Evaluation of philosophy for children: an effectiveness trial

Evaluation Report

March 2021

Pippa Lord, Afrah Dirie, Kelly Kettlewell and Ben Styles

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

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

identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;

evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and

encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education. Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

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

Contents

About the evaluator

The project was independently evaluated by a team from the National Foundation for Educational Research (NFER). The principal investigator for this trial was Dr Ben Styles, Head of NFER's Education Trials Unit. Jack Worth, Lead Economist, led the evaluation team in the initial stages. Pippa Lord, Trials Director, led the evaluation team for the follow-up, analysis and reporting stages of the project.

Contact details:

National Foundation for Educational Research
The Mere, Upton Park
Slough
Berkshire SL1 2DQ
p: 01753 574123
e: p.lord@nfer.ac.uk

Acknowledgements

ONS SRS publication status and statistical results

Executive summary

The project

Philosophy for Children (P4C), (a whole-school programme with levels differentiated as Bronze, Silver and Gold based on school-level engagement) aims to improve pupils' and teachers' capability to think in a caring, collaborative, creative and critical way ('the 4Cs') in order to support pupils' personal, social and educational development. The programme is provided to UK schools by The Society for the Advancement of Philosophical Enquiry and Reflection in Education (SAPERE).

P4C comprises whole-school training and support made available to teaching staff as well as to the school P4C lead. Students take part in weekly one-hour sessions, which are gradually embedded in the school curriculum and approach as the school progresses towards Gold level integration. Sessions are enquiry-based; prompted by a stimulus (for example, a story or a video), pupils participate in group discussions based around a concept such as 'truth', 'fairness' or 'bullying'.

This effectiveness two-arm cluster randomised controlled trial saw 75 schools invited to receive the intervention, whilst 123 schools acted as a control group. The trial evaluated the impact of P4C on Y6 pupils' reading, maths, and social and communication skills, with its primary focus on pupils eligible for Free School Meals (FSM). Additional mixed method research sought to assess compliance and fidelity over the course of the intervention, measured by reference to the achievement of SAPERE's Bronze, Silver and Gold Award scheme. The trial started in October 2016 with programme delivery from September 2017 to July 2019.

Key Conclusions
1. There is no evidence that P4C had an impact on reading outcomes on average for KS2 pupils from disadvantaged backgrounds (i.e. FSM eligible pupils). This result has a high security rating.
2. Similarly, there is no evidence that P4C had an impact on reading attainment at KS2 for the whole cohort of Year 6 pupils. There is also no evidence that P4C had an impact on attainment in maths for KS2 pupils – either for the whole cohort, or for pupils from disadvantaged backgrounds.
3. Whilst teacher feedback on P4C was positive - 96% of intervention teachers felt that pupils had improved their level of respect for others' opinions, and 93% felt that pupils had improved their ability to express their views clearly; there was no evidence of impact on children's social and communication skills, as measured by the pupil survey.
4. Of the 75 intervention schools, after two years from commencement, a substantial minority (35 of 75 schools) were not implementing P4C at the expected level. Of these, six did not implement P4C at all due to other priorities and/or senior leader turnover. The evaluation suggests that it takes time for teachers to become confident with, use and embed the P4C approach and this could have impacted the outcomes.
5. Where schools were implementing P4C, teachers and pupils found it enjoyable, engaging and that it encouraged pupils to share opinions in a non-judgmental way, finding it particularly beneficial for EAL pupils, those who lacked confidence or SEN pupils. Teachers and P4C leads felt that the training and ongoing support was high-quality and that it had enabled them to facilitate P4C sessions effectively in their school. Important factors for sustaining and embedding implementation included: starting with sessions based on standalone topics before incorporating cross-curricular work into sessions; and senior staff support, particularly around understanding and valuing the P4C approach.

EEF security rating

These findings have a very high security rating. This was an effectiveness trial, which examined whether the P4C intervention worked under everyday conditions in a large number of schools. The trial was a well-designed school-level randomised controlled trial and was well-powered (to detect a small effect on attainment for FSM children). The primary and secondary attainment outcomes used data accessed from the National Pupil Database (NPD).

Additional findings

This study found no evidence that P4C has a positive or negative impact on reading attainment for FSM eligible pupils, nor on reading and maths attainment for the entire cohort. This suggests that class time can be directed towards this activity without reducing reading or maths outcomes. The study results do suggest a possible positive greater effect for

lower attaining pupils as compared to higher attainers, however, the evaluators failed to conclude this was not a chance finding.

Social and communication skills are of relevance to the theory of change for P4C. For continuity with the previous trial, the same (non-validated) survey instrument was used, with analysis from two single items: 'I am good at explaining my ideas to other people' and 'I can work with someone who has different opinions' (Siddiqui et al., 2017). The use of these items at a larger scale with randomised groups showed no evidence of effect. These single items did not capture the complex character and metacognitive outcomes highlighted in the theory of change and finding a measurement scale that can better capture these outcomes would be important for any future evaluation.

Results from the process evaluation show that teachers felt P4C had positive impact on pupils' social, thinking and communication skills (96%, felt it helped pupils to respect others' opinions, 91% that it improved their ability to question and reason, and 93% that it improved their ability to express views clearly). Staff across all case studies generally agreed that those pupils who struggle to 'have a voice' in normal lessons have benefited the most from P4C and pupils themselves emphasised that they enjoyed being able to express their own opinions in a non-judgmental way. Often, staff reflected that these were EAL pupils, those who lacked confidence or SEN pupils.

The mixed method process evaluation highlighted that it takes time to develop and embed P4C in schools. P4C requires a change in whole school ethos and curriculum innovation. Just over half of the intervention schools reached the expected level over the two years of programme delivery.

Cost

The average cost of SAPERE's P4C Going for Gold programme is £13.50 per pupil per year when averaged over three years. This estimate is based on the delivery of the intervention across all pupils in an average sized primary school of 282 pupils (Department for Education, 2019).

Impact

Table 2: Summary of impact on primary outcome(s)

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No of pupils	P Value	EEF cost rating
KS2 (Y6) reading, FSM eligible pupils	0.02 (-0.09, 0.13)	0	5	3358 (1248; 2110)	0.68	£ £ £ £ £
KS2 (Y6) reading, whole cohort	0.01 (-0.08, 0.11)	0	N/A	7677 (2878; 4799)	0.81	£ £ £ £ £

Introduction

Background evidence

The Philosophy for Children (P4C) programme was originally developed in the USA, in 1970 by Professor Matthew Lipman, with the establishment of the Institute for the Advancement of Philosophy for Children (IAPC). Since then, the approach has gained interest worldwide, now being implemented in schools across 60 countries (with adaptations from Lipman's original model) (Gorard et al. 2015). In the UK, P4C is promoted by the Society for the Advancement of Philosophical Enquiry and Reflection in Education (SAPERE), who provide schools with the resources and training to practice the programme. Rather than teaching children philosophical knowledge, P4C sessions focus around concepts such as truth, fairness or bullying and encourage children to discuss these issues. The aim of the programme is to develop children's ability to think logically about these issues, generate their own opinions and arguments around such topics, which they can confidently communicate while respectfully listening to and considering the views of others.

P4C has been the subject of a number of evaluation studies since the 1970s, these have used various methodologies, but have consistently shown positive impacts on logical reasoning and reading (e.g. Lipman and Bierman, 1970 in the US; Sasseville, 1994 in Canada). Trickey and Topping's (2004) systematic review of studies that used observational designs and a range of norm-referenced tests/questionnaires to explore the impacts of P4C, highlighted a range of positive outcomes across the studies reviewed relating to reading, reasoning, self-esteem and child behaviour. However, most of these studies were small scale, using pre-post surveys or comparing one class of pupils with another for example, or combining P4C with a reading component in the intervention itself, with no consistent outcome instrument being used across studies. Other recent quasi-experimental studies have reported improvements in social and communication skills, positive changes in pupils behaviour, respect shown towards their peers (including greater ability to collaborate and empathise with others), increased resilience and improvements in reasoning and thinking skills (Siddiqui et al., 2017 and 2019).

A number of organisations and individuals provide P4C support to schools in the UK (for example, SAPERE, Jason Buckley, the Philosophy Foundation, and Philosophy4Children), and although there is no published figure on take up, the approach is used across both primary and secondary schools. In 2015, Gorard et al. (2015) reported SAPERE to have 600 schools registered who were regularly implementing P4C, and more recently SAPERE reported training 5,500 primary and secondary school teachers every year to practice the P4C approach in their classes (SAPERE, 2020). At present, 'philosophy' is not a statutory requirement of the English National Curriculum, however SAPERE offer guidance on how an enquiry-based philosophical approach may be integrated into the curriculum. This includes opening dialogues on relevant issues to subjects, for example in art discussing 'what makes something beautiful; in history, discussing the possibility of having a just war; or the concepts of boundaries, responsibility and sustainability in geography' (SAPERE, 2020). However, a recent focus on character education (Ofsted, 2019b), metacognition (EEF guidance report, 2016) and broader non-cognitive skills (emphasised in the revised Ofsted Inspection Framework, 2019a), means that programmes such as P4C have the potential to support areas of the curriculum relevant to the development of broader skills. Indeed, SAPERE state that schools that have embedded the P4C approach into their practice can now deliver curricula which can address new Ofsted criteria relating to pupils' personal development (SAPERE, 2020).

Building on previous evidence, this effectiveness trial of P4C explored the impact of P4C on both educational attainment and social skills outcomes. A previous study commissioned by EEF, represented the first large-scale evaluation of the impact of P4C on attainment in schools in England (Gorard et al., 2015). That study, conducted in 48 schools reported positive impacts on Key Stage 2 attainment, with pupils in intervention schools making approximately two additional months' progress in reading and maths compared with their control group counterparts. The research suggested that the biggest impact on Key Stage 2 results was for pupils from disadvantaged backgrounds (i.e. those eligible for free school meals). Additionally, qualitative reports from teachers and pupils suggested that participation in the P4C intervention positively influenced pupils' wider outcomes, such as their confidence to engage in discussion, their listening skills and self-esteem. The Gorard et al., (2015) trial was classified as an effectiveness trial, meaning that it sought to test whether the intervention can work at scale. However, had its data been analysed using the methods in this study, it is unlikely to have demonstrated such a positive effect (see Interpretation section). Because of the relatively small number of schools involved, this re-granted effectiveness study aimed to obtain a more secure estimate of the impact of P4C on all children and particularly on children eligible for free school meals.

