have considered these factors together.

Accelerated Reader, Talk for Literacy, Perry Beeches Graduate Coaching Programme and *Improving Writing Quality* were all school group projects, where a group of schools acted as the developer and thus are similar to 'real world' conditions (*Accelerated Reader* report, p5). *Fresh Start* was another efficacy trial of this type which was successful in improving outcomes for pupils who completed the programme. *Fresh Start*, based on phonics instruction, was an intervention for small Y7 groups withdrawn from regular English lessons and, although some teachers found the approach too prescriptive, no significant issues were raised in its implementation in the schools concerned, despite its high intensity.

For *Improving Writing Quality*, *Fresh Start* and *Mathematics Mastery* only English or mathematics lessons were affected by the implementation as was the well received, although lower impact, *Vocabulary Enrichment Intervention*. Although not successful in terms of pupil outcomes, *Chess in Schools* and *Creative Futures: Act, Sing, Play* provide examples of effectiveness trials which were not difficult to introduce as part of the normal timetable and which resulted in trials with high security ratings. However both *Chess in Schools* and *Creative Futures*: *Act, Sing, Play* provide further illustration that interventions which rely on the recruitment of tutors who lack school experience and expertise are likely to experience problems in consistency of quality and of issues with pupil behaviour.

Programmes which require pupils to be withdrawn from their normal lessons for small group work or individual support are more problematic in terms of practical issues arising from implementation. Suitable spaces may not be made available and, particularly when small group or one- to- one sessions are led by TAs, senior leaders may be unwilling to provide sufficient release time for preparation. In the secondary phase, where many teachers are involved and where pupils are often withdrawn from lessons across the curriculum, pupils may feel stigmatised by being taken away from their peers, resent missing all or part of lessons they enjoy and miss core learning in other subjects. Attempts by the school to timetable intensive interventions and to spread withdrawal sessions across many subjects results in pupils forgetting to attend and/or needing to be collected for the intervention session. Subject staff may resist pupils being extracted from their lessons. As well as being issues in *Perry Beeches Graduate Coaching Programme*, *Switch on Reading* and *Nuffield Early Language Intervention*, interventions requiring withdrawal caused similar problems for other efficacy trials *Butterfly Phonics*, *Chatterbooks* (both of these had problems which were further compounded by use of external tutors), *Talk for Literacy* and in both effectiveness trials *Catch Up Literacy* and *Catch Up Numeracy.* In some trials, staff felt that identification of pupils for targeting by

the intervention was inaccurate, as in *Nuffield Early Language Intervention*, which provides one- to-one speech and language activities.

Process evaluation reports note that the availability of external support is a factor which appears to aid implementation. For example, this was reported as being the reason for high fidelity in the *Catch Up Literacy* intervention. The *Mathematics Mastery* programme includes an ongoing programme of professional development led by the developer and membership of a support network. For both *Fresh Start* and *Accelerated Reader*, where schools led on evaluation, the support of an external evaluator was considered to be vital. Support for implementation was lacking in the *Perry Beeches Graduate Coaching Programme* which reported high variability and quality. The evaluation report for the high security trial of *Improving Numeracy and Literacy* is one of a small number that suggests options for replacing the face to face training and in school support provided by the developer team in any wider roll- out of the programme.

Key points

- Programmes developed by groups of schools appear to be relatively successful (also noted in the quantitative evaluation in relation to quality outcomes). Recruitment tends to be less problematic than observed in other types of trials and retention of schools is high.
- Where there are problems of implementation these often appear to be linked to a lack of shared understanding among senior leaders and teachers of what is involved. The amount of preparation required for introducing the interventions is a common issue that occurs across all programmes.
- In some trials, staff felt that the assessment used to identify pupils for targeted interventions did not identify those who would benefit most.
- Programmes which require pupils to be withdrawn from subjects across the curriculum are more likely to experience resistance from pupils and these subject staff. Senior leader support is required for ensuring suitable spaces, preparation time for TAs and timetabling arrangements. Programmes for use with a whole class are easier to implement.
- External support from the delivery team facilitates implementation. It is a feature of nearly all efficacy trials and may be more available in effectiveness trials than in a 'real life' situation.

