



Review: Scale-up of EEF efficacy trials to effectiveness trials

August 2021

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

Hugues Lortie-Forgues (University of York)



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

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

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

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

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

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

- 📍 Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP
- ☎ 0207 802 1653
- ✉ jonathan.kay@eefoundation.org.uk
- 🌐 www.educationendowmentfoundation.org.uk

Contents

About the reviewer	4
Acknowledgements	4
Executive summary	5
Aims and methodology	5
Findings	5
Implications.....	6
Introduction	7
Background and context.....	7
Aims and objectives	7
Research questions	7
Methodology.....	7
Ethics and data protection.....	8
Project team	8
Key findings	10
Primary outcome effect sizes.....	10
The nature of ‘scale-up’	11
Cost-effectiveness.....	13
Structures for supporting scale-up.....	13
Programme features perceived to support high-quality implementation at scale	14
Conclusion	20
References	22
Appendix A: Information sheet for scale-up telephone interviews	23

List of tables

Table 1: Core Sheffield Hallam University project team for the scale-up review	8
Table 2: Efficacy and effectiveness trials of scaled-up interventions—effect size, phase, and outcome measurement	10
Table 3: Seven scale-up interventions—number of schools and pupils for efficacy and effectiveness trials	11

About the reviewer

The 2019/2020 Review of EEF Projects was conducted by a team from the Sheffield Institute of Education, including Sean Demack, Bronwen Maxwell, Mike Coldwell, Anna Stevens, Claire Wolstenholme, Sarah Reaney-Wood, and Bernadette Stiell, together with a senior statistician from the University of York, Hugues Lortie-Forgues. This is one of three publications from the review. The lead evaluator for this report was Bronwen Maxwell.

The Sheffield Institute of Education (SloE) at Sheffield Hallam University is a leading provider of initial and continuing teacher education—undergraduate, post-graduate, and doctoral education programmes—and has a long-established track record in educational research, evaluation, and knowledge exchange. Key areas of research and evaluation expertise span curriculum and pedagogy, policy and professional learning, and diversity and social justice. The SloE has extensive experience of evaluation methodologies and undertaking large-scale evaluations for a range of funders, including the Education Endowment Foundation (EEF).

Contact details

Principle investigators: Sean Demack and Bronwen Maxwell

Sheffield Institute of Education Research and Knowledge Exchange Centre

Sheffield Hallam University

Room 10101, Arundel Building

City Campus

Howard St

Sheffield

S1 1WB

Tel: 0114 225 6066

Email: SloECDARE@shu.ac.uk

Acknowledgements

Lisa Clarkson, Lyn Pickerel, Rosie Smith, Noreen Drury, Charlotte Rowley, Hannah Joyce, Kellie Cook, Coding team

Executive summary

Aims and methodology

The aim of this review was to support the EEF, other funders, and developers and deliverers scaling-up interventions by undertaking, for the seven effectiveness trials that had been completed by 2019:

- a quantitative comparison of effect sizes at the efficacy and effectiveness stages; and
- a qualitative analysis of interviews with developers and deliverers who have been engaged in scale-up from an EEF efficacy trial to an effectiveness trial, to identify any models, supporting structures, and programme features that either support or hinder intervention scale-up to large numbers of schools.

Findings

(a) Factors affecting reported effect sizes

The mean unweighted effect size for the primary outcomes of the seven interventions analysed reduced from 0.25 for the efficacy trials to 0.01 for the effectiveness trials. Data and analyses from the EEF Review of Projects (Demack et al., 2021) indicate that some of this reduction may be explained by variation between the efficacy and effectiveness trials in the trial design, target group, size of study, or the number of geographical locations in which the intervention was delivered. Two of the trials were designed with clustered randomisation at the effectiveness stage and individual-level randomisation (associated with higher effect sizes) at the efficacy stage, two of the efficacy trials were transition projects (associated with higher effect sizes) while all of the effectiveness trials were based in primary schools, and the effectiveness trials were larger studies and more likely to be set in one geographical location (both associated with lower effect sizes).

(b) Factors that facilitate implementation at scale

Analysis of the qualitative interviews indicates that implementation at scale appears to be more effective when:

- the organisational arrangements are characterised by established structures and processes, including quality assurance of training and implementation, and there is access to in-depth expertise on the intervention and how it should be implemented;
- trainers have a deep understanding of how children learn, of approaches to teaching in the subject area of the intervention and experience of teaching the subject and working in schools, and a positive orientation towards the intervention;
- ‘train the trainer’ programmes—
 - provide sufficient time for training, which spans training on the full intervention before trainers begin to deliver CPD to schools and further ongoing trainer CPD and support;
 - involve developers in design and delivery and pay attention to making developers’ tacit knowledge explicit;
 - foster trainers’ ownership of the intervention in order secure fidelity to the intervention;
 - include opportunities for trainers to observe the developers delivering the CPD to direct implementers or opportunities to practice the intervention with full classes; and
 - make no assumptions about trainers’ prior knowledge and understanding; and when
- CPD for direct implementers is sequenced to enable deep understanding and there is ongoing support from trainers including opportunities for reflection on implementation and further development of understanding.

(c) An overly-cautious approach to quality assurance and monitoring

Some developers and deliverers interpreted the EEF’s requirement for achieving ‘real world conditions’ in an effectiveness trial to mean that they should take a ‘hands-off’ approach to quality assurance and monitoring.

Implications

The findings indicate that there is scope for a more systematic and evidence-based approach to supporting developers to scale-up their interventions that includes greater clarity around what is necessary to reduce or change to implement at scale, builds on evidence on effective approaches to training trainers and CPD for direct implementers, and highlights the need for robust quality assurance of training and implementation.

Funders may also wish to consider in more depth how an 'effectiveness trial' should be defined, including whether it is appropriate to change the target group.

Introduction

Background and context

The EEF's mission is to break the link between family income and educational achievement. More specifically, the EEF aims to achieve this mission by:

- raising the attainment of 3 to 18-year-olds, particularly those facing disadvantage;
- developing their essential life skills; and
- preparing young people for the world of work and further study.

This is achieved through summarising the best available evidence in plain language, generating new evidence of 'what works' to improve teaching and learning, and supporting teachers and school leaders to use research evidence to maximise the benefit for young people.

This report supports the achievement of these aims by providing—for EEF effectiveness trials that had been reported by the date of the review—

- a quantitative comparison of the effects at the efficiency and effectiveness stages; and
- a qualitative analysis of interviews with delivery programme leaders who have been engaged in the scale-up of EEF efficacy trials to effectiveness trials to identify any models, supporting structures, and programme features that either support or hinder intervention scale-up to large numbers of schools.