The results from both the previous trial (Gorard et al. 2015) and other work by that evaluation team (underway at the time of that previous trial and published in Siddiqui et al., 2017) informed the design of this effectiveness trial. A similar cluster (i.e. school) randomised controlled trial design was adopted for this effectiveness trial, but on a larger scale. Indeed, this re-granted effectiveness trial was powered for measuring outcomes for children from disadvantaged backgrounds. The primary outcomes were narrowed down from KS2 Reading, Writing and Maths results to only KS2 Reading scores for FSM-eligible pupils, as this subgroup showed the largest effect size (+0.29) in the previous trial (although note that the previous trial was not powered to detect change in FSM pupils). KS2 Maths scores, for all pupils and for FSM-eligible pupils only, as well KS2 Reading scores for all pupils, which showed positive effects in the previous trial[1], were selected as secondary outcomes at the re-granted effectiveness stage. Writing scores were not included as outcomes in the re-granted effectiveness trial as no significant effect was shown in the previous trial. The secondary 'non-attainment' outcome from the previous trial, namely the cognitive ability test (CAT4) instrument (which explores four aspects of reasoning skills – verbal, non-verbal, quantitative and spatial ability) was not used at the re-granted effectiveness stage because it showed inconclusive evidence in the previous trial (Gorard et al., 2015).

At the same time, further quasi-experimental research from the previous trial research team at Durham University showed promising positive impacts of P4C on other non-cognitive outcomes. That evaluation, funded by the Nuffield Foundation, showed positive effects for self-reported communication skills (effect size=+0.10), and teamwork and resilience (effect size=+0.15). The instrument included 11 scaled attitude items, taken from single questions available from 11 established tests of psychological constructs (such as communication skills, and cooperation and teamwork), as well as three scenario-based questions on empathy and generosity, social responsibility and understanding of democracy (Siddiqui et al., 2017). NFER used an adapted version of this survey in the re-granted effectiveness trial; the survey used the same scaled attitude items (with the exception of the item on 'well-being', which was not included as it was considered special data under GDPR), and used only one scenario-question on empathy, with a different scenario. The scenario was changed to make it more suitable for primary school pupils.

As was the case in the previous trial, the control condition was 'business as usual' with a school waitlist design. However, for the re-grant, control schools were asked not to implement P4C with Year 6 pupils for the year following the intervention (2019/20) because they had to represent the control for that subsequent cohort's attainment outcomes[2]. No second cohort attainment follow-up was included in the previous trial.

The previous trial also reached the conclusion that the duration of the programme – which exposed pupils to P4C for just under a year – may not have been long enough for the full impact of P4C to be felt. Consequently, the delivery of the programme was extended to two full academic years at the re-granted effectiveness stage (with a further third year for follow-up).

The re-granted effectiveness trial also involved a process evaluation running alongside the trial as per the protocol. Note that this trial was designed prior to the publication of EEF's guidance on implementation and process evaluations (Humphrey et al., 2017) and the requirement to fully integrate process and impact designs (EEF, 2019). That said, many elements from Humphrey et al.,'s (2017) guidance are included in this evaluation (such as fidelity, responsiveness, adaptations and perceptions of quality). We have been able to focus the IPE on two of the three core areas in EEF's latest IPE guidance (2019), namely compliance and fidelity; but there is less emphasis on usual practice (as this was not an essential area at the time of conducting this study). As P4C is a whole-school intervention that may take time to embed, assessing outcomes at cohort level, and indeed for two successive cohorts, was deemed a suitable design for this effectiveness trial. Powering this study to detect change for disadvantaged children was an important aspect of the design, reflecting EEF's overall remit, and as the previous study suggested the greatest impact was likely to be for these young people. As the intervention is aimed at the whole school, a school-level measure was selected to indicate the level of engagement with P4C and how embedded the approach was in the school. The measure selected was the school-level P4C award status in the Going for Gold programme – from which to then explore any associations with award level and attainment outcomes. The criteria for the awards (bronze, silver, gold) are outlined in the methods section on Implementation and Process Evaluation and can also be found in more detail on the SAPERE website[3].

---

[1] KS2 Maths scores (all pupils) – effect size=+0.10; KS2 Maths scores (FSM-eligible pupils) – effect size=+0.20; KS2 Reading scores (all pupils) – effect size=+0.12.
[2] Due to school closures related to Covid-19 in 2020, no KS2 assessments took place and no KS2 outcomes would be published. Analysis of attainment outcomes for the subsequent cohort was therefore removed from the study.
[3] https://www.sapere.org.uk/Content/Media/P4C%20School%20Award%20Criteria%202018%20for%20website.pdf

Note, as P4C is expected to be delivered as a whole-school approach, the programme does not expect schools to record pupils' exposure to P4C (e.g. attendance in sessions or experience of the approach), and so pupil level exposure data was neither deemed practical nor necessary for this evaluation.

The P4C programme implemented in the re-granted effectiveness trial was the same as that in the previous trial, but with the following additions: schools worked through an award scheme called Going for Gold (see section About the Intervention) (this award scheme was not in full use in the previous trial); the implementation period was for two years rather than one (with an additional addendum year, that was in the event, curtailed due to school closures related to Covid 19); and schools received further training including a session called Tools for Thinking Together (an inset day aimed at improving the facilitation skills of those delivering P4C) and top-up sessions (for example for new teachers who missed the original training, see Table 1 in Appendix C for further details).

Intervention

Why (rationale/theory):

Philosophy for Children (P4C) is an educational pedagogy and a social practice, extending beyond the school community. It emphasises the importance of questioning and enquiry in the development of reasoning. P4C places practical wisdom, reasonableness and good judgement as a goal of education. It focuses upon meaning and value. Prompted by a thought-provoking stimulus, for example a story or a video, children and young people are encouraged to ask philosophical questions (rooted in the branches of philosophy) and engage in dialogue with each other in order to develop their conceptual understanding. The teacher (facilitator) acts as guide to build the Community of Philosophical Enquiry over time, through use of philosophical methods (questioning and reasoning) and the 4Cs (critical, creative, caring and collaborative). The aim of P4C is to help children become more willing and able to ask questions, construct arguments, and engage in reasoned discussion.

As noted above, P4C was originally developed and delivered in the USA in 1970 by IAPC. The Society for the Advancement of Philosophical Enquiry and Reflection in Education (SAPERE), a non-profit society, promotes the use of P4C in UK schools along with developing teaching resources and providing teacher training courses. P4C is practised across both primary and secondary schools. SAPERE's model of P4C differs in some ways from Lipman's original conception. In particular, there is no use of specially written philosophical novels. Materials recommended by SAPERE include stories, poems, scripts, short films, images, artefacts, and picture books. However, Lipman's central aim of creating a classroom 'community of enquiry' is retained along with the broad sequence of activities that constitute a P4C session.

SAPERE provides P4C training and a programme of activity to improve pupils' and teachers' capability to think in a caring, collaborative, creative and critical way ('the 4Cs') to support pupils' personal, social and educational development (including metacognitive skills, often described as 'learning to learn'). The programme comprises whole-school training and support over three years, an in-school P4C lead and resources to support weekly 60-minute P4C sessions with pupils.

A theory of change was developed at the outset of the evaluation setting out the underlying assumptions of the intervention, including what outcomes would be expected from P4C, and the resources and strategies needed in order to support these changes and outcomes. Assumptions from the theory of change were tested through the implementation and process evaluation. More details on this can be found in the research methods section. A copy of the theory of change can be seen in Appendix D.

Who (recipients):

P4C is a whole-school approach delivered by trained teachers within KS1 and KS2. The focus of this evaluation were pupils who were in Year 5 in the 2017/18 academic year and then in Year 6 in 2018/19 (and those in Year 4 in 2017/18 and then in Year 6 in 2019/20 for the addendum cohort4). Each school appointed a P4C lead who was responsible for leading implementation of P4C in the school.

---

[4] Analysis removed due to the cancellation of KS2 assessments in 2020, as a result of school closures relating to Covid-19.

What (materials):

72 intervention schools received training and support over two years from accredited freelance trainers from SAPERE, a package called Going for Gold. The aim of Going for Gold is for the schools to work through the progression indicators, which is a three-level award scheme of Bronze, Silver and Gold which aims to progress schools from first involvement in P4C to P4C being fully embedded in and across the whole school. The aim is for schools to reach working above bronze and at least 50% towards silver after two years of delivery, and to reach SAPERE's Gold Award level of P4C practice by the end of three years of delivery. For each school, the three-year programme consists of the following elements:

- Two days of P4C Foundation Training (Level 1) for all teaching staff (up to 25 staff): this equips teachers to start facilitating P4C enquiries with their pupils, and covers the basic principles of P4C practice, what a standard enquiry should look like and provides an opportunity to experience an enquiry;

- One day of P4C Tools for Thinking Together Training for up to 25 staff; this provides staff with additional facilitation techniques and practical guidance in encouraging stronger reasoning and conceptual thinking among pupils (year 2);

- Four days of Advanced P4C Training (Level 2A and 2B) for 2 staff; Level 2A provides the school's P4C leaders with advanced facilitation techniques so that they can support colleagues who are less advanced in their P4C practice; Level 2B supports the P4C leaders with guidance in how to plan for the development of the school's P4C practice, how to link P4C into the broader curriculum and how to handle sensitive and controversial topics that may arise in an enquiry;

- Seven days of in-school P4C coaching and support (approximately 3 days in Year 1, 3 days in Year 2 and 1 day in Year 3); the SAPERE trainer tailors the content of these days to the school's needs; they may include demonstration, observation or co-teaching by the trainer, or planning with the P4C leader or remedial work with teachers who need extra assistance, or specialist advice on linking P4C to specific subjects;

- Five days of remote administration and planning support; these are for ad hoc support on the implementation of P4C and may include guidance on the Bronze, Silver and Gold award applications;

- Up to four top-up places for schools on open Level 1 courses for new teachers joining the schools during the programme;

- Unlimited access to SAPERE's online P4C resources and practice guides; these include a wide bank of suggested enquiry stimuli, a Getting Started Guide, a Moving On with P4C guide, a range of teaching materials and example enquiry plans and the Award framework which sets out a detailed progression for P4C practice across student, teacher and whole school dimensions;

- Two reference copies of SAPERE's Level 1 and Level 2 handbooks;

- Application and assessment fees for SAPERE's Bronze, Silver and Gold awards.

The initial training (level 1) was delivered as INSET days to up to 25 teaching staff per school between March and October 2017. Trained teachers, teaching assistants (TAs) or higher-level teaching assistants (HLTAs) delivered P4C by facilitating sessions with pupils.

Where, when and how much?

Sessions are usually delivered in classrooms but other areas of school site can be used (as appropriate to the topic/s being discussed). For KS2 pupils (and KS1 pupils), SAPERE recommends one 60-minute P4C session each week to take place within the normal school timetable. As an absolute minimum, one 60-minute session could take place every two weeks but a weekly session is the preferred model. This can be standalone P4C sessions or sessions embedded into the curriculum. If standalone, there is no specific guidance as to what this should replace in the curriculum. Sessions are teacher-initiated but strategies and activities can feature in other lessons/as part of the school day.

How?

P4C sessions take the form of an enquiry which has 10 steps:

The first step in the enquiry (preparation/getting set) aims to ensure pupils are in the right frame of mind for P4C and may focus on building a community, rehearing the aims of P4C, or working on a particular skill, such as listening.

