Schools Partnership Programme

Evaluation Report

March 2023

David Godfrey, Jake Anders, Louise Stoll, John Jerrim,
Bernardita Muñoz Chereau, Ruth McGinity, Toby Greany

The Education Endowment Foundation is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries and colleges to improve teaching and learning for 2 – 19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

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

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

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from the IOE, University College London's (UCL's) Faculty for Education and Society, including: Dr Jake Anders; Dr David Godfrey; Professor Louise Stoll; Professor John Jerrim; Dr Bernardita Muñoz Chereau; and Dr Ruth McGinity, with input from Professor Toby Greany (University of Nottingham).

The lead evaluator was Dr Jake Anders but, due to the changed focus of the evaluation resulting from COVID-19 disruption, Dr David Godfrey led most of the substantive work included in this report.

Contact details

Name: Dr David Godfrey

Address: IOE, UCL's Faculty for Education and Society

University College London,

20 Bedford Way,

London WC1H 0AL

United Kingdom

Email: david.godfrey@ucl.ac.uk

Acknowledgements

Many thanks to all the school leaders and teachers that helped in the evaluation of this programme, participating in the interviews and surveys, and allowing us to observe training and reviews. Also, many thanks to Qa Research, who helped conduct our telephone surveys with the school leaders.

Executive summary

The project

The Education Development Trust's (EDTs) Schools Partnership Programme (SPP) is a structured approach to partnership-based school collaboration and peer review. The programme is intended to build capacity and capability across clusters so they can gradually take more responsibility for their own development and maturity and lead their own improvement. Over time, local areas will own the SPP model, and continue to develop it so it has impact locally. The SPP provides frameworks and tools, training and professional support, and is designed to incorporate and build on schools' existing best practice.

The SPP focuses on three key elements: school self-review; peer review; and school-to-school support leading to follow-up improvement activity.

Each school nominates key members of staff to receive EDT training to become lead reviewers or Improvement Champions (ICs). Workshops and training occur at least once a term to prepare participating senior leaders, reviewers, and ICs, to track progress and assess impact and to provide top-up training for new staff. Schools then work in small clusters and take turns to review one another. The programme involves the use of ICs who lead improvement workshops based on review feedback to assist reviewed schools in deciding on actions from the reviews. Schools are expected to have at least one review and to review two per year and a '90-day' check-in discussion with the school, shared at partnership level to check progress towards the action points agreed in the improvement workshop after the review.

Table 1: Key conclusions

Key conclusions
Due to Covid-19 the primary outcome for this evaluation was not collected and so no measure of impact on attainment is reported. Key conclusions are based on qualitative data from the implementation and process evaluation (IPE). Teacher surveys found self-reported positive impacts on ownership of partnership and school improvement aims and the deepening of trust. While based on small numbers (and therefore to be taken with caution), many of these benefits appeared to be more deeply felt in schools with higher levels of pupil deprivation.
Evidence on perceived pupil-level outcomes is inconclusive—schools had a wide variety of improvement aims, which were directly related to the outcomes of the peer review, many of which changed during the evaluation especially during the second year of peer reviews where a different line of enquiry was agreed. Nevertheless, most participants felt that there had been a positive overall impact on pupils in their schools.
The IPE found that teachers perceived a strong impact of the Schools Partnership Programme (SPP) on leadership development. This came through reports on self-perceived acquisition of important evaluation and reviewing skills, and the social learning elements of peer review. The Improvement Champion role was felt to develop teachers and middle and senior leaders in collaborative school improvement. Where this role was carried out well, participants in host schools reported the coaching approach helped engender full staff ownership of improvement aims. However, capacity issues made the execution of these roles and continuity very difficult, especially in relation to leadership and other staff churn.
The majority of SPP participants regarded the programme as significant to the school, its staff, and students, and felt that it was aided by excellent methodology, support materials, and training. Most survey participants rated the SPP resources, such as the review framework and training materials, as being of very high quality. Participants felt that feedback received in reviews was mostly of high quality and had sharpened their ability to identify, act upon, and follow up on improvements, including being more evidence-informed. A large majority of school leaders surveyed said the time spent on the programme had been worthwhile and felt there had been a positive overall impact on pupils' outcomes at their school.
The complexity of the programme, including shifts in focus of school improvement objectives and the changing nature of the partnerships throughout the trial, caused limitations on capturing evidence on teacher quality and impact on pupils. Pupil impact data was ultimately missing due to the absence of Key Stage 2 Standard Assessent Tests assessments due to the COVID disruption.

This effectiveness trial, which ran between October 2017 and December 2021, was planned as an embedded mixed methods study involving a total of 422 primary schools in 89 clusters. The original evaluation design incorporated a quasi-experimental impact evaluation (school-level matched difference in differences), along with a primarily qualitative implementation and process evaluation (IPE). However, the disruption due to COVID-19 means that impact outcome data were not available, so only the IPE was completed (and this was in an adapted form). The IPE included initial and final surveys of treatment schools, a matched schools survey at the end of the trial, case studies of two clusters of treatment schools, interviews with matched schools, interviews of key stakeholders, observations of EDT training, and the reviewing process.

Additional findings

This trial was affected by COVID-19. As such, impact outcome data were not available, meaning that this study only reports findings from the IPE. Moreover, the impact of the pandemic, the extension of the trial period, the continued emphasis on Key Stage 2 when schools wanted to change the focus, and related lack of capacity resulted in a high proportion of schools that did not complete the full (extended) trial.

However, IPE identified a range of positive teacher-reported outcomes, especially to do with ownership of change and trust. At school level, the programme improved school leadership teams' self-reported abilities to identify its own strengths and weaknesses, to draw on expertise, and share partnership capacities. Interviews and observations showed that research-informed practices were encouraged by the programme and embedded in the training, particularly through the processes used in the improvement workshops.

Surveys of reviewers learned valuable skills of reviewing and self-evaluation; and school leaders learned how to collaborate more rigorously, transparently, offering increasing challenge to each other with time, and enabling more fluid knowledge exchange and shared professional development that benefited their staff and schools.

The IC role was particularly highly valued. The participants, who were often teachers or middle leaders as well as more senior leaders, were taught how to lead improvement workshops in schools other than their own school within the partnership. They worked in pairs and used skills of coaching and facilitative tools to facilitate evidence-based school improvement workshops and developed their ability to lead similar discussions in their own school. There were significant capacity challenges, especially for smaller schools or where the headteacher did not feel there were suitable staff to undertake these roles. This was mitigated by larger schools having more than one IC and sharing them across the partnership.

Participants perceived the programme to have been helpful in developing stronger partnerships, increasing levels of trust, sharing and transparency, and clearer structures to work towards shared improvement objectives. These self-reported benefits were more likely to be found in schools with pupils that had higher levels of deprivation and these benefits were perceived as significantly greater in the SPP schools than matched schools. The majority of participants reported that SPP training, facilitation, its framework, and materials were of high quality. Observations of training indicated that sessions were well-implemented by SPP facilitators.

Considering the current methodological limitations due to the pandemic, and the change in school improvement foci, the programme presents challenges in capturing teacher- and pupil-level impact. One way to address this is to constrain the improvement goals more explicitly across partnerships. However, this to some extent takes ownership away from school participants, since the focus of reviews in the SPP is based on the school's self-evaluation and pre-review meeting. Nevertheless, for the purposes of a trial, with voluntary take up, this may ensure sharper focus for impact measurement.

Future research could look at the role of ICs in leading evidence-informed school improvement and the development of trust-based accountability through peer review and school partnership working.

Introduction

Background

In March 2020, delivery of the Schools Partnership Programme (SPP) was severely disrupted by COVID-19 restrictions and the resulting major disruption to schools. The subsequent cancellation of Key Stage 2 assessments in both 2020 and 2021 meant the outcome data planned for this project were no longer available and, as such, it was not possible to proceed with the planned impact evaluation. In the interest of transparency, this report sets out the planned impact evaluation design (a quasi-experimental matched difference in differences design at school level), reporting of the construction of the matched comparison group, and resultant balance between programme and comparison groups at the outset of delivery. Other than that reporting, this report focuses on findings from the implementation and process evaluation (IPE), which continued in an adapted form.

The need for lateral school-to school partnerships has become apparent in the face of evidence that neither top-down centrally imposed change, nor pure competition can achieve sustained improvement across school systems (Burns and Koster, 2016). The aim, rather, has been to 'unleash greatness' by asking school system leaders to work together in ways, which transfer knowledge, expertise, and capacity within and between schools, so that all schools improve, and all children achieve their potential (DfE, 2010).

This has implications for accountability, with the drive for a 'self-improving school system' leading to an increase in engagement in peer evaluation to promote self-accountability (Greany and Higham, 2018). This is seen as a key step towards self-regulation, in which schools take greater ownership of their quality assurance, not only through self-evaluation, but through exposing their work to the scrutiny and perceptions of trusted peers (Matthews and Ehren, 2017). This accords with the outcomes of an international comprehensive survey of assessment and evaluation in 28 countries by the Organisation for Economic Co-operation and Development (OECD) (2013). In finding little evidence of peer review, the OECD report's authors identified developing school evaluation capacity as a priority, proposing that school leadership teams collaborate to identify common challenges and devise common approaches to peer evaluation.

More recently, research indicates growth in the uptake of peer review approaches, with England being a world leader in this area, but with several notable other international cases across many jurisdictions (Godfrey, 2020a). School peer review can be seen as a form of school internal evaluation and a supplement to the regional and national accountability system, in which formative evaluation and greater levels of professional accountability are emphasised compared with external forms of accountability such as inspections.

The Education Development Trust's (EDTs) SPP is a structured approach to cluster-based school collaboration, through the provision of a coherent and consistent approach to peer review that aims to drive improvement across all schools involved in the cluster. The programme aims to develop a culture of partnership working through school self-evaluation, peer review, and school-to-school support. England's approach to school peer review started principally during the much-vaunted 'London Challenge', leading to the establishment of the pioneering Challenge Partners (CP) organisation (Berwick, 2020). This programme was independently evaluated and shown to have promising outcomes, including providing strong quality assurance and challenge for partner schools through high-quality training and reviews (Matthews and Headon, 2015). This evaluation builds on this research, adding to the empirical data and with a longitudinal element that enables us to track progress and perceptions of the SPP over time.

As with CPs, the EDTs SPP involves elements of self-evaluation, peer review, and school-to-school support. The two programmes (SPP and CP) also have some important differences in their methodology, for instance SPP uses Improvement Champions (ICs), and has its own review framework, not one based on the national inspection criteria. The SPP team developed the framework with reference to practices they considered to be highly effective from the global evidence base.

While it is beyond the scope of this report to discuss these in detail, it is worth noting that findings that apply to this SPP model may not always generalise to other programmes involving peer review. The peer review landscape in England has also moved on since 2015, given the rise in uptake by schools (and later confirmed in our own survey findings). The SPP can be described as a school improvement partnership, within which peer review is both a central feature of its principles and a key modus operandi. Some research, points to concerns over the unintended

consequences of peer review, such as institutional isomorphism and self-policing, particularly in systems such as England with high-stakes accountability regimes (Greany, 2020). No research has been published to date that finds robust evidence of peer-review approaches being linked to improvements in student examination outcomes. Promising findings about the effects of peer review on the improvement of professional accountability, and on leadership development have been cited in previous research (Godfrey, 2020b).

Intervention

The SPP is a peer review model that is intended to build capacity and capability across clusters so they can gradually take more responsibility for their own development and maturity and lead their own improvement. Over time, local areas will own the SPP model, and continue to develop it so it has impact locally. The SPP provides frameworks and tools, training and professional support, and is designed to incorporate and build on, not side line, schools' existing best practice.

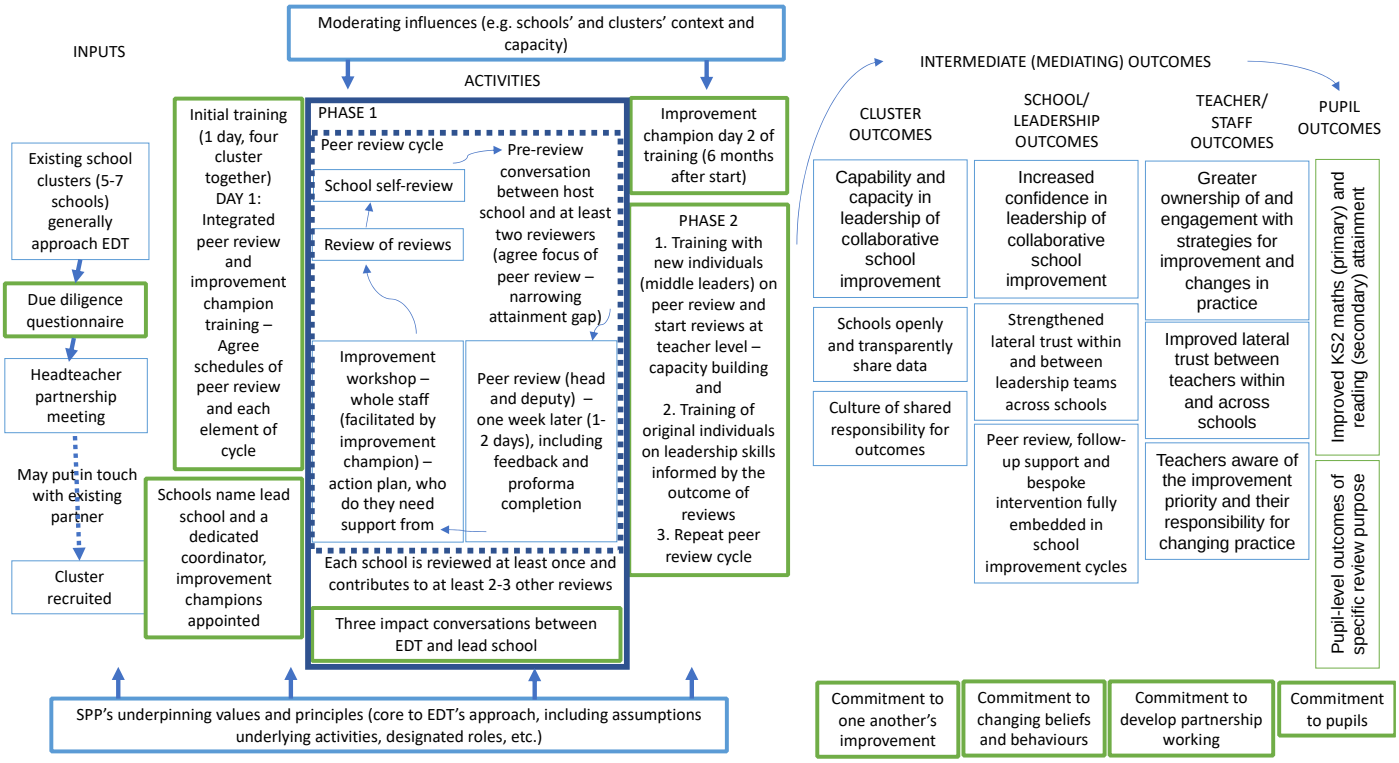

Figure 1: Logic model. EDT = Education Development Trust; KS2 = Key Stage 2; SPP = Schools Partnership Programme.

The SPP focuses on building capacity and capability across partnerships to engage effectively in the three elements:

- school self-review;
- peer review; and
- school-to-school support leading to follow-up improvement activity.

Key features of the model are as follows:

- Self-formed partnerships sign up to the programme and a partnership lead is decided from within this group of schools. The partnership lead is usually one of the headteachers in the partnership. The partnership lead completes a partnership 'readiness audit' to reflect on how established the schools are in their existing work together, including levels of trust, alignment of accountability systems, and impact of prior partnership work. They are encouraged to build on this through the programme.

- School leaders decide, which members of staff to put forward for EDT training in reviewing (the first reviewers put forward were the headteachers themselves and/or senior leaders) and in the ICs roles (these varied from teachers, middle leaders to senior leaders).
- Workshops and training occur at least once a term to prepare participating senior leaders, reviewers, and ICs and to provide top-up training for new staff. Some of these were initially suspended and later switched to online mode due to COVID.
- Schools work in small clusters and take turns to review each other. These were initially suspended due to COVID and later an online or hybrid 'rapid review' option was presented to participants. Many schools waited until face to face reviews were possible again and did not take up this option.
- School cluster activity is coordinated by a partnership lead who was the main contact with the EDT.
- The review process consists of pre-review, review, an improvement workshop, and a 90-day follow up.
- Schools would be expected to receive at least one review per year and reviewers, who work in pairs, will review two schools per year.
- Schools conducted a self-evaluation using the SPP framework that asked participants to consider 'to what extent …' questions. These were designed to set the tone of the review as an evidence- and enquiry-based process.
- A pre-review meeting between lead reviewer and headteacher of the host school starts the process of review, deciding the focus and finalising the timetable for the visiting review team.
- Reviews last 1 day, led by a lead reviewer (normally a headteacher) and assisted usually by another reviewer (a senior leader). Typically reviews involve scrutiny of documents and school data, classroom observations, and interviews with staff and students.
- The review day ends in a feedback discussion between reviewers and the host school leadership team.
- ICs from other schools (usually two) would attend final feedback in the review.
- ICs lead a facilitative coaching discussion with the host school staff in an 'improvement workshop' at this school, approximately 2 weeks later, normally after school and lasting around 1.5 hours. Staff in the host school are encouraged to come up with their own solutions to the review report findings and recommendations. ICs are encouraged to introduce research evidence on the topic of the review in this workshop.
- Schools are expected to have '90-day' progress discussions at school with the lead reviewer level to check progress after the review. This normally occurs between headteachers in the partnership and monitored by the partnership lead.
- EDT workshops encourage half yearly and full year reflections on progress and identifying ways to implement changes and to adjust practices to work more effectively in partnership.
- The cycle would be repeated, usually in the second year (although in this trial there were interruptions and extensions due to COVID).
- The above process seeks to improve cluster-, leadership-, and teacher-level outcomes that ultimately impact on pupil achievement, both in Key Stage 2 numeracy and literacy as well as other self-chosen outcomes (see Figure 1). At cluster level, there should be increases in collaborative leadership, improved transparency and data sharing, and the development of a culture of shared responsibility. At leadership level, there should be increased leadership of collaborative school improvement, strengthened lateral trust between leadership teams, embedding of peer review, and the follow up of school-to-school support. Finally, at teacher level, the programme is designed to increase ownership and engagement in strategies for improvement, improved lateral trust between teachers within and across schools, and improved awareness of improvement priorities and responsibility for changing practice.
- The above changes at cluster, leadership, and particularly at teacher level then should impact pupil-level outcomes, which would vary according to the focus of the reviews (the original outcomes were focused on Key Stage numeracy and literacy for this trial).

Evaluation objectives

The primary objective of this evaluation was to estimate the effect of participating in the EDT SPP for 2 years on pupils' maths attainment. **Due to cancellation of Key Stage 2 National Curriculum tests in the summer of 2020**

and the summer of 2021, it was not possible to provide evidence on this objective. When Key Stage 2 National Curriculum tests were initially cancelled for summer 2020, we considered the possibility of using summer 2021 tests as our post-intervention outcome of interest, but the subsequent cancellation of these tests removed this option.

The planned impact evaluation also intended to answer the following secondary research questions, **which again were not possible for the same reason**:

1. Does participating in EDT SPP have an effect on the reading attainment of young people?
2. Does participating in EDT SPP have an effect on the reading and maths attainment of young people ever identified as eligible for free school meals (FSM)?

Due to two national lockdowns, first in spring 2020 and then again in 2021, many of the IPE activities were curtailed. These re-started from Summer Term 2021 with the Education Endowment Foundation (EEF) support for continued delivery of the SPP until the end of Autumn Term 2021. The IPE questions were modified to reflect the lack of impact pupil attainment data (see above) and subsequently seek to answer the following:

1. In what ways does the SPP influence the capability, culture, and practice of partnerships, leadership, and teachers in involved schools?
2. In what ways do the elements in the SPP theory of change work in achieving participants' perceived forms of impact?
3. What factors influence schools' and clusters' ability to engage in, participate fully in, and successfully implement and sustain their involvement in the programme?
4. What distinguishes schools and clusters that have not continued with the programme?
5. What difference has COVID-19 made to the operation, participant engagement, and perceived forms of impact of the SPP (i.e. the research questions above)?

The final version of the **evaluation protocol** and **statistical analysis plan** (both amended due to COVID-19 disruption) are available from the EEF website (https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/schools-partnership-programme-spp).

Ethics and trial registration

The project's aims, methods, and materials were reviewed through the processes laid out by the research ethics committee of the IOE, University College London's (UCL's) Faculty for Education and Society, and were approved on 20 March 2018.

Schools were informed about the trial through initial information from the developer and formally committed to participation by signing a memorandum of understanding (MoU). A template version of this document is included as Appendix 10 to this report.

The trial was pre-registered at **www.controlled-trials.com** and was assigned an **International Standard Randomised Controlled Trial Number (ISRCTN)** that is **ISRCTN20687346** (although, for avoidance of doubt, it is not a randomised controlled trial)**.**

Data protection

As part of this project, we minimised the amount of personal data processed by working primarily with school-level data, which does not relate to a natural living person and is, hence, not considered personal data—it is also freely available from the Department for Education website. Nevertheless, teachers' personal data was processed as part of the IPE. For this reason, it was important that we processed this data lawfully, following the principles laid out in the Data Protection Act 1998 until May 2018 and the European Union (EU) General Data Protection Regulation (GDPR) from May 2018 until December 2020, and the UK GDPR from January 2021 (the project spanned these three regulatory periods). We explain the lawful basis below with respect to the GDPR, but there are equivalent regulations in the Data Protection Act 1998 for the justifications set out below.

UCL used Article 6(1)e of the GDPR as the lawful basis for processing personal data as part of this project. This is generally known as the 'public task' basis. UCL has reviewed current Information Commissioner's Office (ICO) guidance available here: https://ico.org.uk/for-organisations/guide-to-the-general-data-protection-regulation-

gdpr/lawful-basis-for- processing/public-task/, and has determined that this research forms part of its performance of a task in the public interest, as one of its core purposes provided for in its Charter and Statutes.

We do not believe that any of the data we processed falls within the definition of special category data under the GDPR. This would have required an additional justification under Article 9(2) of the GDPR.

Data will be kept until the end of the research project, including academic paper writing and dissemination (and certainly not longer than 10 years, in line with UCLs policy on data retention). When it is deleted, it will be securely destroyed.

We informed teachers about the proposed data processing as part of providing informed consent to take part in IPE activities, meaning they were free to object to this use of data. The data controllers were named in the privacy information provided as part of this project and contact details provided in case participants had any queries about the data we hold about them, including provision and deletion of their data. The relevant letters and forms have been reproduced in Appendix 1. The information provided to parents explained in clear and plain language: the lawful basis for processing; the purpose to which we put the data; that they did not have to take part in the research, and that they could object to the data processing at a later date and this would be respected; contact details of the organisation; and categories of data that have been processed.

Project team

The impact evaluation was led by Dr Jake Anders at IOE, UCL's Faculty of Education and Society, with support from Professor John Jerrim. The process evaluation was led by Dr David Godfrey (initially by Professor Louise Stoll, who transitioned to a supporting role), along with Dr Ruth McGinity and Dr Bernardita Muñoz Chereau, supported by Professor Toby Greany at the University of Nottingham.

The project team at the EDT comprised:

Programme leads: Anne Cameron; John Cronin (left) / Andrew Ettinger (left).

Assistant programme manager: Jenni Rolls (Left); Sarah Darch had a coordination role, then left in July 2020 and re-joined as programme manager in September 2021.

Administration: Kate Li.

Facilitators: Maggie Farrar; Jane Creasy; Niki Thomas; Vanessa Huws Jones; Nicola Woolf; and Pam Butterworth.

Methods

Evaluation design

The sections below on trial design, participant selection, outcome measures, matching, and difference in differences describe the planned impact evaluation that was cancelled due to the national disruption caused by COVID. These have been retained partly for transparency but also to understand how schools were recruited onto the trial and how we matched case studies and survey participants for the IPE data collection.

Table 2: Trial design

Evaluation design		Matched difference in differences
Unit of analysis		School
Matching variables		Reading and Maths average attainment in the school in 2017, average Key Stage 1 attainment of those who took Key Stage 2 tests in 2017, Number of pupils, Academy status, Most recent Office for Standards in Education, Children's Services and Skills (Ofsted) inspection, income deprivation affecting children index (IDACI) quintile, and government office region
Primary outcome	**Variable**	Maths attainment
	Measure (instrument, scale, source)	Average Key Stage 2 maths scores (mat_average)
Secondary outcome(s)	**Variable(s)**	Reading attainment
	Measure(s) (instrument, scale, source)	Average Key Stage 2 reading scores (read_average)

This effectiveness evaluation was designed as an embedded mixed methods quasi-experimental evaluation, incorporating a school-level matched comparison difference in differences impact evaluation, with an IPE.

This approach to the impact evaluation was chosen because a randomised controlled trial would not have been feasible with this programme. This was partly due to the scale required (because of the grouping of schools into clusters) and because of the practical and ethical difficulties of forming schools into clusters, while expecting them not to cooperate if allocated to a control group. As a result of using a matched design, the control condition is effectively 'business as usual'.

The primary outcome was to be change in school's average Key Stage 2 maths scores between the academic year immediately pre-treatment and 2 years subsequently (further details below), as compared to the change over the same period in those same scores among the matched comparison group. Scores would be taken from publicly published school performance tables on the Department for Education website. The secondary outcome was to be the same but with average Key Stage 2 reading scores.

Participant selection

Recruitment to the study was carried out by the project team (EDT), drawing on their national databases to contact schools. To be considered, schools were required to agree to cooperate with the project and evaluation teams during the trial. The project team advertised the trial and approached schools through their existing networks. Where possible, they aimed to recruit schools that have larger populations of individuals receiving FSM.

Target recruitment was 50 clusters of English state-funded primary schools, to be recruited with an approximate cluster size of 6, making 300 schools in total. The proviso was agreed that if cluster size was smaller than expected, additional recruitment would be undertaken. The project team (EDT) successfully recruited far more schools than anticipated, providing a sample of 437 English state-funded primary schools in 85 clusters (average cluster size of just over 5).

All recruited schools received the intervention, while statistical matching methods were used to identify the comparison group (described further below). No recruitment was carried out for this counterfactual group, as all outcome data was expected to be obtained from publicly available school performance data on the Department for Education website.

Outcome measures

Baseline measures

As this evaluation was designed as difference in differences, there is more than one way to think of baseline measures.

Perhaps most obviously in the context of a difference in differences evaluation, the primary (and secondary) outcome in the year before treatment began is a baseline measure, in that we planned to consider change between this and the same measure in matched schools—these measures are described below. While school-level attainment measures can be 'noisy' in a given year, this is a larger problem where such measures are used for individual accountability. Measurement error should average out between schools for the purposes used here, although it could cause some degree of attenuation bias of the estimated impact. However, the alternative of using a measure averaged across multiple years would mean a less clear attribution of change to the treatment that comes from using a single time point before treatment for baseline, for example.

In addition, (before the impact evaluation had been cancelled) we planned to adjust our analysis for an additional baseline to improve the precision of our impact estimates and, given the context of this as a quasi-experimental evaluation rather than a randomised controlled trial, to reduce bias. Specifically, we planned to include the average attainment of the relevant cohort of each participating school's Key Stage 2 intake,[1] as proxied by average Key Stage 1 pupil attainment (tks1average). This would provide a measure of the average attainment across reading, writing, maths, and science of the relevant cohort of pupils when they were in year two (i.e. 4 years before the primary and secondary outcome measures) based on teacher assessment. Given this gap, these would have been based on assessments carried out prior to the participating schools' involvement in the project. Further details regarding these assessments and summary measures are available from the Department for Education.