In addition to providing insights for the EEF, developers, and delivery partners designing and managing the scale-up of trials, it also contributes to the growing body of evidence on the factors that support wider use of evidence-informed practices in educational contexts.

This review was undertaken in 2019 with the specific aims, objectives, and research questions set by the EEF.

Aims and objectives

The aims and objectives were:

- to carry out qualitative interviews with programme team members from the effectiveness trials to explore models and processes perceived to be associated with effective scale-up; and
- to draw out claims to support EEF grant-making and scale-up as well as more generalisable claims and reports on areas that would benefit from further exploration and analyses.

Research questions

1. Do we see any differences in effects for effectiveness trials compared to other EEF trials?
2. Are there any models for scaling programmes (such as 'train the trainer') that seem particularly successful or unsuccessful?
3. What types of structures need to be in place to help organisations deliver programmes to large numbers of schools (for example, partnerships with other organisations)?
4. What programme features support high-quality implementation at scale?

Methodology

The first research question was addressed by extracting and summarising relevant findings from the main analysis of the Review of EEF Projects. Full details of the methodology for the main analysis and its limitations are provided in Demack et al. (2021).

Data to address research questions two to four was gathered through qualitative interviews with the developers and delivery partners responsible for leading the delivery of the efficacy trials that were subsequently scaled-up to effectiveness trials. At the time this research was commissioned, both efficacy and effectiveness reports had been published by the EEF for eight interventions; seven of these are the focus of this analysis. The delivery partner associated with the remaining intervention did not respond to requests to participate in the research.

An interviewee list was provided by the EEF. In total, nine semi-structured interviews were conducted. At least one developer was interviewed for each of the seven interventions; for one intervention, two developers were interviewed. An interview was also conducted with a delivery organisation that had engaged in the further scale-up of one of the interventions after the effectiveness trial had been completed.

In preparation for the interviews the trial reports were read to provide contextual background. The interview schedule focused on:

- the role of developers in the efficacy and effectiveness trials;
- any changes made in the move from efficacy to effectiveness trial;
- scaling up recruitment—what was learned about what worked well or did not work well;
- scaling up delivery—what was learned about what worked well or did not work well; and
- contextual issues perceived to be impacting on scale-up.

While evaluation design was not an intended focus of this study and developers were not directly asked about this, it is interesting to note that most interviewees felt strongly that the evaluation design of the effectiveness trial had adversely impacted on the primary outcomes. In some instances, they pointed to evaluation design features that had impacted negatively on implementation, such as squeezed time-scales; in others, they questioned more directly the appropriateness of the design and particularly the validity of the impact measure(s).

All interviews were transcribed followed by a thematic analysis of the transcripts. Deductive and inductive approaches were combined. Data was initially organised deductively into a coding frame that was constructed with reference to implementation literature. An iterative process of in-depth reading of all the transcripts was used to refine the main themes and develop sub-themes.

Some caution is needed with regard to the findings of the scale-up analysis as the data is limited to seven interventions and the qualitative analysis is partial in terms of being primarily the perspectives of developers that had limited access to some aspects of implementation. Limitations of the data and findings that are drawn from the **main EEF Review of Projects are set out in Demack et al. (2021)**.

Ethics and data protection

The project received ethical approval from the Faculty of Social Sciences and Humanities at Sheffield Hallam University. All data used in the quantitative analyses that has been summarised in this report has been coded from publicly available sources (the EEF trial reports). The only personal data held for this study was contact information for the qualitative interviews and interview data. The legal basis for processing this data is 'Public Task' as defined in Article 6(1e) of the General Data Protection Regulations (GDPR).

All participants were provided with a project information sheet and given the opportunity to ask questions about the study before the interview. Each interviewee completed a consent form (see Appendix C for further details).

Project team

Table 1: Core Sheffield Hallam University project team for the scale-up review

Team member	Title	Role/responsibilities
Professor Bronwen Maxwell	Head of Commissioned Research	Principal investigator: qualitative fieldwork and analysis.
Bernadette Stiell	Senior Research Fellow	Qualitative fieldwork.
Anna Stevens	Research Fellow	Extraction and summary of quantitative data from main report .

The full project team for the EEF review that generated the quantitative data and analyses used in this report is set out in **Demack et al. (2021)**.

Key findings

Primary outcome effect sizes

For six of the seven interventions in scope for this study, the mean effect size for the intervention was lower for the effectiveness trial than the efficacy trial (Table 2). For one intervention, Improving Writing Quality, an exceptionally high reduction in effect size (-0.73) was found. For the Mathematical Reasoning (MR) intervention, the effect sizes were very similar for both types of trial.

The unweighted mean effect size¹ for the seven efficacy trials in this analysis is 0.25 compared with 0.01 for the effectiveness trials. This compares to an unweighted mean effect size of 0.04 for all trials in the review classified by the EEF as efficacy trials and a mean effect size of 0.09 for all trials classified by the EEF as effectiveness trials. As noted in the main report (Demack et al., 2021) we advise caution due to some ambiguity in these classifications. As might be expected, the unweighted mean effect sizes for the seven scale-up interventions are all higher than the average unweighted mean for all trials.

Table 2: Efficacy and effectiveness trials of scaled-up interventions—effect size, phase, and outcome measurement

Project	Efficacy or effectiveness	Unweighted mean effect size	Educational phase for trial	Outcome measurement
Catch Up Literacy	Efficacy	0.12	Transition KS2 to KS3	Commercial (GL CEM Hodder or Pearson)
Catch Up Literacy (CL)	Effectiveness	0.01	Primary (KS2)	Commercial (GL CEM Hodder or Pearson)
Catch Up Numeracy	Efficacy	0.21	Primary (Multiple Key Stages)	Commercial (GL CEM Hodder or Pearson)
Catch Up Numeracy (CN)	Effectiveness	-0.04	Primary (KS2)	Commercial (GL CEM Hodder or Pearson)
Grammar for Writing	Efficacy	0.17	Primary (KS2)	Official (KS attain, absences)
Grammar for Writing (GRW)	Effectiveness	-0.02	Primary (KS2)	Official (KS attain, absences)
Improving Numeracy and Literacy	Efficacy	0.075	Primary (KS1)	Commercial (GL CEM Hodder or Pearson)
Mathematical Reasoning (re-grant for Improving numeracy) (MR)	Effectiveness	0.08	Primary (KS1)	Commercial (GL CEM Hodder or Pearson)
Switch-on Reading	Efficacy	0.24	Secondary KS3	Commercial (GL CEM Hodder or Pearson)
Switch-on Reading (SR)	Effectiveness	0	Primary (KS1)	Commercial (GL CEM Hodder or Pearson)
Thinking, Doing, Talking Science	Efficacy	0.22	Primary (KS2)	Other or mixed
Thinking, Doing, Talking Science (TDTS)	Effectiveness	0.01	Primary (KS2)	Other or mixed
Improving Writing Quality	Efficacy	0.74	Transition KS2 to KS3	Commercial (GL CEM Hodder or Pearson)
Improving Writing Quality (IPELL)	Effectiveness	0.01	Primary (KS2)	Official (KS attain, absences)

The effects of the seven scale-up interventions discussed here were considered in relation to the quantitative meta-analysis findings in the main review report (Demack et al., 2021) in terms of design, intervention characteristics, and context. The quantitative meta-analysis findings appear to offer at least a partial explanation for the reductions in the effect size observed over six of the seven interventions.