The second step is the presentation of the stimulus. This may be a poem, a picture or a film for example but should always be something that is engaging, meaningful and /or relevant to the pupils.

The third stage allows the pupils time to think on their own about the stimulus (thinking time).

In step 4, pupils work together in small groups to identify questions relating to the stimulus (question making).

Pupils then share their questions which are discussed and analysed by the whole group (question airing).

A question is then chosen , often by voting on it, to take forward for discussion.

first thoughts relates to getting the dialogue on the chosen question started.

Building focuses on building on these first thoughts, working towards better understanding of concepts, opening up the discussion to different perspectives, for example.

Last thoughts draws the enquiry to a close and allows for final words on what has been discussed.

The final step is review, where pupils are able to reflect on the session.

Tailoring:

Teaching staff are trained in the core P4C skills but are encouraged to adapt resources and materials if they wish. Teachers will respond to and be led by the questions and interactions from pupils during the sessions. Accredited trainers provided by SAPERE offer support to schools/P4C leads based on the needs of the school; trainers provide their own resources which may or not be on SAPERE's database of resources.

Strategies to maximise implementation effectiveness include attendance at all training sessions and having a named P4C lead in the school. This is supplemented by accessing support from SAPERE-accredited trainers and resources from either the SAPERE website or individual trainers.

Evaluation objectives

The primary attainment research question for the trial was:

1. Does using the P4C approach in schools improve disadvantaged pupils' attainment in Key Stage 2 reading? (as measured by KS2_READSCORE from the National Pupil Database (NPD) for the 2018/10 cohort, with EVERFSM pupils identified using the EVERFSM_6_P variable from the spring term school census of 2018/19 available from the NPD).

The secondary attainment research questions for the trial were:

Does using the P4C approach in schools improve pupils' attainment in Key Stage 2 reading? (as measured by KS2_READSCORE from the NPD for the 2018/19 cohort).

Does using the P4C approach in schools improve pupils' attainment in Key Stage 2 maths? (as measured by KS2_MATSCORE from the NPD for the 2018/19 cohort).

Does using the P4C approach in schools improve pupils' attainment in Key Stage 2 reading? (as measured by KS2_READSCORE from the NPD for the 2019/20 cohort) (this will no longer be reported due to the cancellation of these assessments).

Does using the P4C approach in schools improve pupils' attainment in Key Stage 2 maths? (as measured by KS2_MATSCORE from the NPD for the 2019/20 cohort) (this will no longer be reported due to the cancellation of these assessments).

The secondary non-attainment outcomes for the trial were:

Does using the P4C approach in schools improve pupils' social skills in terms of their social and communication skills? (as measured by an adapted version of a questionnaire used by Gorard *et al.* (2015) and SAPERE as part of other evaluations of P4C; used with the Year 4 cohort from 2016/17 who were then in Year 6 in 2018/19).

Does using the P4C approach in schools improve pupils' social skills in terms of their team work and resilience? (as measured by an adapted version of a questionnaire used by Gorard *et al.* (2015) and SAPERE as part of other evaluations of P4C; used with the Year 4 cohort from 2016/17 who were then in Year 6 in 2018/19).

The process evaluation was designed to build on the Gorard et al. (2015) study and to assess fidelity to treatment. It focused on the following research questions:

2. What works in implementing the P4C programme in schools (including what appear to be the key factors that facilitate teachers' engagement with and use of the P4C programme)?
3. How are schools using the *recently-developed* P4C progress indicators to disseminate and explain individual and class performance to pupils and teachers?
4. Is there capacity to support the delivery of a scaled-up P4C programme?

Note that the process evaluation was designed and specified prior to both the implementation handbook guidelines published by EEF in 2017 (Humphrey et al., 2017) and the EEF implementation and process evaluation (IPE) guidance published in December 2019. Where possible, we have reported the IPE analysis undertaken in this trial in light of the latest guidance – focusing on compliance, fidelity and usual practice.

The process evaluation also explored the one-off and on-going costs for schools associated with implementing P4C. It was agreed with the EEF that the cost evaluation would report against the cost guidance from 2016 – as the cost data was collected from schools and already analysed prior to the December 2019 cost guidance being published.

The protocol for the trial was published in March 2017; with an update in October 2018 noting staffing changes and an amendment to the teacher survey to be sent to all teachers involved rather than just the P4C lead, and amended to online rather than paper; and a further update in May 2019 clarification that simple randomisation took place, and updating the data security aspects of the protocol in light of the General Data Protection Regulations (GDPR). Links to these protocol documents are outlined below:

Protocol, March 2017:
https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Regrant-_Philosophy_for_children.pdf

Protocol update, October 2018:
https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Philosophy_for_Children_-_Evaluation_protocol_(amended).pdf

Protocol update, May 2019:
https://educationendowmentfoundation.org.uk/public/files/Projects/EEPS_trial_protocol_AMENDED_(2).pdf

The statistical analysis plan (SAP) was published in November 2017. It was updated in May 2019 clarifying that simple randomisation was used, specifying which social skills measures would be analysed at follow-up, and in consultation with the EEF amending secondary outcomes measures from relating to FSM children only to outcomes for the whole cohort. In addition, it was clarified that rather than using the R package in eefAnalytics, .we would use nlme [ref].

Links to these SAP documents are outlined below:

SAP November 2017:
https://educationendowmentfoundation.org.uk/public/files/Projects/Regrant-_P4C_SAP.pdf

SAP update May 2019:
https://educationendowmentfoundation.org.uk/public/files/Projects/EEPS_Statistical_Analysis_Plan_AMENDED_20190509.pdf

Additional exploratory analysis

Note that as a result of school closures related to Covid-19, the summer 2020 KS2 assessments were cancelled and no KS2 outcome results would be available for the 2019/20 cohort. This element was removed from the study. Instead, in agreement with EEF and SAPERE, we carried out additional exploratory analysis using the pupil survey with the aim of

exploring whether P4C improved pupils' social skills in relation to any domains that could be identified through factor analysis.

Ethics and trial registration

The P4C trial received ethical approval through NFER's standard project start up procedures and Code of Practice group on 31st October 2016.

NFER was responsible for recruiting schools to the trial. A letter was sent out to local authorities on 17 October 2016 informing them of NFER's intention to approach schools to take part in this evaluation. A letter for headteachers was dispatched on 1 November 2016 to all schools in the sample asking if they would like to take part in the evaluation (further details on how the sample was drawn can be found in the methods section of this report). The letter gave information on the procedure for randomisation and what would be required from each school prior to and after randomisation. The letter included a reply form asking for a contact at the school for dates they would be available for training (should the school be randomised into the intervention group) and the number of pupils currently in their Year 4 cohort. Also included was an information sheet, which showed the project at its various stages and a memorandum of understanding (MOU) that the school had to sign, setting out expectations for both the intervention and control groups. Headteachers were asked to complete the reply form and sign the MoU and return them to NFER in a provided pre-paid envelope.

Once schools had signed the MoU, schools then provided a list of Year 4 pupils' details (forename, surname, date of birth and unique pupil number (UPN)). At the same time, a parent letter was uploaded to school portals for schools to share with their Year 4 cohort parents. This gave parents the option to withdraw from data processing in relation to their child's survey data.

Copies of the school MoU and parent letter are included in Appendix E.

The trial is registered on the ISRCTN registry at: **https://doi.org/10.1186/ISRCTN11118203** (Evaluation of the effectiveness of philosophy for children at improving the literacy skills of disadvantaged pupils).

Data protection

Data protection statement

All data gathered during the trial was and will be held in accordance with the data protection framework created by the Data Protection Act 2018 and the General Data Protection Regulation (GDPR) 2016/679, and was and will be treated in the strictest confidence by the NFER. No individual or school will be identified in any report.

Legal basis for processing personal data

NFER was the data controller during this RCT evaluation. Our legal basis for processing teachers', pupils' and SAPERE staff personal data was covered by:

GDPR Article 6 (1) (f) which states that 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

We carried out a legitimate interest assessment which demonstrated that the evaluation fulfilled one of NFER's core business purposes (undertaking research, evaluation and information activities) and it has broader societal benefits. We considered and balanced any potential impact on the data subjects' rights and found that our activities will not do the data subject any unwarranted harm. Therefore, it was in our legitimate interest to process and analyse the personal data described below in order to administer the randomised controlled trial.

Personal data processed

The personal data processed for this trial was:

- Names and contact details for headteachers and teachers involved in the trial – so that we could keep in touch with them, carry out the implementation/process evaluation (including case studies, telephone interviews and a teacher/key contact survey), and conduct surveys in schools for the secondary non-attainment outcomes for the trial.
- Names, dates of birth and UPNs of pupils in Year 4 in 2016/17, so that we could ask them to complete a baseline survey (in Year 4) and a follow-up survey (in Year 6), and match their personal data to background data from the National Pupil Database (NPD) for archiving.

In addition, NFER requested anonymised pupil data from the DfE NPD via ONS/SRS including: Key Stage 2 reading and maths assessment results, KS1 attainment data, and background data including EverFSM status, month of birth and EAL status.

No special category data was processed in this trial.

Data security/transfer

All personal data provided electronically was done so using NFER's secure school portal. All personal data provided in hard copy was transferred by secure courier. All NPD data was accessed using the Office for National Statistics secure access procedures. All researchers involved directly with pupils and their data had up-to-date DBS checks. NFER survey administrations obtained personal data in accordance with the GDPR and other applicable legislation.

Data sharing

NFER shared the names, job titles and contact details of the P4C leads in each intervention school with SAPERE, so that they could be contacted about the programme. SAPERE and NFER shared updated teacher/school contact information as needed.

NFER requested **anonymised** pupil data from the DfE as noted above, via ONS. Further matching to NPD and other administrative data may take place during subsequent research.

For the purposes of the research, pupils' data and questionnaire responses from the social competencies/skills survey will be linked with information from the NPD and shared with the EEF's archive manager and, in an anonymised form, with the Office for National Statistics and potentially other research teams. Further matching to NPD and other administrative data may take place during subsequent research.

Personal data collected through telephone interviews, case studies or in the teacher survey was not shared with other organisations. No individuals or schools are or will be named in any report arising from this research.

Data retention and deletion

Within three months of the end of project, NFER will send school and pupil data to EEF's data archive partner. This will include school names, ID and intervention group variable, anonymised pupil data from NPD and pupil survey data. At this point, EEF's data archive partner will keep a copy of the data and EEF will become the Data Controller. Data will also be shared with the DfE, the Office for National Statistics (ONS[5]) and, in an anonymised form, potentially with other research teams. Further matching to NPD and other administrative data may take place during subsequent research.

NFER will retain personal data for one year after report publication in case there are any queries about the report. One year after the report publication, all personal data will be securely deleted.