Intensity of the programme and its bearing on implementation and success

In line with discussion in the quantitative report, we consider programme intensity measured by the number of minutes per week and also its length, measured by the number of weeks on the programme.

Table 19: In- school programmes requiring 50 or more minutes per week.

Project Name	Type	Phase	Focus	Intervention Length (Weeks)	Intervention Intensity (minutes per week)
Improving Literacy and Numeracy in Key Stage 1	Efficacy	Primary	Numeracy	12	60
Vocabulary Enrichment Intervention	Efficacy	Secondary	Literacy	17	220
Butterfly Phonics	Efficacy	Secondary	Literacy	15	80
Chatterbooks	Efficacy	Secondary	Literacy	9	50
Graduate Coaching Programme	Efficacy	Secondary	Literacy	33	150
LIT Programme	Efficacy	Secondary	Literacy	26	180
Rapid Phonics	Efficacy	Transition	Literacy	12	90
REACH	Efficacy	Secondary	Literacy	20	105

Tutoring with Alphie	Efficacy	Primary	Literacy	6	150
Units of Sound	Effectiveness	Secondary	Literacy	18	90
TextNow Transition	Effectiveness	Transition	Literacy	15	100
Physically Active Lessons	Pilot	Primary	Domain-general	20	90
The Visible Classroom	Pilot	Primary	Domain-general	26	120
Word and World Reading Programme	Pilot	Primary	Literacy	37	90

Some in-school programmes required an intensity of 50 or more minutes per week and, as such, required significant timetable modification for schools in the trial. Evaluation reports for these projects show that, apart from the pilot programmes, the school- group *Graduate Coaching Programme* and the Bolton LA *Vocabulary Enrichment Intervention*, there were problems with recruitment of schools and/or drop out of schools from the programme when full implications for timetabling, staffing and resourcing were understood or experienced. In programmes which also required withdrawal for small group or one- to- one work, schools which remained engaged in the programme experienced implementation problems as discussed above. For programmes heavily dependent on the use of IT, such as *Units of Sound* and *Tutoring with Alphie*, IT problems were cited as the principal cause of drop out. The relative success of the whole- class *Improving Numeracy and Literacy* and *Vocabulary Enrichment Programme* suggests that it is the mode of delivery for the programme and/or the programme content rather than the programme intensity which is significant. Recommendations from high intensity pilot trials were generally found to be for adaptations other than that of programme intensity.

Other programmes with high intensity in minutes per week are a number of out- of- school learning projects including summer schools (*Discover Summer School*, *Future Foundations*) parent classes (*Parent Academy*, *SPOKES*) a Saturday school (*Halle Shine*) and after- school activity (*Youth Social Action*) all of which experienced significant implementation problems. In the case of *Youth Social Action*, implementation issues were linked to school capacity to support the initiative and for the other interventions low attendance and high drop out of participants compromised findings from the trial. Evaluation reports suggest that implementation problems were more closely related to factors other than the intensity of the programme.

We then considered a sample of programmes with promising pupil outcomes of 30 or more weeks in duration, but where intensity per week is less than an hour per week.

Table 20: Programmes with promising pupil outcomes of 30 or more weeks

Project Name	Type	Security Rating	Months Progress	Months Progress for FSM	Phase	Focus	Intervention Length (Weeks)	Intervention Intensity (minutes per week)
Catch Up Literacy	Effectiveness	4	2	0	Transition	Literacy	30	30
Catch Up Numeracy	Effectiveness	3	3	.	Primary	Numeracy	30	30
Improving Writing Quality	Efficacy	2	9	18	Transition	Literacy	31	N/A
Nuffield Early Language Intervention	Efficacy	4	4	.	Primary	Literacy	30	30
Thinking, Doing, Talking Science	Efficacy	3	3	5	Primary	Science	37	.