Firstly, in terms of design, the quantitative meta-analysis found that individual randomisation is more likely to yield higher effect sizes. With regards to the seven scale-up projects discussed here, two of the efficacy studies were randomised

¹ This is a standard descriptive mean and is not part of the meta-analysis.

at the individual level while all the effectiveness studies had clustered randomisation. This may have contributed to reducing the effect observed for the two efficacy interventions where individual-level randomisation was used.

Secondly, in terms of intervention characteristics, the meta-analysis revealed that transition projects (projects taking place at the primary to secondary school transition) were more likely to have higher effect sizes; two of the efficacy trials here were transition projects while all of the effectiveness trials were based in primary schools.

Finally, in terms of contexts of the interventions, findings from the meta-analysis showed that smaller studies and those set in one geographical location were associated with higher effect sizes. The average number of pupils in the efficacy trials here is 929 compared with 4,750 for the effectiveness trials, and the efficacy studies are more likely to be set in one geographical location.

It should also be noted that data in trial reports indicated that for all seven interventions there was a direct match between intervention and primary outcome,² which the meta-analysis indicates is likely to lead to larger effects than where there is weaker alignment between intervention and primary outcome.

The nature of ‘scale-up’

At its simplest, scale-up can be observed in terms of increased numbers of participating schools and pupils. The number of schools participating in the trials of the seven scale-up interventions varied between 15 and 85 at the efficacy stage and between 84 and 205 at the effectiveness stage. The number of pupils (intervention and control) varied between 108 and 2,394 for the efficacy trials and between 999 and 8,966 for the effectiveness trials. For many of the interventions there was around a three-fold increase in the number of schools from the efficacy trial stage to the effectiveness stage. The most notable exception here is Switch-on Reading (SR) where there was an increase from 15 to 190 schools. The increase in the number of pupils at the effectiveness stage was, with the exception of SR, proportionately higher than the increase in the number of schools, meaning that generally more pupils were being recruited from each school in the effectiveness trials than in the efficacy trials.

Table 3: Seven scale-up interventions—number of schools and pupils for efficacy and effectiveness trials

Project	Efficacy or effectiveness	No. of schools	No. of pupils
Catch Up Literacy (CL)	Efficacy	85	557
Catch Up Literacy (CL)	Effectiveness	152	1371
Catch Up Numeracy (CN)	Efficacy	54	108
Catch Up Numeracy (CN)	Effectiveness	151	1811
Grammar for Writing (GFW)	Efficacy	53	2394
Grammar for Writing (GFW)	Effectiveness	155	7239
Improving Numeracy and Literacy (MR)	Efficacy	55	2217
Mathematical Reasoning (re-grant for Improving Numeracy) (MR)	Effectiveness	169	7419
Switch-on Reading (SR)	Efficacy	15	314
Switch-on Reading (SR)	Effectiveness	190	999
Thinking, Doing, Talking Science (TDTS)	Efficacy	40	655
Thinking, Doing, Talking Science (TDTS)	Effectiveness	205	8966
Improving Writing Quality (IPEELL)	Efficacy	26	261
Improving Writing Quality (IPEELL)	Effectiveness	84	5444

The Catch Up Literacy (CL) and Catch Up Numeracy (CN) developers did not accept that the EEF effectiveness trials were a scale-up of their respective efficacy trials. Although they accepted that the effectiveness trials involved more schools and pupils, they considered that as an organisation they had already delivered the interventions at scale. They

² Although this was not accepted by all the developers interviewed.

asserted that there were no differences between the two trial stages in terms of how they approached the interventions nor any departures from their established ways of working in delivering their range of programmes nationally.

As well as increases in numbers of schools and pupils, various other changes were made for the move from efficacy to effectiveness trials, some of which may have negatively impacted on the effect size. The main areas of change are set out below. With the exception of the first area, new target groups, these changes can be understood as responses to the challenges of implementing an intervention at scale.

New target groups

In four of the seven interventions, the target group for the effectiveness trial was in a different school phase or Key Stage than for the efficacy trial. This raises the question of the extent to which the effectiveness trials for these interventions—CL, CN, SR, and Improving Writing Quality (IPEELL)—can be considered genuine scale-ups of the efficacy trials.

In the future, where changes in the target group are proposed for the scale-up, it may be appropriate to consider at the outset whether there is sufficient evidence to indicate that the intervention is likely to have a similar effect with the new target group. If this evidence is absent or insubstantial, a further efficacy trial prior to scale-up could be considered.

Changes to the recruitment of participants

For three of the scale-up interventions, there was a shift in who led the participant recruitment. The developers handled recruitment at the efficacy stage but for the effectiveness trials this responsibility fell to trainers—in the case of SR and Thinking, Doing, Talking Science (TDTS)—or to the organisation commissioned to run the intervention at arm's length from the developer (in the case of MR). In the other four interventions (CL, CN, GFW, IPEELL), the delivery partner undertook recruitment at both the efficacy and scale-up stages.

Adoption of a train-the-trainer model

In four of the seven interventions (SR, MR, TDTS, IPEELL), there was a shift from developers being responsible for delivery during the efficacy trials to the adoption of a train-the-trainer model for the effectiveness stage. For two interventions (CL and CN), a train-the-trainer model had also been in place for the efficacy trial. For GRW there already existed an established partnership for programme delivery so the intervention was delivered by a combination of the developer team and a consultancy organisation.

Adaptations to the intervention

A significant change was made to the GFW intervention for the effectiveness study. Training for direct implementers (teachers) in the efficacy trial had focused exclusively on the underpinning principles of the intervention and providing support for teachers to develop their own work based on these principles. In the effectiveness trial, the training focused on supporting teachers to implement two teaching units planned by the developers. For the other six projects, adaptations to the intervention at the effectiveness stage were more limited and based on learning from the efficacy trial or other work.