Right to withdraw

As set out in the ethics section above, participants had the right to withdraw their data or correct any errors in it at any time. The letter to parents at baseline (in 2016 and early 2017) gave parents the opportunity to withdraw their child's data from the study through a withdrawal form. A keep in touch email to all schools in July 2018 included a link to the project Privacy Notice and a Privacy Notice for parents covering the GDPR enacted in May 2018. Schools were asked

---

[5] ONS is scheduled to host the EEF archive from 2020.

to make the Privacy Notice available to parents using their usual channels. The Privacy Notices were available via links on the project pages of the NFER website. Contact details for how to withdraw/correct errors were provided in the Project Privacy Notice and the Privacy Notice for parents (see Appendix E), where further information on Data Protection for this trial could also be found:

https://www.nfer.ac.uk/media/3367/philosophy_for_children_parent_privacy_notice.pdf

Project team

The evaluation was led by a team of researchers, statisticians, project managers and operations staff at NFER. The trial was directed by Dr Ben Styles (Head of NFER's Education Trials Unit). As project director and principal investigator for the trial, he had strategic oversight of the trial. The trial was initially led and managed day-to-day by Jack Worth (Senior Statistician) who also had responsibility for the statistical elements. In June 2018, Jack moved to another role within NFER, and Pippa Lord (Senior Trials Manager and now Trials Director within NFER's Education Trials Unit) took over the day-to-day responsibility for leading the trial and impact evaluation including liaison with SAPERE. Afrah Dirie (statistician) was assigned to the project in September 2018, and she conducted all statistical analysis for the trial. This was quality assured by other members of NFER's Centre for Statistics including Sally Bradshaw and Joana Andrade.

Kathryn Hurd (Head of NFER's Survey Operations) had overall responsibility for leading the recruitment of schools, communications with schools, and survey administration. She was supported by a team of researchers in NFER's Research and Products Operations Department, including Michael Neaves and Guido Miani.

The process evaluation was initially led by Claire Easton (Senior Research Manager), until October 2018 when she left NFER. She had day-to-day responsibility for leading this strand of work, and was supported with quality assurance by David Sims (Research Director), and Kelly Kettlewell (Research Manager) for support with process data collection. Kelly took over the role of process lead from October 2018.

Kam Ahitan provided administrative support to the project in the early stages, and Shila Mistry in the later stages of the project.

The delivery team at SAPERE had the following responsibilities during the evaluation:

- Bob House – Executive Director: school recruitment and project set up (until March 2017)
- Liz Jones – Chair of Trustees: lead on research specification (until March 2019)
- Rod Cunningham – Trustee: support on research specification
- Steve Williams – Project delivery manager: support on training (2016 to December 2019)
- Alison Allsopp – National training manager: trainer allocations
- Amelia Foster – CEO: project oversight/ steering group (September 2016 onwards)
- Jen Simpson – Assistant National Training Manager and EEF Project Manager (January – December 2020)

Methods

Trial design

This was a two-armed cluster randomised controlled trial involving 198 primary schools in England. The two arms were *intervention* (implementing the P4C programme) and a business-as-usual control group that continued with usual classroom practice for two years. The trial included all pupils who were present in Year 6 from participating schools in 2018/19 for the analysis of the primary and secondary attainment outcomes; and pupils who were present in Year 4 in 2016/17 and in Year 6 in 2018/19 from participating schools for the analysis of secondary non-attainment outcomes.

Schools were randomised to the intervention or control group by simple randomisation. The trial was designed with one randomisation of schools. However, two recruitment windows were required in order to recruit sufficient schools to the study, and hence there were two blocks of randomisation of schools; one in January 2017 and another in March 2017. The allocation ratio was adjusted in order to accommodate delivery numbers (maximum 75 intervention schools). The randomisation process resulted in 75 schools randomised to the intervention group and 123 schools to the control group.

The intervention took place over two academic years (from September 2017 to July 2019). A further year of implementation was designed to take place from September 2019 to July 2020, with addendum analysis and reporting planned for those pupils present in Year 6 in trial schools in 2019/20. At the end of the academic year 2018/19, schools in the control group were able to join P4C induction training sessions as part of a 'waitlist' design. Control schools could then deliver P4C from September 2019 (if they wished) with any of their year groups except for Year 6 (2019/2020) – as they would act as a control cohort in the planned addendum analysis. School closures in mid-March 2020 (as part of the UK Government's response to Covid-19) meant that intervention implementation in the addendum year was curtailed. The planned 2020 KS2 addendum analysis has been removed from the study due to the cancellation of KS2 assessments for the 2019/20 cohort.

The results presented here are from the main two-year trial. The main trial ran according to the updated protocol published in 2019 (NFER, 2019). Updates related to staffing, the end-point survey timetable (brought forward to just before the Easter holidays, to avoid any influence of pupils' SATs experience on their survey responses), and the mode of the final teacher survey (amended from paper to online, and from being aimed at the P4C lead only, to all teachers in the schools involved in P4C); these changes are outlined in the protocol update. Table 3 presents the trial design in brief.

In addition, further exploratory analysis was carried out using the pupil survey. Methods and results are presented for this, at the end of each of the methods and results sections of the report, respectively.

Control condition

Schools allocated to the control group continued 'business as usual' and were required to commit to not participating in P4C until September 2019 (and from Sept 2019 – July 2020 to not involve Y6 in any P4C). These schools would receive a financial incentive of £5,700 for participating in the research in July 2019. This financial incentive was equivalent to the saving that the intervention schools received on the cost of purchasing the P4C programme for the purpose of the evaluation. Should a control school wish to sign up to P4C at the end of the main evaluation (i.e. after pupil data collection and KS2 SATs were complete at the end of May 2019) they would be able to use this incentive payment towards to costs of the programme (training and delivery). They would therefore receive the programme for the same reduced costs as those in the intervention group (please note that all schools in the intervention had to pay to take part in the trial).

Table 3: Trial design

Trial design, including number of arms		Two-armed cluster randomised
Unit of randomization		Schools
Stratification variable (s) (if applicable)		N/A
Primary outcome	Variable	Pupil attainment in reading (FSM eligible pupils)
	Measure (instrument, scale, source)	KS2 reading scaled score, min-max80-120, accessed through NPD
Secondary outcome(s)	Variable(s)	Pupil attainment in reading Pupil attainment in mathematics (FSM eligible pupils) Pupil attainment in mathematics Social and communication skills Teamwork and resilience
	Measure(s) (instrument, scale, source)	KS2 reading scaled score, min-max 80-120, NPD KS2 mathematics scaled score, min-max 80-120, NPD KS2 mathematics scaled score, min-max 80-120, NPD Item 1A: 'I am good at explaining my ideas to other people' (social and communication skills), Likert scale, 0-5, bespoke survey (Year 6, follow-up) Item 1C: 'I can work with someone who has different opinions' (teamwork and resilience), Likert scale, 0-5, bespoke survey (Year 6, follow-up)
Baseline for primary outcome	Variable	Pupil attainment in reading (FSM eligible pupils)
	Measure (instrument, scale, source)	KS1 reading attainment point score, min-max 0-51, NPD
Baseline for secondary outcome(s)	Variable	Pupil attainment in reading Pupil attainment in mathematics (FSM eligible pupils) Pupil attainment in mathematics Social and communication skills Teamwork and resilience
	Measure (instrument, scale, source)	KS1 reading attainment point score, min-max 0-51, NPD KS1 mathematics attainment point score, min-max 0-51, NPD KS1 mathematics attainment point score, min-max 0-51, NPD Item 1A: 'I am good at explaining my ideas to other people' (social and communication skills), Likert scale, 0-5, bespoke survey (Year 4, baseline) Item 1C: 'I can work with someone who has different opinions' (teamwork and resilience), Likert scale, 0-5, bespoke survey (Year 4, baseline)

Participant selection

NFER was responsible for the recruitment of schools. To be included in the trial, schools with at least 25% FSM recorded in the 2015 annual school census were included. A list of schools was provided to NFER by SAPERE of schools that had previously purchased the P4C programme; these schools were excluded from the sample. Schools were randomly drawn from NFER's register of schools. Maintained junior and primary schools in England that included pupils in year groups 4, 5 and 6 were considered eligible for the trial.

There was a three-pronged approach to recruitment: 1) approaching the sample of schools drawn by NFER; 2) writing to schools that had directly approached SAPERE expressing interest in the trial (i.e. from SAPERE's list of 'interested' schools, with which they had regular email and twitter contact about recruitment); and 3) tweets and promotion by SAPERE and EEF, including from SAPERE's patron Stephen Fry who tweeted about recruitment to the project, and by EEF who tweeted about recruitment to the trial. SAPERE was also in contact with Liverpool local authority, who supported the recruitment phase of the evaluation.

NFER contacted schools using the following recruitment strategy:

- Letters were sent out to local authorities informing them of NFER's intention to approach schools to take part in this evaluation.

- Letters were dispatched to schools in the sample asking if they would like to take part in the evaluation. Enclosed with the letter was a reply form asking for a contact's details and the number of pupils currently in their Year 4 cohort (i.e. across the whole year). Also included was a memorandum of understanding (MOU).

- When a reply form was received, the contact's information was entered into NFER's administrative systems and the pupil numbers were entered into a spreadsheet.

- Schools were reminded through phone calls, with reminders focusing on SAPERE's list of 'interested' schools. Schools were also sent reminder emails and letters.

One hundred and ten schools went through to the randomisation stage in January 2017 and it was agreed with EEF and SAPERE that a second round of randomisation should take place to try and achieve a sample of at least 200 schools. It was agreed that the second randomisation would take place in March 2017. A further 88 schools went through to randomisation resulting in a total sample of 198 schools. 75 schools were allocated to the intervention group, 123 schools to the control group. The reason for unequal allocation of schools to each group was because the capacity of the developer to deliver training across a large number of schools across different areas of England during two school terms was limited to a maximum of 75 schools. This was taken into consideration when deciding the ratio of intervention to control schools.

Outcome measures

The primary attainment outcome for this trial is FSM eligible pupils' attainment in reading, and secondary attainment outcome measures are: pupils' attainment in reading for the whole cohort, FSM eligible pupils' attainment in mathematics, and pupils' attainment in mathematics for the whole cohort. The two secondary non-attainment outcome measures relate to 'social and communication skills' and 'teamwork and resilience' as measured by single items from an instrument used in previous research into the non-cognitive impacts of P4C (Siddiqui et al., 2017). The items 'I am good at explaining my ideas to other people' and 'I can work with someone who has different opinions' were chosen by SAPERE and EEF as previous research (Siddiqui et al., 2017) suggested that these items from this instrument best reflect the P4C aims. At the time of selecting these outcome measures, NFER recommended an appropriate standardised measure. However, budget constraints and SAPERE/EEF preferences to use the same instrument as had been used previously to evaluate P4C meant a focus on these two outcomes from this instrument.

A further two longer-term secondary attainment measures were planned for analysis. However, schools in England closed in March 2020 due to Covid-19, and the summer 2020 KS2 assessments were cancelled. No KS2 outcome results would be available for the 2019/20 cohort. The longer-term attainment analysis was therefore removed from the study. Instead, in agreement with EEF and SAPERE, we carried out additional exploratory analysis using the pupil

survey with the aim of exploring whether P4C improved pupils' social skills in relation to any domains that could be identified through factor analysis.