Including this baseline measure would be somewhat analogous to including a baseline measure as a covariate in the context of a randomised controlled trial, albeit with some differences. After including the baseline measure as a covariate, our model would have estimated the change in outcome measures we observed, conditional on changes in the intake attainment profile of the school, namely, attempting to compare schools, holding constant the attainment profile of their intakes, particularly if these change over time (matching should largely take care of ensuring we compared schools with similar intake attainment profiles at baseline).

Primary outcome

The primary outcome of interest was school average performance in Key Stage 2 maths tests (mat_average). More specifically, given that this was a difference in differences design, the outcome would be the difference in differences in Key Stage 2 maths test scores between treatment and matched comparison schools in pre- and post-treatment years.

As the name suggests, this is a measure of pupils' maths attainment at age 11, based on statutory Key Stage 2 National Curriculum assessments sat by pupils in all state-funded schools and reported to the Department for Education (DfE). The measures are carried out subject to DfE rules on administration and are externally marked, which should mitigate concerns about blinding. Further details regarding these assessments and summary measures

[1] While this is truly school intake for junior schools, for primary schools this measures attainment part way through the school. However, this is the most appropriate measure to use consistently across all schools with Key Stage 2 phases.

are available from the DfE's website. Raw scores provided in DfE's Compare Schools Performance online database would have been the specific variable used.

The SPPs logic model includes impacts on pupil attainment, including in core subjects of English and maths. Which of these to make the primary outcome measure was discussed at project setup and, ultimately, it was anticipated that a larger impact would be seen on maths based on experience of preponderance of schools' areas of focus during the SPP.

Secondary outcomes

The secondary outcome of interest was school average performance in Key Stage 2 reading tests (read_average). More specifically, given that this was a difference in differences design, the outcome would have been the difference in differences in Key Stage 2 reading test scores. The same specifics regarding administration, measure, and rationale apply as for the primary outcome above.

Additional secondary outcomes would have been the outcomes above for the subgroup within schools of pupils eligible for FSM. These would have been recovered using school-level variables publicly available from the DfE for this purpose (mat_average_fsm6cla1a and read_average_fsm6cla1a), and allowed us to examine differential performance of the programme on those from a disadvantaged background, consistent with the EEFs founding mission.

Matching

Once recruitment was completed, matching was carried out based on baseline characteristics available in schools' records in the DfE's 'Get Information About Schools' and 'Compare Schools Performance' databases, along with linked in data on Office for Standards in Education, Children's Services and Skills (Ofsted) ratings and schools' geographical context. Characteristics that were assessed for inclusion in matching (including interactions and higher order polynomials of these terms) were:

- number of pupils;
- attainment in school measured by Key Stage 2 average points score in each previous year 2010–2015;
- prior attainment of intake measured by Key Stage 1 average points score in each previous year 2010–2015;
- school type (academy vs other);
- Ofsted rating;
- income deprivation affecting children index (IDACI) quintile group;
- geographical location (longitude, latitude);
- local authority area; and
- government office region.

For clarity, matching was planned to be carried out on the basis of a 1:1 nearest neighbour propensity score matching without replacement, including exact matching on key characteristics (likely to include school type and government office region), application of a caliper on the propensity score matching, and imposition of common support.[2] Since clusters are observed in the treatment group but not in the pool of potential matched comparators, it is necessary to match at school level, rather than at cluster level; the importance of the clustering of treatment schools was planned to be recognised in the analysis.

The day that the matching exercise was completed and the resulting sample of successfully matched schools and comparators was identified and confirmed was considered the project's pseudo-randomisation date, with the final

[2] Imposing common support ensures that treatment schools with a treatment propensity above that of the highest propensity comparison schools, and comparison schools with a treatment propensity below that of the lowest propensity treatment schools, are removed from the matched, and hence, analysis sample. As such, it helps to ensure internal validity, albeit with some trade-off to the representation of the full sample of treated schools.

treated and matched samples considered fixed at this point, before any outcomes data were expected to be available. This was to ensure that outcomes could not affect the matching process in a way that could bias the results.

Ultimately, in our preferred specification, matches were identified using a nearest neighbour algorithm[3] with no replacement (in practice, allowing replacement makes no difference in this application, seemingly because there are plenty of potential matched comparators available) using the MatchIt package in R, based on a treatment propensity score estimated using the following characteristics:

- average Key Stage 2 reading score of the school in 2017;
- average Key Stage 2 maths score of the school in 2017;
- average Key Stage 1 score of the school's intake (among those taking Key Stage 2 tests in 2017);
- the number of pupils in the school in 2017;
- whether the school is an academy in 2017;
- school's most recent Ofsted rating in 2017;
- the quintile group into which the school falls in terms of the average IDACI of its intake; and
- and the government office region in which a school is located.

This model was based on iterative testing of model fit of the matching variables proposed in the evaluation protocol, with an important exception. In the evaluation protocol (and as noted above), we proposed the potential use of lagged performance variables to improve the probability of achieving common trends in our matched sample. However, further reading has suggested that, while this would appear to improve the plausibility of this assumption, in fact it may cause problems with regression towards the mean in the future trends that we will use to estimate the treatment effect (Daw and Hatfield, 2018).

A 1:1 nearest neighbour matching based on the estimated propensity score was carried out without replacement and enforced exact matching on the school's IDACI quintile, on urban/rural classification, and on school region. These characteristics were chosen for exact matching to ensure we compare between schools with similar contexts and intakes, but without resulting in large reductions in the sample for whom we can find a match: exact matching on Ofsted rating was explored but rejected as it reduced the sample of schools that could be matched without offering an obvious benefit (see details of alternative specifications in the project's Statistical Analysis Plan). We used a caliper of 0.2 in line with the advice of Austin (2011). Schools outside the range of common support were excluded, although when we tested removing this restriction, this revealed that it was resulting in only one treatment school being excluded.

Difference in differences

While matching attempts to ensure that the treated and comparison group are comparable based on observable characteristics, there is still the potential for confounding due to unobservable differences between the treated and matched comparison schools. We planned to take a difference in differences approach to deal with remaining time-invariant unobservable characteristics. This means that our impact estimate assumes common trends, namely that, in the absence of the treatment, the change in our outcomes of interest between the pre- and post-treatment period would have been the same between our treatment and our matched comparison schools (Anders *et al.,* 2017, Ch. 4). We aimed to improve the plausibility of this assumption by allowing for the possibility of considering previous years' average attainment as part of our matching (although see discussion above).

We planned to compare the differences in outcomes between treated and matched comparison schools in 2017/18 (pre-treatment) with the difference in outcomes between treated and matched comparison schools in 2019/20 (post-treatment)—2 years after the intervention began. In the wave of the initial COVID-19 disruption, we proposed to change the outcome time period to be 2020/21 given the cancellation of Key Stage 2 National Curriculum assessments in 2019/20—the subsequent cancellation of these assessments in the 2020/21 academic year precipitated the abandonment of this project's impact evaluation.

[3] A nearest neighbour algorithm was chosen to prioritise the intuition and transparency of the selecting comparison schools.

School-level compliance categories and criteria

The school-level compliance categories and criteria were defined as follows. Please note that some changes from the originally proposed criteria are documented in a revised Evaluation Protocol.[4]

Table 3: Compliance criteria

Categories	Attendance to training/workshops*	Review visits**
Non-compliant schools	No attendance/dropped out	No review visits/dropped out
Minimally compliant schools	Attendance to less than 75% of training sessions/workshops	Hosted one peer review visit across programme
Fully compliant schools	Attendance to 75% or more of training sessions/workshops	Hosted two review visits across programme (if they had not completed two before COVID-19 lockdown) or at least three (if they had completed two before COVID-19 lockdown)

*Source of data: *Education Develoment Trust (EDT) database; **EDT database.*

To be categorised as minimally or fully compliant, both 'Attendance to training/workshops' and 'Review visits' criteria had to be met.

Cluster-level compliance categories and criteria

Cluster-level compliance categories were based on aggregation of school-level compliance of all schools within a cluster. This aggregation was carried out as follows:

- **Minimally compliant**: A minimally compliant cluster contained no non-compliant schools and at least one fully compliant school.
- **Fully compliant**: A fully compliant cluster contained a maximum of one school that is minimally compliant and the rest are fully compliant.

All clusters that failed to meet either of these criteria were considered to be **non-compliant**.

Since the properties of using a method such as instrumental variables to conduct compliance analysis within the scope of our evaluation design are unclear, we instead planned (before COVID disruption curtailed this) to repeat our primary analysis among the following subgroups:

- all treatment schools judged to be at least minimally compliant and school-to-school matched comparators;
- all treatment schools judged to be fully compliant and school-to-school matched comparators;
- all treatment schools in clusters that are judged to be at least minimally compliant and school-to-school matched comparators; and
- all treatment schools in clusters that are judged to be fully compliant and school-to-school matched comparators.

Research questions and data collection

The original purpose of the process evaluation was to establish fidelity and to assess the factors, which affected impact from the different phases of the SPP on the stakeholders within the project, and which may explain the findings of the quantitative evaluation. Due to the absence of the impact data caused by COVID-19, the research questions were updated to look at *perceived impact*. We also wanted to examine factors leading to sustaining or dropping out of the programme and at the impact of the pandemic on the programme. Thus, our research questions were:

[4] Specifically, compared to the originally defined criteria in the evaluation protocol, we clarified the school-level compliance criteria to remove ambiguity, particularly in the context of disrupted and extended delivery due to COVID-19 restrictions. Beyond simple clarification, we also varied the review visits criterion to be considered a fully compliant school in order not to exclude schools whose year two visits were disrupted by COVID-19 restrictions, but successfully resume these in year three. Likewise, we clarified the same criterion to ensure that those who did conduct visits in year two had to continue these in year three in order to be considered fully compliant.

1. In what ways does the SPP influence the capability, culture, and practice of partnerships, leadership, and teachers in involved schools?
2. In what ways do the elements in the SPP theory of change work in achieving participants' perceived forms of impact?
3. What factors influence schools' and clusters' ability to engage in, participate fully in, and successfully implement and sustain their involvement in the programme?
4. What distinguishes schools and clusters that have not continued with the programme?
5. What difference has COVID-19 made to the operation, participant engagement, and perceived forms of impact of the SPP?

We were interested in factors affecting implementation at partnership, leadership, teacher, and school levels, and perceptions of impact (intermediate outcomes). Intermediate outcomes were influenced by EDTs own Theory of Change, shown in their learning map in the SPP materials (see Appendix 11):

Partnership intermediate outcomes:

- capability and capacity in leadership of collaborative school improvement;
- culture of shared responsibility; and
- open and transparent systems and processes.

Leadership intermediate outcomes:

- increased confidence in leadership and collaborative improvement;
- strengthened lateral trust within and between leadership teams; and
- peer review, follow-up support, and bespoke intervention fully embedded in school improvement cycles.

Teacher intermediate outcomes:

- greater ownership of and engagement with strategies for improvement and changes in practice;
- improved lateral trust within Key Stage 2 teams; and
- Key Stage 2 teachers aware of the improvement priority and their responsibility for changing practice.

School-specific outcomes:

- Unspecified as this will be different for each school.

The above intermediate outcomes are implicit in research questions 1–3, albeit the specific focus on Key Stage 2 teams was mostly due to the initial logical model that had Key Stage 2 impact outcomes (see Appendix 2). We also generated factors related to engagement and impact from our own prior research, and Vaughan and Albers (2017) such as:

Contextual and capacity issues at the school level

Ofsted rating, stability of staffing, commitment of school headteacher, time, quality of Key Stage 2 leaders, deprivation indices, length of time as headteacher, teachers' lack of engagement with partnership/reticence, capacity to improve funding, and readiness, for example prior participation in peer review, peer review embedded in school improvement cycle, and nature of previous experience of self-evaluation.

Contextual and capacity issues at the partnership level

Shared vision, partnership systems, partnership status, drawing on support, commitment to other schools' success/each other's improvement and pupils, trust, open sharing of data, quality of partnership leadership, length and maturity of partnership, commitment to changing beliefs and behaviours, commitment to developing ways of partnership working, governance/partnership—multi-academy trust (MAT), local authority, and readiness at partnership level.

In addition to the above research questions we also looked at: fidelity (if the intervention was delivered as intended); adaptation; (how the intervention was changed to meet circumstances); dosage (how much of the intervention was delivered); reach (if the intervention was delivered to the intended recipients); responsiveness (recipient engagement during the delivery of the intervention); quality (how well the intervention was delivered); and differentiation (how different the intervention was from existing practices). Most of these are addressed separately from the research questions in the findings, although adaptation during COVID-19 is addressed explicitly under research question 5.

Overall, our approach was to strengthen the validity of our findings by triangulating our data collection from multiple data collection methods (interviews, observations, and surveys); multiple sources of data—including from the SPP permanent staff, associated staff (facilitators), partnership leads, headteachers, senior and middle leaders, across different schools and regions; and over various time points. Validity was further strengthened by having multiple perspectives of researchers through visiting the same kinds of programme activities and through holding regular team meetings in which we discussed emerging findings. We also explored the counterfactual, using interviews and surveys with matched schools. Table 4 below gives an overview of IPE methods. The programme was initially planned for 2 years but extended into 3 years due to the COVID-19 school lockdowns and two pauses to SPP activity, therefore: Year one (2018/19), year two (2019/20), and year three (2020/21). Our data collection covered each of these phases. The data collection instruments were informed by the logic model developed in partnership with the SPP (see Figure 2), using the experience of team members on comparable projects and with reference to published work on IPE, especially for complex interventions (Humphrey *et al.*, 2016; Vaughan and Albers, 2017; Anders *et al.*, 2017). Detailed notes, discussions and early reports from earlier data collection also contributed to the development of year two/year three interviews and end surveys to the treatment group and matched sample.

Table 4: IPE methods overview

Research methods	Data collection methods	Participants / data sources	Data analysis methods	RQs addressed	Implementation / logic model relevance
Baseline surveys	Telephone surveys	Headteachers and senior leaders at treatment schools n=339 (out of 422*). No specific target except to achieve a high completion rate	Descriptive statistics, cross-tabulations, and thematic coding of open responses	RQ1, RQ2, RQ3	Business as usual, context, readiness
Final survey of treatment schools	Telephone surveys – routed survey to distinguish completed schools with those that withdrew before November 2021 and after December 2018	Headteachers and senior leaders at treatment schools n=157: 134 active and 23 early withdrawn (out of 422 treatment schools total). The initial target was 200 treatment schools, including 150 completed schools and 50 withdrawn	Descriptive statistics, cross-tabulations, and thematic coding of open responses	RQ1, RQ2, RQ3, RQ4, RQ5	Context, implementation, and factors affecting implementation
Matched schools survey	Telephone surveys	Headteachers and senior leaders at matched schools n=44 (out of 374 potential matches). Initial target was 220 schools, revised down to 50	Descriptive statistics, cross-tabulations, and thematic coding of open responses	N/A	Business as usual/matched school comparisons
Qualitative case studies	Interviews and observations	Thirty-four interviews with headteachers, senior and middle leaders at eight schools (out of ten) in two case study clusters (telephone and in person). Observations of three review processes	Thematic analysis of interview transcripts and observational notes	RQ1, RQ2, RQ3, RQ4, RQ5	Business as usual, context, readiness, implementation, and factors affecting implementation
Matched schools case studies	Interviews	Seven telephone interviews with headteachers in four matched schools	Thematic analysis of interview transcripts	N/A	Business as usual/matched school comparisons
Attendance at EDT training events	Observational notes	Headteachers/EHTs/SLT/middle leaders and teachers at 24 training events	Thematic analysis of observational notes	RQ1, RQ2, RQ3, RQ5	Context, implementation, and factors affecting implementation
Shadowing review processes	Observations	Observations of six review processes: four full review processes, one partial (no improvement workshop) and one online	Thematic analysis of observational notes	RQ1, RQ2, RQ3, RQ5	Context, implementation, and factors affecting implementation
Stakeholder interviews	Group and individual interviews in person and online. Short survey	Partnership leads, ICs, permanent and associated SPP staff n=86 participants, from 4 individual interviews and 14 group interviews	Thematic analysis of interview transcripts	RQ1, RQ2, RQ3, RQ5	Context, implementation, and factors affecting implementation
EDT evaluations and records	EDT partnership initial audit, records of school withdrawals, records of attendance to training and review completions	Attendee feedback in after EDT workshops and training. Partnership leads' records. Facilitators of training providing records of reviews. Initial partnership audits on readiness to engage completed by 64 partnership leads	Descriptive statistics and thematic coding of open responses	RQ4	Readiness, compliance, implementation, and factors affecting implementation

**The original 437 number reduced early on to 422 schools as some essentially did not start any part of the programme or its activities, so the 422 number is used herein. EDT = Education Development Trust; EHT = executive headteacher; IC = Improvement Champion; N/A = not applicable; RQ = research question; SLT = senior leadership team; SPP = Schools Partnership Programme.*

Surveys

Surveys were conducted by Qa Research (an independent research company) via telephone, following construction of the survey items by the IOE, UCL IPE team. The first stage of the baseline survey involved a pilot conducted by the Qa Research team on a small number of treatment schools. This did not result in substantive changes to the questionnaire items but helped provide clearer prompts to the telephone operatives. Most respondents in the baseline survey were headteachers (81%), and 11% were executive heads. Around 88% of participants had 11 or more years' experience in teaching and 91% had over 7 years senior management experience. The survey had a mixture of closed responses, Likert scales, and open-ended questions. The baseline (treatment schools only) survey (see Appendix 3) was conducted in June 2018 to July 2018, by which point schools had attended Spring Term training and preparation workshops and would have scheduled reviews for Autumn Term 2018. This survey was designed to provide baseline questions based on school and partnership variables such as trust and capacity to improve, information on other interventions that schools were engaged in regarding numeracy and literacy at Key Stage 2, as well as previous experience with peer reviewing, and a description of other partnership sources of school improvement and support.

The baseline survey had 339 responses from the treated schools total of 422 primary schools (80%). The final treatment schools survey was originally intended for June 2020 to July 2020, but due to the two lockdowns and extensions to the funding period of the programme, this was moved to the last week of November 2021 following the last online SPP celebration. Two of these four celebration events were attended by members of the IPE team to thank participants of the EEF evaluation and to promote completion of the end survey. In the other two events, a video message was played. The final survey was aimed at schools that had completed the full 3 years of the trial (up to November 2021) as well as schools that had withdrawn before this date, however excluding schools that had withdrawn prior to December 2018 as we decided that: a) participation from this group would be difficult to achieve; and that b) data from this group would be less interesting than ones that had at least finished the first phase of reviewing. This was a routed survey with overlapping sections and separate ones specific to whether they had withdrawn from the trial or had completed it.

This final survey matched several Likert scales on the baseline survey and asked additional questions based on revised research questions, including about perceived impact of the programme, the effect of COVID-19 on programme activity, and about withdrawing from the programme, if relevant. This also had other closed and open-ended response items. A member of the IPE team carried out the first telephone survey on one of the case study schools and this resulted in minor changes to wording of questions and prompts for telephone operators. The initial target agreed following discussions with the EEF team was 200 treatment schools, including 150 completed schools and 50 withdrawn, and for this to be completed before Christmas 2021. In the context of school leaders feeling over worked and tired from the pandemic, it was clear that this target was very difficult to achieve and, although the Qa Research team adopted protocols to ensure multiple calls to each school to maximise participation, the period of calls was extended to get closer to the target. This survey achieved a response rate of 157 schools (37% of total schools in the treatment school group) (134 active and 23 early withdrawn) where 'withdrawn' or 'completed' status were self-declared based on one of the survey responses rather than data provided to us by the EDT (there were minor discrepancies). Of these 157, 141 respondents had come from schools that completed both the baseline and final surveys.

Surveys of headteachers in matched schools (see Appendix 4) were conducted in parallel to the final treatment schools' survey (see Appendix 5). These were designed to compare business as usual and provide comparisons to equivalent questions about working in partnerships. For the matched schools' survey, we also constructed 'equivalent' worded questions to the final treatment schools' survey and asked participants to give their judgements based on 'their most significant school improvement partnership over the last 3 years'. Respondents were prompted with a definition of partnership as: 'School improvement partnerships are strong alliances or affiliations either between two or more schools, or schools and organisations, aimed at increasing participant schools' capacity to provide high quality education for their pupils'. We also asked treatment school survey respondents to state whether they judged the SPP as the most significant partnership over the last 3 years.

In anticipation of low motivation to participate in this survey due to lack of incentive and the pandemic, the initial response target of 200 schools was revised down to 50. This survey achieved a response rate of 44 from a matched schools' database of 374 schools (12%).

Case studies

Alongside the surveys, our case studies provided the most substantial data collected and in-depth contribution to answers to our research questions. We selected case study clusters based on having a cluster size of five schools; one cluster with predominantly high-deprivation indices and one predominantly with low deprivation indices, and one cluster in the South of England and one in the North of England (see Table 5, below).

Table 5: Case study school characteristics and labels

Partnership/ school case label	School type	IDACI quintile	Urban/rural	Location	Ofsted	Compliance
1a	Community	2	Rural town and fringe	North	2 (2019)	Fully
1b	Academy converter	5	Urban minor conurbation	North	3	Non-compliant*
1c	Community	5	Urban city and town	North	2	Minimally
1d	Community	5	Urban city and town	North	2	Minimally
1e	Community	1	Rural town and fringe	North	1	Fully
2a	Community	1	Urban city and town	South	2	Non-compliant*
2b	Community	2	Urban city and town	South	2	Minimally
2c	Voluntary-controlled	1	Urban city and town	South	2	Minimally
2d	Community	1	Urban city and town	South	2	Minimally
2e	Voluntary-aided	1	Urban city and town	South	2	Minimally

**Withdrew start of second year (pre-COVID). IDACI = income deprivation affecting children index; Ofsted = Office for Standards in Education, Children's Services and Skills.*

In year one (January 2019–February 2019), case study partnerships in the North (Cluster 1) and the South (Cluster 2) were identified, and interviews carried out to include the headteacher, a member of the senior leadership team (SLT) involved in reviewing, or the Key Stage 2 leader and a member of staff who had been trained as an IC (if the school had one). Interview schedules overlapped for triangulation purposes while pursuing unique lines of enquiry with Key Stage 2 leaders, other SLT, and those designated as ICs at the school (see Appendix 6). One school in the South (Cluster 2) did not participate in year one due to issues with staff illness. In year two, interviews in just two schools in the South (Cluster 2) were conducted before the first lockdown. One school from each cluster dropped out after 1 year. The remaining interviews in Cluster 1 were carried out remotely in year three during 2020/21, identifying two key people to follow up with. Cluster 2 largely paused activities on the programme in year three and we were not able to secure the remaining two schools in interview. Nevertheless, further data on the case study schools was collected from other data sources throughout the 3 years. We used this to triangulate the analysis for an internal report on the case studies (see Table 6, below).

Table 6: Data collected on case study schools

Partnership/ school case label	Interviews	EDT partnership audits	Initial survey	End survey	Reviews observed	Partnership leads group interview	EDT training and workshops
1a	Headteacher (also partnership lead) Year one (January 2019) and year three (March 2021) Total = 4	x			Year one focus on TAs: pre-review meeting on 17 January 2019, Review on 24 January 2019 IW: 31 January 2019 February 2021 (rapid online review)		Year three
1b	N/A	x	x				
1c	Year one: five interviews: headteacher (trained as peer reviewer); peer reviewer (AHT); IC (AHT), two KS2 leaders Year three (June 2021): two interviews: headteacher (trained as peer reviewer); peer reviewer (AHT) Total = 7	x	x				Year three
1d	Year one: two interviews: headteacher (trained as peer reviewer), DHT (peer reviewer) Year three (June 2021): two interviews: (new) headteacher (peer reviewer), new DHT (peer reviewer) Total = 4	x	x	x			Year three
1e	Year one: two interviews: headteacher (peer reviewer); AHT (peer reviewer) Year three (June 2021): two interviews: EHT (peer reviewer), acting head of school (peer reviewer) Total = 4	x	x	x			Year three
2a	Year one interviews (30 January 2019): interim head of school (IHoS) KS2 English leader (KS2) (IC) – interim assistant head Total = 3	x	x	x			Year one
2b	(30 January 2019) executive head teacher (EHT) and head of school together KS2 English leader, IC Total = 3	x	x	x	Pre-review (12 November 2019) Review 12 March 2020 IW (scheduled 17 March 2020): Did not happen due to pandemic		Years one/ two/ three
2c	Unable to gain participation of school	x	x				Years one/ two/ three
2d	28 January 2019 with headteacher at school. AHT via Skype interview on Friday 1 February 2019 Year two interview only with headteacher in person on 28 February 2020 Total = 3	x	x				Years one/ two/ three
2e	24 January 2019 28 February 2020 Headteacher (also partnership lead) DHT ICs: Maths teacher (interviewed year one) New IC for year two: Maths lead Total = 6	x		x	Pre-review: 21 September 2018 (20 mins recording) Review date: 25 September 2018 IW: 3 October 2018, 3.30 p.m. at school	x	Years one/ two/ three

AHT = acting headteacher; DHT = deputy headteacher; EHT = executive headteacher; IC = Improvement Champion; IW = improvement workshop; KS = Key Stage; TA = teaching assistant.

Matched schools case studies

A matched purposive sample returned 84 primaries, of which 44 were approached via email by the researcher. Of these 44 schools, 3 agreed to an interview with the headteacher (matched schools 1, 2, and 3). After a pause in the project as a result of the COVID-19 pandemic, the matched school sample of 84 schools were re-approached for either first or follow-up interviews. This resulted in a further one new participant interview (matched school 4) and a follow-up interview with two of the original three participant headteachers (matched schools 1 and 3). In total, seven interviews were undertaken across four schools. All participant schools were offered a £200 discretionary payment for their school funds. Table 7 contains a brief synopsis of each of the four participant schools, their school improvement priorities at the time of interview participants detail, and dates of interviews.

With the matched schools' interviews, we re-wrote the research questions to become 'equivalent' to our SPP research questions:

1. In what ways do matched schools engage in school-to-school support and how effective are these activities perceived to be for school improvement priorities and what impact were they perceived to have?
2. What factors influence schools' ability to engage in, participate fully in, and successfully implement and sustain their involvement in school-to-school support, partnership, and/or collaboration?
3. How do matched schools perceive the benefits and potential limitations of collaborative partnerships?
4. What difference has COVID-19 made to the school-to-school partnerships?