Development of resources for trainers and direct implementers

The most significant changes to resources involved producing PowerPoints and guides for the trainers. In creating these resources some developers were actively seeking to make their tacit knowledge explicit, as was explained by the TDTS developers:

'Things that we would know how to say because of our experience and our specialism, that we'd know to say if we were training staff, we tried to make sure that they were written down as part of the training materials or as part of the resources.'

The TDTS developers reported that working with the new trainers was particularly beneficial in helping them to 'really identify what were the strategies, the clear strands, and how they could keep coming back to them and make that explicit to teachers so that teachers connect with the underlying strategies and not the specific examples'.

In most of the scale-up interventions, resources for teachers were 'sharpened' in the light of experience from the efficacy trial.

CPD for direct implementers

For two interventions there was a reduction in the amount of CPD provided for direct implementers. In the case of the GFW effectiveness trial, all CPD was delivered in three full-day sessions to large groups of teachers whereas in the efficacy trial the developers had worked with schools in clusters over time, responding to their needs and providing coaching and mentoring. The rationale for the revised CPD model was perceived cost. In the case of TDTS, the reduction in the CPD for teachers from five days to four was to take account of the difficulties schools encounter in releasing staff to attend CPD.

Developers reported only limited changes from efficacy to effectiveness stages in relation to the theories of change underpinning their interventions, although the GFW developer noted the influence of cumulative learning from the efficacy trial and other projects they had led.

Cost-effectiveness

More surprisingly, for five of the seven projects there was little difference between the effectiveness and efficacy stages in terms of the cost-effectiveness of the intervention. Five of the interventions were rated by the EEF as very low cost (less than £80 per pupil per year) for both efficacy and effectiveness trials. For CL the cost reduced from high (more than £1,200 per pupil per year) for the efficacy trial to very low for the effectiveness trial, while SR's cost went from moderate (up to £700 per pupil per year) to low (up to £200 per pupil per year). As noted in the main review, there is inconsistency in measuring costs across reports so caution is needed in interpreting this finding.

Structures for supporting scale-up

This section outlines key features of the different structures that were deployed to support scale-up across the seven interventions, together with interviewees' perceptions of which structural features that worked well and less well.

The Catch Up (CL and CN) delivery partners reported that their existing organisational framework worked well in supporting effective delivery at scale. Catch Up is an established charitable organisation that was already experienced in delivering nationally and at scale prior to the EEF CL and CN effectiveness trials. It is useful to note that this experience meant that a core set of features were already in place, unlike the case in most of the other effectiveness trials where existing structures to support scale-up were absent or less well developed. The pertinent features of the Catch Up projects included a core team with established procedural systems and a team of self-employed trainers who had been trained and accredited prior to the effectiveness trials, together with a rigorous system for quality-assuring the work of the trainers.

Although Catch Up had established recruitment processes, it found recruitment at scale to be challenging and it attributed the difficulties encountered to insufficient resource and time to recruit rather than a structural issue. The interviewees perceived that this was an issue for the EEF to address rather than an organisational constraint.

Scaling-up through a national organisation was found by the MR delivery partners to be beneficial in terms of having an established organisational framework as well as a ready-vetted pool of high-quality trainers from which to recruit. Nonetheless, challenges were encountered in recruiting trainers in some geographical locations and the delivery team perceived that this impacted negatively on the quality of delivery of the training in those locations. The MR delivery partners deployed an arm's length approach to scale-up by appointing the National Centre for Excellence in Mathematics (NCETM) to deliver the scale-up. NCETM was responsible for recruiting participants and trainers and managing implementation. The MR developers worked with NCETM in designing and delivering the training. NCETM was selected by the delivery partner because there was seen to be a good match between the aims of the MR intervention and NCETM's own objectives, NCETM had the structure to enable further scaling, and the developers had links with key NCETM personnel.

The partnership model of working employed by GFW was perceived by the delivery partners to work well and combined the strengths of GFW and its partner organisation. While GFW maintained responsibility for delivery of the effectiveness trial, including recruiting schools, it worked with an education consultancy it had previously partnered with to provide sufficient trainers. GFW brought in-depth expertise in relation to the intervention as well as a very effective multi-strand strategy for participant recruitment that capitalised on their well-established network of contacts. The education consultancy was accustomed to working at scale and could provide experienced, tried and tested trainers who had previously delivered similar training programmes to schools.

In contrast, the model of directly recruiting a new team of trainers to implement at scale was, to varying degrees, considered challenging by each of the delivery partners (SR, TDTS, and IPEELL) that took this approach. The key issue encountered was recruiting sufficient tutors with the appropriate knowledge, understanding, skill sets, and attitudes towards the intervention. However, other projects with different structural arrangements were not immune to problems of recruiting and training adequate trainers at scale. These are reported in the next section.

A further issue evident in the implementation of the SR, TDTS, and IPEELL scale-ups by the delivery partners was the limited quality assurance and monitoring of both training and implementation. This was largely reported to be a response to the EEF's instruction to adopt a 'hands-off approach' to mirror 'real world conditions'. However, what is meant by 'real world conditions' needs to be considered. Typically, organisations that are experienced in delivering at scale will have in place routine processes for monitoring the quality of trainers and training. Some of these organisations will also directly monitor implementation and impact: for example, the National Literacy Trust regularly surveys pupils participating in interventions to assess impact. There are potential implications here for researchers and funders around determining what constitutes 'real-world conditions' for the purposes of effectiveness trials and a need to ensure that delivery partners understand what is intended.

Working with an organisation with a large-scale infrastructure was considered essential by the IPEELL delivery partner to successfully recruit sufficient participants. IPEELL reported that the difficulties encountered in trying to recruit at scale for the effectiveness trial had subsequently led them to partner with the National Literacy Trust for further scale-up. The programme was then 'traded' in the same way as other National Literacy Trust CPD programmes with recruitment of participants being undertaken mostly online.

In summary, given the small number of scale-up interventions reviewed, the data from this analysis is neither sufficient nor conclusive regarding the best structural arrangements to scale-up interventions. The limited data does, however, tentatively indicate that organisations that possess the following characteristics are more likely to be successful in leading implementation at scale—those with:

- established and tested organisational structure and processes;
- access within the organisation, or through partnership, to in-depth expertise on the intervention and how it should be implemented;
- established processes and networks for recruitment and selection of trainers and recruitment of participants; and
- robust quality monitoring systems for ensuring the quality of training and implementation in schools.