Primary outcome

The primary outcome was measured using KS2 reading scaled scores as the outcome measure and KS1 reading point scores as the baseline measure. Key Stage 2 assessments are normally administered by class teachers, marked by markers recruited and trained by the Standards and Testing Agency's (STA's) marking supplier, and converted to scaled scores before the results are entered onto the NPD.

The reason for using Key Stage tests was to reduce the burden of additional testing on teachers and pupils. Using NPD data meant that the research team were able to obtain outcome data if a pupil attended a different school at baseline (i.e. the DfE were asked to use the school names and URNs to find and match the information of all the pupils who were in that school and cohort at that time). Whilst this means there may have been a few pupils included who joined school during intervention delivery, this was far outweighed by the inclusion of all infants school results. NFER provided DfE with a list of school names, URNs, their randomisation group allocation and their randomisation block (January or March) which was used to obtain KS1 and KS2 attainment data from the NPD for the Year 6 2018/19 cohort. Variables used were KS2_READSCORE and KS1_READPOINTS. As EEF has a strong focus on improving outcomes for disadvantaged pupils, and the results of the previous trial (Gorard et al., 2015) suggested greater impact for children eligible for FSM, the pupils of interest for the primary outcome were FSM-eligible Year 6 pupils from participating schools in 2018/19. It is worth noting that the design was for a cohort analysis, of FSM-eligible pupils in Year 6 in trial schools in 2019, to explore learning outcomes in schools after two years of teachers' implementing and embedding the P4C approach in the school.

Secondary outcomes

Attainment measures: The secondary attainment outcomes were measured using variables accessed from the NPD. These were collected at the same time as the primary outcome and consisted of KS1 and KS2 reading and KS1 and KS2 mathematics attainment data for the whole cohort, as baseline and endpoint measures. For the reading outcome, variables used were KS2_READSCORE and KS1_READPOINTS, for the whole cohort. For the mathematics outcomes, variables used were KS2_MATSCORE and KS1_MATPOINTS, for both the whole cohort, and for FSM eligible pupils only.

Non-attainment measures: In order to investigate whether the intervention had an impact on pupils' social and communication skills, the same questionnaire used by Durham University as part of a separate evaluation of P4C was used (Siddiqui et al., 2017). The survey consisted of two questions: Question 1 of the survey was made up of 12 separate items each with five categories of responses ranging from 1 ('not at all true') to 5 ('completely true'); and Question 2 involved a scenario where pupils selected one response from three options – this measured pupils' empathy. There was lengthy discussion between EEF, evaluator and developer regarding the selection of non-attainment outcomes. The NFER team recommended that a standardised measure be used, and suggested the Social Skills Improvement System (SSIS), but it was considered too costly to purchase and administer. This was only resolved by choosing items from the Siddiqui et al. (2017) study, despite the clear psychometric limitations of using single items to measure outcomes. To be consistent with the previous research, the items used for analysis of the non-attainment measures were item 1A ('I am good at explaining my ideas to other people') and item 1C ('I can work with someone who has different opinions'). Note, the instrument is made up 12 separate items; they do not form a validated scale. All remaining items that did not form a secondary outcome measure were analysed through cross tabulations and results are presented in Appendix K. A copy of the survey is found in Appendix F.

In 2017, pupils in Year 4 in participating schools were asked to complete the above 'survey' instrument on social skills, forming our baseline measure. This baseline survey was administered by teachers and completed by pupils. The administration of the same survey was repeated in March-April 2019 with Year 6 pupils (who had completed the survey at baseline) in all schools participating in the trial (i.e. both control and intervention schools)[6]. NFER test administrators administered the follow-up survey in schools. This helped to ensure that the administrators were blind to randomisation

---

[6] Note, spare questionnaires were also administered, to any other pupils in Year 6 (i.e. who had joined the school since Year 4), where this aided administration logistics in the school.

allocation. Additionally, this reduced the burden placed on schools by ensuring teachers were not administering the surveys. Pupils that had completed the survey in both Year 4 and Year 6 were included in the non-attainment outcome analysis.

Sample size

Sample size from the protocol

Initially, the number of schools in the trial was driven by sample size calculations. NFER proposed that a sample size of 150 schools required for the trial would achieve an estimated minimum detectable effect (MDES) of 0.13 for all children. Given EEF's focus on disadvantage pupils, EEF wanted to explore the possibility of powering the trial to detect an effect of 0.1 on FSM pupils or at least an MDES of 0.13 on FSM pupils, as opposed to all pupils. This meant that the estimated number of schools in the trial increased to 200. It was decided to have an unequal allocation with 75 schools in the intervention group and 125 in the control group. This was because the capacity of the developer to deliver training across a large number of schools across different areas of England during two school terms was limited to 75 schools and so this was also taken into consideration when deciding the ratio of intervention to control schools.

At protocol stage, we used a power calculation with the following two assumptions obtained from the EEF's pre-testing paper (EEF, 2013); correlation between pre-test and post-test (the lower of reading) would be 0.73, and the intra-cluster coefficient as 0.137. While writing the protocol, it was assumed that the average number of FSM eligible pupils per eligible school was 14.4. Based on these values, the MDES for FSM eligible pupils was 0.125 at 80.5% power. Balancing the proportion of intervention and control schools (equal number of schools in each group) was also considered and this resulted in a slightly higher power (83%), but had potential to cause delivery issues.

Randomisation

NFER recruited 208 junior and primary schools in England. Schools were required to fill out a MoU with their reply form to be accepted onto the evaluation. Of these schools, nine schools did not return their MoU or administrative pupil data and one school did not complete the baseline questionnaire. As mentioned in the Participants section of the report, 110 schools went through to the randomisation in January and a further 88 schools went through to the randomisation in March, giving a total of 198 schools for randomisation.

Schools were randomised to the intervention or control group using simple randomisation. This differed to what was initially reported in the protocol where a process of stratifying the randomisation by region was considered (to aid the practicalities of delivery by avoiding geographical 'clumping'). Simple randomisation was preferred as fewer degrees of freedom are lost to control for the stratification. Simple randomisation was used once it was established that stratification by region was unnecessary for intervention delivery.

An NFER statistician carried out the randomisation using SPSS with a full syntax trail. The syntax is included in Appendix G (January block) and H (March block). Table 4 presents the number of schools and year groups randomised to each trial arm.

Table 4: Number and proportion of schools at randomisation

		Intervention (%)	Control (%)	Total
Randomisation Block	January	41 (37%)	69 (63%)	110
	March	34 (39%)	54 (61%)	88
Total	Number	75 (38%)	123 (62%)	198

Source: NFER P4C randomisation data, 2016/7

Statistical analysis

The analysis followed the EEF's Statistical Analysis Guidance (EEF, 2018) and the trial SAP (NFER, 2019). This section provides an outline of the analysis undertaken.

Primary intention-to-treat (ITT) analysis

The primary outcome analysis was an intention-to-treat (ITT) analysis, and was conducted at the pupil-level, comparing reading attainment between FSM-eligible pupils in intervention schools with FSM-eligible pupils in control schools. Pupils who were eligible for free school meals made up the analysis of the primary outcome, using the EVERFSM_6 variable obtained from the NPD. As the pupil-level data was clustered in schools, a multilevel model with two levels (school and pupil) was used.

The primary outcome measure 'KS2 reading scaled score' was the dependent variable in the model and the following covariates were added in the model:

- A variable that indicated the intervention group of a pupil (reference category was the control group)

- A variable that identified when the pupil was randomised (reference category was January)

- Pupils' KS1 reading point score

Analysis of the primary outcome was carried out using the nlme (Pinheiro et al., 2019) package in R (R core team, 2019).

Estimation of effect sizes

As mentioned in EEF's statistical analysis guidance (EEF, 2018), for comparability between EEF projects and with the wider literature, EEF requires effect size calculations to be standardised. In multilevel models, variations in post-test outcomes are assumed to be due to different sources which need to be accounted for in statistical models.

To do this, the following formula was used to calculate the effect size.

$$\mathbf{ES} = \frac{(\overline{Y}_T - \overline{Y}_C)\mathbf{adjusted}}{\sqrt{\sigma^2}}$$

Where the numerator for all effect size calculation was the coefficient of the treatment group from the regression model adjusted for the baseline measure and other variables in the model. All effect sizes were calculated using the square root of the population variance of the two groups, as the denominator, which was calculated using total variance from the multilevel model without any covariates.

Confidence intervals for each effect size were derived by multiplying the standard error of the interaction coefficient by 1.96. These were converted to effect size confidence intervals using the same formula as the effect size itself.

Estimation of ICC

The intra-cluster correlation for the primary outcome was calculated using the following formula:

$$\frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\epsilon^2}$$

Where the numerator is the variability between schools and the denominator is the total variance i.e. the sum of the variance between and within schools. An ICC score of 0 implies no variation between schools and a high ICC score close to 1 implies high similarity between pupils' KS2 reading scaled scores within the same school.

Analysis in the presence of non-compliance

The main analysis was followed by a Complier Average Causal Effect (CACE) analysis in order to investigate the effect of level of compliance on the primary outcome measure. Compliance information about P4C was obtained from SAPERE's awarding scheme to categorise schools according to how far through the 'Going for Gold' programme

schools had progressed at summer 2019. The criteria for the different levels are outlined in the methods section on Implementation and Process Evaluation. SAPERE were responsible for collecting information on the progress made and any award level achieved. Schools are assessed against the criteria when they make an application for an award. As stated in the protocol, to ensure that every school had a measure of compliance for the analysis, schools that had not recently submitted an application for an award were assessed against the criteria by a SAPERE trainer. It is worth noting that SAPERE's trainers work closely with schools and are in a position to assess the level each school is working at, even if the school has not completed an award application. Table 5 below shows the categories derived from the SAPERE bronze/silver/gold scheme which was agreed upon by NFER and SAPERE. By the summer term of 2019, SAPERE intervention schools were expected to have completed the Bronze award and met 50% of the criteria for the Silver award (or above) (i.e. Level 5 or above).

Compliance for this intervention therefore relates to whole-school progress with engaging with, delivery and embedding the P4C approach in the school. The section on Compliance in the IPE findings presents further information on the requirements at each of the levels (Bronze, Silver and Gold).

Table 5: Description of engagement levels derived from the SAPERE bronze/silver/gold scheme

Engagement/compliance level	Description
0	No P4C sessions/withdrew before P4C activity started
1	Not actively working towards a level
2	Below bronze but some P4C activity and/or P4C activity by a few staff only
3	Working towards Bronze
4	Bronze, but not at 50% Silver
5	Bronze and 50% silver or above
6	Silver, but not at 50% Gold
7	Silver and 50% Gold, or above

Source: SAPERE compliance data 2019/20

We used a two-stage least-squares model to calculate the CACE estimate (Angrist and Imbens, 1995). In the first stage of the model, we regressed the compliance level on the covariates that were used in the primary outcome model and included (as an instrument variable) a binary variable that indicated a pupil's school's pre-treatment allocation.