Catch Up Literacy, *Catch Up Numeracy* and *Nuffield Early Language Intervention* required withdrawal for one- to- one tuition by trained teaching assistants. In the primary phase, issues related to withdrawal from a range of curriculum subjects did not arise, although there were issues of lack of suitable space and the provision of release time for teaching assistants to prepare for the one- to- one sessions in adherence with the programme. The *Catch Up Numeracy* evaluation

compared pupil outcomes to a 'time equivalent' intervention group, who received the same amount of one- to- one teaching by TAs, but did not use *Catch Up Numeracy* and concluded that "the effect is likely to be a result of regular and sustained one- to- one teaching, rather than an intrinsic benefit of Catch Up Numeracy" (p 5). However, it does not comment on the preparation time or delivery methods in the alternative tuition programme.

Like *Improving Writing Quality*, *Thinking, Doing, Talking Science* is designed for flexible use in normal whole- class science lessons by teachers over the course of the school year and process evaluation outcomes similarly suggest that the approach is welcomed by teachers, experiences few problems of implementation and has a positive impact on practice with use over the whole year developing confidence and familiarity. Lengthy projects also include several programmes that were general in focus, for which academic attainment gains in a relatively short period may be considered unlikely, such as *Philosophy for Children* and *Promoting Alternative Thinking Strategies.* For all of the long- term interventions, the length of the programme does not seem to have been a contributing factor to its success or otherwise. However, it would appear that programmes which require teachers to change their pedagogy require a long term approach. Evaluation reports for one- to- one tuition, as with the *Perry Beeches Graduate Coaching Programme*, suggest that success is dependent on the skills and quality of the tuition and the engagement of the pupils developed through relationships, which perhaps may require time to be developed. However, shorter one- to- one and small group tuition programmes show similarly promising outcomes, suggesting that further research might be needed in this area.

Key points
- Intensive interventions requiring significant timetable adjustments find it more difficult to recruit and retain schools for the trial.
- Intensive interventions requiring attendance by pupils or parents out- of- school hours experience problems with attendance and attrition, with little evidence for impact on attainment.
- Interventions which require a change in teacher's whole- class pedagogy may benefit from a long term programme.
- The success of one- to- one tuition appears to be related to the skill and quality of tuition, but may also be linked to sustaining the programme over time.

The success of domain- general versus curriculum- specific programmes

Table 21: Efficacy and effectiveness trials for General programmes

Project Name	Type	Security Rating	Months Progress	Months Progress for FSM	Attrition Rating	Intervention Length (Weeks)
Act, Sing, Play	Effectiveness	4	0	.	4	45
Affordable Individual and Small Group Tuition: Primary	Efficacy	0	-1	-1	0	39
Affordable Individual and Small Group Tuition: Secondary	Efficacy	0	0	0	3	111
Changing Mindsets	Efficacy	2	2	2	2	12
Chess in Schools	Effectiveness	5	0	0	5	30
Engaging Parents through Text Messaging	Efficacy	3	0	-1	3	30
Hallé SHINE on Manchester	Efficacy	.	0	-1	5	24
Hampshire Hundreds	Efficacy	2	0	1	2	26
Increasing Pupil Motivation	Effectiveness	2	1	1	5	33
Mind the Gap	Effectiveness	1	-2	-4	0	24

Parent Academy	Effectiveness	4	0	0	2	13
Philosophy for Children	Effectiveness	3	1	3	5	39
Promoting Alternative Thinking Strategies	Effectiveness	4	-1	0	5	35
Youth Social Action: Secondary	Effectiveness	3	-1	-3	5	24

Domain- general programmes are intended to increase pupil attainment through an intermediate outcome, for example, increased confidence or increased parental engagement. A direct consequence of this is that any potential attainment gains are less likely to be measurable in the timeframe for the project. The tracking of pupils to measure long term gains is frequently noted in recommendations made in evaluation reports. However, it should be noted that EEF trials, through their focus on academic attainment, may provide a valuable contribution to an evaluation programme which considers other outcomes over a longer term, for example as for *Philosophy for Children* and *Promoting Alternative Thinking Strategies*.

We focused on a small sample of general programmes to provide comparison with points made in earlier sections. Only three of the domain- general trials, *Changing Mindsets*, *Philosophy for Children* and *Increasing Pupil Motivation* found any evidence of attainment gains. The last of these is not discussed because the mechanism, of offering pupil rewards for increased effort, is not comparable to any other programme.