Table 7: Characteristics of matched schools interview sample

Matched school	School characteristics	School improvement priorities	Participant details	Dates of interview(s)
Matched school 1 (IDACI 1; Town) **(MS1)**	**2019:** • Single form entry village school • Ofsted: Good • Voluntary-aided • 188 pupils on roll **2020:** • 198 pupils on roll	**2019:** • Progress in maths • Mental health and well-being **2021:** • Progress in maths • Mental health and well-being	• Headteacher • Female • 2021 – 5 years in post **(HF1)**	• 10 June 2019 • 19 June 2021
Matched school 2 (IDACI 2; Urban) **(MS2)**	**2019:** • Locally maintained • Two-form entry • 580 pupils on roll • Ofsted: Good • Will join a local MAT with the feeder secondary school (August 2019)	• Writing (specifically boys) • Embedding a mastery approach across the curriculum • Assessment	• Interim headteacher • Male • 2019 – 10 months in post **(HM2)**	• 21 June 2019
Matched school 3 (IDACI 2; Urban) **(MS3)**	**2019:** • Locally maintained • 275 pupils on roll • 72% English as an additional language • Ofsted: Good **2021:** • 233 pupils on roll	**2019:** • Writing • Pakistani childrens' outcomes • Closing the gap for disadvantaged children • Embedding leadership team and impact on school improvement **2020:** • Writing • Closing the gap for disadvantaged children • Developing the curriculum	• Headteacher • Female • 2021 – 3 years in post **(HF3)**	• 11 July 2019 • 22 July 2021
Matched school 4 (IDACI 5; Urban) **(MS4)**	**2021:** • Newly academised (2018) and joined with a large MAT (26 schools) in the area – previous Ofsted was RI now awaiting first inspection as an academy school • 300 pupils on roll • 32 home languages spoken	**2021:** • Closing the attainment gap in number and reading • Parental engagement with learning • Embedding a broader curriculum	• Headteacher • Female • 2021 – 7 years in post **(HF4)**	• 19 July 2021

IDACI = income deprivation affecting children index; HF = headteacher female; HM = headteacher male; MAT = multi-acacdemy trust; MS = matched school; Ofsted = Office for Standards in Education, Children's Services and Skills; RI = requires improvement.

Attendance at EDT training events

EDT events were a mixture of training on conducting the process and workshops to share feedback, evaluate progress, and refine approaches. Repeat training also came into years two and three for new reviewers or ICs. In year one, each of the five training and workshop sessions run by EDT for the SPP were observed by each of the IPE team, sampling different regions. This also enabled the team to develop a high level of familiarity with the programme and how participating schools were being trained and how they were engaging with the process. In years one and three, we adopted an approach of observing each type of training and workshop once by a member of the IPE team. In year three, we also attended and spoke at two of four scheduled online celebration events in the last week of November 2021, see Table 8, below for full details:

Table 8: Observations of EDT training and workshop sessions

Type of training / workshop	Participants	Date	Year	Locations	Observer
Impact session 1	Headteachers	8 March 2018	1	Barnet	DG
Peer review workshop	Headteachers, reviewers, and ICs	23 March 2018	1	Barnet	DG
Impact session 1	Headteachers	26 March 2018	1	Halterworth Teaching School Alliance (TSA), Hants	LS
Peer review workshop, including IC workshop 1	Headteachers, reviewers, and ICs	27 March 2018	1	Halterworth TSA, Hants	LS
Impact session 1	Headteachers	26 April 2018	1	Market Harborough	TG
Peer review workshop	Headteachers, reviewers, and ICs	27 April 2018	1	Market Harborough	TG
Impact and support workshop	Headteachers and reviewers	2 October 2018	1	Sussex Coast	LS
Impact and support workshop	Headteachers and reviewers	8 October 2018	1	East Sussex	DG
Impact and support workshop	Headteachers and reviewers	31 October 2018	1	Leicestershire/Derbyshire	TG
ICs Day 2: Developing evidence-informed actions	ICs and partnership leads (also conducted short group interviews with ICs)	3 October 2018	1	Sussex Coast	LS
ICs Day 2: Developing evidence-informed actions	ICs and partnership leads (also conducted short group interviews with ICs)	18 October 2018	1	East Sussex	DG
ICs Day 2: Developing evidence-informed actions	ICs and partnership leads (also conducted short group interviews with ICs)	1 November 2018	1	Leicestershire/Derbyshire	TG
Partnership development and growth	Headteachers	March 2019	1	Webinar: Mixed regions	DG
Review of reviews	Headteachers, reviewers, and ICs	24 June 2019	1	West Berkshire	LS
Review of reviews	Headteachers, reviewers, and ICs	5 June 2019	1	East Sussex	DG
Review of reviews	Headteachers, reviewers, and ICs	17 June 2019	1	Leicestershire/Derbyshire	TG
Peer review and IC training phase 2	Headteachers and ICs	20 October 2019	2	Oxford	RM
Leading collaborative school improvement	Headteachers, reviewers, and ICs	12 November 2019	2	East Sussex	DG
Progress and sustainability workshop	Headteachers and reviewers	March 2020	2	East Sussex	DG
SPP rapid response training session	Headteachers and reviewers	8 October 2020	3	Webinar: (147 attendees, including outside of the EEF trial, nationwide)	DG
The power and potential of peer review	Headteachers and reviewers	24 November 2020	3	Webinar: (South training, included Cluster 2 case study)	DG
Summer progress check-in workshops	Headteachers and reviewers	May 18 2021	3	Webinar: (South – two groups combined, including case study cluster 2, area)	DG
Ownership and sustainability: What next for the peer review community?	Headteachers, reviewers, and ICs	24 November 2021	3	Webinar: The EEF partnerships: Kent; East Sussex; Brighton & Hove	DG
Ownership and sustainability: What next for the peer review community?	Headteachers, reviewers, and ICs	18 November 2021	3	Webinar: The EEF partnerships: Trafford; Bury; Bolton; Warrington; Sheffield; Doncaster; and York	LS

Observations of review process

School reviews were central to the SPP and consisted of: a pre-review conversation between the headteacher of the school to be reviewed and the lead reviewer from another school in the cluster, usually another headteacher; a review over one school day and then, usually around 2 weeks later the school that had been reviewed would host an improvement workshop led by typically two ICs from other schools in the cluster. Following the review, schools were encouraged to have 90-day discussions to discuss progress at the level of headteachers in the cluster (we did not plan to observe this part). The IPE teams observed the following over the 3 years:

- Year one: Three full peer reviews (pre-review, review, and improvement workshop) across three separate partnerships. Two of these partnerships were also case study school partnerships.
- Year two: One full review (pre-review, review, and improvement workshop) in a non-case study school partnership. One pre-review and review in a case study school partnership (the improvement workshop was cancelled due to the first lockdown).
- Year three: One online review (review and feedback session only) in a non-case study school partnership.

For each of these reviews, detailed process notes were kept by an evaluation team member, together with key artefacts, such as the agenda for the review day, and slides used in the Improvement Champion workshop, etc. In one case, notes from the 90-day follow-up call were also collected. Observations of these processes enabled a first-hand view of key processes in action, helping us to understand issues of implementation, fidelity, quality, reach, and so on as well as contributing further to our understanding of change processes in our research questions and adaptation during COVID-19.

Interviews with stakeholders

These interviews served several purposes: they helped us in the early stages to gain a clearer understanding of the theory of change embedded in the SPP and of specific mechanisms; they gave us multiple perspectives of these processes, their impact, and challenges of implementation; and they helped us to construct interview and survey questions for later data collection. Over the 3 years, we conducted the following interviews, having added further group interviews to our evaluation protocol for year three. The further interviews with ICs, partnership leads, and SPP associates gave additional sources of impact; the SPP associates were able to give an idea of how they viewed the programme through a difficult period of national disruption, our IPE preliminary analysis showed it could be particularly interesting to explore the IC role, and by interviewing partnership leads we wanted to discover more about the complexities of the programme at partnership level.

Year one:

- March 2018 (EDT SPP manager);
- July 2018 EDT lead and lead associate (joint);
- July 2018 chief executive officer (CEO) of a MAT with experience since 2016 of using the SPP model;
- June 2018 two headteachers (individual);
- October 2018–November 2018 (three IC group interviews, with short survey, after IC training phase 2); and
- EDT associates (July 2019).

Year three:

- IC group interviews (two interviews): Completed (nine participants);
- associates group interview (one interview): Completed (four participants); and
- partnership lead group interview (six interviews): Completed (37 participants).

EDT data on partnership readiness, school dropout, feedback on workshops and training, and compliance

The EDT team were responsible for collecting data on training attendance by school and completion of reviews in clusters. They used a variety of sources including online surveys, feedback from SPP facilitators, and feedback from partnership leads. Before the first training took place, leads from participating clusters were asked to complete a

'partnership audit', which asked about issues of readiness to engage in peer review, prior levels of partnership work, maturity, impact, motivation, and trust. In total, 65[5] of these were completed and the results of these were shared with us. EDT also shared the results of nine evaluations by participants of training events with us.

Analysis

Our approach to analysis was largely deductive, driven by the IPE variables mentioned above and coding within these pre-existing categories and under the research questions, identifying child nodes in relation to the data.

Most interviews were recorded and then transcribed by a third-party transcription service, while for others (mostly group interviews with stakeholders) we used our own detailed notes, then coding was carried out for the above research questions and IPE variables as they applied to each data source, identifying key quotes, examples, and vignettes. Internal reports were generated for the baseline treatment schools' survey, EDT training observations, review observations, matched school interviews, and cross-data set analysis of the case study schools and on ICs, as the role of the latter emerged as a particularly interesting feature of the programme.

Descriptive statistics on the survey data included frequencies of responses on fixed and Likert scale response items and on coded open-ended responses in our surveys. Most of these were conducted by the Qa Research data analyst under guidance of the IPE team. Other inferential statistics involved cross-tabulating factors within the baseline survey such as against prior experience of peer review, motivation to take part in the SPP, and partnership structure. On the final treatment and matched schools' surveys we compared responses to survey items against IDACI quintiles, urban/rural setting, and Ofsted grade. To the treatment schools final survey response frequencies, we also compared responses according to compliance category (fully, minimally, or non-compliant) and also five self-ratings from the initial partnership audits (maturity of partnership, motivation, effectiveness of partnership, capacity to peer review, and alignment with accountability systems). We also looked for differences in responses between completed and withdrawn schools. All analyses were carried out using the latest version of the Askia software package (Askia SAS, London) and significant differences reported according to the Classical Student's t-test ($p < 0.05$, two-tailed).

[5] Note, some of these divided up for the purposes of reviewing into smaller clusters, hence the larger number of partnerships identified by the end of the project by EDT is 93.

Timeline

Table 9: Timeline

Dates	Activity	Staff responsible / leading
October 2017 – February 2018	Recruitment	Education Development Trust (EDT) staff
October 2017 – February 2018	Pre-matching data collection and preparation	EDT staff, then JA and JJ
March 2018	Matching carried out and pseudo-randomisation date	JA and JJ
September 2018 – November 2021	Schools Partnership Programme delivery	EDT staff
January 2019 – June 2021	Interviews of school leaders and teachers in schools in two clusters	LS, DG, and TG
January 2019 – February 2021	Observations of pre-review, review, and improvement workshops	LS, DG, TG, and RM
May 2018 – July 2018	Baseline interview surveys with treatment schools	DG with Qa Research conducting telephone interviews
March 2018 – October 2021	Interviews with stakeholders	LS, DG, and TG
June 2019 – July 2019 and June 2021 – July 2021	Interviews of school leaders and teachers at matched schools	RM
November 2021 – February 2022	Final interview surveys with treatment schools	DG and BM with Qa Research conducting telephone interviews
November 2021 – February 2022	Interview surveys of headteachers in matched schools	DG, RM, and BM with Qa Research conducting telephone interviews
March 2018 – December 2021	Observation of EDT training events	LS, DG, TG, and RM
January 2022 – May 2022	Analysis and report writing	Evaluation team (led by DG and JA)

Results

Impact evaluation

Given the absence of all planned outcome data due to the disruption to end of Key Stage National Curriculum testing as a result of the COVID-19 pandemic, minimal impact evaluation results are available to be reported. In the interest of balancing transparency of reporting and focus of this report, most aspects of impact evaluation reporting have been moved to Appendix 2.

Participant flow including losses and exclusions

We report the flow of clusters and schools through the trial in the adapted Consolidated Standards of Reporting Trials (CONSORT) diagram in Figure 2. Recruitment of clusters was carried out by the EDT in Winter/Spring Term 2017/18, with the aim of starting the intervention in September 2018, drawing on their wide network of schools and wider advertisement. Given this approach to recruitment, it is only possible to provide a rough estimate of the number of schools approached. Target recruitment was set at 50 clusters of English state-funded primary schools with an average cluster size of 6, making 300 schools in total. This was with the proviso that if cluster size was smaller than expected, additional recruitment would be undertaken to bring the number of schools recruited up to 300. In the event, the project team (EDT) successfully recruited far more schools than anticipated, providing a sample of 437 English state-funded primary schools in 85 clusters (average cluster size of just over 5).

Missing administrative data on the matching variables meant that the sample of schools considered for matching was 383 schools. In addition, a further 9 treatment schools were dropped as part of the matching process due to the imposition of common support and a caliper, resulting in a final analysis sample of treatment schools of 374; these will be compared to 374 matched comparison schools identified as described above. We regard both of these sources of sample loss as essentially random processes marginally reducing the initial sample (and, hence, statistical power of the analysis would have been able to achieve) rather than being attrition from the trial.

There has been no follow up or analysis phase of the impact evaluation in this project due to the cancellation of Key Stage 2 National Curriculum tests in 2019/20 and 2020/21 as a result of the COVID-19 pandemic. As a result, the remaining sections of the CONSORT diagram (Figure 2) are blank.

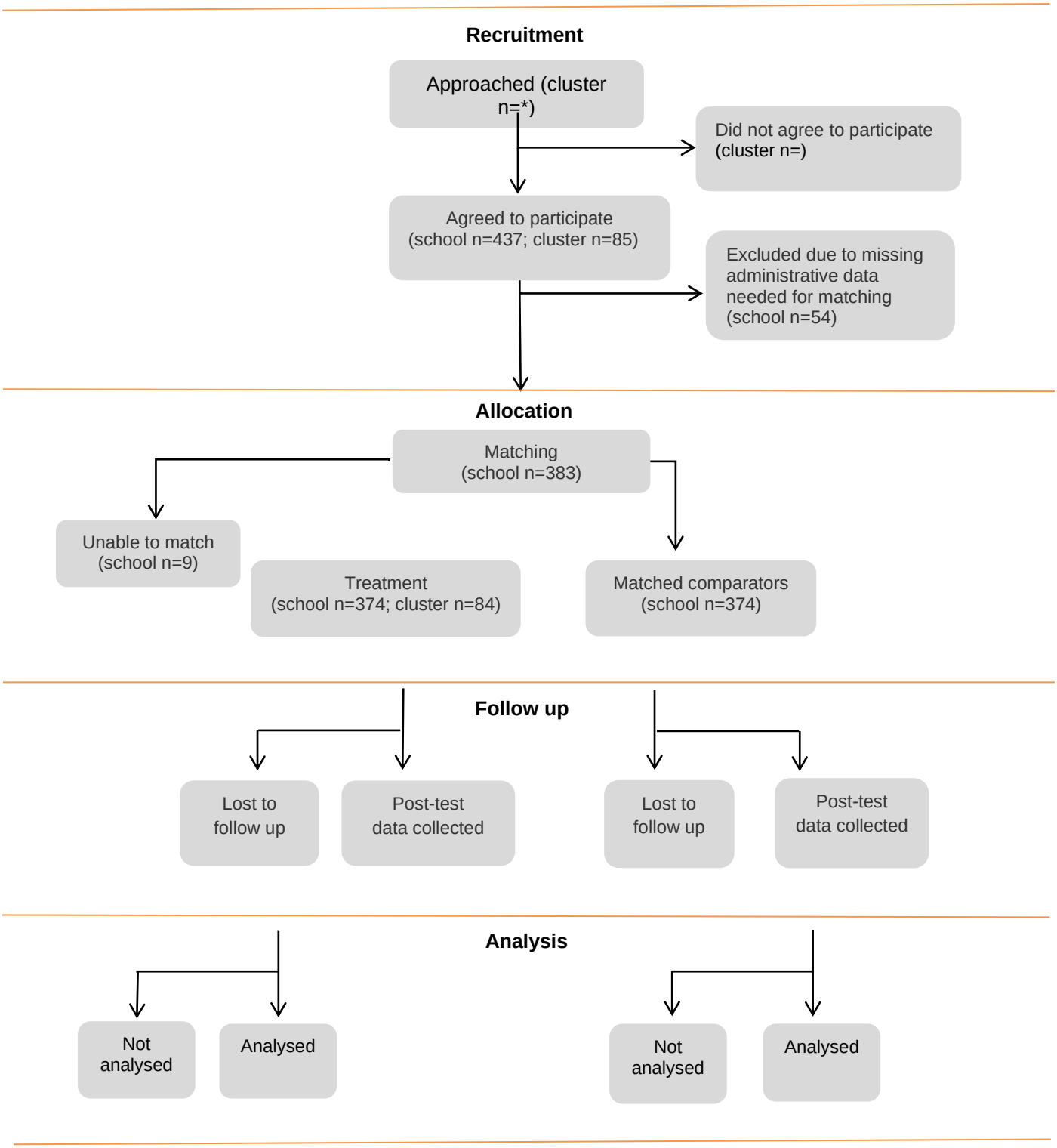

Figure 2: Participant flow diagram (two arms)

*NB Approached clusters data is not available form EDT. Their approach to recruitment was to hold a briefing about the trial and then interested clusters of schools contacted them.

Pupil and school characteristics

The process of propensity score matching led to a well-balanced sample of schools in terms of the characteristics considered, with the largest residual imbalance being the average Key Stage 2 maths score of the schools in our baseline year (2017), which is still fairly low at 0.06. Exact matching on IDACI quintile groups (among other characteristics) explains the exactly 0 standardised differences in the case of these.

We demonstrate that the similarities in the means of continuous measures after matching is not hiding large differences in the distributions of the samples by plotting the full distribution of these variables in the treated and matched comparison samples. This is reported in Appendix 2 for Key Stage 1 average points score of intake, average Key Stage 2 maths score, average Key Stage 2 reading score, average Key Stage 2 maths score among FSM pupils, and average Key Stage 2 reading score among FSM pupils. Unsurprisingly, given that they are not explicitly included in the matching model, the distributions are not quite as closely matched among FSM pupils, but still perform acceptably.

Table 10: Baseline characteristics of groups as at pseudo-randomisation

School level (categorical)	National level %	Intervention group		Matched comparison group		Standardised difference
		N (missing)	%	N (missing)	%	
Academy	22	68 (0)	18	74 (0)	20	-0.04
Non-academy	78	306 (0)	82	300 (0)	80	0.04
IDACI quintile 1	18	78 (0)	21	78 (0)	21	0.00
IDACI quintile 2	17	91 (0)	24	91 (0)	24	0.00
IDACI quintile 3	28	91 (0)	24	91 (0)	24	0.00
IDACI quintile 4	18	72 (0)	19	72 (0)	19	0.00
IDACI quintile 5	19	42 (0)	11	42 (0)	11	0.00
Ofsted rating: Outstanding	15	64 (0)	17	70 (0)	19	-0.04
Ofsted rating: Good	62	267 (0)	71	262 (0)	70	0.03
Ofsted rating: Requires improvement	20	37 (0)	10	38 (0)	10	-0.01
School level (continuous)	**National level mean**	**N (missing)**	**Mean (SD)**	**N (missing)**	**Mean (SD)**	
KS2 reading score in 2017	104.4	374 (0)	104.9 (2.96)	374 (0)	104.8 (2.93)	0.02
KS2 maths score in 2017	104.2	374 (0)	104.4 (2.75)	374 (0)	104.5 (3.02)	-0.01

KS2 reading score in 2017 (FSM)	101.6	374 (0)	101.7 (3.24)	374 (0)	101.8 (3.39)	-0.02
KS2 maths score in 2017 (FSM)	101.7	374 (0)	101.4 (3.05)	374 (0)	101.6 (3.44)	-0.06
KS1 intake score in 2017	15.9	374 (0)	16.1 (16.08)	374 (0)	16.1 (16.07)	0.02

FSM = free school meals; IDACI = income deprivation affecting children index; KS = Key Stage; Ofsted = Office for Standards in Education, Children's Services and Skills; SD = standard deviation.

Outcomes and analysis

Analysis in the presence of non-compliance

All outcome data was missing due to the disruption to end of Key Stage National Curriculum testing as a result of the COVID-19 pandemic. As such, it was not possible to consider impact in the presence of non-compliance. However, we can report on rates of compliance at school and cluster levels.

Compliance was calculated at both school and cluster levels according to the evaluation protocol. Table 11 below shows only 16% were fully compliant. We need to consider the effect of COVID, and of the extension to the trial into a third year, on school withdrawal on the high non-compliant numbers. Our final survey asked participants to state the most challenging aspect of their involvement in the programme and COVID was given as the most frequent response (44%). Of those from the withdrawn group surveyed, 30% withdrew in 2020. A total of 231 schools also continued 'post-COVID', leaving 182 schools deciding not to continue into a third year of the trial, according to EDT data.

Around 56% of schools were therefore, minimally or fully compliant (combined) over the full period of the intervention. In the final treatment schools' survey, we compared responses between minimally and fully compliant groups and found only two significant differences in responses between these groups, suggesting some logic in combining these. The two exceptions were: 'Our teachers feel a greater sense of ownership and engagement with our school's improvement aims' (minimally compliant [13%] schools disagreed more than fully compliant schools [3%]) and 'My school draws on expertise and support from the partnership on a regular (i.e. monthly) basis' (minimally compliant [19%] schools disagreed more than fully compliant schools [5%]).

Table 11: School-level compliance

Treated schools' compliance	Treated schools N	%
Fully compliant	70	16
Minimally compliant	168	40
Non-compliant	184	44
Total	**422**	**100**

Table 12: Cluster-level compliance

Cluster-level treated schools' compliance	Clusters N	%
Fully compliant	3	3
Minimally compliant	10	11
Non-compliant	76	86
Total	**89**	**100**

The calculation of compliance at the cluster level is very low, as the formula penalises clusters with non-compliant schools. It is enough to have one school non-compliant in a cluster to be considered a non-compliant cluster. From EDT data, we understand that of the 422 schools who started the trial, 295 of these dropped out after the COVID disruption and 57 prior to this.

IPE

The IPE results is divided overall into three sections:

- usual practice;
- research questions; and
- implementation factors.

The first section sets out a detailed analysis of '**usual practice'** (albeit some of this occurred through a global pandemic). The purpose of this is to understand what kinds of partnership practice occurred in the SPP schools prior to the commencement of the trial. Furthermore, the nature of partnership working during this period is contrasted with the matched schools.

Next, the **findings according to the five research questions** are presented, the first section combines results in relation to research questions 1 and 2:

1. In what ways does the SPP influence the capability, culture, and practice of partnerships, leadership, and teachers in involved schools?
2. In what ways do the elements in the SPP theory of change work in achieving participants' perceived forms of impact?

The next section combines results for research questions 3 and 4:

3. What factors influence schools' and clusters' ability to engage in, participate fully in, and successfully implement and sustain their involvement in the programme?
4. What distinguishes schools and clusters that have not continued with the programme?

And the final section looks at findings for research question 5:

5. What difference has COVID-19 made to the operation, participant engagement, and perceived forms of impact of the SPP?

Following that, the following **further implementation factors** are addressed:

- fidelity (if the intervention was delivered as intended);
- adaptation (how the intervention was changed to meet circumstances);
- dosage (how much of the intervention was delivered);
- reach (if the intervention was delivered to the intended recipients);
- responsiveness (recipient engagement during the delivery of the intervention);
- quality (how well the intervention was delivered); and
- differentiation (how different the intervention was from existing practices).

Usual practice

This section describes multiple data sets that give a detailed picture of the nature of partnership working in both the SPP treatment schools and matched schools. Understanding the structures and patterns of partnership work are key to interpreting the later data we have in relation to the research questions, such as: How SPP clusters were either embedded in or working alongside other partnerships? We can also see to what extent matched school partnerships compared to SPP activity. We start by looking at the four matched schools where we interviewed staff and then move to the data we have on the two SPP case study clusters. We also show EDT data from initial audits of SPP schools to do with readiness to engage and prior experience of partnership work, and other records to do with geographical spread of clusters. The survey data for both SPP and matched schools allows for further comparison between these groups in terms of improvement priorities for partnership and also further data on prior experience of peer review in the matched schools. The purpose is to show what kind of partnership work was happening, what improvement priorities were, what motivations existed, and how these were affected by COVID.

Usual practice in matched case study schools

We asked matched school participants in four schools about the school-to-school support they engaged in and how effective these activities were perceived to be for school improvement priorities and their impact. We also wanted to know what difference COVID-19 had made to their school-to-school partnerships.

Interviews at matched schools revealed a wide range of areas of school improvement aims from the start, engagement with evidence (primarily the EEF sources cited), and how their focus shifted due to COVID, see Table 13 below:

Table 13: Summary of matched case study schools' areas of focus and partnership working

Matched school	Partnership working	Evidence-informed	Difference as a result of COVID
MS1	**2019:** • Bought into a self-evaluation programme called 'Two Eskimos' • Local authority advisor undertakes frequent visits to support school improvement • Bought in consultant to work specifically on moderation • Informal network of eight schools across the locality – described as a local cluster	• The EEF toolkit • Learning Without Limits course by the IOE • Various shared resources from local clusters such as books, discussions, and links to research	**2021:** • Importance of context and values for what headteachers perceive to be important in partnership and collaborative working (particularly for more experienced headteachers) • Still using Two Eskimos for Self-Evaluation – and finds it very effective • Key Stage 1 expected progress suffered due to COVID • Subject leads and moderation and learning walks, etc. have suffered • Priority on post-pandemic internal collaboration before external • Pre-pandemic cluster group may not return in original format due to the range of experiences (differential) of the schools involved as a result of the pandemic
MS2	• Local link group with cross phase moderation • MAT support with curriculum review (prior joining) • Local authority intervention with mastery approach (Twilight CPD)	• The EEF toolkit • Research generated in the field by practitioners and part or ResearchEd movement • Twitter for resources and blogs (mentioned David Didau and Tom Sherrington)	• N/A
MS3	• Local authority school improvement advisor • Moderating with a local primary school • Assistant head is in a local cluster – undefined • Attends city headteacher briefings	• The EEF toolkit • Ambition School Leadership programme	
MS4	• Regular trust head meetings (weekly primary board meetings) • Support and Challenge Partners (from the Trust) • Collaborative groups for subject leaders across the trust • Trust-level peer review system • Collaboration with special school on behavioural approaches as well as peer-to-peer support	• The Great Teaching Toolkit • Trust Core Principles for Teaching and Learning drawing on Rosenshine (Tom Sherrigton) • Early Reading	

There were a wide range of interpretations of the idea of a partnership; some were *mutually* supportive networks, brought together informally, for example:

> *It's interesting because the consultant that I pay separately for, she was working for the local authority. She could see a group of head teachers that were quite like minded and quite enthusiastic probably. So she put us in touch with each other. So she could see an opportunity for this group of head teachers to work together and so that's how it was formed.* (headteacher female 1 [HF1], 2019)

This collaboration involved headteachers who 'dipped in' from time to time but also with governors and teachers at times. Another example, headteacher male 2 (HM2) reported that partnership working in matched school (MS) 2 was 'sporadic' and dependent on the 'incumbent head' (HM2 is an interim headteacher, 10 months in post and was due to leave at the end of the Summer Term 2019 just after this interview took place), reflecting the fragility of partnerships in line with staffing, especially at leadership level. MS2 also explained that there were a lack of agreed definitions, approaches, and activities towards partnership working.

Other examples from the matched schools, such as buying into Self-Evaluation Framework (SEF) tools online did not seem to fit the traditional sense of a mutually supportive network that might be considered indicative of a 'partnership'.