Programme features perceived to support high-quality implementation at scale

This section draws together findings on the features of programmes that appear to support or inhibit high-quality implementation at scale. The richest data concerns the recruitment and training of trainers as most of the developers had fairly limited involvement in the monitoring of the training or implementation, as noted above. Some form of train-the-trainer model was deployed to scale-up five of the seven interventions. For the Catch Up interventions, a workforce of self-employed trainers, who were trained and experienced in delivering the interventions, had been established prior to the EEF effectiveness trial.

Trainer characteristics and recruitment

A key theme across most interventions was the need for all trainers to have a deep understanding of how children learn and an understanding of approaches to teaching in the subject area of the intervention. The challenge of recruiting sufficient trainers in a short period of time meant that most developers had to go beyond their existing professional networks and recruit some trainers who, it transpired, did not have sufficient knowledge and understanding. The reported consequence (particularly in relation to SR, TDTS, and IPEELL) was that the trainers delivered training to direct implementers who did not adhere faithfully to the intended intervention and, in turn, those direct implementers (teachers or teaching assistants) did not implement the intervention as intended.

The MR developer suggested a process of accreditation:

'If the trainers felt that they had to be accredited in some way by presenting what they had observed ... So [in the current effectiveness trial] irrespective of how they did it in the training phase, they'd go on to the next phase. But I would think that you have some incentive ... I think that "you will be accredited to deliver

this training afterwards” would have made a difference. And people who are not accredited cannot deliver it.’

Related to the previous themes is the need for all trainers to have the relevant experience both of teaching the subject and working in schools. For example, the SR developer reported that trainers who were already established teacher leaders for the Reading Recovery programme had both the necessary subject specialist knowledge to deliver SR and the leadership experience necessary to effect change in schools where the intervention was not being delivered as intended:

‘I think the understanding of the teacher leaders was greater—not just the understanding during training, but also when they were going back into school—their expectation of how the programme would be delivered was much higher. So when things weren’t being done the way we’d written it down or the way they’d been trained, they were keener to go back in and to talk to the staff than maybe the non-teacher leaders were. I think that’s about an understanding of how interventions should be delivered and why, and also just that deeper understanding of how children learn to read, as well.’

Similar issues were encountered in scaling-up TDTS where trainers who lacked experience of primary science, or whose experience was in the informal learning sector rather than schools, were perceived to be less likely to implement the training with fidelity.

Recruiting trainers with appropriate knowledge and experience was more difficult where trainers were expected to recruit schools as well as deliver training in schools. The SR and TDTS developers both reported that expecting trainers to recruit schools had the effect of obliging the developers to recruit trainers who had established networks but who may not have been the highest-quality trainers. The SR developer recommended that in future scale-ups recruitment should be separated from delivery commitments.

Positive trainer attitudes towards the intervention were reported to be essential. This required recruiting trainers that had values and beliefs about how children and young people learn in the subject area that were compatible with the intervention. In addition, it was found necessary to recruit trainers who were willing to set aside any of their own established practices that reduced fidelity to the intervention. A lack of belief in the intervention or a lack of commitment to it was reported to lead in some instances to disruption to the training and negative impacts on implementation:

‘There were a couple of places where NCETM wanted to have trainers and nobody in that region wanted to be trained, so I think they pushed a bit those regions and those didn’t work very well’ (MR developer).

‘It’s almost like identifying a mindset. Some of them were able to sort of get it—they’d understand what we were saying: “This is what you’re got to hold true to, and to a large extent this has already been successful, so you’ve got to stick to it, because we’re trying to replicate something that already has success. We’re trying to replicate it at scale.” And some of them understood that and had respect for what the limitations were, and some of them, it’s like they couldn’t help but want to change it, to kind of go “I know best” because I’ve got this other bit of experience and this other bit of experience’ (TDTS developer).

A further issue linked to prior experience and attitude is the mode of CPD that trainers are used to delivering. Interestingly, the SR developer reported that trainers who were experienced Reading Recovery trainers often came back to seek clarification or to discuss issues observed in schools, reflecting a way of working that was embedded within Reading Recovery. In contrast, other trainers rarely made contact with the developer after the train-the-trainer sessions. These trainers included those with experience of delivery of National Strategies where the mode of CPD primarily involved no further interaction with the schools after the delivery of the training.

A final issue raised related to trainer recruitment was the need for a selection process rather than just a recruitment process. As the developers for TDTS pointed out, time constraints meant that it was not possible to train and then observe the trainers and then select only those who demonstrated that they fully understood the programme and delivered the CPD with fidelity. This issue was particularly pertinent when a cadre of trainers was being established for the first time. The CL, CN, and GFW projects already had in place a set of trained and ‘tested’ trainers so did not report either this difficulty or the other issues related to trainer characteristics as barriers to implementation.

Training for trainers

Sufficient time for training the trainers and not making assumptions about trainers’ prior knowledge were considered important. The training for trainers to deliver the IPEELL consisted of only one half-day because the developer had assumed that the trainers would have the necessary background knowledge. However, this assumption proved not to

be case and the drawback was perceived to undermine the fidelity to the intervention in the training for teachers with the consequence that teachers did not implement the programme with fidelity:

'We made some assumptions ... First, that because [the trainers] were members of the local authority improvement services that they would have full understanding of concepts of self-regulation, and also knowledge of the content of English teaching in particular with respect to writing, and this proved not to be the case. ... nobody said, "Well, what is self-regulation?" So we assumed that they all understood it. And later on [when the trainers delivered CPD to teachers], it was blatantly not the case.'

Scheduling the training so that training on the full intervention is completed before the trainers begin to deliver CPD to schools was reported to be necessary. Due to squeezed timelines, TDTS trainers had to begin delivering CPD in school after just two of their four days of training. While the developers stopped short of claiming a direct connection between this scheduling and its impact on implementation, they did consider it problematic as the trainers 'never got a chance to see the course in its entirety, to experience it in its entirety, before they started delivering the course to their own teachers'.

Making the tacit knowledge of the developers explicit was challenging for some developers but recognised as important. The IPEELL developer was able to trace a link from the training for trainers not making tacit knowledge explicit through to teachers in schools not implementing with fidelity:

'In the training, initially, we didn't make clear enough the negotiables and non-negotiables. Later on, as we went around to monitor this in the schools, it became obvious that some aspects should have been more clearly emphasised as non-negotiables [in the training of trainers]. Some teachers didn't ... we said from the beginning that this was a strategy to improve writing. It was not a straitjacket. Therefore, we didn't expect them to follow everything to the last dot and cross. But there were certain points that we did say ... there were certain stages that were part of the strategy and that the whole strategy depended upon these things being carried out properly. They were the non-negotiables. In some schools, teachers didn't fully take these on.'