Attainment gains in the Portsmouth LA *Changing Mindsets* appear more remarkable in light of the short duration of the project. Overall security of the trial was low, largely due to high attrition. However, further examination found that the reasons for high attrition were related to points made earlier and were not related to the general focus of the intervention. There were two versions of the intervention programme, one of which used pupil randomisation within each school to identify intervention and control groups. In this version, a series of workshops for pupils were led by trained university students, using materials for the intervention pupils based on Dweck's growth mindset (Dweck, 1999) and the control group receiving general study skills workshops. One of the six schools involved dropped out because of dissatisfaction with the behaviour management strategies of the university students. The other version used school- level randomisation and recruited 30 schools. Two of the 15 intervention schools and four of the 15 control schools dropped out before the final test, and when reasons were provided, these were related to capacity issues or other priorities. This version provided training for teachers in the intervention schools, who could use this at their discretion in the course of their normal teaching. The effect was measured by English and mathematics tests of pupils ten months after the intervention, thus allowing, at least to some extent, for the time required for intermediate outcomes to influence pupil learning.

Philosophy for Children, in which Y4/5 children received teaching following the approach over the course of a year was measured in relation to key stage 2 results, thus allowing for the intended intermediate outcomes (being more confident in asking questions or discussion and improved reasoning) to be assimilated. The evaluation report comments that a year of implementation "may not have been long enough for the full impact of P4C to be felt" (p 4) and further suggests that assessment material is provided for measuring the programme's intended outcomes. Schools could choose to incorporate the *Philosophy for Children* approach within the normal curriculum or to use dedicated lessons. Although it was suggested that one lesson of one hour a week would be needed, teachers were given flexibility in use of guidance and materials. The evaluation report notes the care that was taken when recruiting schools, "it was felt that it was more important to have schools that were committed to the trial than to have a few more but run a real risk of school dropout" (p12) and it is thus perhaps surprising to note that the main barrier to introduction was reported as being its lack of direct relevance to literacy or numeracy targets. However, 14 out of 22 schools in the trial were judged by developers to be implementing the programme satisfactorily or well with no schools dropping out. Where there were issues, these were reported as being linked to the context issue of capacity, including changes in school or project leadership, changes in teaching staff and changes in priorities related to poor Ofsted outcomes.

The lack of effect of some other programmes would seem to be related to factors other than the length of the programme. *Creative Futures: Act, Sing, Play* and *Chess in Schools* were provided through the use of external tutors and took place over the course of a school year or longer. Despite concerns about variability in quality of external tutors, schools and pupils were positive about the interventions and reported impact on outcomes such as confidence and a broadening of opportunities for disadvantaged pupils. However, teachers were concerned about the amount of time taken from the statutory curriculum for these programmes. *Parent Academy*, *Halle SHINE* and *Youth Social Action* involved out- of- school learning with implementation issues discussed earlier.

The quantitative review found that literacy programmes gave a larger effect size, on average. No reason for this could be found when we compared programmes and examined those which used similar approaches. The peer coaching approach *Durham Shared Maths Project* had zero effect, whereas the equivalent literacy programme *Paired Reading* had a small negative effect. In the whole- class, *Improving Numeracy and Literacy* trial, the positive effect was for numeracy only, with nil effect for the literacy intervention. The two one- to- one tuition *Catch Up* trials produced similar results. The two science trials were both intended to lead to long- term, fundamental, change in teachers' pedagogy. As with *Mathematics Mastery*, the effect may be more likely to be demonstrated over a longer time frame than the trial allowed. It would appear that the higher mean effect seen in literacy trials is unrelated to the curriculum focus.

Key points
- The measure of pupil academic attainment and the timeframe of trials may not be appropriate for domain- general programmes.
- Domain- general programmes may have to compete for space in the curriculum.