The second stage of the model regressed the primary outcome on all covariates in the primary outcome model and a covariate that represented the estimated compliance level which was obtained from the first stage of the model. The coefficient of compliance level was the CACE estimate. The package ivpack (Jiang and Small, 2014) in R (R core team, 2019) was used to carry out the analysis.

Missing data analysis

As per the SAP (NFER, 2019), we envisioned that the number of pupils with missing outcome data would be small and these cases would be removed from the analysis without risk of bias. It was anticipated that the level of missing data would not exceed 5% at either the school or pupil-level.

As seen in the participation flow diagram (see Figure 1) in the impact evaluation section of the report, there is a missing level of 0% at the school-level. This was because de-identified pupil-level data from all schools were collected from the NPD even if a school had withdrawn from the intervention (or from taking part in the survey).

At the pupil-level, there is 6.7% missing data which can be seen in the attrition table (Table 12) in the impact evaluation section of the report. The reason for missing data at the pupil-level is unrelated to the randomised group as the outcome is a statutory test and so it is highly unlikely that the reason for a pupil to miss the test is related to the group a pupil was allocated to. Whilst the statistical analysis plan indicated that missingness analysis would be carried out should attrition

be greater than 5%, there is no reason to suspect bias in our attainment data, and so balancing this reason with the reasonably small amount of missing data, the evaluation team felt it unnecessary to carry out a missing data analysis.

Sub-group analyses

As defined in the SAP (NFER, 2019), subgroup analyses took place to explore the differential impact of the intervention when pupils' FSM, EAL, and prior attainment were taken into consideration. Since the SAP was written, gender has also been included as a variable to consider, as this variable was considered in the previous trial. Data on gender and EAL were also obtained from the NPD and collected at the same time as attainment data.

Analysis was carried out using interaction models that were identical to the primary outcome model and included the variable of interest (FSM, EAL, prior attainment, and gender) and an interaction term between the variable of interest and the treatment group as additional covariates. All pupils were included in the model rather than FSM eligible pupils only.

All analyses were carried out using the same package as the primary outcome model.

Secondary analysis

Attainment measures: Overall there were three multilevel models (two levels; school and pupil) for the analysis of the secondary attainment outcomes; pupils' attainment in reading for the whole cohort, FSM eligible pupils' attainment in mathematics, and pupils' attainment in mathematics for the whole cohort. Each model consisted of the following covariates:

- The outcome as a dependent variable (described in the outcome measures section of the report)

- A variable that indicated the intervention group of a pupil (reference category was the control group)

- A variable that identified when the pupil was randomised (reference category was January)

- The baseline measure (described in the outcome measures section of the report)

Analyses of all secondary attainment outcome measures were carried out using the same package as the primary outcome model.

Non-attainment measures: The two non-attainment outcomes were analysed using two separate multilevel models (each with two levels; school and pupil). Each model consisted of the following dependent variable (treated as a categorical variable) and covariates:

- Pupil's response to the item at endpoint as a dependent variable (described in the outcome measures section of the report)

- A variable that indicated the intervention group of a pupil (reference category was the control group)

- A variable that identified when the pupil was randomised (reference category was January)

- Pupil's response to the item at baseline as a dependent variable (described in the outcome measures section of the report)

Analyses of all secondary non-attainment outcome measures were carried out using the ordinal (Christensen, 2019) package in R (R core team, 2019).

Analysis of the remaining 10 items and the scenario based question were carried out using cross tabulations and the results are presented in Appendix K.

Additional exploratory analysis

As mentioned earlier in the report, schools in England closed in mid-March 2020 as part of the Government's response to the Covid-19 pandemic. Summer 2020 KS2 tests were cancelled, therefore no outcome data would be available for the addendum cohort. The EEF, SAPERE and NFER discussed the possibility of using the 2021 cohort's KS2 data, but

agreed this would be problematic because schools may not have continued with usual P4C practice as they would have done under normal circumstances. The EEF, SAPERE and NFER therefore agreed to remove the addendum report from the project plan.

Instead, NFER carried out an exploratory analysis based on pupil survey data already collected in the study (including the two items used in the secondary analysis). The instrument consists of twelve items; two of which were analysed for the pre-specified secondary outcomes. A factor analysis of these items at baseline was carried out with a view of extracting a factor (or factors) that could then be used in an exploratory model with the follow-up data.

Factor analysis is a statistical technique that summarises information from a number of survey items into a smaller set of reliable outcome measures. It combines survey items that assess the same underlying latent construct by grouping together question items that have similar patterns of responses. This enables a more robust analysis. The factors derived through this analysis were used as outcome measures (factor scores at endpoint) in multi-level models to report the survey findings.

All survey questions were answered on a Likert scale (e.g. an 5-point scale with responses ranging from 'not true'-'completely true' scale). Any pupil that answered a third or fewer of the items entered into the factor analysis were removed from the analysis for the purpose of constructing the factors on a consistent set of responses.

Factors were selected that met the following criteria:

- internal consistency of each factor which indicates reliability (indicated by a Cronbach's Alpha statistic on a range from 0 to 1) (the threshold for inclusion was a reliability of greater than or equal to 0.5)

- loadings above 0.3 which indicate an association between items and the underlying factors. The relationship of each item to a factor is expressed by a factor loading. Factor loadings are similar to correlation coefficients – a higher value on a range from -1 to 1 indicates a stronger correlation with the factor

- Eigenvalues greater than 1 which indicate strong validity of the factors (the additional variance explained by bringing items together into a single factor)

- moderate levels of correlation between factors, indicating that each factor is measuring something slightly different.

The analysis of all twelve items in the survey produced two factors which explained 25.8% of the variation in the data, the results are presented in Table 6 and 7. Some statements did not fit the modelling of the factors, and so are not included in the factor measure ('I like meeting new people', 'I want to try and make my local area a better place', 'I like to be told exactly what to do', 'I am often afraid to try new things').

The first factor measured 'Teachers' caring approach' (Cronbach's Alpha=0.668, 2 items) and forms the first exploratory outcome measure. It relates conceptually to one of the 4Cs in P4C – developing a 'caring' approach to teaching and learning. The second factor forms another exploratory outcome measure and relates to a number of concepts within the theory of change (collaboration, resilience, problem-solving, confidence). It has been labelled 'A collaborative and confident approach to learning' (Cronbach's Alpha=0.568, 6 items)[7].

Factor scores were calculated by summing the constituent items for the baseline and the follow up surveys. These outcome measures were analysed using multi-level models, with endpoint scores forming the dependent variable. The baseline scores were included as covariates in the multi-level models. The endpoint scores formed the dependent variable in the multi-level models. The score distributions of these factors are presented in Appendix L. Factor scores are presented in the following tables (6 and 7) as scale means.

---

[7] Note that the order of the factors reflects the amount of variance they explain i.e. the first factor explains more variation in the constituent responses than the second. The first factor was comprised of two items. However, we deemed this to be acceptable as a two-item factor provides a more robust measure of a concept than two separate items.

Factor measure tables (development and results)

Table 6: Factor 1: Teachers' approach to caring

Factor 1: Teachers' approach to caring			
Cronbach's Alpha = 0.668 (2 items)			
Item-Total Statistics			
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted[8]
Teachers treat children fairly at my school	1.37	1.172	0.502	.
Teachers and other grown-ups at school care about me	1.30	1.335	0.502	.
This measure was scored on a scale of -4 to 4. All items had response options on a scale of 1 – 5, from not at all true to completely true. The mid-point was re-scaled to zero, so that each item was scored from -2 to 2.			
At baseline scores from -4 to 4 were observed; with a mean of 2.7			
At endpoint scores from -4 to 4 were observed; with a mean of 2.2			

Table 7: Factor 2: A collaborative and confident approach to learning

Factor 2: A collaborative and confident approach to learning			
Cronbach's Alpha = 0.568 (6 items)			
Itam – Total Statistics			
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
I am good at explaining my ideas to other people.	5.77	11.439	0.331	0.511
I can work with someone who has different opinions	5.53	11.120	0.299	0.527
I can do most things if I try	4.99	12.437	0.274	0.536
Once I have started a task I like to finish it	5.21	11.293	0.324	0.514
I try to understand other people's problems	5.21	11.465	0.314	0.519
I know where to go for help with a problem	4.92	11.955	0.301	0.525
This measure was scored on a scale of -12 to 12. All items had response options on a scale of 1 – 5, from not at all true to completely true. The mid-point was re-scaled to zero, so that each item was scored from -2 to 2.			
At baseline scores from -12 to 12 were observed; with a mean of 6.33			
At endpoint scores from -12 to 12 were observed; with a mean of 6.12			

In order to understand more about the proportion of students whose views have been impacted by the programme, we also ran an analysis of the distribution of the students' scores – whether they were on the negative side of the scale, indifferent, or positive, and the change in this distribution from baseline to follow-up. This was carried out for both factors. Results of this analysis is presented with plots of score distributions in Appendix L.

---

[8] In order to calculate Cronbach's alpha, a minimum of two items are needed.

Implementation and process evaluation

Research methods

The purpose of the implementation and process evaluation was to provide information on and insights into the delivery of P4C in schools under 'real world' conditions. The implementation and process evaluation had three main research questions, namely:

1. What works in implementing the P4C programme in schools (covering implementation factors, fidelity, and engagement of stakeholders)? What are the key factors that facilitate teachers' full engagement with, and use of, the P4C programme (for example, how does peer and management support, school and curriculum innovation, and leadership, equip teachers with the drive and direction to deliver the programme)?

2. How are schools using the P4C progress indicators (the Bronze, Silver Gold award criteria), and how useful are these indicators to schools? (This would provide insights into this new dimension9 of the P4C programme, by establishing how useful schools find the indicators for providing feedback on the progress made by pupils in their thinking and reasoning skills).

3. What is the capacity for a scaled-up P4C programme?

In addition to these main areas of focus, the IPE also covered perceived outcomes of P4C on staff and pupils.

In order to address these areas of investigation, the methodology design consisted of five strands and a mix of quantitative and qualitative methodologies in order to collect both a breadth and depth of evidence. The five strands were:

4. Qualitative telephone interviews with 16 senior leaders and/or P4C leads in intervention schools in November 2017.

5. In-depth case-study visits to five schools at the end of the academic year 2017/18 and a repeat visit to the same schools at the end of the academic year 2018/19.