Between 2019 and 2021 peer review was used as a tool for school improvement for two schools, initiated either by local authorities (MS3), or in the case of MS4, both the local authority and the MAT. In the follow-up 2021 interview, HF3 reported that a peer review partnership had been established by the local authority in the post-July 2019 interview period and the first pre-COVID lockdown. MS4 also undertook a formalised peer review process, which occurred as part of the MAT-wide approach to school improvement. In addition, to this MAT-led approach to peer review, MS4 had a longstanding school-to-school support arrangement (as their joining the MAT was relatively recent – 2018) led by a local special needs school in coordination with two local authorities, in which they adopted an informal approach to peer review.

Where local clusters existed (MS1) these were generally weakened rather than strengthened as a result of the COVID disruption. However, HF3 commented specifically on the benefits of the post-COVID approach, where partnership working that had started more in earnest after the 2019 interview was carried on virtually. HF3 reported:

> *I think there's been lots of benefits actually, you know, in terms of being able to use things like Teams, like we're doing right now. It's opened up lots of different ways of working with each other. So, for example, my head buddy relationship, if you like, came through Covid. I meet with my head buddy over Teams. I'm unable to go into her school. We did our moderation over Teams and if we had to do a cross-moderation with her school, it's on the other side of the local authority, it would have taken us half an hour to get there.* (HF3, 2019)

What seemed important to the three schools that were involved in some sort of more formalised partnership arrangement (MS1 – local cluster; MS3 local authority supported peer review; MS4 local authority and MAT supported peer review) was that there was a similarity in the types of schools that they worked with in these arrangements for these to be seen as productive. So for MS1, post-COVID it was the decision to move away from a local cluster because there was not that shared 'journey' in terms of where they were and where they wanted to go and led HF1 to look further afield for a school to collaborate with. For MS4, she felt similarly about the importance of shared school characteristics in making partnership working productive.

> *The problem for me in the trust is that the vast majority of schools in the trust are nothing like us in terms of socio-economic status and diversity, and so yes, I mean, it's always useful for people to come in and have a look at what you're doing, but we sometimes find…Support and Challenge Partners, if they haven't actually worked in a similar school, the advice is not as targeted as it could be* (HF4, July 2021).

The importance of values and similarity of characteristics and contexts were significant for both HF1 and HF4 in terms of seeing benefits from partnership working.

HF4 also expressed reservations about a more 'done to model' from a centralised trust where, for HF3 the peer review model, which was established by the local authority was reported upon favourably. There is no evidence here

to suggest that centralised MAT approaches are generally an issue in comparison to local authority approaches, but rather reveals a specific shared experience of two such instances in this sample.

While we need to caution against generalisation, in our small sample, there was a mixed picture in terms of both benefits and limitations to partnership working—where benefits appeared to be acknowledged by all participants in that having external professional colleagues collaborating on areas of school improvement was seen as a positive and important activity, but the methods and the execution of these collaborations, as well as the localised experiences, contexts, individuals, and extent of shared values, seemed to have an important bearing on how well they were viewed. How individual school leaders engaged with these collaborative processes appeared to be an important part of how successful these partnerships were deemed to be, and staff churn could have significant impact on their sustainability. There seemed to have been a growth in peer review activity in these schools, but there was generally disruption to this during COVID. In one case, the benefits of working remotely in partnerships were cited. There did not seem to be a fixed interpretation of what constituted collaborative or partnership working—the activities reported upon in the research included more tightly coupled and traditional interpretations (such as peer review), as well as the identification of tools or platforms for activities such as the EEF toolkit, where these tools are used to support internal review of school improvement priorities but not with external partners.

Usual practice in SPP case study schools

In this section, we look at two case study school clusters (one in the North of England and one in the South of England) to understand about prior partnership working, motivations for taking part in the SPP trial, areas of planned school improvement focus and how these evolved, and other interventions that schools had been using or continued to use alongside SPP for school improvement.

Partnership working and motivations for taking part in the SPP trial

Both partnerships had worked together for a period of a few years and in multiple ways. From Cluster 1 in the initial partnership audit, (Spring Term 2018), the partnership lead stated that they had worked together previously between 1–5 years. All schools had separate governing bodies but information from partnership work was shared with governors through headteacher reports. Headteachers met at least termly. Other partnership meetings included: moderation; joint training; sharing of resources; and information. Previous work had been at all staff levels, including headteachers, senior leaders, middle leaders, and teachers. They had not taken part in peer review activity prior to the SPP trial. The partnership lead assessed that there was ‘some evidence’ that schools in their partnership were equally committed to and had a clear motivation for joining the SPP, and that they had made earlier enquires about doing a National Association of Head Teachers (NAHT) peer review. They had also talked about forming a MAT.

In terms of impact from partnership work to date, the partnership lead said there was ‘some evidence’ citing: clarity of judgements through moderation activities; the impact of Continuing Professional Development (CPD) being evident in schools; and a positive impact on development of Newly Qualified Teachers (NQTs); and retention. The partnership lead also claimed there were economies of scale, to do with CPD costs. Headteacher support was both informal and formal, including support with recruitment of senior staff, and NQTs were jointly observed across the partnership.

The partnership lead perceived their partnership to have ‘some evidence’ of capacity to conduct peer review too, due to the time already committed to the partnership over a number of years and commitment to place the SPP as a key element of the School Improvement Plan (SIP) for 2018/19. Financial and other resources had also been committed.

In terms of ‘clear alignment [of the SPP] with current accountability systems’ the partnership lead felt there was as yet no evidence for this, but that they would strengthen their readiness by attending the training in March 2018, build a capacity plan into the school budget and staffing plans, brief all staff, create a timetable of reviews and meetings, and develop impact measurement tools.

One of the reasons cited for taking part in the SPP trial in Cluster 1 was the role of the EEF, as the executive headteacher (EHT) of school 1e (year three interview) mentioned:

> *When we came to SPP, I think the fact that it was EEF and that particularly (school x) but all of us would trust in an organisation that has got a lot of evidence-based research and are in it for the right reasons really, to support schools, to support disadvantaged groups, and to improve standards generally for children.* (EHT, school 1e, year three interview)

In Cluster 2 (from EDT initial partnership audit, Spring Term 2018), the partnership lead stated that the partnership had been working together for over 5 years. All had their own governing bodies, but they met every half term as headteachers and as heads/governors. Previous work had been at all staff levels, including headteachers, SLT, and middle leaders. They had engaged in some previous work with peer review, where triads had reviewed each other's schools with a particular focus.

In terms of motivation, the partnership lead felt that there was significant evidence for schools in their partnership being equally committed to the SPP. In terms of evidence of impact of partnership working to date 'some evidence' was selected, which was from 'action plans and reviews over years'. The partnership lead also felt they had some evidence of capacity to conduct peer review due to prior commitment of significant time working together, experience of meeting and reviewing their action plan, and experience of conducting peer review in the past. The partnership lead therefore, also felt that there was some evidence of alignment of the SPP with current accountability systems, leaving them only needing to 'carefully plan our peer review' to strengthen this readiness. Therefore, taking part in the SPP was in order to have more carefully planned work that would have even greater impact.

In terms of motivations for taking part, the headteacher in school 2e (year one interview) stated that their local authority had faced extensive cuts, so schools saw the greater need to work with partner schools for school improvement. The diocese wanted them (school 2e) to work in a Catholic schools' MAT but they were not keen, due to geographical isolation, so this partnership strengthened already existing semi-formal arrangements with local schools. Headteacher of school 2e (also partnership lead) also mentioned (year one interview) that she wanted to help other schools improve in the partnership too, although lamented that one local school was not a member of the partnership and would have benefited from joining it.

As with Cluster 1, the EHT of school 2b also mentioned the excellent reputation of the EEF as a motivation for taking part in the trial and potentially a lack of distinction between the EEF and the EDTs SPP, namely, between trial funders and the programme delivery organisation. Therefore, comments later in our findings, about schools being encouraged to be 'evidence-informed' seemed to be attributed to the EEF as much as the EDT (perhaps participants were also confused due to the similarity of the EEF/EDT).

Areas of school improvement focus

As with the matched case study schools, there was a wide range of areas of focus. The following list summarises these areas, gained from year one case study interviews with headteachers and EHTs:

School 1a: Re-shaping of the curriculum.

School 1c: Progress and greater depth in maths, efficient methods of calculation; and in reading and writing, a focus on using quality, and challenging texts. Also, a big focus on Special Educational Needs (SEN), as well as developing the curriculum.

School 1d: Several areas: reading; subject leadership; teaching assistants (TAs); welfare and safeguarding; and Ofsted-related priorities, handwriting, problem-solving in maths, and the early years 'increase and exceed'. In year three, they focused on consolidating teaching and learning due to all the staff changes and CPD to support this, using an improvement cycle from the EEF, focusing on science, technology, engineering, and mathematics (STEM) in the first term.

School 1e: They focused on 'Maths Mastery' for their reviews, but the school was also looking at thinking/reasoning, adopting the 'philosophy for children' approach. In year three, the focus changed to re-organising the curriculum.

School 2a: Spelling/editing of student work.

School 2b: Spelling/editing of student work.

School 2d: Greater depth for maths. In year two, the school was developing its own maths curriculum, borrowing from various sources. Also, for students to develop a love of reading and to improve teaching of reading. A new priority for year two was also developing subject leadership.

School 2e: In year one, the review focus was differentiation for girls in maths—wanting to stretch girls to take on more challenging tasks. In year two, this changed to looking at the curriculum, although with a specific focus on the use of TAs across the school (partly driven by budget and how to deploy them effectively).

Case studies initial survey responses

Appendix 7 has a full comparison table on nine areas of the initial survey summarising and comparing cluster responses to items about context, school leader experience, approaches to self-evaluation, and interventions introduced prior to the start of the programme, in summary:

- Cluster 2 had much less experienced school leaders (all less than 2 years), with all of Cluster 1 having 7–15 years in role in school.
- School self-evaluation (SSE) rating: Cluster 1, three schools' SSEs involved senior and middle leaders, two involved all staff. In Cluster 2, there was a similar picture (two and two).
- Cluster 1, four schools had introduced new approaches to literacy (varied), and two schools were substantially different to what they were doing before. In Cluster 2, two new approaches to literacy (within last year) but which were an incremental change.
- Cluster 1, one school had introduced a substantial new approach to numeracy, others had not introduced any new approaches to numeracy (within the last year). In Cluster 2, four schools said their new approach to numeracy was substantially different to what they had been doing before (i.e. before March 2018) (Maths Mastery).
- Reason for joining—both clusters—all said to add structure to existing partnership. All but two schools (1a and 1d) also wanted support for school improvement. All but two schools (1a and 1b) mentioned reputation of the SPP; only three schools (1b, 1c, and 1d) wanted to be 'Ofsted ready'. Four schools from Cluster 1 mentioned 'cost' as a reason—none from Cluster 2.

<u>Other differences between Clusters 1 and 2 from initial survey</u>

School leaders in both clusters showed high average ratings of confidence on items in Section 4 to do with readiness to engage, levels of transparency, capacity to improve, and ownership of school improvement priorities, for instance. There were minor differences in seven items (see Table 14, below). Cluster 1 was more confident it had sufficient funding for staff costs, disagreed more strongly that there was a lack of trust between schools in their area, and had peer review more firmly embedded in their school's improvement cycle (this appears to contradict responses to earlier partnership audit). Cluster 2 agreed more strongly that all schools were able to benefit from external support, rated Ofsted a higher priority, were able to recruit high-quality teachers, and were more confident that the SPP would help school achieve its aims.

Table 14: Mean confidence ratings from initial survey of treatment schools where differences were shown* (n=339)

Question	Average Likert scale ratings for case study schools, 2018 survey (5 = strongly agree, 1 = strongly disagree)	Cluster 1	Cluster 2
4.1	My school will have sufficient funding over the next 3 years to employ the staff it needs	**3**	**2**
4.1	Within my locality, all schools that would benefit from external support are currently able to access appropriate help	**3**	**4**
4.1	Making sure my school does well in Ofsted inspections is one of my top priorities as a leader	**4**	**5**
4.1	A lack of trust between schools in my area hinders meaningful collaboration	**2**	**1**
4.3	Peer review is firmly embedded in our school's improvement cycle	**3.75**	**2**
4.4	My school is usually able to recruit high-quality teachers	**3**	**4.5**
4.5	I am very confident that our partnership will help this school to achieve the aims it has set in the Schools Partnership Programme	**4**	**5**

**Where mean rating between the cluster differed by at least a value of 1*

Overall, the case study schools (and this was typical from our wider survey and other data on SPP schools, see below) had already worked together and had built a good level of trust at various levels of activity and staff over years. They were looking for more structure to this work and for it to have a greater impact on pupil outcomes. School improvement areas varied, with some school leaders naming multiple objectives and others singular and very specific ones, and years two and three changes showed shifts in some schools due to the Ofsted framework changing (e.g. towards subject leadership and a curriculum focus). Schools in both clusters were highly motivated to work together, to use SPP alongside other, sometimes very significant interventions to improve numeracy or literacy, and showed a high level of readiness to engage in peer review and partnership work, and confidence that this would lead to them achieving their improvement aims.

EDT data on SPP schools

EDT data shows that schools came from 25 local authority areas geographically spread across England, with the largest representations coming from Kent (60 schools), then East Sussex (52), followed by Hampshire (26), Milton Keynes (22), West Berkshire (21), and Wokingham (18). These 199 schools made up 47% of the total number of SPP schools in the trial (total of 422). Southern counties are disproportionately represented in the trial. Table 15 below shows how local authority areas were distributed into 93 clusters (usually five out of six schools per cluster)

Table 15: Regional distribution of SPP school clusters

Local authority area	Number of clusters
East Sussex	10
Kent – Ashford	7
Hampshire	6
Wokingham	6
Barnet	5
Dorset	5
Kent – Canterbury	5
Kingston upon Thames	5
Milton Keynes	5
Brighton & Hove	4
Bury	4
Kent – Dartford	4
Sheffield	4
West Berkshire	4
Northamptonshire	3
South Leicestershire	3
Oxford	2
Salford	2
Trafford	2
York	2
Bolton	1
Doncaster	1
Manchester – Rochdale	1
South Derbyshire	1
Warrington	1
Total	**93**

Initial partnership audits

EDT asked partnership leads to complete responses for each partnership. These partnerships varied in maturity: 47% had been together for less than 1 year; 32% had been together 1–5 years; and 21% for over 5 years.

Partnership leads also rated their partnership's motivation to the programme, impact to date of their partnership work, capacity to conduct peer review and school-to-school support, and alignment of school's existing accountability systems and processes with the SPP. Table 16 shows the percentage responses.

Table 16: Likert scale responses to February 2018 initial partnership audits (n=62)

Partnership audit questions	No evidence yet %	Some evidence %	Significant evidence %
To what extent are schools in your partnership equally committed to, and have a clear motivation for joining the SPP?	2	45	53
To what extent has your partnership working had impact to date (e.g. school improvement, impact on pupil outcomes)?	37	53	10
To what extent do schools in your partnership have the capacity needed to both receive and support peer reviews and school-to-school support?	5	45	50
To what extent is the SPP aligned with your existing accountability systems and processes?	45	39	16

SPP = Schools Partnership Programme.

We can see that there is strong evidence particularly of motivation to join the SPP and capacity to peer review and support each other. Lower levels of agreement were shown for alignment with SPP processes (perhaps unsurprisingly prior to starting the programme). Agreement on evidence of impact was lower, with 37% having no evidence yet for this.

Treatment schools initial survey

Motivation to join the SPP

We asked schools in the 2018 survey (n=339 out of 422 total treatment schools) what their motivations were for joining the SPP. Table 17 below shows the reasons selected (when asked to select all that applied) in descending order of frequency. Adding structure to existing partnership work was the top reason, followed by support on school improvement. More senior school leaders tended to want to take part in the SPP to add structure to existing arrangements and the less experienced school leaders tended more towards using the SPP for improvement purposes. We also found that respondents from partnerships that were MATs were more likely to regularly draw on expertise and support from other schools in their partnership than other partnership structures such as local authority-based clusters.

Table 17: What was your school's main reason for getting involved in the Schools Partnership Programme (SPP)?

Reason selected	N	%
To add structure to existing partnership work	229	68
To find support on school improvement	224	66
Reputation of SPP e.g. recommended by others	106	31
Other	99	29
We had prior experience of peer review, but wanted to switch to the SPP model	79	23
Cost	67	20
In order to ensure we are Ofsted ready	63	19
Wanted to work collaboratively/increase collaboration	34	10

Type of SPP partnership

In the initial survey we asked: Which of the following best describes your SPP peer review partnership?

- The most frequent description of the SPP peer review partnership was of an informal local cluster or partnership (38%), followed by a local authority cluster (24%), or a mixture of elements

(MAT/local authority/Teaching School Alliance [TSA] etc.,15%). A total of 13% described their partnership as a MAT and 6% as TSAs, and non-local clusters were only ten schools (3%).

Strength of partnerships

Part of the SPP methodology was for school participants to come in ready-made clusters. It is perhaps not surprising then that the majority of our initial survey respondents confirmed that their partnerships were strong:

- Around 85% agreed or strongly agreed that levels of trust were high between schools in their partnership;
- Around 82% either strongly agreed or agreed that their partnership was well led, they had a shared vision and values, and they understood how collective decisions were made;
- Around 74% agreed or strongly agreed that their partnership was well managed and that they met regularly and communicated well, supported by systems and processes that were fit for purpose;
- Around 54% agreed or strongly agreed that their school drew on expertise and support from other schools in the partnership on a regular (i.e. monthly) basis; and
- Around 64% either agreed or strongly agreed that they felt responsible for the success of all schools and pupils in the partnership.

Therefore, the SPP activity was designed to build on partnerships that were already seen to be working well.

Partnerships for school improvement

Most schools had used more than one source of support for school improvement in the last academic year: the median number of sources of support used was four (99 schools). Nearly a third of schools (117) had used five or more sources of support. See Table 18 for most common forms of school improvement partnerships cited by SPP participants in the year prior to the trial commencement.

Table 18: To support your school improvement efforts, in this or the last academic year, has your school made use of any of the following types of support?

Source of support	N	%
Support from or work with a local cluster, family of schools, or network	320	94
A local authority advisor or improvement partner	281	83
Peer review undertaken by another school or schools	170	50
An Ofsted inspection or HMI monitoring visit	149	44
A teaching school	140	41
A commercial consultant or service	129	38
A national or local leader of education	103	30
A multi-academy trust or sponsor	77	23

Focus for the reviews

- More schools had chosen literacy (26%) than numeracy (12%) as their (initial) improvement focus. However, the top answer was both numeracy and literacy (159, 47%). Quite a large proportion did not know at the time of the survey (15%).

Recently introduced new approaches to literacy and or numeracy

The trial was originally set up to look at the impact at Key Stage 2 for literacy and numeracy, and therefore we asked whether schools were already engaged in interventions that may have an influence on achievement in these two areas. Below is a summary of what the survey showed:

Literacy

- Two-thirds of schools had introduced significant new approaches to teaching or assessing literacy skills in Key Stage 2 during this or the last academic year.

- The top answer was the introduction of a specific programme or curricular approach for literacy teaching, such as 'talk for writing' (38%). The second answer was the introduction of new principles or ideas, such as meta-cognition (36%). The third top answer was changes to assessment, monitoring, or testing (34%).
- Around 43% said that the changes to teaching literacy were substantially different to previous practice and 6% stated that the new approach was radically different to what they were doing before.

Numeracy

- Two-thirds of our sample (68%) had introduced new approaches to teaching or assessing numeracy in Key Stage 2 during this or the last academic year.
- The top answer was the introduction of new ideas or principles such as Maths Mastery (69%). The second top answer was the introduction of specific programmes or curricular approaches for teaching numeracy (22%). The third top answer was changes to assessment, monitoring, or testing (21%).
- Most changes to teaching numeracy that were introduced were described by respondents as either 'an incremental change' (43%) or 'substantially different to what we were doing before' (41%). Around 16% describe their new approach as 'radically different'.

These findings suggest that involvement in the SPP trial would, for around two-thirds of the participants, be building on strategies they were already undertaking to improve results in numeracy or literacy.

Previous experience with peer review[6]

For a substantial minority of the schools involved, peer review was not a new approach:

- A third of our sample (111 schools or 33%) said that they had been involved in a peer review programme other than the SPP over the 2 years prior to the survey.
- The most popular selection was 'a model developed by ourselves in partnership with other schools' (59%).

Ongoing peer review involvement

Peer review, although central to the SPP methodology, was also happening in other ways within many partnerships. We asked survey respondents if they intended to participate in any other peer review programmes over the next 2 years (i.e. alongside the SPP). A significant minority were planning to do so:

- Around 22% (n=75) of our sampled schools were intending to participate in another peer review programme over the next 2 years alongside the SPP.
- In fact, 14 of these 75 schools said that they intended to participate in more than one other peer review programme.

Final surveys—comparing matched to treatment schools

We wanted to know if matched schools had experienced peer review during the same trial period and what kind of significant partnerships they had (peer review or otherwise). It is worth stating here that on the matched school survey, participants were given a series of statements about their most significant partnerships over 3 years prior to being surveyed. School improvement partnerships were defined in the introduction to this section of the survey as:

> School's experiences of partnership working, which could include peer review or other kinds designed to support school improvement. School improvement partnerships are strong alliances or affiliations

[6] By (school) peer review programme, we mean programmes involving evaluations carried out of schools (or parts of schools such as departments, subject areas, or year groups) by colleagues from other schools. These include review visits, during which they collect and analyse data and feed back findings and recommendations to the host school.

either between two or more schools, or schools and organisations, aimed at increasing participant schools' capacity to provide high-quality education for their pupils.

The above definition was used to equate to the SPP which, while including peer review visits, positions itself as a school improvement partnership, involving self-evaluaton and school-to-school support in addition to review visits.

For both SPP and matched schools, we also wanted to know what had been the most beneficial aspects of these partnerships and also the most challenging. We also wanted to know what their improvement priorities had been and how their partnership work had been affected by COVID. In terms of characteristics of the treatment and matched schools' surveys, we found no significant difference in terms of urban/rural category, IDACI quintile score, and Ofsted grade (Classical Student's t-test, $p < 0.05$, two-tailed).

Matched school participation in peer review

We asked school leaders from matched schools retrospectively about involvement in partnership and peer review activities in the 3 years prior to the survey to see how they compared with the SPP group. Participation in peer review was not unique to the SPP group during the 2018–2021 period. This may also suggest a general increase in peer review activities in primary schools in England since 2018, since the initial survey of SPP schools had only 33% having experienced peer review in the year prior to starting the trial.

Significantly, ***61% of the matched schools surveyed had participated in peer review*** during this time, with the top answer being one they had developed themselves with partner schools (67%), followed by one developed by a MAT (33%), or local authority (30%).

Top motivations for taking part in peer review were: feedback to improve on best practices (56%); to provide internal/external comparative assessment (37%); and for networking and collaboration (37%). Unlike in the initial treatment schools' survey, giving structure to existing partnerships and school improvement was not cited as a reason, nor was the reputation of a particular provider. Feedback on best practice to improve practice was also slightly different as a motivation to the idea of 'finding support' for school improvement in SPP schools.

Most significant partnerships in matched schools

We also asked matched schools what their most significant partnership had been over the 3 years prior to the survey. Top answers here were: 'a school cluster or hub partnership' (43%); a local authority partnership (20%); a MAT (18%); or a federation (14%). We do not know how many of these used peer review as, at least, part of their partnership methodology. Asked to say what the most beneficial aspects were in taking part in these partnerships (up to three responses), top responses given are shown in Table 19, below:

Table 19: Most beneficial aspects of taking part in most significant partnership over the last 3 years for matched schools (n=44)

Most beneficial aspects	N	Response %
Sharing ideas and best practices	28	64
Support	13	30
Continuity and consistency across schools	10	23
Networking and collaboration	10	23
Regular meetings and contact	9	20
Raising issues and challenging one another	8	18
Continuing Professional Development	7	16
Specific programme provided	2	5
Training	1	2
Safeguarding pupils welfare	1	2

We also asked what the most challenging aspects were of involvement in these partnerships (up to three responses). Table 20 below shows the responses:

Table 20: Most challenging aspects of taking part in most significant partnership over the last 3 years for matched schools (n=44)

Most challenging aspects	N	Response %
Hard to cover staff time for participation in programme	24	55
COVID pandemic	14	32
Unable to meet face to face	12	27
Differing demographics or values	7	16
No challenges	4	9
Differences between school participation levels	3	7
Getting started with the model of review and support	2	5
Online training	1	2

Other peer review in SPP schools

For a significant minority of SPP participants, peer review was a methodology they used in other projects and partnerships and was not confined to SPP work. In the final survey, we asked if they had taken part in any other peer review besides the SPP in the last 3 years and found that 18% of them had, and these were a mix of learning alliances (21%), MAT processes (21%), internal processes (14%), Special Educational Needs and Disability (SEND), and local authority peer reviews (21%). Given the nature of some of these examples, it is unclear whether participants interpreted 'peer review' strictly in the sense that it could be appled to the SPP, namely, as a programme involving evaluations carried out of schools, including review visits to collect and analyse data and feed back findings and recommendations to the host school. Findings should be interpreted with a little cauation as it may be that a looser sense of using peers for some form of feed back or support was implied.

Most significant partnership in SPP schools

Around 61% of SPP respondents said that the SPP had been their most significant partnership over the last 3 years.

Those that said SPP was not their most significant partnership said that they had other longstanding partnerships that had been more significant (72%) or that COVID had been an issue (19%). A wide variety of partnerships were specified, including local learning alliances and partnerships (43%) and MATs (12%). Those that had stated SPP as the most significant partnership said so because it had added structure for processes and improvement (44%) and it added for good team insight and networking with staff in other schools (44%) as top answers. Around 20% said it had been their only partnership. Treatment schools in the survey that had completed the programme (n=134) were asked what the most beneficial aspects of involvement had been (up to three responses possible), Table 21 below shows responses in descending order:

Table 21: Most beneficial aspects of involvement in the SPP (n=157)

Most beneficial aspects	N	Response %
Building networks and partnerships	70	52
Improvement Champion identification and role	57	43
The frameworks and models used	54	40
Honest and open external feedback and support	54	40
Being a part of peer review and developing skills	52	39
Training and workshop sessions	44	33
Opportunity to visit other schools	28	21
Review process	28	21
Continuing Professional Development	27	20
Improvement of school issues	22	16
Other	8	6
Partnership lead was great help	5	4
General positive comment	2	1

Since many of these points are specific to the programme itself or to peer review, it is difficult to compare to the matched schools' table 19, above, however CPD was highlighted on both to some degree, as well as networking and collaboration and support. The idea of challenge/feedback was evident for both SPP and matched schools.

We also asked SPP schools that had not withdrawn, what the most challenging aspects of their involvement had been (up to three responses possible). Table 22 below gives the top four responses:

Table 22: The top four most challenging aspects of involvement in the SPP (n=134)

Most challenging aspect	N	Response %
COVID-19	59	44
Short staffing or availability	37	28
Timing and workload issues	34	25
Commitment issues or changes within partner schools	18	13

Here we do see similar responses to the matched schools in their citing of COVID, staff time and workload, and variable commitment of other schools as particular challenges. Other issues mentioned included: the 90-day review (n=9, 7%); apprehension about allowing other schools in (n=8, 6%); other (n=7, 5%); workshops and training not effective (n=5, 4%); reception when delivering negative feedback (n=3, 2%); and expectatons of IC role (n=3, 2%).