Models of professional development and training associated with more successful programmes

The majority of comments made in process evaluations about the quality of **pre- programme training** are positive, both with regard to the quality of delivery and the response of trainees, whether teachers or teaching assistants. In- school coordinators may receive training together with staff responsible for delivery, as in *Catch Up Literacy* and *Catch Up Numeracy*. In most cases, training is provided face to face, although a small number of programmes use on- line training modules. The number of these is too low to form conclusions about the relative effectiveness of online versus face to face pre- programme training. However, one of the barriers to implementation reported in the low security and negative effect *Units of Sound* effectiveness trial was that many teachers failed to complete the lengthy on- line training modules (this was exacerbated by in- school IT issues). At training events, participants were introduced to the supporting materials, such as handbooks and resources. In most cases, participants are those who are directly responsible for programme delivery to pupils, although in some cases, a 'cascade' model is used, as in the programme discussed earlier, *Response to Intervention*, such that participants must return to their schools and train other members of staff who will be involved in delivery. As this example shows, the lead- in time for the programme may be insufficient for this and might be considered carefully when planning programmes which rely on this model.

On the whole, participants felt that training had prepared them well for implementation, but this was not always the case and was noted in some less successful trials. For example, in the low security trial *REACH*, a programme of one- to- one literacy tuition, some TAs who were expected to deliver the intervention reported feeling overwhelmed by the quantity of material and poorly prepared. However it is not clear whether this is primarily related to the quality of the training or the lack of background knowledge of the TAs involved as the evaluation report also notes that implementation was of higher quality with more experienced TAs. The manual was found to be unhelpful. Because this trial failed to recruit enough schools initially, it took place in two phases and it may be revealing that fewer problems were reported in the second phase, with the developers learning from the first implementation. This suggests that a pilot trial might have been helpful. In the effectiveness trial *Let's Think Secondary Science* the training programme for teachers included inter- session tasks and collaborative preparation time, but time allowed was rarely sufficient for this.

Following initial training, nearly all efficacy programmes provide **additional support** from the developer, through a programme of school visits and/or support on request. The extent to which additional support is important to the programme varies. For example, in *Switch On Reading* each school was assigned a trainer who visited the school regularly to provide extra support to members of staff and frequent online communication. In *Nuffield Early Language Intervention*, a single visit was found sufficient in most schools. The extent to which support might realistically be provided at an affordable cost in extensions to the programme is commented on in some process evaluations, for example, the *Improving Numeracy and Literacy* report.

Problems with support may be linked to school capacity issues. The *Units of Sound* process evaluation reports that, "Despite clear lines of communication set up between school staff … (the developer) and the group running the online tests, schools often struggled to communicate with the appropriate organisation to resolve their particular problems, such as accessing the training, running the programme, or administering the online assessments" (p21) and that the developer found it extremely difficult to schedule the offered support visits to schools. For the effectiveness trial of *Let's Think Secondary Science* the quality monitoring and support expected from some senior leadership teams was not provided in many schools, although this only seems to have affected the fidelity of implementation in a minority of classes.

There were frequent comments about the **quality of resources** and this was the case in successful as well as less successful programmes. In *Improving Numeracy and Literacy*, teachers complained about basic errors in materials, including in the computer games that were an integral part of the programme. In this programme, teachers also complained about the heavy reliance on worksheets, which do not match their usual teaching practice and that materials did not allow for sufficient differentiation. Materials prepared for use in another context may not be suitable. This applies to those developed for use in a different cultural context, as for the U.S. originated *Improving Writing Quality*, or for a different age group, as in the phonics strand of *Vocabulary Intervention Programme.*

However, the most common issues reported were, firstly, that guidance on prescribed session content was unrealistic with too much expected for the time allotted. Secondly, the time required for preparation was too high and was expected in addition to staff's normal workload. Both of these were reported by TAs (as in *Talk for Literac)y* and by teachers (as in *Improving Numeracy and Literacy)..*

Lack of sufficient preparation time was reported even when the resources provided were flexible, as in the *Vocabulary Enrichment Programme*. However, it is not clear whether the predominant influencing factor in high preparation time is unfamiliarity with materials at first use or inherent in the programme itself. For example, the evaluation report for *Philosophy for Children* noted that "Teachers reported that additional time was necessary for them to prepare P4C lessons, although it was not clear whether this was simply the extra time involved because P4C was new to them" (p4). Research evidence has found that there is typically a gap between professional development for a new initiative and its smooth implementation in everyday practice (Hall and Hord, 2011)

With the exception of the small- group literacy intervention *Rapid Phonics*, where experienced, specialist teachers travelled between schools to deliver the programme, process evaluations for all programmes which involved external staff coming into school reported variability in the quality of delivery, due to some tutors' inexperience of working with school- age children. In the case of *Chess in Schools*, chess tutors were unfamiliar with working with whole classes and lacked differentiation strategies. In *Perry Beeches Graduate Coaching Programme,* insufficient training and support in some cases meant that tutors did not have the knowledge and skills to cater for learning needs of tutees. In all programmes which relied on external or recently appointed staff without school experience, there were reports of behaviour issues and recommendations that this should be considered in the training programme.