6. A quantitative online survey of P4C leads and teachers who facilitate P4C sessions within intervention schools.

7. A meeting with SAPERE in autumn 2019 to explore sustainability and capacity for scale-up.

8. A short survey of control schools in summer 2019 to check what 'business as usual' entailed.

In Strand 1, experienced researchers from NFER conducted nine telephone semi-structured interviews with headteachers, three with deputy headteachers assigned the P4C lead role and four interviews with classroom teachers who were the designated P4C lead within the school. This totalled 16 interviews across 12 schools (see paragraph below for selection criteria). The purpose of the interviews was to investigate the reasons for becoming involved with the intervention, intensity of P4C delivery in schools, the approach to implementation, the role of the P4C lead, support and training, and the use of resources. These interviews were also used as a method for identifying schools to take part in the more in-depth case-study element of the IPE.

Strand 2 (case-study visits) built on the findings from Strand 1 by investigating implementation factors in greater depth, models of delivery, the knowledge and use of the Going for Gold award criteria (used as the measure of compliance), perceived outcomes and sustainability of the programme. Case-study schools were selected from the 12 schools involved in Strand 1. Selection criteria included geographical location (North England, South England and Midlands), and school context (size and FSM levels) . We aimed to get a mix of different locations, contexts and models used (such as either embedded in curriculum or standalone sessions) within the confines of visits to a small number of schools. The purpose was to visit schools in different situations to provide a breadth of experience, rather than to represent all school contexts and models within the schools in the intervention.

---

9 NB – these progress indicators were not in place during the previous trial.

Case-study visits included interviews with a range of staff within the school and a focus group with pupils. The numbers of interviews with members of school staff varied by institution due to the multiple roles held by members of staff within primary schools. This meant that often the P4C lead was also a member of the senior leadership team and/or a coordinator of literacy or numeracy. The selection of pupils' for the focus groups was opportunistic – we asked teachers to organise a mix of FSM and non-FSM children involved in the trial. These pupils therefore cannot be taken as fully representative of pupils involved in the trial. Table 8 below summarises the numbers of research activities undertaken in case-study phases 1 and 2:

Table 8. Numbers of interviews and focus groups completed in case-study schools

	Number of schools visited	Total number of staff members[10] interviewed	Total number of interviews	Observations of P4C sessions	Total number of focus groups	Total number of pupils in focus groups
Initial case-study visits 2017/18	5	13	12	2	4	29
Follow up case-study visits to the same schools 2018/19	5	10	8	1	4	25

Source: NFER IPE of P4C (2017 – 2019).

In the summer term of 2019 an online quantitative survey was sent to all intervention schools (see Appendix M). The survey was aimed at the P4C lead in schools and teachers who were facilitating P4C sessions within the school. The online survey was open for 6 weeks between May and June 2019. Intervention schools received a weekly reminder email and in the final week non-responding schools were called to encourage participation. The survey covered questions related to provision (which year groups the school has delivered P4C to), frequency of sessions, reach (how many pupils the programme reached), implementation fidelity, perceived outcomes of teachers and pupils, and training and support. The survey was routed so that some questions were only asked of the P4C Lead, particularly related to dosage and reach. In total, there was 221 responses to the survey from 56 of the 75 intervention schools. Fifty three P4C leads and 148 teachers responded. The section on IPE findings presents the results from this survey.

Control schools received a brief online survey in the summer term 2019 which asked them to report on any philosophy-based or enquiry-based programmes they had used during the trial period, and to check that they had not engaged in any P4C training or activities since the beginning of the trial (see Appendix N) This was to assess 'business as usual' in these schools. In total we received responses from 105 of the 123 control schools control schools.

In September 2019, a discussion was carried out with the management team at SAPERE to understand their future plans and the costs associated with P4C.

In addition, SAPERE provided NFER with school-level award data in July 2019 (using a pre-agreed template listing school ID, school name and then columns for award level – see Appendix O) – for the purposes of CACE analysis. SAPERE also provided NFER with updated award level data in July 2020 – despite school closures in mid-March 2020, SAPERE retained close contact with trainers and schools where possible in order to provide the relevant data for awards achieved up until mid-March 2020. The EEF, SAPERE and NFER agreed to use this 2020 data for descriptive purposes only, and not conduct any analysis on associations with outcomes or practice.

Costs

Cost data was collected and analysed in line with the then cost guidance provided by EEF (cost guidance, 2016), on direct and indirect costs to schools. We discussed the costs included in the Going for Gold model with SAPERE.

---

[10] Including a range of senior leaders, P4C leads, and literacy and numeracy coordinators.

SAPERE have a standard model for Going for Gold which comprehensively covers training, resources and support for the three years it is expected to take for a school to achieve the gold award. However, this model is flexible and adaptable and schools are able to remove items they do not feel they need and schools are able to share the package across schools so that the price per pupil is reduced. This is particularly useful for small primary schools. For the purposes of this evaluation, we have taken the standard cost of the Going for Gold model as our basis for the cost calculations, as this is the standard model offered to schools.

In addition to discussing costs with SAPERE, we also asked case-study schools about the indirect costs, if any, that were associated with running P4C. Interviews with the P4C leads in case-study schools also identified the amount of additional time teachers spent on P4C (such as time spent preparing for the sessions and identifying resources).

IPE Analysis

In the analysis of telephone and case-study interview data and survey data we examined responses by theme across all data collection methods, using both inductive and deductive techniques. Overarching themes were identified through the research questions and logic model, with sub-themes identified inductively after data collection. By analysing the data in this way we were able to combine the quantifiable findings for any given theme, identified through the survey, with the qualitative data from the telephone and case-study interviews, in order to provide a holistic understanding of each theme investigated. In order to test assumptions in the logic model, specific questions were added to both the survey and the qualitative interview schedules which were then analysed using a deductive approach to test the assumptions. This enabled us to provide both a broad and in-depth evidence-based account of how P4C is being delivered in schools within the trial, and to what extent the logic model and these findings aligned.

Table 9. IPE methods overview

Research methods	Data collection methods	Participants/ data sources	Data analysis methods	Research questions addressed	Implementation/ logic model relevance
Interviews with P4C leads	Telephone interviews	16 interviews in 12 schools	Thematic qualitative analysis	What works in implementing the P4C programme in schools (covering implementation factors, fidelity, and engagement of stakeholders).	Assumptions Facilitating factors
Case study visits to 5 schools (first visited in 2017/18 and visited again in 2018/19)	Qualitative interviews, pupil focus group, observation	20 interviews with 23 members of staff; 8 focus groups with 54 pupils	Thematic qualitative analysis. Comparative analysis of responses to common questions across participant groups between 2017/18 and 2018/19 data.	What works in implementing the P4C programme in schools (covering implementation factors, fidelity, and engagement of stakeholders). The use of the P4C progress indicators (the Bronze, Silver Gold award criteria), and how useful these indicators are to schools (covering compliance). Costs and time associated with implementing P4C in schools.	Assumptions Facilitating factors Impact
Survey of teachers facilitating P4C	Online survey	221 respondents from 56 schools	Descriptive statistics and cross-tabulations of results by role of respondent.	What works in implementing the P4C programme in schools (covering implementation factors, fidelity, and engagement of stakeholders). The use of the P4C progress indicators (the Bronze, Silver Gold award criteria), and how useful these indicators are to schools (covering compliance).	Assumptions Facilitating factors Impact

| Proforma of control schools | Online proforma | 105 control schools | Descriptive statistics | Contamination and business as usual | |
| Interview with SAPERE | Informal meeting | 2 participants | Thematic qualitative analysis | The capacity for scale up of P4C. Costs associated with running P4C. | |

Source: NFER IPE of P4C (2017 – 2019).

Timeline

Table 10: Timeline

Dates	Activity	Staff responsible / leading
October 2016	Draw school sample; prepare recruitment materials	NFER
October 2016	IDEA workshop	NFER and SAPERE
November – December 2016	Schools sign MoUs Schools share Year 4 pupil data with NFER	NFER
November – December 2016	Finalise theory of change	NFER and SAPERE
January 2017	Year 4 pupils complete social skills questionnaire (baseline)	NFER and schools
End January 2017	Randomise first batch of schools; inform schools and SAPERE	NFER
January – March 2017	Second batch of schools sign MoUs and provide Year 4 pupil data Year 4 pupils complete socials skills questionnaire (baseline)	NFER
End March 2017	Randomise second batch of schools, inform schools and SAPERE	NFER
February – September 2017	Intervention schools book training sessions SAPERE provides initial full day of P4C training for all participating staff	SAPERE
September 2017 – July 2019	P4C programme is implemented in whole school with weekly P4C sessions and further SAPERE training and support up to bronze and silver levels NFER carry out exploratory telephone interviews (autumn 2017) and case study visits (spring/summer 2018 and spring/summer 2019)	SAPERE
November – December 2017	Exploratory telephone interviews with schools	NFER
March – July 2018	Case study visits to schools Schools and parents informed post-GDPR	NFER
March – June 2019	Follow-up case study visits to schools	NFER
March – April 2019	Year 6 pupils complete social skills questionnaire (follow-up) (administered by NFER test administrators)	NFER
May 2019	Statutory Key Stage 2 reading and maths tests	Schools
May 2019	Teacher survey (including P4C leads)	NFER
July 2019	Control schools received financial incentive (which could be used to put towards P4C from Autumn 2019 onwards)	SAPERE
July – September 2019	School-level award data (e.g. bronze, silver, gold) collated by SAPERE and provided to NFER	SAPERE
October 2019 – May 2020 (extended from December 2019)	Analysis (including accessing NPD data via ONS/SRS) and draft reporting	NFER

Dates	Activity	Staff responsible / leading
September 2019 – July 2020	Intervention schools continue with P4C up to gold level Control schools take up Level 1 P4C training (if they wish), and deliver P4C to all year groups except Year 6	SAPERE, schools
May 2020[11]	Key Stage 2 reading and maths tests for addendum cohort - cancelled	Schools
July 2020	School-level award data for 2019/20 (e.g. bronze, silver, gold) collated by SAPERE and provided to NFER Additional factor analysis using pupil survey data	SAPERE
Summer/Autumn 2020	Reporting	NFER

---

[11] Due to school closures in relation to Covid-19, these tests were cancelled.

Impact evaluation results

Participant flow chart including losses and exclusions

Figures 1 and 2 present details of the participants' flow through each stage of the trial, where Figure 1 represents the flow of particpants for the primary outcome and Figure 2, the non-attainment secondary outcomes.

As mentioned in the selection and recuitment section of the report, NFER was responsible for the recruitment of schools. Initally 208 schools agreed to particpate, but 10 of these schools either did not complete the baseline questionnaire, pupil data file, or return their MoUs leaving 198 schools that were put forward to be randomised.

Schools provided administrative pupil data (pupil's name, DOB, UPN, SEN and EAL status) to NFER via a secure portal. This was used in the analysis of the social skills questionnaire (note, the analysis of the attainment outcomes required de-identifed pupil data which was accessed through the NPD).

NFER sent a list of administrative data from schools (URN, school name, randomisation group, and randomisation block) to DfE which was used to match in KS1 and KS2 attainment data. NFER rquested KS2 data for all pupils in the Year 6 2018/19 cohort from participating schools and asked the NPD team to match in their KS1 attainment data even if a pupil did not attend the same school. Overall, for the analysis of the primary outcome, 3601 FSM-eligible pupils' data were collected; 117 of these pupils had missing KS2 reading data and 126 had missing KS1 reading data , these pupils were excluded from analysis.