Developers did provide detailed resource packs for trainers and refined resource packs to be used by the direct implementers to help make tacit knowledge explicit in order to enhance fidelity. However, while making tacit knowledge explicit was considered important by developers, it should not necessarily be prioritised by itself. Some developers saw the need to find a way to balance the maintaining of fidelity to the intervention on the one hand and the building of trainers' ownership of the intervention on the other. This dilemma is illustrated in the TDTS developers' observations on the delivery of training:

'We noticed things like sometimes some of the tutors would say things like "[the developers] said I had to say this", rather than making it their own. Some of the trainers probably gave too much emphasis to areas that we wouldn't emphasise as much... trainers would spend a lot longer on some sections than I would have done ... and it's how you facilitate the trainers to make it their own, so they really feel it's theirs but it's still true to the original.'

This issue cannot be addressed simply by tighter codification of an intervention. The TDTS developers pointed to 'the complex nuances of a really effective train-the-trainer programme and the amount of effort that it takes to replicate something at scale through that mechanism'. Illustrating this point, they explained:

'You want to give them a little bit of flexibility to play to their strengths, to insert their own anecdotes. That is quite an art—I think one of the things we learned was actually how challenging that is, to tread that line between making sure that there is the consistency [about] the core messages [and] strategies and the connection all the time to opportunities for higher-order thinking remains really clear and strong, whilst at the same time, you do not want people to be feeling like they're standing there with their training folder following a script, because it's not natural and actually that makes the training less effective. Even with the very experienced trainers we were working with, that was really challenging.'

The direct involvement of developers in delivering training to trainers was considered important across the interventions. Taking this a step further, the evaluators of the IPEELL effectiveness trial recommended that trainers should observe the developers delivering the CPD direct to teachers. This recommendation was taken up by the National Literacy Trust in the subsequent scale-up activity and was regarded by both the Trust and the IPEELL developers as key to ensuring fidelity of CPD delivery by the trainers.

The opportunity for trainers to practise delivery, by delivering a complete programme of CPD to direct implementers once before an effectiveness trial begins, was mentioned as important by the TDTS and GRW developers. While this is more of an issue of evaluation design than a factor supporting effective implementation, it does highlight the complex process and the time and practice needed for trainers to develop the understanding and skills to implement CPD and, if appropriate, support implementation. The SR developer highlighted issues of cognitive overload in train-the-trainer sessions and pointed to the need for trainers to be able to try out CPD delivery and then have the opportunity for further reflection and training.

Ongoing communication and support for trainers through the delivery phase was perceived to be important. This appeared to be working effectively in the SR intervention:

'Our [previous] experience is that cascading training doesn't work ... So, we tried very hard to have as much training and contact with the trainers as possible. I did have a lot of communication with trainers when things were happening that were unexpected. ... a lot of the trainers were really shocked at how the intervention was being delivered. I had a lot of emails and phone conversations with trainers who were saying "I don't know why they're doing that, because I didn't say that", or "I don't know why they're missing this thing out, because it's really obvious from the training materials".'

The following features were perceived to support effective train-the-trainer programmes, with each item mentioned by one of the developers interviewed.

- Trainers should implement the intervention with a full class prior to training delivery. Although the MR developer noted that they did not have empirical evidence, they stressed the importance of trainers delivering the intervention to a full class of children prior to delivering any CPD to teachers so that the trainers were better able to address teachers' questions and concerns:

'We believe very strongly ... [trainers] must do it, on the same scale that they are going to ask the teachers to do it ... If it's a classroom intervention, deliver it in the classroom so you know what happens in a classroom, so when a teacher asks you, you can actually answer.'

- CPD should be ongoing. The SR developer highlighted the need for ongoing CPD for trainers and contrasted the approach in the effectiveness trial to the Reading Recovery programme where:

'You have ongoing CPD throughout every single year that you're in practice, and we know that works because it brings that fidelity of the programme, and it deepens the understanding behind the programme that is being delivered or trained. And that wasn't possible because of the short amount of time that the [effectiveness trial] was delivered over.'

Training for direct implementers

The need to sequence CPD for direct implementers over time to enable deeper understanding was highlighted by the GRW and SR developers who both felt that their effectiveness trials were not conducive to the development of the deeper understanding required. As the SR developer explained:

'It's only when they go through CPD on a continual basis rather than the odd, ad hoc training that they start to realise how much more they need to know, and we're always learning from that ... [this intervention] is a day's training and a bit of support and that's quite limited. What I've learned over these projects, but also other things, is that it's not about the approach you use, it's about the knowledge and understanding of how children learn to read that is going to have a real impact.'

This issue, together with the lack of opportunity to build relationships with direct implementers, was expanded on by the GFW developer. Instead of their usual 'drip-feeding, slower, developmental approach' with smaller groups of teachers, for the effectiveness trial the training for teachers was delivered using 'a much more didactic, approach—here it is, do that' in 'three monolithic inputs' and with no opportunity to build relationships with teachers in the very large training groups. The developer perceived this as 'somewhat unnatural' with negative impacts on teacher development and consequently pupil outcomes. While the developer acknowledged the cost issue of implementing their usual approach at scale, they pointed out that if there is willingness to accept the cost of delivery of CPD to direct implementers, implementation can be more effective if CPD is developmental and involves establishing ongoing relationships between trainers and direct implementers.

An effective practice reported by TDTs was using two trainers to deliver to direct implementers; this was perceived to address some of the challenges of a train-the-trainer approach and, going forward, the developers wanted one of the two trainers to be a member of the core developer team:

'Another thing I feel strongly is, if we were scaling up without an effectiveness trial, there is no way we would have, in retrospect, have trained 14 trainers and said, "OK now you go and deliver the course." We would have trained people initially to do co-delivery with us. And that would have improved the ... What we've been doing since the effectiveness trial result is building up our core team ... If we do this again as a future trial, we would want to do co-delivery: one of us with one of the new trainers on an intervention, because that way [ensures] fidelity and allows scalability.'

Inevitably this approach has a cost implication and may not be feasible in all effectiveness trials.

Quality assurance of training for direct implementers

There was a general consensus that quality assurance of training is important. While this was an established process for the Catch Up interventions, some of the other developers expressed concern that in their effectiveness trials it was too light-touch and insufficiently rigorous or it focused only on checking fidelity rather than being a quality improvement process. As the MR developer explained:

'This idea of hands-off is not very good. We should have a role in monitoring the quality of the trainers' implementation. We should go and visit the trainers in the same way that we visited the teachers in the efficacy trial. And we also think that there should be a formal evaluation of their implementation.'