We asked both matched schools and SPP schools about improvement priorities for their most significant partnership versus SPP, respectively. Table 23 below compares the responses:

Table 23: Comparison of improvement priorities in the SPP and matched schools from surveys

Improvement priorities – SPP schools	N	Response %	Improvement priorities – matched schools	N	Response %
Sustained focus on teaching quality and improvement	39	29	Curriculum development	17	39
Reading	36	27	Writing	14	32
Maths	32	24	Reading	12	27
Writing	31	23	Sustained focus on teaching quality and improvement	11	25
Curriculum development	25	19	Focus on learners and well-being	11	25
Disadvantaged or vulnerable pupils	13	10	Continuing Professional Development	9	20
Upkeep of peer review process	12	9	Remote teaching and post-COVID recovery	8	18
Other	10	7	Maths	6	14
Networking and collaboration	8	6	Disadvantaged or vulnerable pupils	5	11
Remote teaching and post-COVID recovery	8	6	Special Educational Needs	5	11
English	8	6	Support staff involvement and recruitment	4	9
Core subjects	7	5	Networking and collaboration	3	7
Parental engagement	6	4	Upkeep of peer review process	3	7
Behaviour management	6	4	Inclusion	2	5
Pupil Premium progress	6	4	Pupil Premium progress	2	5
Special Educational Needs	6	4	Attendance	1	2
Support staff involvement	6	4	English	1	2
Inclusion	3	2			
Continuing Professional Development	1	1			
Attendance	1	1			

We can see very similar categories of responses in both groups, albeit some ranked in different places, and which may simply reflect the differing nature or scope of the SPP versus the most significant partnerships in the matched groups. Both groups agreed in the majority (66% SPP schools and 57% matched schools) that their improvement priorities had changed over the last 3 years, both citing multiple reasons, with COVID being highest among the reasons stated in the matched schools (SPP schools 32%, matched schools, 41%).

Response to COVID

In both matched and SPP (treatment) schools (completed), we asked comparable questions about how they had adapted partnership working during the pandemic and Table 24 below shows frequencies of the responses in descending order.

Table 24: Comparison of top responses to partnership working due to COVID between SPP (treatment) schools and matched schools

SPP schools	N	Response %	Matched schools*	N	Response %
We stopped all activities on the SPP programme and have not resumed them	49	37	We shifted to online working but have now resumed face to face working	19	43
We paused activity on the SPP programme during lockdown but then resumed face to face reviews once schools re-opened	39	29	We shifted to online working and have continued with this until now	17	39
We shifted to conducting online reviews but have now resumed face to face reviews	18	13	We continued with partnership activities but adapted them, e.g. shortening them or missing out aspects	3	7
We shifted to conducting online reviews and have continued with these until now	11	8	We paused all activities in the partnership or alliance but resumed them as normal once schools re-opened	3	7
We continued with face to face reviews but adapted them, e.g. by shortening the visit time or missing out part of the process	5	4	We stopped all partnership activities and have not resumed them	2	5
Other	12	9	Other	0	0

*****Note: For matched schools, the percentages do not add to 100% due to rounding*

SPP respondents were more likely to have paused activities and not resumed them due to the pandemic and matched schools were more likely to have continued working online than SPP schools. This may reflect the nature of the SPP, which requires school visits to conduct reviews, and although online reviews were an option, uptake was low on the programme.

Key conclusions about usual practice and implications for analysis

Many SPP schools were relatively new to collaboration prior to starting SPP reviews but the majority rated their own partnerships as strong, high in trust, and to have good standards of self-evaluation.

Peer review was very common in matched schools as were other kinds of partnerships that offered support, feedback, and evaluation. This means it is only the unique elements of the SPP that are distinct from much of usual practice, such as the use of the SPP framework, its training, structure of clusters with leads, ICs, and 90-day check-ups. A minority of SPP schools also participated previously and continued to participate in peer review (besides the SPP) during the trial.

We found both matched schools and SPP schools to be involved in a wide range of partnership activity in terms of formality and regularity and inclusiveness of engagement. Motivations for joining the SPP focused on adding structure to existing arrangements and support for school improvement.

Both matched and SPP schools had similar priorities for partnership work with some differences in order of priority, which may reflect the scope of the SPP trial (initially numeracy and literacy focused) and the peer review methodology. The latter may lend itself more to certain areas of focus than others. We also found similar challenges to partnership work, such as time, workload, commitment issues, and especially COVID. Responses to partnership work during COVID were more likely to indicate a complete pause in activity during COVID for the SPP than other partnerships, which may reflect how participants valued live review visits to other schools. Other interview data tends

to support that this was seen as an important element of peer reviews, despite the possibility of a (usually second best) virtual review experience.

Our detailed (treatment) case study schools showed similarly high levels of commitment to engage, reasons for engaging, and capacity to improve, with minor differences such as ability to recruit high-quality teachers. One cluster had introduced more numeracy-related changes prior to starting the programme (Cluster 2) and had longer experience of partnership working than Cluster 1, but less experienced headteachers. Both clusters (some in year one and some in later data collection) had instances of staff churn, illness, and retirement of key senior staff that provided further challenges to participation and full engagement in the project. Each cluster had one school drop out in year one. These contextual factors come into our analysis of findings in research questions 1–5 below.

Around 61% of SPP schools surveyed rated the programme as their most significant partnership. It is not clear if this is a positive endorsement of the partnership overall or a negative one, however there are some potential complexities too, which will be discussed later in the interpretation section and that arise from other data sources. We know that case study schools 1d, 1e, 2a, 2b, and 2e all responded to the final survey and all but school 2b rated the SPP cluster as their most significant partnership. However, even school 2b headteacher commented in this survey that their local cluster (which almost exactly mirrors SPP Cluster 2) was their most significant partnership. Therefore, the SPP may be seen by some participants as about learning 'ways of working', embedded within existing (and more permanent) partnerships, rather than existing as a partnership itself. In any case, saying that a partnership is 'significant' does not equate to it being positive since partnerships may be seen as influential but in sometimes unhelpful ways.

Finally, although we asked matched schools to name their most significant partnership (over the last 3 years) in the end survey, we cannot be certain in many cases if these involved peer review or other elements of school evaluation and support. Comparing the equivalent responses on the survey about impact also needs to consider that the bar was set quite high, namely, the impact of matched schools' most significant partnerships is compared to the impact of SPP in treatment school respondents. The latter group also includes many schools that withdrew before the end of the trial and were considered 'non-compliant'.

Findings according to our research questions

Research questions: 1. In what ways does the SPP influence the capability, culture, and practice of partnerships, leadership, and teachers in involved schools?; 2. In what ways do the elements in the SPP theory of change work in achieving participants' perceived forms of impact?

In most cases, the changes to capability, culture, and practice of the partnerships, leadership, and teachers can be conceived of as intermediate outcomes or moderating variables (and hence they were seen as key to achieving the forms of impact as perceived by participants of the programme and other stakeholders, making research questions 1 and 2 somewhat overlapping). Nevertheless, we can make some distinction between changes to capability, practice, and culture and the processes that led to these changes at partnership, leadership, and teacher levels, and also what mechanisms might be expected to have influenced changes at pupil level. While the original logic model had numeracy and literacy outcomes for pupils, we focus here on any reported outcomes, according to the theory of change used in the SPP learning map (see Appendix 11) and reflecting the revised protocol in response to COVID.

For an overview of the changes perceived to have occurred, we can compare changes from the initial survey in 2018 with the final survey in 2021/22, see Table 25, below (NB, percentage agreement combines strongly agree and agree, and disagreement combines strongly disagree and disagree, unless stated). Analysis was conducted using the Classical Student's t-test ($p < 0.05$, two-tailed). As in other sections, the final survey treatment group responses combine those that completed with early withdrawn schools unless stated otherwise. We should also remember that perceived impact is from the perspective of the headteacher or the SLT that responded to the survey.

Table 25: Comparison of perceived outcomes at partnership, leadership, school, and teacher levels between initial baseline (n=339) and final survey (n=157)

No shading = no significant difference, brown shading = significantly lower for end survey, and yellow shading = significantly higher at end survey.

Initial survey item	Equivalent final survey item	Analysis*
4.2.1 Our partnership is well led we have a shared vision and values and we understand how collective decisions are made	5.1.1 Our partnership was well led, we had a shared vision and values and we understood how collective decisions were made	Significantly higher at end survey (n=140, 89% vs n=278, 82%) with more don't knows in initial survey too (9% vs 1% in end survey)
4.2.2 Our partnership is well managed we meet regularly and communicate well, supported by systems and processes that are fit for purpose	5.1.2 Our partnership was well managed, we met regularly and communicated well	SigDiff – higher in end survey (n=134, 85% vs n=253, 75%)
4.2.3 All school leaders in our partnership have an equal level of status	5.1.3 All school leaders in our partnership had an equal level of status	SigDiff – more disagreed in initial survey (n=23, n=4, 7% vs 3% at end)
4.2.4 My school draws on expertise and support from other schools in the partnership on a regular (i.e. monthly) basis	5.2.12 My school draws on expertise and support from other schools in the partnership on a regular (i.e. monthly) basis	SigDiff – higher in end survey (n=106, 68% vs n=182, 54% initially)
4.2.5 I feel responsible for the success of all schools and pupils in the partnership	5.1.4 I felt responsible for the success of all schools and pupils in the partnership	NSD (n=216, 64% vs n=97, 62%)
4.2.6 Levels of trust are high between the schools in our partnership	5.1.5 Levels of trust were high between the schools in our partnership	NSD (n=143, 91% vs n=291, 86%)
4.2.7 All the schools in our partnership openly and transparently share data, systems, and processes	5.1.6 All the schools in our partnership openly and transparently shared data, systems, and processes	SigDiff – higher for end survey (n=143, 91% vs n=229, 68%)
4.3.1 I am confident in my school leadership team's capacity to make improvements to our school	5.2.1 Participating in the SPP has increased my confidence in my school leadership team's capacity to make improvements to our school	SigDiff – but lower for end survey (n=121, 77% vs n=322, 95% initially)
4.3.2 I am confident in my school leadership team's ability to lead improvement in our partner schools	5.2.2 Participating in the SPP has increased my confidence in my school leadership team's ability to lead improvement in our partner schools	SigDiff – but lower for end survey (n=116, 74% vs n= 306, 90%)
4.3.4 Staff at our school feel a sense of shared responsibility for the success of all the pupils in our school	5.2.3 Participating in the SPP has helped staff at our school feel a greater sense of shared responsibility for the success of all the pupils in our school	SigDiff – but lower for end survey (n=105, 67% vs n=318, 94%)
4.4.2 Teachers in our school feel a sense of ownership of and engagement with of all the schools in our partnership	5.2.4 Teachers in our school feel a greater sense of ownership of and engagement with the improvement aims of all the schools in our partnership	SigDiff – higher in end schools survey (n=54, 34% vs n=52,15%)
4.4.1 Teachers in our school feel a sense of ownership of and engagement with our school's improvement aims	5.2.8 Engagement in the SPP has helped teachers in our school feel a greater sense of ownership of and engagement with our school's improvement aims	SigDiff – but lower in end survey (n=101, 64% vs n=299, 88% initially)

**Note: results compared using Classical Student's t-test ($p < 0.05$). SigDiff = Signficant difference, NSD = no significant difference.*

Table 26: Comparison of impact items on SPP final survey (n=157) and matched schools survey (n=44)

Unshaded = no significant difference, orange shading = significantly higher agreement for matched schools, and yellow shading = significantly higher agreement for SPP schools.SPP schools end survey	**Matched schools survey**	**Significance***
5.1.1 Our partnership was well led, we had a shared vision and values and we understood how collective decisions were made	4.1.1 Our partnership was well led, we had a shared vision and values and we understood how collective decisions were made	NSD (n=140, 89% vs n=36, 82%)
5.1.2 Our partnership was well managed, we met regularly and communicated well	4.1.2 Our partnership was well managed, we met regularly and communicated well	NSD (n=134, 85% vs n=38, 86%)
5.1. All school leaders in our partnership had an equal level of status	4.1.3 All school leaders in our partnership had an equal level of status	SPP schools significantly higher (n=140, 89% vs n=33,75%)
5.1.4 I felt responsible for the success of all schools and pupils in the partnership	4.1.4 I felt responsible for the success of all schools and pupils in the partnership	SPP schools significantly more (n=97, 62% vs n=22, 50%)
5.1.5 Levels of trust were high between the schools in our partnership	4.1.5 Levels of trust were high between the schools in our partnership	NSD (n=143, 89% vs n=39, 96%)
5.1.6 All the schools in our partnership openly and transparently shared data, systems, and processes	4.1.6 All the schools in our partnership openly and transparently shared data, systems, and processes	SPP schools significantly more (n=143, 91% vs n=34, 77%)
5.2.1. Participating in the SPP has increased my confidence in my school leadership team's capacity to make improvements to our school	4.2.1 Participating in this partnership has increased my confidence in my school leadership team's capacity to make improvements to our school	Matched schools significantly more (n=40, 91% vs n=121, 77%) NSD compared to completed schools (82%)
5.2.2 Participating in the SPP has increased my confidence in my school leadership team's ability to lead improvement in our partner schools	4.2.2 Participating in this partnership has increased my confidence in my school leadership team's ability to lead improvement in our partner schools	NSD (n=116, 74% vs n=32, 73%)
5.2.3 Participating in the SPP has helped staff at our school feel a greater sense of shared responsibility for the success of all the pupils in our school	4.2.3 Participating in this partnership has helped staff at our school feel a greater sense of shared responsibility for the success of all the pupils in our school	Matched schools significantly higher (82% vs 67%) NB: NSD compared to completed schools (n=97, 72%)
5.2.4 Teachers in our school feel a greater sense of ownership of and engagement with the improvement aims of all the schools in our partnership	4.2.4 Teachers in our school feel a strong sense of ownership of and engagement with the improvement aims of all the schools in our partnership	NSD (n=54, 34% vs n=20, 45%)
5.2.5 Peer review is firmly embedded in our school's improvement cycle	4.2.5 Peer review is firmly embedded in our school's improvement cycle	NSD (n=98, 62% SPP vs n=22, 50% Matched)
5.2.7 Engagement in the SPP has helped my school to identify its strengths and areas for development more rigorously	4.2.6 Engagement in peer review has helped my school to identify its strengths and areas for development more rigorously	NSD (n=125, 80% vs n=35, 80%)
5.2.10 Engagement in the SPP has enabled my school to become more evidence-informed	4.2.7 Engagement in this partnership has enabled my school to become more evidence-informed	NSD (n=120, 76% vs n=30, 68%)
5.2.13 My school draws on expertise and support from other schools in the partnership on a regular (i.e. monthly) basis	4.2.8 My school draws on expertise and support from other schools in this partnership on a regular (i.e. monthly) basis	NSD (n=106, 68% vs n=34, 77%)
5.4.1 I am very confident there has been a positive overall impact on pupils in our school	4.3.1 I am very confident there has been a positive overall impact on pupils in our school	Matched schools significantly higher (n=114, 73% vs n=38, 86%) (NSD for completed schools 80%)
5.4.3 There has been a positive impact on the ability of our school's leaders to improve partner schools	4.3.2 There has been a positive impact on the ability of our school's leaders to improve partner schools	SPP significantly higher (n=122, 78% vs n=27, 61%)
5.4.4 There has been a positive impact on senior leaders' professional development	4.3.3 There has been a positive impact on senior leaders' professional development	NSD for net treatment schools (n=137, 87% vs n=34, 77%) but SPP completed schools significantly higher than matched (n=123, 92% vs n=34, 77%)
5.4.5 There has been a positive impact on the professional development of middle leaders and teachers	4.3.4 There has been a positive impact on the professional development of middle leaders and teachers	Matched schools significantly higher (n=100, 64% vs n=35, 80%) NSD when comparing completed (n=92, 69%)

**Note: results compared using Classical Student's t-test ($p < 0.05$)*

We can see from Table 25 above that there were many perceived positive changes to how well their partnerships had been led, the use and sharing of data, expertise, and other types of support. There was also a perception by respondents (mostly headteachers) that teachers in their schools felt a greater sense of ownership (compared to before the SPP) of and engagement with the improvement aims of all the schools in their partnership, albeit with only 34% agreement on this statement in the final survey. There was no change in levels of trust in the partnership and no change in the feeling of responsibility for the success of all schools and pupils in the partnership (although the majority continued to feel this responsibility).

However, there were decreases in the confidence in their SLTs capacity to make improvements to their own school or in partner schools. There was also a decrease in agreement that staff at their schools felt a sense of shared responsibility for the success of all the pupils in their school or for the school's improvement aims. It is unclear if this represents over-optimism initially and/or a sense of greater understanding of the challenges in retrospect. The very different context of schools due to the pandemic may also have played a significant role in how respondents viewed these matters. Overall, there is a trend towards greater perceived impact at partnership rather than leadership or teacher level.

Comparison of end treatment school survey and matched schools

Below we look at responses from the final treatment group survey compared with matched schools in relation to research questions 1 and 2. Matched schools that participated in the survey were not found to have any significant differences on Ofsted, urban/rural, or IDACI characteristics (Classical Student's t-test, $\boldsymbol{p} < 0.05$). We first conducted an analysis of responses between the completed and withdrawn schools on the SPP end survey and found only Section 5 (perceived impact) contained significant differences in responding, with all but five items out of 31 items showing higher agreement (as would be expected) for completed schools and the remaining items showing non-significant differences (Classical Student's t-test, $\boldsymbol{p} < 0.05$). On this basis, we made the decision to include *all treatment schools* surveyed in the comparison below, namely, combining completed and withdrawn schools (see Table 26 above), unless otherwise stated.

The majority of the above statements showed no statistically significant differences in agreement between matched school perceptions of their most significant partnership and SPP school responses. It is worth noting that peer review was seen as ***firmly embedded in 50% of matched schools***, further highlighting that these practices were prevalent in primary schools in general, so any differences in responses are to do with specifics of the SPP rather than peer review per se. We should note that matched schools may be skewed in that they were interested in responding to a survey on 'partnership working and peer review'.

However, we can see that SPP respondents were significantly more likely to say that school leaders in their partnership had an equal level of status; that they felt responsible for the success of all schools and pupils in the partnership, that they openly and transparently shared data, systems, and processes, and that there had been a positive impact on the ability of their school's leaders to improve partner schools.

Matched school respondents rated their most significant partnership more highly compared to SPP respondents in terms of confidence in SLTs capacity to make improvements to our school; staff feeling responsible for the success of all pupils in their school; having a positive overall impact on pupils in their schools and on the professional development of middle leaders and teachers.

Again, we can see high levels of agreement on all items for both SPP and matched groups and that all significant differences (Classical Student's t-test, $p < 0.05$) vanish if we only look at schools that completed the SPP trial, apart from the confidence of pupil-level impact in their own schools. Therefore, the SPP schools stand up very well and sometimes exceed confidence in improvements at partnership, leadership, and teacher levels against perceptions of matched schools' most significant school improvement partnerships. However, for SPP schools, there is a tendency towards higher confidence in impact/ownership of change at partnership level rather than at school or pupil level.

Changes at partnership, leadership, and teacher levels and their impact

The SPPs theory of change anticipates change at: partnership level; school level; and teacher level, with all these ultimately contributing to impact on outcomes for pupils. Here we look at various sources of data—the case studies, observations of training, shadowing of school reviews and stakeholder interviews, and a few questions that were specific to the end SPP survey participants. We describe changes to capacity, culture, and practice (research

question 1) and consider the impact of each of these types of changes on other levels, including the last section, which describes some examples given of pupil-level impact (research question 2). This quote from a headteacher in one of our partnership lead group interviews sums up many of the feelings about involvement in the SPP and how the culture and practices changed in partnership working compared to usual practice:

> *I think one of the most important effects for us really has been the stimulus for collective working really. There's another expression, isn't there, about teachers being islands or whatever, being isolated and this has been an opportunity for teachers to work together more honestly and openly because the structure is external but the structure of the process is something that everybody understands and people buy into to [sic] a different degree at the beginning but when they see just the impact of having opportunities to talk to each other, to explore ideas in a low stakes environment, then it builds people's confidence. Because it is so completely different from monitoring and moderation or anyone else, like a trust body or an LA [local authority] coming in to look at the quality of teaching, it's just a purely enabling activity but it's also quite freeing because it doesn't matter if you get things wrong.* (Group D, p5).

Partnership level

We saw from the survey results above that many participants had very positive experiences working in partnership. The majority of partners felt that they worked as equals (peers), had shared values and vision, used collective decision-making, felt responsible for the success of all schools and pupils in the partnership, and openly shared practices. Most partnerships were seen to be well run, with regular engagement, they shared expertise, and had high levels of trust. Other data sources (in particular our case studies) revealed further subtleties about how the SPP strengthened and deepened capabilities, practices, and partnership culture. These are interpreted below.

<u>Greater structure</u>

Headteachers reported more consistent, wider, and structured engagement in their partnership practices, as these two instances reveal:

> *It's given us a structure because we worked quite closely together beforehand. We had quite good relationships, we were quite open but we didn't have a structure of school improvement in each other's schools. (*Group A, p1, partnership leads group interview)

> *It has changed the way in which we work. We used to really just meet up once a term. And, you know, we might do some work scrutiny or we might do some writing moderation. It has widened it. [Pause] I can't really see us going back to the way we worked before. I think we will probably carry on in this vein, in some format.* (Headteacher, school 1e, year one).

Taking part in reviews helped further develop trust and openness; in our observations of peer reviews, participating schools talked about how the SPP process had helped them to develop a culture of openness (partly because the reviews were seen as 'low stakes') and had enhanced how schools and partnerships work together to identify and evaluate practice.

These reviews also served to develop much more intimate knowledge of each other's working environments:

> *Now having had three reviews with the same partnership, peer reviewer leading it, we kept the same teams all the way through and that's really helped. There's such an openness of sharing of data and when the reviews happen, there's a really good triangulation of data, it's not just looking at one thing, you're gathering evidence from all sorts of aspects of school life. They know our schools inside out now and it's very open and very transparent.* (Group E, p1, partnership leads interview)

SPP partnerships helped to develop a collaborative culture of school improvement:

> *So just setting that up at the beginning, the shared joint values, everyone knowing what we were trying to achieve and going into the reviews with our eyes open completely, getting to the point where you're just developing and enabling a school team to come up with its own solutions has been phenomenal at lots of different levels. It really has created the opportunities for genuine culture and practice for self-improvement. It's been brilliant.* (Group D, p5, partnership leads interview)

A few participants cited evidence that partnerships were engaging more stakeholders and becoming more inclusive, beyond the involvement of headteachers, for instance, involving more teachers and middle leaders:

- Within the last year they have started subject leader meetings to share good practice, look at where the subject is in different schools, what's working and how they are monitoring. (Interim head of school [IHoS], school 1a, year one)

And school governors:

> *So, they have decided to think about governor-to-governor peer review. We took a theme—how each school used its sports grant—it made it very straightforward in the first instance.* (Observation of workshop, leading collaborative school improvement [South], November 2019, headteacher Cluster 2)

Training and practice in both reviewing and in the IC role was also valued in the case studies by participants in senior leadership roles below the level of headteacher.

Evidence from our case study schools showed how, with increasing trust, schools were able to be more challenging to each other, for instance one headteacher (school 1c, year three) said: *'We're not afraid to challenge one another. I think we used to be a little bit. I think that we did have that caution early days, but I think people are much more likely to be outspoken and say what they really think now'.*

Nevertheless, in a review of a school in Cluster 2, year one, the EHT that led this, spoke about how she did not want to challenge the new headteacher too much, in the context of a school where morale might be lower given recent visits from Ofsted and a local SIP. Thus, the level of challenge was something gauged by reviewers in the position of greater knowledge about the context. In the partnership leads interviews, one headteacher commented that an *'inward looking'* headteacher in the partnership became more open to working collaboratively, attending more meetings, and asking more open questions (Group B, p1).

Over time, we can see from accounts in the case study schools and in partnership leads' interviews, that SPP reviewing practices became embedded into existing partnerships. In an observation of the workshop, 'leading collaborative school improvement' (South), November 2019, a headteacher from Cluster 2 suggested revisiting their local MoU and adding peer review to this. They also wanted to use SPP reviewing methodology in a forthcoming local authority review of SEND in their cluster, as they saw it as more effective than a previous peer review they had taken part in.

There were examples, particularly from case studies of knowledge exchange being more fluid. The acting head of school 1e said in a year three interview:

> *Now as time has gone on, we've had training days together, so all the staff have spent days together and lots of things like that, like staff from the other schools have delivered CPD for the other schools, not just the improvement workshops but completely unrelated CPD … It's not formal, we're not charging each other, we'll just pay it back with another favour, but I do think that's got deeper now, I'm not sure, I think all of our teachers, if they couldn't contact me and [EHT name] for whatever reason, all of our teachers would know that the Heads at one of our partnership schools would be the people that they would ask for advice.* (Acting head, school 1e, year three)

The EHT of the same school and cluster, confirmed this picture, adding that this was aided by greater knowledge of the needs of each school due to the review visits.

<u>Processes involved in partnerships that led to impact</u>

In addition to the reviewing practices that helped identify problems or validate/challenge schools' existing approaches, the SPP methodology focused on school-to-school support. We saw examples of this from multiple data sources. In our shadowing of reviews, there were examples of how schools were cross-pollinating ideas through visiting each other's schools. Observations of training revealed examples of shared CPD across clusters that had been stimulated through the process of mutual reviewing. Schools in our case studies were also able to identify where expertise was located and share this with each other. Observations of training workshops also showed that schools could help each other by conducting progress checks, the latter became more emphasised in training sessions by EDT and this seemed to have been picked up in our case study schools too.

The IC role in the improvement process was particularly valued, as we can see from the final survey, see Table 27 below.

Table 27: The Improvement Champion role has been important to our school's capacity to engage successfully in the Schools Partnership Programme (n=157)

Scale	*N*	*%**
Strongly agree	43	27
Agree	54	34
Neither agree nor disagree	27	17
Disagree	16	10
Strongly disagree	16	10
Don't know	1	1

***Note: Percentages do not sum to 100% due to rounding.*

Interviews with partnership leads and headteachers, confirmed that this role was particularly valued as it not only led to a concrete follow-up process in the reviewed school, to ensure actions were initiated in response to the review findings, but also as it provided leadership opportunities for those trained in the IC role.

Leadership

Multiple accounts indicate that the SPP provided excellent CPD, particularly for leaders—for teachers to become successful middle leaders, for senior leaders to move into headship, and for forms of system leadership, including leading collaborative change across the partnership. Above we saw in the survey that, compared to matched schools, SPP participants agreed more strongly that there had been a positive impact on the ability of their school's leaders to improve partner schools. In some cases, this had led to the promotion of staff who had then moved elsewhere. This headteacher interviewed at a partnership lead group interview expresses this, thus:

> *I think one of the biggest things that's come out of it though is the way that it has developed and empowered deputies and middle leaders because I think they got little chance to go into schools to do that sort of work as either improvement partners or peer reviewers. The programme has developed in our school, those middle leaders and deputy leaders. They've been into other schools. I've watched them grow and blossom. I've lost members of staff because they've gone off and applied to be deputy heads but that's actually great and that's what we should be doing for our staff.* (Group B, p2)

Of particular note is the role of ICs. This unique leadership role in the programme had several important benefits. It gave opportunities for both middle and senior leaders to lead in other schools. When averaged over the 2 years of training (more were trained in year two, to increase capacity), around a third of ICs were teachers, a third were middle leaders, and a third were more senior leaders (see Table 28, below).

Table 28: Who were the Improvement Champions?