Key points

- To achieve ideal conditions in efficacy trials, the preparation time required for introducing any new initiative into a school should be considered and costed into the trial.
- Lack of time to assimilate training and prepare for implementation may be more significant than the model of training prior to the programme.
- The quality and relevance of programme guidelines, protocols and teaching resources is variable. The amount of expected content is frequently found to be too high for times allotted for programme delivery. Resources provided are rarely ideal for direct use with pupils, and considerable time may be needed to prepare additional material and make adaptations. In some cases, resources did not allow for sufficient differentiation. Some resources, such as pupil worksheets, are not well- matched to current expectations of good teaching practice.
- External or inexperienced tutors need training which enables them to use subject or academic expertise with the pupils in the trial. Those who have not previously worked in schools need training in behaviour management.
- Training that relies on cascading within the school may require additional time and resources.
- The extent to which the contribution of external monitoring and support provided in efficacy trials is essential to the success of the programme might be taken into account when planning further trials.

Differences between programmes delivered by charities, schools and universities.
Differences between projects in the early years, primary and secondary phases.

We found the deliverer or the phase is less likely to be related to the success of the programme than factors considered earlier. Qualitative analysis supports the finding from the quantitative analysis that school group trials are more likely to be successful, with noticeably fewer issues about recruitment or retention of schools. This group of trials includes the two, *Accelerated Reader* and *Fresh Start*, where schools independently bid for funding and where commitment of schools is pre-determined. In contrast, the senior leaders in schools that withdrew from the efficacy trial of the university- developed literacy programme *REACH* stated that in early stages they had not engaged sufficiently with the information and material and had failed to realise the complexity of what was involved.

It may also be more likely that materials for use in school group trials being trialled are school ready, with *Mathematics Mastery* being trialled in the ARK group of schools and the *Improving Writing Quality* trial having a working group of teachers supporting implementation. The quantitative analysis notes than trials delivered by a charity had lower security, on average, than other trials. It is noticeable that programmes delivered by charities are more likely than other programmes to involve external tutors working with school pupils and to be general than curriculum specific.

School group trials

Project Name	Type	Security Rating	Months Progress	Months Progress for FSM	Rating Limiting Factor	Phase	Focus	Level of Randomisation	Intervention Length (Weeks)	Intensity (min) per week)	Training	Targeted/Whole Year/Whole School
Accelerated Reader	Efficacy	3	3	5	Power	Secondary	Literacy	Pupil	22	N/A	Teachers	Targeted
Improving Writing Quality	Efficacy	2	9	18	Power	Transition	Literacy	School	31	N/A	Teachers	Targeted
Graduate Coaching Programme	Efficacy	3	5	5	Attrition	Secondary	Literacy	Pupil	33	150	None	Targeted
Talk for Literacy	Efficacy	4	3	4	Power	Secondary	Literacy	Pupil	23	115	TAs	Year

Fresh Start	Efficacy	3	3	3	Power	Secondary	Literacy	Pupil	22	180	Teachers	Targeted
Mathematics Mastery: Primary	Effectiveness	3	2	.	Attrition	Primary	Numeracy	School	37	N/A	Teachers	Year
Mathematics Mastery: Secondary	Effectiveness	4	1	1	Balance	Secondary	Numeracy	School	37	N/A	Teachers	Year

Only one of the efficacy or effectiveness trials was in an Early Years context, with the efficacy trial for *Nuffield Early Language Intervention* run in primary schools with attached nurseries, with no nursery schools or private providers. Quantitative analysis indicated that trials in the transition phase between primary and secondary school are least likely to be successful and reasons for this have been discussed in section 4.2. Qualitative analysis indicates that the success of the programme is more likely to be related to factors other than the phase for delivery. A reason for the higher success rate of secondary- phase trials, identified in quantitative analysis, is likely to be that a higher proportion of primary- phase trials were general, rather than curriculum specific, and that these included the two parenting interventions and one of the summer schools, which experienced significant problems.