The most extensive quality assurance of training took place for the IPEELL intervention where the IPEELL developer observed all training sessions, fed back to the trainers, and carefully intervened where the information being given to teachers was incorrect. Providing feedback to trainers was perceived by the IPEELL developer to play a role in improving the fidelity of subsequent delivery of the CPD. However, it is noteworthy that this level of monitoring was only put in place because it became apparent that teachers were not implementing the intervention with fidelity.

Support for implementation

Establishing a strong ongoing relationship and continued support for direct implementers by trainers was identified as important. This factor was perceived by some developers to support the development of teachers' understanding and enabled issues that arose during implementation to be addressed. The SR developer reported that the process of trainers going back into schools following the training to support implementation 'can have a massive impact' and as noted earlier is likely to be even more effective where the trainer has leadership experience.

This approach aligns with the views of some of the other developers who stressed the need for direct implementers to be provided with opportunities to reflect on the intervention and implementation during the implementation period. For example, the IPEELL developer reported that network days for direct implementers run by the developer and supported by the trainers part-way through the intervention period were effective in improving fidelity of implementation. These network days were put in place as it became apparent that fidelity of implementation was weak. Similarly, the use of gap tasks as an integral part of the CPD programme in TDTs was reported to be beneficial in both monitoring implementation and enabling a cyclical process of reflection and sharing of experiences.

Another method of providing ongoing support for direct implementers was ensuring that there was in-school support. For example, the TDTs delivery team worked with more than one teacher per school so that those teachers could share ideas and experiences. The SR developer acknowledged that implementation could have been enhanced if the trainers had been tasked with establishing stronger communication with class teachers so that teachers could support the teaching assistants who were the direct implementers:

'Communication ... We know that is an issue. What's happening with Mrs so-and-so down the corridor on a one-to-one basis, the class teacher doesn't necessarily know about. And I think what we would have wanted to do was the trainers to have some input with class teachers and for them to come to some kind of training just to hear about ... the strategies that were being taught, the principles that were being taught, so they could back that up back in the classroom.'

Quality assurance of implementation by direct implementers

The need for quality assurance of implementation by direct implementers and its more immediate impacts on pupils was perceived as an important basis for improving fidelity. Developers who mentioned the importance of this quality

assurance tended to juxtapose this with the reality of the limited quality assurance that was possible during the effectiveness trial. As the GFW developer pointed out, the 12 sample visits they conducted provided a very incomplete picture:

'I think it was 12 visits, I can't remember ... we visited once just to see what they were doing, but that's tiny. And it becomes terribly dependent on which lesson in the unit that you happen to see. ... because in a proper unit of teaching narrative writing or argument writing, you could go into a lesson where there wasn't much of the grammar stuff going on because that was a lesson that was looking at story structures or something. So, it was very, very tiny.'

An important lesson from the IPEELL scale-up, where monitoring was undertaken in most schools (in the form of observations of lessons and interviews with children), was that the monitoring of implementation must be undertaken by people who have a full understanding of the strategy and have attended the train-the-trainers training. As the developer explained, not all individuals conducting the monitoring had this knowledge or experience:

'Over the two years of the project, there were changes in personnel at both venues—of trainers and therefore of people who were also expected to monitor what was being delivered, so the people who were monitoring what was being delivered were not necessarily totally au fait with what the project was about ... [and that is where it fell down].'

Conclusion

For the seven interventions that were the focus of this analysis, scale-up was not confined to increasing the numbers of pupils and schools. Most of the other changes observed in the progression from efficacy trial to effectiveness trial—for example, changes in recruitment methods, adopting a train-the-trainer approach, developing new resources or refining resources for trainers and direct implementers, and adaptation of the training for direct implementers—such changes can be fairly regarded as processes needed to manage delivery at scale. However, a change in the target group in four of the seven interventions does raise questions about whether those studies were genuine scale-ups from the efficacy trials and whether either an effectiveness trial with the same age group or a second efficacy trial with a different age group may have been more appropriate.

The explanation for the reduction in the mean unweighted effect size for the primary outcomes for these interventions from 0.25 for the efficacy trials to 0.01 for the effectiveness trials may be explained in part through findings from the quantitative meta-analysis in the Review of EEF Projects (Demack et al., 2021) in terms of variation between the efficacy and effectiveness trials in the trial design, target group, size of study, or the number of geographical locations in which the intervention was delivered. Specifically, (a) two of the trials were designed with clustered randomisation at the effectiveness stage and individual-level randomisation (associated with higher effect sizes) in the efficacy stage, (b) two of the efficacy trials were transition projects (associated with higher effect sizes) while all of the effectiveness trials were based in primary schools, and (c) the effectiveness trials were larger studies and more likely to be set in one geographical location (both associated with lower effect sizes).

The qualitative analysis of interviews with intervention developers and deliverers enabled an in-depth exploration of the structural and programme factors and provided deeper insights into the factors that either supported or impeded successful scale-up. While it is not possible to make definitive claims about the effectiveness of different structural models for scale-up, the data indicates that implementation is more likely to be supported where there are established and tested organisational structures and processes, which include processes for the recruitment and selection of trainers and the recruitment of participants and for monitoring the quality of training and implementation. The structural arrangements also appear to be most effective where they incorporate or facilitate access to in-depth expertise on the intervention and how it should be implemented and, in the case of scale-ups that are not led by the developer, where there is an alignment between the aims of intervention and the aims of the organisation implementing at scale.

The richest data on programme features that are perceived to support implementation at scale relates to train-the-trainer aspects of scale-up. The data was more limited on other features of programmes that support implementation at scale as the developers had less engagement with these aspects of the programme implementation. In summary:

- In relation to the recruitment and selection of trainers, the data indicates that trainers require a deep understanding of how children learn and of approaches to teaching in the subject area of the intervention, experience of teaching the subject and working in schools, and a positive orientation towards the intervention. Developers perceived that the pool of high-quality trainers was reduced where trainers were also expected to have networks to recruit participants. Developers also felt that the quality of trainers would be enhanced by robust monitoring of trainers' work and its impact. A process of accreditation was suggested to support this.
- In relation to training the trainers, the data suggests that training programmes are more likely to be effective when (a) sufficient time is allocated for the training, (b) the developers are involved in training delivery, (c) no assumptions are made about trainers' prior knowledge and understanding, (d) training for the full intervention is given before the trainers begin to deliver CPD to schools, and then (e) there are ongoing opportunities for trainer CPD and support. Training may be further enhanced where there are opportunities for trainers to observe the developers delivering the CPD direct to teachers or for trainers to implement the intervention with a full class prior to training delivery. A cycle of quality assurance of training delivery and feeding back to the trainers was considered very important but was reported to be limited in most of the scale-ups analysed. Indeed, some developers partially attributed the limited monitoring possible during the effectiveness trial to reduced fidelity of training and, in turn, weak fidelity as the intervention was implemented with pupils.
- The data also indicates that two possibly under-recognised issues in scale-up are the complexity of making the developers' tacit knowledge explicit and the need to balance fidelity to the intervention alongside building trainers' ownership of the intervention.