School role	Year 1		Year 2	
	N	%	N	%
Class teacher/teacher governor	130	36	81	33
Any kind of teacher/middle leader – coordinator/phase/SENCO/lead for anything, etc.	80	22	100	41
Assistant/deputy headteacher/SLT/lower school lead	135	38	62	25
Head of school	15	4	2	1
Total	360	100	245	100

Most staff were nominated by their Heads to take up this role (74% of our small survey of ICs in year one, n=31) and this was a way of growing leadership capacity at the school. Most of the headteachers we interviewed in case studies and in partnership leads groups were complementary about how their staff had taken up this role and how they had blossomed. Some of the ICs we interviewed commented on how their communication skills had improved and how their knowledge of good practice in other schools had increased and their abilities to lead coaching style workshops were valuable.

There were a few challenges mentioned by headteachers too, including staff being too introverted or not having other personal qualities needed to undertake this role effectively. There were schools in our case studies who did not nominate anyone to be IC due to capacity issues too.

The ICs' training gave a methodology, the skills and tools to lead an improvement process in a partner school. Headteachers and ICs we interviewed were highly complementary on the training they received for this role. We observed the training in year one and participants were shown a wide range of group tasks they could set up to lead a coaching style workshop that would be focused on enabling school staff to come up with their own solutions. EDT evaluations show this training was valued very highly and as observers we regarded this very well too. One partnership lead commented that an external SIP evaluation had complemented the leadership skills of their middle leaders and the headteacher felt that this was due to the leadership experiences they had gained from the SPP training. Another headteacher in our partnership leads' interviews suggested that this methodology provided a model that could be used to lead improvement in their own schools too. Headteachers and partnership leads we interviewed said that the IC coaching model used in improvement workshops led to a greater sense of ownership of the change process by teachers and leaders, confirming the survey results.

Learning to evaluate and review

Leaders in the programme were able to develop skills of evaluation, reviewing other schools, and providing feedback. The SPP methodology was regarded very highly in our end surveys:

Table 29: The SPP framework for reviewing has been highly effective (n=157)

Scale	N	%*
Strongly agree	48	31
Tend to agree	76	48
Neither agree or disagree	18	11
Tend to disagree	7	4
Strongly disagree	6	4
Don't know	–	–
Not applicable	2	1

*****Note: Percentages do not sum to 100% due to rounding*

Table 30: The benefits that accrue from undertaking SPP reviews are not sufficient to justify the time and effort involved (n=157)

Scale	N	%
Strongly agree	2	1
Tend to agree	7	4
Neither agree or disagree	15	10
Tend to disagree	51	32
Strongly disagree	81	52
Don't know	1	1
Not applicable	–	–

Focus

The self-evaluation and pre-review process was designed to hone leaders' and school staff's ability to sharpen the focus of their improvement activities.

The reviewer training we observed emphasised getting this focus right, with examples, although this was not always successful in practice in year one reviews. After one round of reviews, there was some evidence in our case studies and in observations of training and reviews, that participants had reflected on this and were understanding more what a good 'focus' question would consist of. In the review of school 2d in year one, the EHT leading the review talked of needing to advise the headteacher of the school to focus on a more specific area, since he wanted to review all teaching and learning. However, in school 2b, the lead reviewer felt that a focus on spelling and grammar had been too narrow and that it had been steered this way by the EHT who had already formed a view of the school's progress in this area and therefore was not open to learning sufficiently during the review. The ICs that went to this school also confirmed that this area did not give much leeway to discuss strategies other than ones already in place. There was evidence from two observations of reviews of progress workshops at the end of year one that more schools were

adopting a shared focus approach, so that they could consolidate their learning in one area, although it is not clear how widespread this was.

The role of the improvement workshop in helping to focus minds on the issues was also considered significant:

> *... that person that goes in and opens the staff team's eyes a little bit to the missing bit of the puzzle ... diving in with the staff teams and saying to them: "What are you doing really well? There are strengths here, how are we going to tackle this now, what do you want to achieve?" It's that breaking it down a little bit because sometimes when you see a problem you almost put up a wall, don't you, and you think: "Goodness me, there's so much to tackle here", but when you strip it right back to the basics, like we do when we do our RAG [Red Amber Green] ratings and all of the different affective models that we've been given on our training, it really does help the staff to break it down.* (IC interview, year three)

The development of peer reviewing and evaluation skills helped some schools we observed to become more 'Ofsted ready', in terms of understanding their own strengths and weaknesses and identifying their own strategies to deal with them. Furthermore, reviews gave recommendations and strategies for how school leaders could continue to evaluate the impact of their improvement approaches (seen in observation of training, June 2019).

<u>How reviewing led to change</u>

Feedback:

Feedback provided by reviewers and followed up by ICs was generally valued highly and this is seen in our survey results above and supported by our case study interviews. The reviews gave an additional, external 'pair of eyes', and where the reviewer themself was regarded as being high in 'credibility', such as the EHT in case cluster 2, the feedback and recommendations they gave was taken on board more readily. However, as the focus was decided on principally by the school, in discussion with the lead reviewer, this usually built on existing school self-evaluations. Reviews sometimes were simply validating a school's existing approaches to what was reviewed, and therefore it is unclear what further added value the review or the improvement workshop made. However, usually there was some sense that the reviewer had refined their understanding somewhat. The validation was also welcomed as an endorsement for the headteacher to continue or consolidate their approach. At other times, as we saw in observations of reviews, there was a distinct celebration of good practice that could enable an approach to be shared across schools in the cluster.

Social learning:

There are many examples, especially in our case studies of social learning occurring, for example through observation, reflective learning, social influence, and modelling. By discussing the outcomes of reviews in schools, headteachers and reviewing teams, along with later whole staff in improvement workshops, participants were able to engage in a comparison of professional judgement in a way that would be lacking from an Ofsted inspection.

> *So because it's your peers coming in, people don't see it as a judgemental process. We're quite clear about that when we go into things; you're not here to pass judgement. You're here to I suppose … I talk about this phrase a lot but in a way you air your dirty laundry to people and actually you want another person's perspective on this particular problem that you identify in the enquiry approach.* (Group D, p1)

The involvement of partners in analysing school data also had a motivational influence on schools, coupled with positive peer pressure, to see these improvements through. For instance:

> *I knew that I wasn't quite satisfied with how reading was going on and there was a slightly mixed picture but having the review in and having everybody's eyes on what was going on, all through school and looking at all of the data, different kinds of data and talking to children, talking to staff as well, it [the review process] gave us the impetus to carry it forward.* (Partnership lead, Group E, p1)

The training and reviewing elements of the programme were valued by one of our partnership leads interviewed, who said that it gave valuable space to reflect and learn that schools may not otherwise find the time to create.

The coaching methodology of ICs was also seen as a stimulus to change, and we saw this in reflections from ICs after year two training. The elements that were seen as successful to this were: being 'neutral' to the judgements made in the review report, since they were not involved directly in the review itself; being 'one of them', often teachers or middle leaders with similar challenges; and being facilitators of change rather than as an 'expert' who was dispensing advice. Furthermore, by working in pairs, usually with another IC from another school, they were seen to model the peer working process itself.

One of the great values of the reviewing approach was that many participants involved in this were able to learn from the visit, 'pinch' or share ideas seen elsewhere, and feed these back into their schools. We saw evidence of this across our case studies, where approaches to teaching maths, and spelling and grammar were shared in Cluster 2, for instance, and in our observations of several review processes.

Teachers

The partnership lead group interviews confirmed that leadership opportunities were created through the IC role. The improvement workshops also helped facilitate ownership of the improvement process and we saw an increase in agreement that teachers felt a sense of ownership throughout the programme from the end survey compared to the initial one. Our observations of reviews also revealed how some schools were able to develop a whole school understanding of issues, including by involving TAs in reviews. In interviewing ICs (most of whom were teacher and middle leaders), we know that they felt a strong sense of collective responsibility to ensure all schools in the partnership improved. This was particularly noticeable among those in larger partnerships such as MATs and TSAs. In terms of improving teaching quality, it is harder to ascertain the role of the SPP, and depended mostly on the strength and commitment of the follow-up process, and there is insufficient evidence here to comment.

<u>Developing research-informed practice/interest in this (teachers and leaders)</u>

After the second ICs' training, which included use of research evidence into the process, we saw a greater awareness of including research in improvement practices in our observations of improvement workshops. These were confirmed by headteachers in interviews:

> *I think people are more prepared to move forward and make changes because it's very evidence-based and the Improvement Champion workshops, using the EEF research have been instrumental in that because ... it's hard to argue with when that's been carefully researched. So, I think that sense of reflective practice and moving forward together as a team, is very well established.* (Headteacher and partnership lead, 6-school cluster, year three interview)

As the account above shows, and this was echoed in our case study schools (review of school 2e), that by introducing published research, ICs could introduce a stimulus for change that had some intrinsic 'authority'. Another partnership lead described the use of such evidence as: *'Very powerful because it's very persuasive, it's difficult to argue with, it does generate an awful lot of buy-in'* (Headteacher and partnership lead, 7-school cluster, year three interview).

We came across examples of school leaders, as well as ICs, now giving staff research articles to read: *'I selected them mutually with the Head'* (IC interview), and how research was influencing change more widely in many of the SPP schools: *'I wanted some concrete evidence, not just me saying thi*s' (Senior leader and IC, interview).

<u>Negative consequences of IC roles</u>

For teachers taking on the role of ICs, there were also negative consequences, as they were taken out of the classroom on several occasions. The process, the involved EDT training (two half days), attending a school review feedback session, usually towards the end of a school day, then meeting with a colleague to plan the workshop (a day) then conducting the workshop (half a day), was very time consuming as well as requiring considerable skills of coaching and facilitation, often working with someone from another school who was unknown to them. We also know from our case studies that in at least two cases, the school staff were not very receptive to the workshop and in one case, attendance was low. Where there were shortages of ICs, for instance in case study Cluster 2 in year one, one IC had to conduct up to four of these workshops in year one. As these staff were often considered to be high-calibre teachers and middle/senior leaders, the loss to the school of these days is worth factoring in alongside the benefits.

Follow up

We heard many accounts of headteachers, ICs, and reviewers from training, shadowing of reviews and case studies that they were unsure of follow up, after the review/improvement workshop. Although the 90-day follow up was increasingly emphasised in year two, this generally involved only the headteachers discussing progress in partnership meetings. Even among headteachers there were doubts about the follow-up process:

> *Who is responsible for follow up? Depends on brokerage. Hard to impose structure on follow up.* (Headteacher, review of reviews workshop, June 2019).

We heard from ICs in our case studies that they were not clear how the impact of reviews was meant to be ascertained and a lack of awareness about whether the improvement workshop they facilitated, led to any substantial changes in the recipient school. Whether teachers at their own school were aware of progress from reviews was something we cannot be clear about either.

Perceived forms of impact at pupil level

While 76% of respondents were very confident there had been a positive overall impact on pupils in their schools, specific pupil-level improvements were not substantially documented in this survey. In the context of 2 years with no Standard Assessment Test (SAT) results to drawn upon, this is perhaps unsurprising. We also know that schools had a wide range of areas of focus, particularly online learning, well-being, and the recovery curriculum, after the COVID pandemic (see more below in research question 5 findings), so determining the impact of these approaches, which only came into year three of the programme, would in any case be too soon. The analysis of pupil outcomes was also problematic due to the wide variety of areas of focus of our treatment schools and the difficulties of systematically capturing this. It is likely that impact would have been largely down to efforts made by teachers in day-to-day classroom practice, guided by senior and middle leaders. We did hear examples of this, for instance of a teacher-led development trial in one school based on lesson study, however, each school focus would have its own unique requirements, requiring significant investments of time and to have a clear strategy and buy-in from teachers, who may also need support.

A few more examples of pupil-level impact were given when we interviewed partnership leads in year three about impact at pupil level. Seven responses were recorded for which they felt there was some evidence, either in their own school or in one they had reviewed in their partnership. Two indicative responses are paraphrased from field notes below (letter = interview group ID and p = participant number, see Appendix 8 for full details of group interviews):

> *Work with EAL [English as an Additional Language] and Pupil Premium children in one school has had impact by year 2, e.g. better use of vocabulary and outcomes.* (Group B, p2)

> *I haven't got this in front of me, I'm pretty sure it was 80% of children made accelerated progress. Obviously, we haven't got the final data because the SATs went pop didn't they but certainly before we got to that data point, we were looking at 80% accelerated progress for those children, in our focus group that we'd focused in on.* (Group E, p2)

However, overall, from the case studies, there were many more examples to do with changes at partnership, leadership, school, and teacher levels then at pupil level. This quote from an EHT (school 1e) in a year three interview also captures the complexities in gaining evidence of pupil-level impact:

> *What the reviewers found when they did our Maths Mastery review was that they were very, very complementary about what we were doing … but they didn't think that we could always have evidence of the learning journey, because a lot of the Maths Mastery is using manipulatives and isn't actually recorded. So, we worked with the Improvement Champion to consider ways of capturing the pupil's learning journey, and you don't want to stilt the children's thinking processes by asking them to record … Hand on heart, I'm not sure whether that's still happening because pandemic stopped a lot of things and I'm not sure, but I know we did start to do that, and yes, we moved away from maths after that.* (EHT, school 1e, year three interview)

The quote above shows that part of the recommendation of this review was to do with ways of capturing the impact of the change. In the IC interviews, ICs mentioned that there was insufficient follow up to know whether or not there

had been impact at the school reviewed (and where ICs facilitated an improvement workshop). Partnership lead interviews in year three suggested that more emphasis was given to the SPP '90-day follow up' compared to year one, but that nevertheless establishing clear impact was a challenge.

At an observation in June 2019 of a 'review of reviews' workshop, one headteacher noted that:

> *The outcome of the review was that we realised we weren't doing enough scaffolding. So it has kept it high on the agenda—and we are thinking more about how we support those learners—e.g. via lesson study, coaching, etc. How to support and scaffold. We have seen some green shoots of low prior attainers shifting, especially for Y1 and 3 and 4—not so much 5 and 6. Would it have happened anyway? Maybe, but by tying it in to this has helped. Teachers perhaps not thinking hard enough about how to ensure that progress.* (Headteacher, observation of training, June 2019)

The examples given to us by participants in interviews reveal the subtleties in the ways that schools reported 'impact', sometimes through observing classes and particular pupils, other times by noting more consistent ways that classes were taught and how these suggested promising outcomes for pupils. It was not clear that improvements were entirely due to the programme from the above examples either, but that the reviews may have shone a further spotlight on the area and provided a stimulus to change.

Research questions: 3. What factors influence schools' and clusters' ability to engage in, participate fully in, and successfully implement and sustain their involvement in the programme?; 4. What distinguishes schools and clusters that have not continued with the programme?

In this section, we consider what our evaluation shows about schools that were more engaged, more successful, and who completed the programme compared to ones who were less engaged, and had less success in implementing aspects of the SPP. It also looks at reasons why schools withdrew from the programme. The data is taken initially from end surveys and relationships on items with other factors are considered, such as IDACI score, Ofsted grade, and urban/rural context. Further data from interviews and case studies is used to give further nuance.

Survey relationships

We wanted to know if participants' perceptions about the programme were related to deprivation levels of students at the school, the prior Ofsted grade, the urban or rural context of the school, or how much the school had engaged in the programme. We looked at levels of deprivation affecting children as indicated by IDACI quintiles, Ofsted grade of school, compliance (fully, partially, non-compliant) variables, and ran Classical Student's t-test ($p < 0.05$, two-tailed) to check for significant differences in survey responses within these categories.

We also wanted to find out if initial levels of partnership working, trust, and capacity related to how positively the programme was perceived. The five variables taken from an initial partnership audit were: how long the partnership had been working together; strong buy-in and clear motivation for joining the SPP; partnership effectiveness; capacity to conduct peer review; and alignment with current accountability systems (Scale: significant evidence; some evidence; or no evidence) (Classical Student's t-test, $p < 0.05$, two-tailed).

We start by reporting relationships between partnership audit ratings and the end survey. Overall, there were relatively few statistically significant differences between groups when responses were compared on initial audit ratings, see Table 29 below.

<u>Improving teacher and staff partnership work in mature partnerships</u>

Those in longer pre-existing partnerships (from the initial audit) were more like to see improvements in: 'Teacher sense of ownership and engagement with the improvement aims of all the schools in the partnership' and 'effective staff collaboration with staff in other schools across the partnership'.

<u>Puzzling relationhips</u>

The statement: 'Our partnership was well led; we had a shared vision and values and we understood how collective decisions were made' was somewhat surprisingly inversely related to prior estimation of evidence for partnership capacity to conduct peer review and partnership alignment with accountability systems (from the initial partnership audit). Similarly, the statement: 'Our partnership was well managed, we met regularly and communicated well' was

inversely related to prior estimation of evidence of partnership capacity from the initial audit. It's hard to interpret these findings apart from speculating that where partnerships had a low starting estimation of their capacities for working in partnership, meant that they were comparatively very happy with the SPP experience in relation to a low level of expectation.

There was a significant inverse association between partnerships' prior motivation levels and whether the partnership members were seen to have an equal level of status in the final survey. In other words, stronger and more consistent motivation to join the SPP was related to lower agreement on equal status.

To the statement: 'The SPP partnership has been a key source of support for me during the pandemic', there was a significant assocation between strong agreement here and lower levels of agreement about evidence of alignment with current accountability systems.

It is unclear what these associations might signify apart from being either chance occurrences or an artefact of some other factor.

Table 31: Significant assocations between initial partnership audit (n=64 audits* and final treatment survey n=157)

SPP end survey	Partnership (together)	Partnership (motivation)	Partnership (effectiveness)	Partnership (capacity)	Partnership (alignment)
5.1.1 Our partnership was well led, we had a shared vision and values and we understood how collective decisions were made				Significantly more agreement where previously had stated no (n=9, 100%) or some evidence (n=63, 93%) compared to significant evidence (n=55, 86%)	Significantly more agreement where previously had stated no (n=58, 98%) compared to significant evidence (n= 53, 87%)
5.1.2 Our partnership was well managed, we met regularly and communicated well				SigDiff: Agreement: some evidence (n=63, 93%) compared to significant evidence (n=51, 80%)	
5.1.3 All school leaders in our partnership had an equal level of status		Previous equal motivation inversely related to status (no evidence, n=6, 100%; some evidence 86%; significant evidence, n=69, 91%)			
5.2.4 Teachers in our school feel a greater sense of ownership of and engagement with the improvement aims of all the schools in our partnership	1–5 years (n=23, 40%) vs < 1 year (n=13, 23%) agree				
5.2.9 Engagement in SPP has enabled staff in my school to collaborate with staff in other schools across the partnership more effectively	Over 5 years (n=21, 78%,) < 1 year (n=32, 59% agree)				
5.2.14 The SPP partnership has been a key source of support for me during the pandemic					Higher agree where significant evidence (n=12, 57%) vs some (n=26, 43%)

*(Classical Student's t-test, $p < 0.05$). *Note: Each partnership lead at the start of the programme completed one partnership audit. Some of these contained schools that never started the programme and schools not taking part in the trial, such as special schools, infants, or junior schools. SigDiff = significant difference; SPP = Schools Partnership Programme.*

Compliance assocations

As might be expected, we found many significant relationships between compliance level and perceptions of impact of the programme. In other words, fully compliant schools were more likely to agree that partnership was well run, the school was able to rigorously identify its strengths and weaknesses, and that the programme had had a positive impact on teachers, leaders, the partnership, and pupils.

Ofsted associations

We wanted to know if Ofsted grade related to survey responses, since schools may feel great capacity to engage in peer review at higher grades, for instance. However, for SPP schools, all but one survey item showed non-significant associations for Ofsted grade. In the one case there was a significant relationship, it was for the item: 'Peer review is firmly embedded in our school's improvement cycle'—specifically, schools with Ofsted grade 1 agreed more strongly (n=19, 83%) than schools with Ofsted grade 2 (n=69, 57%). There were few respondents with grade 4 in any case (3) and only nine schools had a grade 3, which would likely make numbers too small to detect significance (NB: 23 schools surveyed were grade 1 and 122 schools were grade 2, 9 schools were grade 3, and 3 schools were grade 4).

We found no significant relationships on the matched survey responses on Ofsted grade.

Rural/urban context

We found no significant associations with end survey responses in our treatment group with the urban/rural context of the school.

IDACI quintile

Schools were rated 1–5, where 5 shows a higher IDACI. Of the 157 schools (i.e. school leaders) that completed the final treatment survey, 41 were in IDACI quintile 1 (least deprived), 41 were in IDACI quintile 2, 39 were in IDACI quintile 3, 22 were in IDACI quintile 4, and 14 were in IDACI quintile 5. The associations below (particularly related to quintile 5) are therefore based on low numbers and must be treated with caution.

That said, we found a number of interesting associations on the SPP end survey and IDACI quintile score that are worthy of further exploration in future research, see Table 33, below. ***These significant relationships may suggest several perceived benefits of the programme were felt more strongly by schools with higher levels of deprived children.*** These included levels of trust in the partnership, headteacher feelings of collective responsibility for pupils in the partnership, of having a well-run partnership, confidence in the benefits of peer review, and embeddedness of peer review in school improvement. Also, respondents in high-deprivation areas gave higher agreement ratings on the ownership of teachers of their school's improvement aims and of their school being more evidence-informed because of SPP participation.

By contrast, in the matched survey, we did not find any significant associations against IDACI score.

These associations suggest that among participants from schools with more deprived intakes, the SPP was seen to fulfil a particularly important role in collective school improvement. The explanation for this is not entirely clear; partly because we are unable to track responses on this survey back to particular clusters. However, Table 32 below suggests that this may be the case. In our overall treatment schools database (n=422), 48 IDACI quintle 5 schools came from 23 partnerships (135 schools in total). From Table 32, below, we can see a tendency towards 'bunching' of lower IDACI quintile scores in treatment schools' partnerships, where, if one school was in IDACI quintile 5 only 20% (of the 135 schools) contained schools in IDACI quintiles 1 and 2 combined, compared to 46% in the overall treatment group.

All but one of the IDACI quintile 5 treatment schools were in urban areas, and of the 48 total schools that were in this category, 31 came from the North of England or Midlands: Sheffield (n=14); Bolton (n=4); Manchester (n=6); Bury (n=1); Northamptonshire (n=2); Doncaster (n=3); and Salford (n=1). While 17 came from the South: Brighton (n=2); Havant (n=3); Milton Keynes (n=3); Dover (n=3); Canterbury (n=1); Margate (n=1); East Malling (n=1); Chatham (n=1), Bournemouth (n=1); and Edgware (n=1). It is unclear to what extent being a school in the North of England had an impact on the significant assocations we found in Table 33 below.

Table 32: Comparison of IDACI quintile distributions between overall treatment group (n=422) and where one school in partnership had an IDACI quintile score of 5 (n=135)

Quintile for IDACI deprivation score	Overall treatment schools counts (n=422)	Overall treatment schools % (n=422)	Counts where one school in cluster is in IDACI 5 (n=135)	Percentages where one school is in IDACI 5 (n=135)
IDACI 1	87	21	6	4
IDACI 2	106	25	20	16
IDACI 3	95	23	27	20
IDACI 4	80	19	33	24
IDACI 5	48	11	48	3
Missing data	6	1		

IDACI = income deprivation affecting children index.

We found no statistically significant associations between IDACI and compliance level, whether they had withdrawn from the programme, or whether the SPP was considered their school's most significant partnership overall. So, deprived schools respondents were no more or less likely to complete the SPP during this trial compared to less deprived ones, and were no more or less likely to find the SPP partnerships significant compared to other partnerships.

Table 33: IDACI quintile score and final treatment significant survey associations (n=157)

SPP end survey item	IDACI*
5.1.1 Our partnership was well led, we had a shared vision and values and we understood how collective decisions were made	IDACI 5 more agree (n=14, 100%) than IDACI 1 (n=35, 85%) (although not a consistent upward trend)
5.1.4 I felt responsible for the success of all schools and pupils in the partnership	IDACI 5 more agree (n=11, 79%) than IDACI 1 (n=21, 51%)
5.1.5 Levels of trust were high between the schools in our partnership	IDACI 5 more agree (n=14,100%) than IDACI 1 (n=36, 88%), or IDACI 2 (n=36, 88%)
5.1.6 All the schools in our partnership openly and transparently shared data, systems, and processes	IDACI 5 higher agree (n=14, 100%) than IDACI 2 (n=37, 90%), and IDACI 1 (n=36, 88%)
5.2.2 Participating in the SPP has increased my confidence in my school leadership team's ability to lead improvement in our partner schools	IDACI 5 higher agree (93%) than IDACI 1, IDACI 2, and IDACI 3 (n=29, 71%; n=30, 73%; and n=28, 72%, respectively)
5.2.5 Peer review is firmly embedded in our school's improvement cycle	IDACI 5 higher agree (n=11,79%) than IDACI 2 (n=21,51%) (also higher than IDACI 1 but nsd)
5.2.7 Engagement in the SPP has helped my school to identify its strengths and areas for development more rigorously	IDACI 5 agree (n=13, 93%) vs IDACI 1 (n=29, 71%)
5.2.8 Engagement in the SPP has helped teachers in our school feel a greater sense of ownership and engagement with our school's improvement aims	IDACI 5 higher agree (n=11, 79%) than IDACI 1 (n=21, 51%)
5.2.10 Engagement in the SPP has enabled my school to become more evidence-informed	IDACI 5 (n=13, 93%) higher than IDACI 1 (n=28, 68%)
5.2.13 My school draws on expertise and support from other schools in the partnership on a regular (i.e. monthly) basis	IDACI 5 (n=12, 86%) higher than IDACI 2 (n=25, 61%), or IDACI 1 (n=24, 59%)
5.3.3 The benefits that accrue from undertaking SPP reviews are not sufficient to justify the time and effort involved	IDACI 5 higher disagree (n=13, 93%) than IDACI 1 (n=27, 66%)

**Note: Classical Student's t-test ($p < 0.05$, two-tailed)*

Other factors affecting successful implementation of the SPP

Looking across a range of data sets we can see patterns that affected the ability of schools to successfully engage, implement, and sustain SPP activities. These are divided below under the categories of: Partnership conditions; leadership conditions; school conditions; and capacity: having the right people in key roles and developing them.

Partnership conditions

We found in both our case study clusters, that headteachers were motivated to find partnership arrangements that supported school improvement efforts. This was in part due to the policy direction where heads of schools still not in MATs were feeling pressure to join them and to convert to academy status. Not all headteachers were committed to this idea, valuing their autonomy, and wanted to find partnership arrangements that were beneficial but less permanent and formally binding. Furthermore, the lack of local authority support in the local area was providing a further incentive to take part in the SPP, and the funded trial made involvement even more opportune:

> *I mean we came together as a [SPP] cluster originally over disenchantment with [local authority] about NQTs because their programme was costing hundreds. We didn't have the money. I mean financially, we're not too bad off, I'll be honest, but a lot of our group couldn't afford it so we thought, well why are we paying when we can do it between us?* (Headteacher, school 1d, year one)

As time went on, the headteacher in school 1d felt the reasons for continuing in their local partnership arrangement. He felt that wider structures for supporting schools and school improvement across the town had continued to develop in line with national policy (e.g. Teaching School Hubs replacing TSAs) and that, while the school had engaged with some of this provision it was 'too segmented' and said: *'it hinders our children'* (Headteacher, school 1d, year one).

In the case of Cluster 2 in the South, the local authority itself was very supportive of the programme and apparently was financially supporting more schools, including two-thirds of secondary schools who had joined up to the SPP, which helped motivate ongoing involvement during and beyond the trial period. Given the existing partnership in Cluster 2 virtually mirrored the SPP cluster, they also had the potential to use their partnership coordinator to help facilitate future school reviews and improvement workshops.