4.4   Additional themes identified in qualitative review

Some evaluation reports refer to sustainability and this is an important school outcome indicative of the success of a programme. In *Nuffield Early Language Intervention*, senior leader and teaching assistant responses on sustainability were linked to:
   "• the perceived need for the programme in the school—including whether other language interventions were already running;
   • the perceived or anticipated impact of the programme;
   • experiences of delivering the programme—including the use the resources accompanying the programme;
   • the eventual financial costs of the programme; and
   • the control schools would have in selecting pupils for the programme after the trial" (p 39).

Other process evaluations ask delivery staff for their views about sustainability but the views of senior leaders are less likely to be included. As the following comment in the *Creative Futures: Act, Sing, Play* evaluation notes, the staff responsible for delivery are not the decision makers in the school.

   "Some teachers described how the aspirations they held for continuing the programme were constrained by practical and financial considerations. The decisions about the sustainability of the programme were discussed by class teachers but the ultimate decision about further delivery of the Act, Sing, Play programme would be made at a school level by headteachers".(p40)

In the current policy climate in England, external accountability through examinations and to Ofsted has a strong influence on schools' priorities and choices. Successful schools may be unwilling to make long- term, radical changes in teaching and intervention approaches which have proven successful for their pupils, with less successful schools unwilling to take risks and facing lack of capacity for managing the requirements of a trial. Time pressures may make schools reluctant to engage or continue with programmes which are not directly related to examination targets (as for example in *Philosophy for Children*).

Comments in relation to some programmes indicate that they enhanced, rather than competed, with schools' usual approaches and this seems to have contributed to comments in relation to programme sustainability. In the *Response to Intervention* process evaluation:

"Most teachers said they would definitely continue with the intervention the following year… The head teacher of School B was so confident in the intervention, having seen the results it had produced among their Year 6 pupils, that she said she would roll out part of RTI to other years " (pp 35).

Several evaluation reports note an intention, or make a recommendation, to track intervention pupils to assessment at the end of their current phase of schooling (for example, the *Thinking, Doing, Talking Science*), with the aim of identifying if interventions may show impact in the longer, rather than the short term. What does not seem to be considered is the importance of identifying if individualised, expensive interventions intended to help low prior attainers catch up with their peers are sustained when they return to the whole class, or if they again fall behind and require further intervention. Schools need to know if the high investment made in one- to- one or small- group, short- term interventions will have a lasting effect.

Key points
Schools base decisions on continuing with an intervention on:
- The fit with school improvement priorities, current approaches and the needs of pupils.
- Perceived impact when compared with other interventions in place in the school.
- Ease of implementation.
- Control on the selection of pupils for the intervention.
- Quality of materials.
- Cost
- Curriculum time for programmes that are not directly linked to accountability targets.

References

Centre for Effective Education, Queens' University Belfast (2015) *Tutoring with Alphie: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Centre for Effective Education, Queen's University, Belfast (2014) *Quest: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Clay, M.M. (1985) *The early detection of reading difficulties: a diagnostic survey with recovery procedures.* 3rd edition, Auckland, New Zealand, Heinemann.

Dweck, C.S., (1999) Caution--Praise Can Be Dangerous. *American Educator*,*23*(1), pp.4-9.

Gorard, S. (2016) *Youth Social Action: Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Gorard, S., See, B.H. and Siddiqui, N. and See, B.H. (2014) *Switch On Reading: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Gorard, S., Siddiqui, N. and See, B.H. (2014) *Response to Intervention: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Gorard, S., Siddiqui, N. and See, B.H. (2014) Future Foundations*: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Gorard, S., Siddiqui, N. and See, B.H. (2015) *Accelerated Reader: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Gorard, S., Siddiqui, N. and See, B.H. (2015) *Fresh Start: Evaluation Report and Executive Summary.* Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Gorard, S., Siddiqui, N. and See, B.H. (2015) *Philosophy for Children: Report and Executive Summary.* Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Hall, G.E. and Hord, S.M., 2011. Learning builds the bridge between research and practice. *Standards for Professional Learning*, *32*(4), pp.52-57.