- In relation to the delivery of training to direct implementers, the data highlights the importance of sequencing CPD over time to enable deeper understanding of the intervention and principles of implementation. Establishing strong, ongoing relationships between trainers and direct implementers and providing continued support for direct implementers by trainers were also identified as important, as was the need for direct implementers to be provided with opportunities to reflect on the intervention and implementation during the implementation period. Ensuring that direct implementers had support in school, for example, by involving more than one teacher per school or by trainers working with key stakeholders, was also highlighted.
- In addition to robust quality assurance of training, the data suggests that implementation is likely to be more effective if there is also quality assurance of implementation, including measurement of its more immediate impacts on pupils, as a basis for providing further support to direct implementers and improving fidelity.

The findings also highlight the complexity of making the tacit knowledge of developers explicit in ways that can be owned by trainers. Trainers were reported to be less inclined to revert to their preferred approaches and fidelity to the intervention was reported to be stronger where trainers both had good knowledge of the intervention and felt a sense of commitment and ownership, indicating a possible area for further research so guidance can be provided to delivery partners.

Taken holistically, the findings indicate that it may be beneficial for trial designers and funders to consider in more depth how an effectiveness trial should be defined. At its simplest level, this would involve considering instances where the target group has changed to determine whether another efficacy trial would be more appropriate before wider scale-up. The more challenging task would involve, firstly, developing greater clarity around what is necessary to reduce or change in order to implement at scale—this will be influenced by both practical and costing constraints. Interestingly, there were notable variations across the analyses of the seven scale-ups about what was considered necessary to change and what was not. Secondly, greater clarity must be developed about those features that need to be present for an intervention to be implemented effectively at scale. These findings offer a starting point by (a) identifying particular approaches to training trainers and CPD for direct implementers that appear to offer the greatest likelihood of leading to effective implementation and (b) highlighting the need for robust quality assurance of training and implementation. Inevitably these features will need to be ‘traded-off’ with practicality and cost, but it appears from this analysis that there is scope for this to be done in a more systematic and evidenced way.

References

Demack, S., Maxwell, B., Coldwell, M., Stevens, A., Wolstenholme, C., Reaney-Wood, S., Stiell, B. and Lortie-Forgues, H. (2021) 'Review of EEF Projects':
https://educationendowmentfoundation.org.uk/public/files/Publications/Review_of_EEF_Projects.pdf

Appendix A: Information sheet for scale-up telephone interviews

PROJECT INFORMATION SHEET

Review of EEF grants:

Learning from scale-up from efficacy to effectiveness trials

Sheffield Institute of Education at Sheffield Hallam University has been commissioned by EEF to conduct an external review of their reported projects in order to identify overarching themes, patterns and other lessons beyond the headline impact on pupil outcomes.

Purpose of the study

The overall aim of this work is to provide an analysis of the published EEF RCT trials to identify patterns related to **who** and **what** has been successful and **why**, particularly in terms of scale up.

The study will identify overarching themes, patterns and other lessons beyond the headline impact on pupil outcomes. The intention is to look 'beneath the surface' of single projects to identify cross cutting issues which will be useful in EEFs future grant-making work and inform their approach to scale-up.

Our methods involve looking at both the quantitative impact results and implementation and process evaluations through: an exploratory review of EEF reports; meta-analyses to explore the relationship between programme characteristics and outcomes; and conducting telephone interviews with developers who have scaled up trials.

Why have I been chosen to take part in a telephone interview?

EEF has passed on your details as you have had a leading role in scaling up a project from an efficacy to effectiveness trial. Your experience and reflections will inform our learning on what works in scale up projects and improve the approach EEF take to these projects in future.

What will it involve?

Your participation in a telephone interview is voluntary. It will take up to an hour and will be arranged at a time that is convenient to you over the next few weeks. The interview will be recorded with your consent and transcribed.

What will be the focus of the telephone interview?

The telephone interview will explore what you have learned about the following areas in relation to moving from efficacy to effectiveness trial:

- Learning from efficacy trials - what did you change and why?
- Recruitment of schools
- Scaling up delivery approaches - what you did, and what you have learned about what worked well/didn't work well including:
 - Your approach to designing the intervention overall
 - your approach to managing training/CPD at scale
 - The CPD teaching and learning approach

- implementation in schools
- continuing engagement/monitoring in schools
- key contextual issues
- Any further general reflections on scaling up interventions

How will my data be used?

We will analyse your responses alongside others interviewed who were engaged in similar trials. Data collected from interviewees will be anonymised as far as possible and stored confidentially. All data will be stored securely on password protected computers, and will be held in compliance with the **SHU's Privacy Notice for Research Participants** and General Data Protection Regulations (GDPR). Ethics approval has been obtained through **Sheffield Hallam University's Faculty Ethics Committee**.

Will the project or organisation be identified?

The output from this study is an internal report to EEF. Given the small number of efficacy trials and the knowledge that EEF has of them, we cannot guarantee that you or your organisation will not be identifiable to EEF. The EEF may use the information to produce externally-facing reports, but in these cases your organisation would not be identifiable.

We will ask you to check your data prior to reporting. In any subsequent use outside of EEF neither you nor your organisation will be identified.

If you agree to take part, please complete the **EEF Review Consent Form**

Any questions?

Please contact Bernie Stiehl [b.stiehl@shu.ac.uk] or Louise Glossop [louise.glossop@shu.ac.uk] if any further information is needed about the telephone interviews.

The overall Project Directors are Dr Bronwen Maxwell [b.maxwell@shu.ac.uk] and Prof Mike Coldwell [m.coldwell@shu.ac.uk].

Centre for Development and Research in Education (CDARE), Sheffield Institute of Education, Sheffield Hallam University, Sheffield S1 1WB

Tel: 0114 225 6060

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v3.0.

To view this licence, visit <https://nationalarchives.gov.uk/doc/open-government-licence/version/3> or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third-party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

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



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

<https://educationendowmentfoundation.org.uk>

 [@EducEndowFoundn](https://twitter.com/EducEndowFoundn)

 [Facebook.com/EducEndowFoundn](https://www.facebook.com/EducEndowFoundn)