In some cases, the partnership context hindered participation as there were conflicts of interest. We saw from case Cluster 1 that one school left the SPP after moving into a local MAT and in partnership lead interviews we heard an account from a headteacher who said that he wanted to do more in the SPP but felt obliged to dedicate more time to the schools in the MAT, most of whom were not part of their SPP cluster.

Many schools had reconfigured their clusters in year two and were planning to reform new SPP clusters beyond the trial period, which somewhat muddies the picture regarding drop out of the programme in this trial. We encountered two cases of this in our partnership leads' group interviews in year three who were reforming to continue after the trial with the SPP and another example where the headteacher was promoting the model more widely within their MAT:

> *Yes. In our multi-academy trust there are other schools who are also part of the programme. One of our senior improvement advisers on the team, he was involved in the programme as well. So we're trying to push it out as a model. Whether we'll have any success with that I don't know. We're a large multi-academy trust if you know what I mean but because the model worked for us or has been working for us, we just felt we're trying to push it out maybe wider.* (Headteacher, Group A, p4)

We know of at least one other case of a MAT where the intention had been to take part in the SPP for 2 years and then to use this as a period of learning after, which they would appropriate the partnership working practices within their trust but without subscribing to the EDT programme. In our interviews with EDT stakeholders in year one, they had stated that this was in some ways encouraged as part of their principles of handing over the idea of partnership-led school improvement into the system albeit this proved challenging to their own business model. In a stakeholder interview in 2018, we also saw how trusted previous participants of the SPP were able to take up a 'sustainability' model of the SPP to continue to receive some top-up training and materials but without full membership to keep down costs. A CEO of one MAT remarked how this enabled them to embed the SPP practices at multiple useful levels in their annual cycle of evaluation and improvement planning.

The size of clusters was an interesting variable that affected engagement too. In the case of one headteacher at the partnership leads interview in year three, the small size of their cluster (3) was indicated to be one factor in the lack of motivation to collaborate and when one further school dropped out it led to inertia. In another case, a headteacher said they had split into two groups of six, having started with 12 in their cluster but even then, it was somewhat unwieldy in terms of completing the review cycle so had decided to split into three. In another group, a partnership

lead spoke of pairing up schools within their cluster of six, so they could feedback to each other and then reconfiguring pairs for years two and three.

In terms of the 'cultural' conditions of effective partnerships, we saw, from survey responses, how trust had developed further through SPP involvement. However, there was a recognition from our case study participants that this trust should also be combined with a level of 'challenge' to ensure relationships did not get too 'cosy'. The head of school (HoS) of school 2b (year one) said that she found their own review *'very comfortable—a bit too comfortable'*—perhaps since they knew the reviewers very well. She added that reviewing another school *'wasn't as comfortable—I was quite nervous about it'*. Generally, though she thought the process got more rigorous as it went on, once trust had become further developed and people had overcome the nerves about the process.

Leadership conditions

According to some of our partnership leads, commitment by school leaders was seen as key to the successful engagement in the programme:

> *… commitment as well and really from the get go. I've got a memory of really the training helping us with this at the outset, really being quite action focused and saying, "Okay, we're going to walk away from this meeting with a series of dates that we're all going to commit to", and we've stuck with that template quite rigidly actually. Even last week we met to plan into next year and the format of having a pre-meeting, doing the review, feedback, improvement workshop, check-in, that rhythm has been really supportive in helping maintain that commitment.* (Partnership leads interviews, Group F, p5)

Another participant in the partnership leads' interviews commented that leadership commitment ideally extended to governors and that this would also ensure the longevity of the school's commitment to the programme; this was especially important if there were changes to the school's SLT. Part of this commitment, according to another partnership lead was a 'willingness to help other schools' too. However, expectations were also important, as was committing to the process wholeheartedly:

> *It's felt to me like some schools have come into this perhaps with different expectations to what the programme actually delivers. Perhaps they were expecting more of a giving you knowledge and giving you training as opposed to this is very much more I think a coaching approach isn't it and maybe there it just didn't meet their needs in that respect. I think I observed a few schools that tried to tweak the framework … but then [to] add their own bits to it or take bits away. I don't think the outcomes were as good quality as they could have been because of that and they were disappointed and then they've dropped out. So certainly I've always recommended to schools that want to get involved really go for it in a purist way at least initially so you really understand the structures and can get the full benefit from it.* (Partnership leads interviews, Group F, p3).

Changes to leadership were challenges to some schools' investment in the programme. We saw how in case Cluster 2 one school dropped out when a new headteacher came into post and wanted to focus on other priorities. We also saw in year one of case Cluster 2 schools (school 2a) was unable to fully participate in reviews due to long-term illness of the headteacher. So, although the HT of school 2b became the EHT for both schools, she stated that morale had suffered in school 2a, which made it more difficult to implement changes. This also appeared to lead the EHT to 'limit' the scope of this school's review to an area that was more confined and less controversial (spelling and grammar).

School conditions

School engagement in the SPP was partly determined by how receptive the school staff were to learning from each other and receiving feedback in improvement workshops. We saw in a couple of cases in our case study schools that improvement workshops were either not well attended or staff were a little 'hostile' to the approach. However, later interviews with school leaders in the same schools showed that they were still conscious of the outcomes of the review and working on implementing changes as a result, so perhaps the effect of receptivity of staff on a particular day should not be overstated. Overall, we saw several comments about 'ownership of school improvement' by partnership leads—that getting 'buy-in' from teachers and leaders at the school was seen by most as a recipe for success and a particular feature that was engendered by participation in the SPP.

The Ofsted category of the school influenced the nature of involvement in the SPP. However, this was not a straightforward effect. In our partnership leads interviews, one headteacher mentioned the effect of a declining grade on one school in their cluster:

> *We've had one school that had a tricky time with Ofsted and at the time were then plunged into a category of, requires improvement and had so many people coming in and telling them what to do and I've been there and it's not pleasant. You just don't want another set of people coming in and telling you how to do it differently. So, that head did take part in all of the other reviews but didn't want to review at his own school for that year.* (Group E, p1)

The above headteacher did say that the school continued in the programme however, and had subsequently fully engaged in activities. In one of our case study schools, a recent 'requires improvement' grade had added new pressure on staff, and this led to the headteacher of school 1a, who conducted their review to be sensitive to this and avoid giving them 'another bashing' and focus on 'TLC (tender loving care)'.

In two of our Cluster 2 schools, having *'got Ofsted out of the way'* meant that they were able to now focus on the peer review and *'focus on what they needed to do as a school'* (Headteacher, school 2e). In the case of one of the reviews we observed, we also saw evidence that the review was used to prepare staff for an imminent Ofsted inspection.

Capacity: Having the right people in key roles and developing them

Many factors that determined the extent of successful and full engagement in the SPP were to do with staff capacity. This was a mixture of having enough people to fulfil roles, while being able to maintain the core business of the school (quantity) and also the type of people with the skills able to fulfil SPP leadership roles (quality).

One of the issues we saw in our case studies (both clusters), was that a few individuals had very high credibility in their schools and in partnerships. These people were, well respected and had a lot of experience. In the case of Cluster 2, it was clear that the EHT took on a coaching role with one inexperienced headteacher who was reviewed and who was new to the school and in another case, the EHT was influential in steering a review to be more focused and therefore 'actionable'. In a further case, the same EHT was minded to 'protect' a school IHoS where morale was low by choosing an area that the school could be fairly confident that they were succeeding in and therefore they were able to celebrate and validate this in the review. Experienced staff were also more credible when they reviewed partner schools and headteachers were more receptive to feedback from them. One of our partnership leads interviewed also talked of the advantages of matching expertise of the reviewer to the needs of the reviewed school.

Finding staff with the right qualities to undertake the IC role was considered particularly challenging. In the March 2018 impact session, the SPP facilitator described the type of people who make really good ICs to be ones who would make *'good SLEs [specialist leaders of education], ASTs [advanced skills teachers]—get what great teaching and learning is like, understand what it takes to change schools, becoming evidence literate'*. Key to the IC role in practice was having someone with the personal qualities to lead a facilitative, coaching session in a partner school. We saw in our school 2e that the headteacher mentioned how in year one, the teacher had struggled in this role due to his introverted nature, but that the year two person had really thrived. School 1a, HoS and EHT (year one interview), illustrates the qualities of a successful IC in another of their schools in the cluster:

> *They selected a "brilliant" IC in their school, already doing school-to-school support for a local educational partnership. She had been coaching, mentoring and supporting staff in schools with issues with standards. She was used to working with staff meetings and groups of people. She already had the skills. And they had 1–2 others who could have been ICs.* (EHT, school 1a, year one)

While achieving the above qualities would seem difficult across all schools, there was also plenty of evidence from evaluations and within our interviews that people were able to 'grow' into this role. The training was crucial and additional top-up training was offered in year two when several partnerships, including our case Cluster 2, were found to be lacking in IC numbers to cope with the improvement workshops that were needed following reviews. However, in school 2d, the headteacher was not confident, even during year two that any of his staff were quite ready to take on this role. In the partnership leads interviews, one headteacher commented that it was very hard in one-form entry schools to find time to release staff to take up the training and undertake the tasks required. Several partnership

leads we interviewed mentioned the importance of the year two top-up training for both reviewers and ICs and in at least some cases, some staff took on both roles.

Withdrawal from the trial

Compared to the start of the trial (422 treatment [SPP] schools), by the end of the evaluation 239 remained; of these, 57 withdrew before the COVID lockdown in March 2020 and 182 withdrew after this date (EDT records). EDT records showed the following reasons for withdrawal, see Table 34 below:

Table 34: Reasons for withdrawing from programme from EDT records (one reason given) (n=137)

Reasons given for withdrawal given to EDT	N	%
COVID	90	66
Lack of capacity	29	21
We decided it was not working well enough to dedicate so much time to it	10	7
New school leadership and/or different priorities	5	4
We joined a MAT	2	1
Complex changes to school or partnership	1	1

Using initial EDT records, we further explored reasons for withdrawal and our survey respondents' top reasons were given as: changes in priority due to the pandemic; or due to school leadership changes; or lack of capacity. A substantial minority either lacked confidence in elements of the programme or doubted its impact (31%). See Table 35 below:

Table 35: Reasons for withdrawing from the programme in the final survey (up to three reasons) (n=23)

Reason for withdrawal	N	%
We wanted to focus on other areas due to the pandemic	13	57
New school leadership	8	35
Changes in the school meant we no longer had capacity to commit to this programme	5	22
Elements of the programme did not work (e.g. numbers of reviewers, Improvement Champions, partnership lead, organisation etc.)	5	22
Change in school's priorities	5	22
Other	4	17
Changes in our SPP partnership	3	13
We doubted the impact of the programme and did not feel it was worth investing so much time in it	2	9
Changes in other alliances, which conflicted with the SPP	2	9
COVID restrictions	2	9
Gotten everything they could from the programme	2	9
Changing in circumstances due to Ofsted inspection	1	4
Joining a multi-academy trust	1	4

It is hard to interpret these differences—perhaps schools used COVID as an excuse to withdraw rather than report criticisms of the programme to the EDT team, however, since few withdrawn schools responded to our survey we cannot make firm conclusions about this. It may also be that the survey simply gave greater opportunity to reflect on the relative impact of various factors (up to three reasons), so these should be seen in combination.

Research question: 5. What difference has COVID-19 made to the operation, participant engagement, and perceived forms of impact of the SPP?

In this section, we explore four main themes in relation to the operation, engagement, and impact of the SPP that emerge from across the data sets: disruption of activity; mutual support; change of school improvement/reviewing focus; and changes to reviewing practices.

Disruption of activity

We know from earlier sections in this report that COVID was the principal reason given for withdrawal from the programme and was seen as the biggest challenge to the conduct of SPP work. By contrast, COVID was given as the second most frequently cited challenge for matched schools (for whom time to participate in partnership work was seen as more important). SPP respondents were also more likely to have paused activities and not resumed them due to the pandemic and matched schools were more likely to have continued working online than SPP schools. This may reflect the nature of the SPP, which requires school visits to conduct reviews, and although online reviews were an option (see below), uptake was low on the programme.

From other data, such as stakeholder interviews with partnership leads and case study interviews, we know that some schools felt unable to continue their SPP work, despite wanting to, due to the stress of the previous months due to COVID.

And at a November 2020 'progress check' workshop one headteacher gave an example of what they had been dealing with:

> *We were catching up with what is going on in partnership. We lost 2 schools (from 8 to 6). One interim head is doing some work in [states the region] and another coming from the X Academy and don't want to be part of the group anymore. Another school Head is on long term sick. Couple of school staff are stepping out as all SLT are isolating. Our school has people self-isolating—8 cases.* (Headteacher, progress check workshop, November 2020)

The effect of staff morale/churn was felt in our case studies and this had an adverse effect on the ability of schools to conduct their reviews. In a year three interview with a headteacher of school 2e, the headteacher mentioned that well-being of both staff and pupils had suffered tremendously, and that some staff had 'reviewed their lives' and decided to move out of their post.

At the partnership leads' interviews, the following quote expresses the whole period well and was met with lots of assenting nods:

> *It just feels like we've been continually firefighting and battling for fifteen months. We're all really weary now. We're ready for the end of term.* (Group A, p1)

Mutual support

Three statements appeared on the final survey relating to the role of the programme in providing mutual support during the pandemic. See Tables 36–38, below.

Table 36: The Schools Partnership Programme has been a key source of support for me during the pandemic

Scale	N	%*
Strongly disagree	30	19
Tend to disagree	27	17
Neither disagree or agree	36	23
Tend to agree	26	17
Strongly agree	37	24
Don't know	–	–
N/A	1	1

*****Note: Percentages do not sum to 100% due to rounding*

Table 37: The Schools Partnership Programme has been a key source of support for members of my team during the pandemic

Scale	N	%
Strongly disagree	36	23
Tend to disagree	42	27
Neither disagree or agree	52	33
Tend to agree	22	14
Strongly agree	3	2
Don't know	–	–
N/A	2	1

Table 38: The Schools Partnership Programme helped our school get through the circumstances arising from the COVID-19 pandemic

Scale	N	%*
Strongly disagree	33	21
Tend to disagree	39	25
Neither disagree or agree	48	31
Tend to agree	23	15
Strongly agree	10	6
Don't know	–	–
N/A	4	3

***Note: Percentages do not sum to 100% due to rounding*

The tables above show that respondents of the survey (mostly headteachers) were more positive about the role of the SPP in supporting them than for the rest of their staff. This suggests a more direct influence of the programme on school leaders than other senior leaders, teachers, or other support staff. As might be expected, on all three above statements, agreement was statistically significantly higher for completed than withdrawn schools. For the first two statements (sources of support) we found matched schools respondents agreed more to the equivalent statement than SPP school respondents with 75% strongly agreeing/agreeing that their most significant partnership was a key source of support to them and 48% strongly agreeing/agreeing that it supported their team. Given the multiple layers of partnership work involved with both matched and SPP schools, it may be that SPP schools found similar support in other partnerships they belonged to in addition to SPP or that SPP overlapped with other partnerships. We saw an example from our case study Cluster 2, that the partnership lead spoke of how the SPP work had brought them closer together in their local alliance (that overlapped with the SPP cluster) and that this local alliance had been key in getting through the pandemic, by agreeing approaches to fast changing rules, and sharing resources.

From other sources, such as partnerhip leads' interviews and interviews at case study schools, we got a sense of a rich picture of ways in which SPP clusters were offering support to each other. These included being able to share concerns in EDT training workshops, account of how adversity was building further trust and collaboration; that the combination of the SPP and pandemic had helped school/staff to be even more open with each other, and generally being able to offer moral support in cluster work, providing reassurance about 'being in the same boat' and helping each other to get through COVID/lockdowns, this quote gives a flavour:

> *I suppose we've developed a stronger relationship, we met regularly, just as much for headteacher well-being as anything. It is the group that does the peer review that was meeting rather than other schools in the town. We just met regularly on Teams and just offloaded and chatted and said: "What are you doing about this? What are you doing about that?" It was somebody else to bounce ideas off and things.* (Group A, p1)

Overall, given the central methodology of SPP revolves around reviewing each other's schools, it is perhaps not surprising that the SPP itself did not stand out among other partnerships in terms of offering help and support during the pandemic but that it may have complemented the picture.

Change of school improvement/reviewing focus

Around 41% of respondents in the final treatment survey said that the need for a COVID-19 recovery plan was one of the main reasons that they had changed their school improvement priorities over the last 3 years (this was the top answer also on the matched survey); 15% of treatment survey schools also said they had shifted to a 'well-being' focus. In a November 2020 training observation, participants used an online 'Jamboard'[7] activity to post their new areas of focus, these included: using blended/virtual reviews; focusing reviews on well-being; workload; and recovery curriculum. However, a few were about reading/phonics and some wanted to simply continue with prior foci. We know from the end survey (Table 23 above) that most schools wanted to pursue a 'non-COVID-related' long-term focus—only 6% responses were specific to this, namely, 'remote teaching and post-COVID recovery'. We saw an example from our partnership leads group interviews:

> *We actually just wanted to get back to looking at our schools and knowing our schools and looking at what we wanted to improve. So we all decided that we were going to share a theme for the year of Maths Mastery and challenge for all pupils. So we were quite adamant that we weren't going to let COVID derail what we wanted for our schools.* (Group B, p1)

Changes to reviewing practices

In October 2020, EDT ran a webinar on conducting rapid reviews. These reviews were designed to be carried out wholly online, partly online, or wholly in-person but with a truncated methodology that allowed for the process to be quicker and more streamlined. An online review had been carried out by EDT with a pilot school and the headteacher fed back the lessons from this, citing some advantages to do this. We do not have accurate figures about how many schools carried out online reviews after this but from conversations with EDT staff, we suspect they were very few in number and this was confirmed in our survey that schools were more likely to have simply paused activity and waited to go back to full in-person reviews. We did observe one virtual review. This followed the EDT suggested process; the reviewers completed the pre-review papers and questions; interviewed a range of middle leaders and pupils; and provided feedback to the host school team. This feedback session was attended by an IC from the host school. During a brief discussion at the end of the session, the host school leaders and review team members all stated that this 'rapid' approach had been successful, although they would—ideally—have liked to be in school to observe teaching. This view that the virtual reviews had worked but 'not as good as in-person reviews' was shared by one of our case study headteachers too (Cluster 1). Comments from ICs interviews where they had taken part in virtual reviews showed that they felt that while they had worked well overall, the focus had to be chosen carefully, since collecting evidence from lesson observations was difficult online; the process worked better for larger schools, where breakout rooms and 'Jamboard' could be used collaboratively to good effect more than where there were few staff; but that there were concerns about the shortened process having less impact overall on staff compared to the longer in-person improvement workshops where staff were engaged for two or more hours.

One of the partnership leads talked about a hybrid form of reviewing in which they visited another school but sat on different tables to socially distance and instead of live observations of classes they used cameras:

> *We met with a real-life maths leader and talked about her experiences and journey in maths which was really great. We looked through actual books, having sanitised before and after obviously. So it felt a lot nicer than it would have if we were on screen. It wasn't the same as going into school classes and talking to real children, etc., but it was more normal than we thought we were going to have. It was a really positive day and it was really refreshing to get back to having real discussions with people.* (Group B, p1)

In one cluster from our partnership leads interviews, a headteacher said that as they were in an area (in the North of England) with growing transmission of a new variant (Delta), they had decided to proceed with virtual reviews instead. They chose a shared theme of subject leadership, conducting three reviews and used remote interviewing of pupils, teachers, and subject leaders, wanting to check what kids had missed out during COVID (Group B, p5).

We also observed a facilitator of a summer progress check-in, (May 2020 webinar) who mentioned that a 'number' of schools had tried virtual/rapid reviews and commented that they were in some ways better, as they enabled cross

[7] Jamboard is an online tool for sharing virtual 'post-it' reflections.

regional reviews; she gave the examples of six clusters in Wales where it had not been easy to get around. We also heard from partnership leads that if parts of reviews could be online, for example pre-review meetings, this could be advantageous beyond COVID too.

The overall impression about virtual/rapid reviews was that they were a welcome additional tool with some advantages, but that the majority were looking forward to getting back to in-person reviewing as soon as their situations allowed.

Other IPE factors

In this section, we look at what the data shows us in relation to: fidelity (if the intervention was delivered as intended); adaptation (how the intervention was changed to meet circumstances); dosage (how much of the intervention was delivered); reach (if the intervention was delivered to the intended recipients); responsiveness (recipient engagement during the delivery of the intervention); quality (how well the intervention was delivered); and differentiation (how different the intervention was from existing practices). NB, the COVID-specific issues have already been addressed above, under research question 5 findings.

Fidelity and adaptation

Stages of the review process

In terms of the review process, the first four reviews we observed (pre-pandemic) were all conducted in line with the SPP handbook—namely, pre-review, review, and improvement workshop. Within this three-part structure, the reviews all adopted the EDT paperwork (e.g. the template for pre-reviews) and proposed approach (e.g. enquiry questions) and adhered closely to the proposed design (e.g. in terms of agenda for review days, and involvement of pupils and staff etc.).

We either observed, or heard about, several variations to the ICs role or improvement workshop process. In some cases, these variations appeared to be driven by specific contextual issues (e.g. lack of trained ICs or logistical/travel issues). For example, the improvement workshop in one school was largely led by the IC who was the assistant head at the school (i.e. not external, although a second, external IC did also attend). In the review of reviews workshop, June 2019 (West Berkshire) one IC said they had shadowed the peer review process as they thought it would help (IC only supposed to attend the debriefing at end of review). We also heard from our case study headteacher at school 2e (year two) that it had been hard to follow up their year two review with an improvement workshop in the usual 2 weeks after the review as they had decided it was too close to the summer break but then was finally delayed until November 2020.

We do not have reliable data across many schools to know how much the improvement workshop or role was faithfully applied, but there were numerous remarks about their value from interviews and training sessions. After the second IC training session, the headteacher of school 2e mentioned that there was an increasing use of research evidence and a clearer understanding of the 'coaching/questioning' role of ICs, namely, not going in and telling the school what to do. We heard this from partnership leads and improvement champion interviews too, so this all paints a picture of how the training was effective at shaping the involvement of ICs in the partnership improvement process following reviews, as intended in the model.

We know from our case study schools and also from observing the year two progress and sustainability workshop, that the 90-day follow up to reviews was not always adhered to formally in year one. In the collaborative leadership sessions in year two training, this was re-emphasised and both our case study schools and some headteachers in the partnership lead interviews confirmed that schools saw the importance of following these up more diligently and embedding follow-up discussions in partnership meetings.

In terms of the original focus of reviews, we saw already in pre-pandemic training, such as the 'progress and sustainability workshop', March 2020 (South), that some schools had a 'non-literacy- or numeracy-related' goal for reviews. One school observed in a group discussion mentioned that their focus was on feedback and questioning, another on 'oracy'. Other schools were involving whole school or Key Stage 1 instead of Key Stage 2 as well. While schools were initially steered towards Key Stage 2 literacy or numeracy as part of the trial, part of the methodology was in conflict with this in a sense, namely that the focus allowed for self-driven improvement objectives.

In observations of EDT workshops and training, we saw consistent approaches by facilitators in equivalent sessions across regions, with only slight adaptations, for example to ensure timings worked with larger groups, but otherwise they were very faithful to agreed methods, using the same slides across the programme (these were available to all participants in online resources section).

Dosage and reach

In general terms 'dosage' was heavily affected by the COVID pause in delivery and the extension to the trial beyond its originally intended 2 years (see research questions 4 and 5, and the compliance figure). To quantify this (using EDT data):

- Around 210 (50%) of SPP schools attended 75% or more of the compulsory EDT training over the entire trial period (i.e. the figure needed for full compliance).

In terms of reviews:

- Around 121 schools had hosted one review before the COVID lockdown in March 2020, 104 had been reviewed twice, and 16 had hosted three times.

Post-COVID reviews:

- Around 139 schools hosted no more reviews, 99 received one review and three had hosted two reviews. Of the 121 schools that had only had one review before the COVID lockdown, 64 received a further review (53%).
- Therefore, in terms of reviewing, 171 schools (41%) met the criteria for full compliance.

From our compliance estimates and EDT data, we do not know whether the school sent all staff to these core sessions or just one representative (which was enough to meet the criterion for school attendance at a session). For instance, we know that school 1d (year one interview), headteacher attended one of the EDT training sessions, but not the deputy headteacher, due to capacity issues (maternity leave).

We have already seen, in terms of reach, that the engagement in the programme was seen more readily by those that took active parts in the reviewing processes, namely, headteachers, reviewers, and ICs. Responses to the survey's statements to do with engagement and ownership of improvement aims also to allude to a more direct involvement of leaders than teachers. Capacity issues also limited the extent of staff involvement in some schools (especially one-form entry) as we have seen (research question 3, above). For instance, school 2d only had the headteacher and deputy directly involved in the SPP as reviewers, but they did not have an IC.

There were 66 schools that were not in the treated schools list but participated in treatment school clusters (EDT records). Some started from year one and many others joined later (joining date records are incomplete). These comprised 27 infant schools, 21 more primary schools, 12 secondary schools, 5 special schools, and 1 sixth form college. In a sense, the intervention had 'reach' beyond the trial. We also heard from case Cluster 2 that the school leaders wanted to add another local school to their SPP partnership but the headteacher of that school did not want to take part and this was lamented by several of the other schools in this cluster. We do not know how common the latter picture was, but we heard in partnership leads interviews that local partnerships were seen as particularly valuable, and, unlike the SPP ones, this was not always the case in their other partnerships, which could be more geographically dispersed.

Other stakeholders

There was nothing explicit in the SPP logic model about direct engagement of pupils in the process, but we do know from our observations that pupils were often observed in classrooms and interviewed as part of reviews. We also heard from one of our case studies that governors had been encouraged to take part in a peer review. In the year two review of school 2e, we know that TAs did not take part in the improvement workshop at their own school, despite this review being about their deployment/use. They were however, interviewed on the day of the review, to give their perspective.

These examples suggest there is potential for involving a wider group of stakeholders in reviews and the improvement process than was consistently seen.

Responsiveness

Generally, observations of workshops and training showed good levels of engagement by school participants, who actively reflected on their experiences, challenges, and successes while showing eagerness to pick up the skills they were learning about reviewing and leading improvement workshops. In a minority of the EDT progress sessions, school partnerships showed a lack of depth of analysis in their discussion, so that they were unable to provide convincing and specific evidence of the impact of the programme in their schools and on pupils.

We do not have systematic data on engagement shown by school staff during improvement workshops across the programme, but respondents to our final survey suggest this was probably very good since staff ownership of the school and partnerships' improvement aims were rated highly. However, in school 2d in year one, the deputy headteacher had apparently booked another CPD course on the day of their improvement workshop, despite this being a key part of the follow-up process from her own school's review and for which she would be responsible for overseeing. Another key member of staff was absent from this review and the IC that led this workshop felt the staff were not receptive, perhaps due to broader staff morale issues after a recent SIP visit. Nevertheless, our ICs' interviews suggest that these cases of more hostile or disengaged staff were more exceptions than the norm.