Hanley, P., Bohnke, J.R., Slavin, B., Elliott, L. and Croudace, T. (2016) *Let's Think Secondary Science: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Hanley, P., Slavin, R. and Elliott, L. (2015) *Thinking, Doing, Talking Science: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Hayward, S., Griggs, J., Lloyd, C., Morris, S., Kiss, Z. and Skipp, A. (2015) *Creative Futures: Act, Sing, Play: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Husain, F., Jabin, N. Haywood, S.. Kasim, A., Paylor, J. (2016) *Parent Academy: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Humphrey,N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (n.d.) *Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook.* Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Humphrey,N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (n.d.) *Implementation and process evaluation (IPE) for interventions in education settings: A synthesis of the literature.* Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Jerrim, J., Austerberry, H., Crisan, C., Ingold, A., Morgan, C., Pratt, D., Smith, C., Wiggins, M. (2015) *Mathematics Mastery: Secondary: Evaluation Report*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Jerrim, J., Macmillan, L., Micklewright, J., Sawtell, M., and Wiggins, M. (2016) *Chess in Schools: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

King, B. and Kasim, A. (2015) *Rapid Phonics: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Lloyd, C., Edovald, T., Kiss, Z., Morris, S., Skipp, A. and Ahmed, H. (2015) *Paired Reading: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Lloyd, C., Edovald, T., Morris, S., Kiss, Z., Skipp, A. and Haywood, S. (2015) *Durham Shared Maths Project: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Lord, P., Bradshaw, S., Stevens, E. and Styles, B. (2015) *Perry Beeches Graduate Coaching Programme: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Manchester Institute of Education, University of Manchester PATHS to Success Research Team (2015). *Promoting Alternative Thinking Strategies (PATHS): Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

McNally, S. (2014) *Hampshire Hundreds: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Merrell, C. and Kasim, A. (2015) *Butterfly Phonics: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Menzies, V., Kasim, A.. Kokosatsi, D., Hewitt, C., Akhter, N., Collyer, C., Younger, K., Wiggins, A., Torgerson, C. (2016) Halle SHINE*: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

National Foundation for Educational Research (NFER)(2014) *Catch Up Numeracy: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Pawson, R. and Tilley, N., (1997) *An introduction to scientific realist evaluation*. London, Sage Publications.

Rienzo, C., Rolfe, H. and Wilkinson, D. (2015) *Changing Mindsets: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Rutt, S. (2015) *Catch Up Literacy: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Sibieta, L. (2016) *REACH: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Sibieta, L., Greaves, E. and Sianesi, B. (2014) *Increasing Pupil Motivation: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Sibieta, L., Kotecha, M.and Skipp, A. (2016) *Nuffield Early Language Intervention: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Stoll, L., 2015. *Three greats for a self-improving school system: pedagogy, professional development and leadership: teaching schools R&D network national themes project 2012-14: Spring 2015*. Nottingham, National College for Teaching and Leadership.

Styles, B. and Bradshaw, S. (2015) *Talk for Literacy: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Styles, B., Clarkson, R. and Fowler, K. (2014) *Chatterbooks: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Styles, B., Clarkson, R. and Fowler, K. (2014) *Rhythm for Reading: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Styles, B., Stevens, E., Bradshaw, S.and Clarkson, R. (2014) *Vocabulary Enrichment Programme: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Torgerson, D. and Torgerson, C. (2014) *Improving Writing Quality: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Torgerson, D. and Torgerson, C. (2014) *Discover Summer School: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Tracey, L., Chambers, B., Bywater, T.and Elliott, L. (2016) *SPOKES: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk

Worth, J., Sizmur, J., Ager, R. and Styles, B. (2015) *Improving Numeracy and Literacy: Evaluation Report and Executive Summary*. Education Endowment Foundation available at www.educationendowmentfoundation.org.uk