Quality

EDT training and workshop sessions

EDT own evaluations of sessions by participants showed consistently high ratings. We saw nine sets of workshop evaluations, and the outcomes from the first peer reviewer and ICs training were typical of later sessions, see Table 39 below:

Table 39: EDT evaluation of first peer reviewer and Improvement Champions training: Overall, how would you rate this training? (n=738 participants from treatment schools)

5 (Excellent)	4	3	2	1 (Poor)
363	340	32	3	0

Without exception, our observations of EDT sessions showed that facilitators were knowledgeable, had high levels of credibility due to their extensive leadership experience, and were skillful in leading highly participative and engaging sessions. In our final survey, we had two statements relating to the quality of SPP training and materials that showed high levels of respondent satisfaction, see Tables 40 and 41 below:

Table 40: Final treatment survey: The training provided by the programme has been very high quality (n=157)

Scale	N	%*
Strongly disagree	2	1
Tend to disagree	6	4
Neither disagree or agree	11	7
Tend to agree	60	38
Strongly agree	76	48
Don't know	1	1
N/A	1	1

**Note: Percentages do not sum to 100% due to rounding*

Table 41: Final treatment survey: The materials provided by the programme to support the SPP have been very high quality (n=157)

Scale	N	%
Strongly disagree	–	–
Tend to disagree	2	1
Neither disagree or agree	10	6
Tend to agree	55	35
Strongly agree	88	56

Don't know	1	1
N/A	1	1

Even among withdrawn schools, agreement was very high, and there was no statistically significant difference between active and withdrawn groups on the two statements above, suggesting that quality of EDT training or materials were not important factors in schools' decisions to withdraw from the programme.

Partnership leads, reviewers, and ICs

EDT training prepared people well for the key roles they needed to undertake during the programme, although one headteacher we interviewed in our case study schools (2e) remarked that it was important to conduct a review before really understanding the process. We do not have reliable data about the consistency in quality of partnership leads, reviewers, or improvement champions, however the Executive Head teacher and Head of School we interviewed (together) after year one reviews at school *2b* both told us that there was a variation in quality of ICs across the partnership and that selecting the right people was crucial:

> EHT: "It's very important that those people [ICs] are very strong in their role."
>
> HoS: "And maybe they weren't always as strong as each other would be the polite way of putting it. If they had known this, they would have pushed harder to have some of their other teachers in the role because it makes a difference who does the school improvement process."
>
> EHT: "Definitely. It's very important"
>
> HoS: "The head of the school with a weak IC thought it would be good for his development. But that's not the time to develop him."

The same two school leaders above also mentioned that with reviewers there had been 'possible power issues around more experienced heads', where they even felt they had little to learn from others. So, it was not clear if the reviewer was not doing a good job or whether the receptiveness of reviewed schools was a function of preconceptions about the credibility of the reviewer.

Differentiation

Many participants compared the experience of the SPP to other external reviews, and in particular, the Ofsted inspection process. Such comparisons were invariably extremely favourable, with several remarks from interviewees across data sources about the advantages of the SPPs 'less judgemental' review process compared to Ofsted; reviews were seen to be more motivational, and more formative than summative:

> *I'm glad it's not given a sort of, grading because I don't think that that's helpful. I think that it's more around giving positive messages … if we give positive messages to staff, wherever possible, it always motivates them to do better. Whenever there's something negative, people can become very defensive and yes, you do have to have sometimes difficult conversations, but I think it's around the way you do it and I think that's about knowing staff well enough to do that.* (Headteacher, school 1a, year one)

The SPP process was seen by many in our interviews, as instilling a form of empowering professional accountability and collaborative school improvement:

> *I think it's replacing Ofsted. It's showing that we can do it for ourselves, and do it in a critical way, and a way that gives accountability, rather than previous models which are just possibly either too extreme, Ofsted-based, or not enough rigour. It's that middle ground of, you know, us being in control of our school improvement, but doing it with other professionals in a way that gives us critical accountability.* (HoS, school 2b, year one)
>
> *On top of all of that, the beauty of that is that you are then in position to support people after they've had their peer review. So, when you're working in close partnership, they will then come to you and say,*

> *"Well, can you support us with that? Could our subject leaders come to work with your subject leaders?"* (EHT, school 2b, year one)

The headteacher of school 2e (year one), who had had experience with local school improvement partners, said that by contrast, the SPP involved less paperwork, involved more time with reviewers seeing the school and talking with teachers, and that staff were very positive about the peer review and felt more directly engaged in the process.

Nevertheless, we did hear from our case study schools that reviews could still place pressure on teachers, and that it took the personal qualities of the reviewer to try to alleviate this and for staff to gradually realise it was not the same as an inspection:

> *I didn't expect it to be as formal as it was … you felt a bit like you were Ofsted, and I was a bit like, "We are normal. We are humans." And you could tell when you walked into classes that everyone was looking at you, to think, "Oh no, they're here." And the dread on teachers' faces, and I didn't want it to be like that at all because you want to get a truer picture, don't you, and a true reflection. But then, I suppose, it depends how you observe people's practice, isn't it, if you just stand there with a clipboard and look at them, sit down and give them a big stare, they probably do think it's quite daunting, but once they realise that you are a teacher yourself and you're then talking to children and finding out what they're learning about in the classes, it's a little less intimidating, I feel, from the teacher aspect.* (Deputy headteacher, school 1d, year one)

And teachers could feel the need to impress visiting reviewers, despite reassurances from their own headteachers:

> *I think when the first review happened, everyone was spending loads of time tidying up the sides and making sure the school was really tidy, which is lovely, but we were saying, just do what you normally do, we want them to see us as we are, it's not all about whether the sides are tidy, it's just be the teacher you are because we get good results.* (Acting head, school 1e, year three).

Conclusions

Table 42: Key conclusions

Key conclusions
Due to Covid-19 the primary outcome for this evaluation was not collected and so no measure of impact on attainment is reported. Key conclusions are based on qualitative data from the implementation and process evaluation (IPE). Teacher surveys found self-reported positive impacts on ownership of partnership and school improvement aims and the deepening of trust. While based on small numbers (and therefore to be taken with caution), many of these benefits appeared to be more deeply felt in schools with higher levels of pupil deprivation.
Evidence on perceived pupil-level outcomes is inconclusive—schools had a wide variety of improvement aims, which were directly related to the outcomes of the peer review, many of which changed during the evaluation especially during the second year of peer reviews where a different line of enquiry was agreed. Nevertheless, most participants felt that there had been a positive overall impact on pupils in their schools.
The IPE found that teachers perceived a strong impact of the Schools Partnership Programme (SPP) on leadership development. This came through reports on self-perceived acquisition of important evaluation and reviewing skills, and the social learning elements of peer review. The Improvement Champion role was felt to develop teachers and middle and senior leaders in collaborative school improvement. Where this role was carried out well, participants in host schools reported the coaching approach helped engender full staff ownership of improvement aims. However, capacity issues made the execution of these roles and continuity very difficult, especially in relation to leadership and other staff churn.
The majority of SPP participants regarded the programme as significant to the school, its staff, and students, and felt that it was aided by excellent methodology, support materials, and training. Most survey participants rated the SPP resources, such as the review framework and training materials, as being of very high quality. Participants felt that feedback received in reviews was mostly of high quality and had sharpened their ability to identify, act upon, and follow up on improvements, including being more evidence-informed. A large majority of school leaders surveyed said the time spent on the programme had been worthwhile and felt there had been a positive overall impact on pupils' outcomes at their school.
The complexity of the programme, including shifts in focus of school improvement objectives and the changing nature of the partnerships throughout the trial, caused limitations on capturing evidence on teacher quality and impact on pupils. Pupil impact data was ultimately missing due to the absence of Key Stage 2 Standard Assessent Tests assessments due to the COVID disruption

Evidence to support the logic model

In relation to the original logic model (see Figure 1), the lack of SATs data due to COVID disruption mean that we cannot comment on this. Furthermore, our change in research questions in the revised evaluation protocol, changed from impact on Key Stage 2 numeracy and literacy outcomes to 'perceived forms of impact'. Thus, the conclusions we report here are based on the findings reported above in relation to research questions 1 and 2:

1. In what ways does the SPP influence the capability, culture, and practice of partnerships, leadership, and teachers in involved schools?
2. In what ways do the elements in the SPP theory of change work in achieving participants' perceived forms of impact?

The above questions deal with the part of the original logic model that is in the programme's theory of change:

Partnership intermediate outcomes:

- capability and capacity in leadership of collaborative school improvement;
- oulture of shared responsibility; and
- open and transparent systems and processes.

Leadership intermediate outcomes:

- increased confidence in leadership and collaborative improvement;
- strengthened lateral trust within and between leadership teams; and

- peer review, follow-up support, and bespoke intervention fully embedded in school improvement cycles.

Teacher intermediate outcomes:

- greater ownership of and engagement with strategies for improvement and changes in practice;
- improved lateral trust within Key Stage 2 teams; and
- Key Stage 2 teachers aware of the improvement priority and their responsibility for changing practice.

School-specific outcomes:

- Unspecified as this will be different for each school.

Participants in the SPP showed consistently strong perceptions about partnership-level factors, particularly to do with having an equal and shared status; feeling responsible for the success of all schools and pupils in the partnership, and transparent sharing of data, systems, and processes. Other partnership factors were at least as positive as the 'most significant partnerships' of the matched schools. Partnership factors somewhat overlapped with leadership factors, in particular with senior, and to an extent, middle leadership roles in leading improvements in the partnership. Evidence that peer review was firmly embedded in school improvement cycles was also strong. The evidence that skills of evaluation and peer review increased among lead and other reviewers was strong. The evidence that staff trained in the improvement champion role benefited, and that this was likely to have a positive knock-on effect in the host school and across the partnership was also very strong. While the strength of evidence is weakened by low numbers (14 IDACI schools represented in final survey), many of the perceived benefits, especially at partnership and leadership levels, appear to have been felt strongest by schools in the most deprived areas.

At teacher level, our data suggest there was an increase in ownership of and engagement with strategies for improvement and changes in practice, however this may be less of a feature of the SPP specifically since evidence among matched schools was comparable. We have limited evidence about the impact of trust or awareness of improvement priorities for changing practice of Key Stage 2 teams specifically. However, there is strong evidence that staff increased their awareness of improvement priorities across the partnership. There is also strong evidence that staff who took part in the reviewing and improvement champions process felt greater responsibility for improvements in schools across the partnership.

At school level, there is strong evidence that the programme improved SLTs' abilities to identify its own strengths and weaknesses, was able to draw on expertise and share partnership capacities that would lead to school improvement, and relatively strong evidence that research-informed practices were encouraged by the programme, particularly through the improvement workshops.

In terms of perceived pupil-level outcomes, our evidence is inconclusive; and schools had a wide variety of improvement aims, many of which changed during the evaluation. Nevertheless, most respondents in the final survey (73%) felt that there had been a positive overall impact on pupils in their schools (albeit this was less than matched schools reported benefits).

Interpretation

In this section, we discuss the interpretation of our evaluation, considering these in the context of partnership-led school improvement at national policy level. Lessons learned about the role of the SPP during the COVID pandemic will also be discussed. We consider the sustainability and scaling implications for this model of school improvement. Finally, we suggest further areas of research for the EEF and lessons for impact evaluations of similar programmes.

The national context

In the background section we mentioned that school-to-school partnerships had been prioritised due to the evidence that neither top-down centrally imposed change, nor pure competition could achieve sustained improvement across school systems (Burns and Koster, 2016). Government policy to encourage school system leaders to work together to transfer knowledge, expertise, and capacity within and between schools (DfE, 2010) has continued in more recent legislation that supports 'trust-led' school improvement, including the ability of local authorities to set up their own MATs and in which: 'We expect all actors in the system, including trusts and local authorities, to collaborate to ensure the best

outcomes for their communities' (HM Government, 2022). This approach has implications for accountability, where schools take greater ownership of their quality assurance through self-evaluation and exposing their work to the scrutiny and perceptions of trusted peers (Matthews and Ehren, 2017). There is evidence from various sources that school-to-school peer review is on the increase in several jurisdictions, but especially in England (e.g. Greany and Higham, 2018), and given the dearth of evidence to date that evaluates the effectiveness of large-scale peer review approaches, that makes this evaluation particularly important (Godfrey, 2020a).

The various nuances of improving partnership work and peer review

While partnerships and peer reviewing practices were highly valued by most schools in this programme, the picture in our evaluation is a complex one. First, 36% of SPP schools did not perceive the programme to be their most significant partnership. Second, most schools (68%) joined the SPP to add structure to *existing* partnerships and because many SPP clusters also mirrored or substantially overlapped with these partnerships. Furthermore, the matched schools we studied also commonly had multiple partnerships, and many had experiences of peer review.

A further complexity comes from the organic and fluid notions of partnership. For one thing, the operational definition of 'partnership' changed during the programme, as the partnerships started off in larger units and many subdivided into smaller clusters. In these sub-clusters, the existing partnership lead sometimes oversaw all clusters and other times new partnership leads were assigned for each sub-cluster. Clusters also worked within larger organisational structures, such as the LA, the MAT, or the Teaching School; these larger structures were at times perceived as 'the partnership' as much as/more than (or in addition to) the reviewing cluster/sub-cluster. We know that membership of clusters was fluid too, with new schools entering and some leaving during the trial and we know that existing schools sometimes moved across sub-clusters to get experience of reviewing different schools. This made tracking partnership changes very difficult to capture precisely.

The shifting sands of the notion of partnership do not merely pose methodological limitations on our evaluation, but they also influence the way in which we may need to understand partnership-level improvement or success. While not all SPP participants considered these to be their most significant partnerships, the SPP was perceived to be complementary to, and strengthening of, existing (or newly formed) partnerships (with some exceptions who saw involvement to be in conflict with other partnership work). An early stakeholder interview with a CEO of a MAT showed how the SPP approach could be embedded across the trust and help with schools' peer and self-evaluations, in a cycle of annual trust improvement planning. The potential for trust-based and other partnership incorporation embedding of SPP methodology should therefore be further explored. This quote from one of the partnership leads we interviewed expresses this well:

> *We do want to be outward looking but part of me thinks I would love almost to put my full energy into school improvement by using a model like this with my partnership schools within the trust rather than possibly spreading myself thin and having that additional group through SPP because at the end of the day, we've got a commitment to the children within our trust and the model is so fantastic it makes sense to me to maybe steer myself down that road and the other schools down this road. (*Group C, p1)

Evaluation and accountability

Peer reviewing practices were common among all the schools we surveyed; 61% of our matched survey participants had participated in a peer review in the prior 3 years and half said peer reviewing was embedded in their school improvement cycle, and 18% of SPP schools took part in other peer reviews alongside the SPP.

Many participants valued the reviewing skills they learned from the SPP and some then employed them in other collaborations for specific purposes (e.g. to review SEND). Peer reviewing was seen as a methodology, or set of practices, that could be used in other partnership work. Around 47% of SPP schools were going to continue with peer review (other than the SPP) and the main reasons were: to use in alliance work; because they felt it was effective; and that they felt it was beneficial for their school's development. Most (62%) said that this would be a model they had developed themselves working in partnership with other schools.

The SPP was also viewed as an alternative form of accountability. Many participants valued the more participatory nature of the peer review process when contrasted with Ofsted inspections, the value of getting feedback from respected professionals and their greater role in ensuring ongoing support to implement desired changes. Several participants valued a process in which they were not being 'done to', supporting the perceived value and need to increase

professional accountability in the education system, as advocated by others (e.g. Gilbert, 2017). Given the layering of peer evaluation on top of existing, external inspections (see Hadfield and Ainscow, 2020), the conditions for effective alignment between external evaluations and peer review would be a fruitful area for future research.

Leadership development

Our evaluation gives very strong evidence that the SPP was perceived to have many positive benefits for developing leaders. Reviewers learned valuable skills of reviewing and self-evaluation; and school leaders learned how to collaborate more rigorously, transparently, offering increasing challenge to each other with time, and enabling more fluid knowledge exchange and shared professional development that benefited their staff and schools. Leaders were also able to reflect on their own and other's schools in workshops and training, developing a moral sense of shared objectives in their partner schools. The role of peer review in developing leaders has been considered elsewhere and merits further attention (Godfrey, 2022a; Matthews and Headon, 2015).

The IC role was particularly highly valued. The participants, who were often teachers or middle leaders, as well as more senior leaders, learned how to lead coaching sessions, to facilitate evidence-based school improvement discussions and developed their ability to lead in their own school. However, there were substantial capacity challenges, especially for smaller schools or where the headteacher did not feel there were suitable staff to undertake these roles. There were unintended consequences of teachers being away from their own classrooms, often excellent teachers who had been selected for this work. Staff churn made continuity of improvement objectives very difficult for many schools. Further support to cover release time, costs, and ongoing training for reviewers and ICs would be strongly recommended at the middle-tier level to support this as a sustainable model.

Filling a school improvement/evaluation gap in deprived areas especially

Our evaluation found some very interesting findings regarding levels of trust in SPP schools. These included having a shared vision and collective decision-making, transparent sharing of data, increased sharing of expertise, and regular communications between schools. These perceptions increased over the course of the programme. We also found interesting associations between levels of student deprivation (IDACI) and trust that were not found in the matched schools. Prior research suggests that relational trust tends to be more difficult to foster in areas of high deprivation and that in schools where high relational trust can be achieved, academic achievement of pupils in these schools is much more likely than where trust is low or not sustained (Bryk and Schneider, 2002). There is also evidence from social network analysis, showing that levels of trust are related to increased research use in schools (Brown *et al.*, 2016).

The kinds of shared, collective partnership work engendered by the SPP appear highly conducive to building trust:

> *Trust is fostered in schools where educational policies and authorities support professionalism and collective work of teachers and when they offer opportunities and resources, in a broad sense, to enable school communities to conduct transformative processes.* (Weinstein, 2022, p.5).

The COVID context

While a significant minority in our survey found the programme supportive during COVID disruptions, compared to matched schools, this did not stand out as a particular feature of the SPP, and mostly the COVID pauses led to pauses in activity on the programme. The programme may have been of help for schools to evaluate and implement new areas of focus to do with online learning and may also have helped to show innovative ways of partnership work online. However, the evidence mostly shows that participants were eager to get back to in-person visits and seeing the schools in action.

Sustainability and scale

The SPP provides a coordinating structure and highly valued facilitation and training for schools that are involved. Many of the schools were apparently keen to continue with the SPP beyond the trial and many others were keen to continue either with a franchised version or embedding what they had learned within existing partnerships. However, involvement in the SPP or SPP-like partnerships comes at a cost to the involvement of reviewers and ICs in release time and extra work. Where local authorities or other middle-tier elements are able to support cover time and costs of a continual schedule of staff development, this looks much more likely to be sustainable and with low resource implications otherwise.

Implications for the EEF

The role of the ICs was particularly interesting in this study. The participants gained tremendously in leadership skills, including leading change in partner schools. The coaching approach to change was also seen as particularly important in developing trust and collaboration from host schools. This IC approach builds on other studies evaluated by the EEF, such as research learning communities, which employed 'evidence champion' roles (Rose *et al.,* 2017). With the EEF research schools, the use of SPPs ICs' methodology to stimulate research-informed change should be considered.

Limitations and lessons learned

The biggest limitation was COVID disruption and in particular the lack of quantitative measures of pupil impact. This notwithstanding, we think the early signs were promising for the impact evaluation design—with good balance of key characteristics at baseline and encouraging pre-trends providing encouraging suggestive evidence for the identifying assumptions of our design. The suspension of Key Stage 2 National Curriculum tests for 2 years was beyond anything for which we could have reasonably planned at the outset.

COVID disruption also strongly affected compliance, dosage, withdrawal, and changed the focus of many schools away from Key Stage 2 outcomes. The latter change was already the case before COVID disruptions, given the way in which schools were able to choose their focus. This was seen as a strength of this school-led approach by participants, but makes it difficult to track multiple school-level outcomes.

The impact of the programme on teaching quality was not measured systematically. As this is a key part of the logic model that would lead to changes at pupil level, it is not possible to say if the programme was effective in this sense. The wide range of areas of foci also mean that no single measure would have been appropriate for this purpose.

Response rates on the final surveys was relatively low. This was a very difficult time for school leaders to be asked to respond to quite a lengthy telephone survey as many felt exhausted having coped with many changes and disruptions. Future research of this kind may need an incentive for matched school survey participants too.

We took an exploratory approach to look for survey relationships, such as comparing responses according to IDACI, Ofsted, and partnership audits. These were not a priori analyses, rather were exploratory, and taking full advantage of multiple data sets. While this could be criticised as being speculative, we felt justified in doing so in order to look for new potential areas of assocation in the process evaluation that could be explored more rigorously in future research.

Comparisons with the matched school sample are very problematic. This is because two-thirds of respondents in the survey had indicated that they had taken part in peer reviews over the period of the SPP trial and around half felt that peer review was embedded. Although the SPP had other elements of self-evaluation, ICs, and it's unique training and framework, peer review is clearly central to the work of the SPP and therefore may not be very distinct from the matched sample.

Multiple belongings to partnerships and fluid membership to SPP clusters (see above) made tracking changes at partnership level very difficult to ascertain. The success of the programme as a methodology to be used in existing partnerships also means that it would be better used as part of an implementation architecture than the intervention itself.

Future research and publications

Given the combination of many promising elements of the SPP, further research may wish to explore how peer review, and an IC approach could be combined in a trial in which an existing and proven high-impact teaching programme or strategy could be implemented across a group of schools.

This programme contains elements of successful and sustained implementation of innovations, such as improving ownership of change at both school and partnership level and developing the evaluative and accountability dimensions necessary to embed practices at scale. The use of an 'audit' approach (combined with coaching) has also been mentioned as promising in a recent cross sector systematic review on encouraging evidence-use (Gorard *et al.,* 2019).

The EEFs recent guides, such as on implementing meta-cognition, provide excellent ready-made materials to this aim that could be used with partner schools and to evaluate intermediate outcomes at teacher level. The use of tailored pupil

impact measures would also likely be more likely to detect effect sizes than standardised, global measures like Key Stage 2 data allow (see Lortie-Forgues and Inglis, 2019).

Further work that looks at trust as a specific outcome variable of the SPP or other peer review-related partnership work also looks very promising, particularly looking at these partnerships as a sustainable model in particularly deprived areas.

References

Anders, J., Brown, C., Ehren, M., Greany, T., Nelson, R., Heal, J., Groot, A., Sanders, M. and Allen, R. (2017) *Evaluation of Complex Whole-School Interventions: Methodological and Practical Considerations.* Report to the Education Endowment Foundation. **https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/Evaluation/Setting_up_an_Evaluation/EEF_CWSI_RESOURCE_FINAL_25.10.17.pdf** (accessed 26 April 2022).

Austin, P.C. (2011) Optimal Caliper Widths for Propensity-Score Matching When Estimating Differences in Means and Differences in Proportions in Observational Studies. *Pharm Stat,* 10, 2, 150–61. doi:10.1002/pst.433.

Berwick, G. (2020) 'The Development of a System Model of Peer Review and School Improvement: Challenge Partners'. In: Godfrey, D. (Ed.), *School Peer Review for Educational Improvement and Accountability: Theory, Practice and Policy Implications*. Cham: Springer Nature, pp. 159–80.

Bryk, A. and Schneider, B. (2002) *Trust in Schools. A Core Resource for Improvement*. New York: Roussel Sage Foundation.

Burns, T. and Koster, F. (2016) Governing Education in a Complex World. Educational Research and Innovation. OECD Publishing.

Brown, C., Daly, A. and Liou, Y-.H. (2016) 'Improving Trust, Improving Schools: Findings From a Social Network Analysis of 43 Primary Schools in England'. *Journal of Professional Capital and Community,* 1: 1, 69–91.

Daw, J.R. and Hatfield, L.A. (2018) Matching and Regression to the Mean in Difference-in-Differences Analysis. *Health Service Research,* 53, 6, 4138–56. doi:10.1111/1475-6773.12993.

DfE. (2010). The Importance of Teaching. Cm 7980. London: Department of Education.

Gilbert, C. (2017) *Optimism of the Will: The Development of Local Area-Based Education Partnerships. A Think-Piece.* London: London Centre for Leadership in Learning, UCL Institute of Education. **https://discovery.ucl.ac.uk/id/eprint/10065147/** (accessed 26 April 2022).

Godfrey, D. (2020a) 'From Peer Review to Collaborative Peer Enquiry: Action Research for School Improvement and Leadership Development'. *London Review of Education,* 18, 3, 373–89.

Godfrey, D. (Ed.). (2020b) School Peer Review for Educational Improvement and Accountability: Theory, Practice and Policy Implications. Cham: Springer Nature.

Gorard, S., Griffin, N., See, B.H. and Siddiqui, N. (2019) How Can We Get Educators to Use Research Evidence? A Review of the Best Ways to Get Evidence Into Use From Many Areas of Public Policy. Morrisville, NC: Lulu Press, pp.1–193.

Greany, T. (2020) 'Self-Policing or Self-Improving? Analysing Peer Reviews Between Schools in England Through the Lens of Isomorphism'. In: Godfrey, D. (Ed.), *School Peer Review for Educational Improvement and Accountability: Theory, Practice and Policy Implications*. Cham: Springer Nature, pp. 71–94.

Greany, T. and Higham, R. (2018) Hierarchy, Markets and Networks: Analysing the 'Self-Improving School-Led System' Agenda in England and the Implications for Schools. London: UCL Institute of Education Press, pp. 1–116.

Hadfield, M. and Ainscow, M. (2020) 'Layering' Peer Enquiry as a System Change Strategy: Some Lessons From Wales. In: Godfrey, D. (Ed.), *School Peer Review for Educational Improvement and Accountability: Theory, Practice and Policy Implications.* Cham: Springer Nature, pp, 51–67.

HM Government. (2022) *Opportunity for All: Strong Schools With Great Teachers for Your Child*. London: HM Government. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1063602/Opportunity_for_all_strong_schools_with_great_teachers_for_your_child__print_version_.pdf (accessd 26 April 2022).

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (20XX) *Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook.* Report to the Education Endowment Foundation.
https://educationendowmentfoundation.org.uk/public/files/Evaluation/Setting_up_an_Evaluation/IPE_Handbook.pdf (accessed 22nd February 2023).

Lortie-Forgues, H. and Inglis, M. (2019) 'Rigorous Large-Scale Educational RCTs Are Often Uninformative: Should We Be Concerned?' *Educational Researcher*, *48,* 3, 158–66.

Matthews, P., & Ehren, M. (2017). Accountability and improvement in self-improving school systems. In: Earley and Greany (Eds) School leadership and education system reform, 44-55. London: Bloomsbury Academic

Matthews, A.P. and Headon, M. (2015) *Multiple Gains: An Independent Evaluation of Challenge Partners' Peer Reviews of Schools*. London: Institute of Education Press, pp. 1–72.

OECD. (2013). School evaluation: From compliancy to quality. In Synergies for better learning: An international perspective on evaluation and assessment (pp. 383–485). Paris: OECD Publishing. **https://doi.org/10.1787/9789264190658-10-en** (acessed 22 February 2023).

Rose, J., Thomas, S., Zhang, L., Edwards, A., Augero, A. and Roney, P. (2017) *Research Learning Communities*: *Evaluation Report and Executive Summary.* London: Educational Endowment Foundation. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Research_Learning_Communities.pdf (accessed 22 February 2023.

Vaughan, T., & Albers, B. (2017). Research to practice–implementation in education. Teacher Magazine, 20. **https://www.teachermagazine.com/au_en/articles/research-to-practice-implementation-in-education** (accessed 22 February 2023)

Weinstein, J. (2022) 'Building Trust: A Challenge for School Leaders and Teachers'. In: Peters M.A. (Ed.) *Encyclopedia of Teacher Education*. Singapore: Springer, pp. 1–5. https://doi.org/10.1007/978-981-13-1179-6_483-1 (accessed 22 February 2022).

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

@EducEndowFoundn

Facebook.com/EducEndowFoundn