For the analysis of the non-attainment measures, twelve schools withdrew at the time of collection of the endpoint survey. Additionally, pupils were absent at endpoint or had left the school and these pupils were excluded from the analysis. A total of 7919 pupils completed the baseline survey. As mentioned earlier, only pupils that completed the baseline survey were asked to complete the endpoint survey12, this resulted in a total of 6232 pupils completing both surveys.

---

[12] As mentioned previously, spares were included for any Year 6s who had joined the school since Year 4, to aid administration. These were not analysed.

Figure 1: Participant flow diagram for the primary outcome (FSM eligible pupils' attainment in reading, 2 arms)
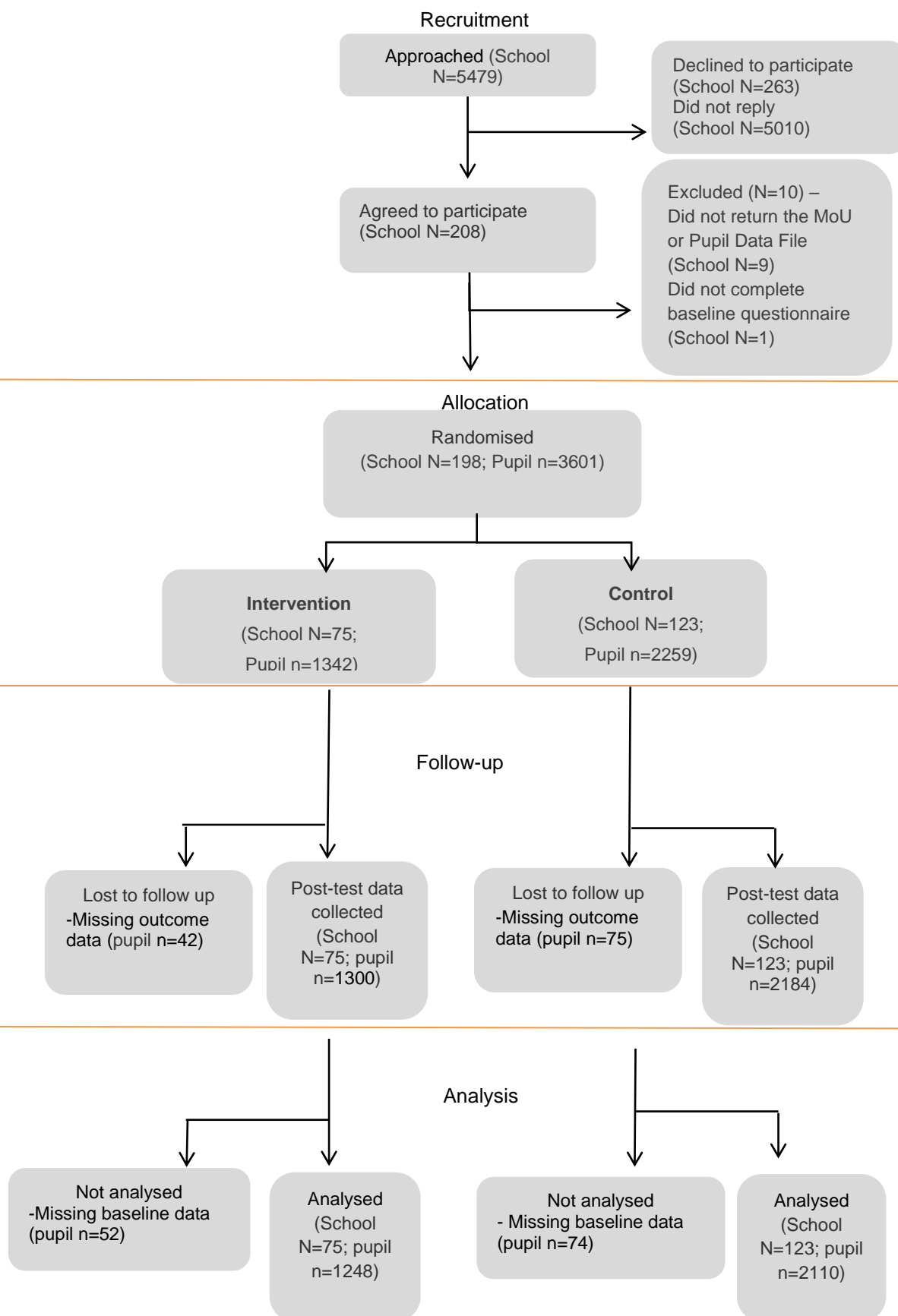
Source: NFER P4C randomisation data 2016/17, NFER pupil data 2016/7 and 2019/20

Loss to follow-up refers to pupil data being missing at the point of follow up (i.e. missing KS2 data). Not analysed pupils are those that did not have data for any of the covariates of the model ,as the multilevel models were carried out on complete cases (pupils with data for all covariates).

Figure 2: Participant flow diagram for the secondary non-attainment outcome (2 arms)

Item 1A: 'I am good at explaining my ideas to other people'
(social and communication skills and)
Item 1C: 'I can work with someone who has different opinions'
 (teamwork and resilience)

Recruitment

Approached (School N=5479)

Declined to participate (School N=263)
Did not reply (School N=5010)

Agreed to participate (School N=208)

Excluded –
Did not return the MoU or Pupil Data File (School N=9)
Did not complete baseline questionnaire (School N=1)

Allocation

Randomised (School N=198; Pupil n=7919)

Intervention (School N=75; Pupil n=3029)

Control (School N=123; Pupil n=4890)

Follow-up

Lost to follow up (Pupil n=675)
- School withdrawal (school N=8; Pupil n=253)
- School leavers after baseline (Pupil n=306)
- Absent at endpoint (Pupil n=116)

Post-test data collected (School N=67 Pupil n=2354)

Post-test data collected (School N=119; Pupil n=3878)

Lost to follow up (Pupil n=1012)
- School withdrawal (School N=4; Pupil n=182)
- School leavers after baseline (Pupil n=576)
 - Absent at endpoint (Pupil n=254)

Analysis

Not analysed Item 1A -Pupil n=37

Analysed Item 1A (School N=67; Pupil n=2317)

Not analysed Item 1C - Pupil n=57

Analysed Item 1C (School N=67; Pupil n=2297)

Analysed Item 1C (School N=119; Pupil n=3792)

Not analysed Item 1C - Pupil n=86

Analysed Item 1A (School N=119; Pupil n=3820)
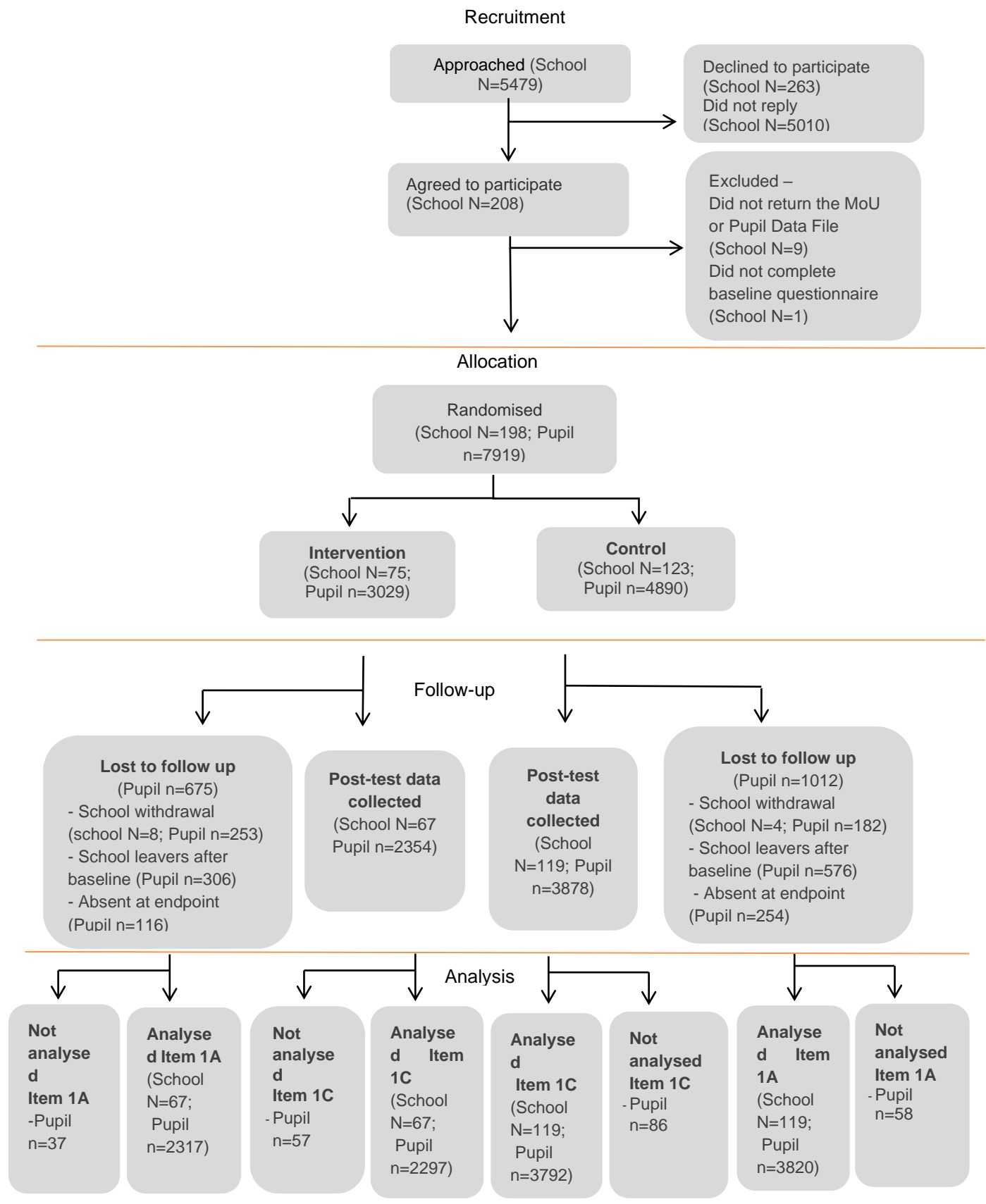
Not analysed Item 1A - Pupil n=58

Table 11 provides details of the minimum detectable effect size at different stages in the trial. The intra-cluster correlations (ICCs) and pre-test/post-test correlations at the protocol and randomisation stages were obtained from EEF's pre-testing paper (EEF, 2013). At the analysis stage, the pre-test/post-test correlation was calculated at the pupil-level when analysing the primary outcome. Calculation of the ICC at the analysis stage is covered in the methods section of the report.

The assumed parameter values were based on older key stage assessment systems. We have used KS1 and KS2 performance data at the analysis stage after the change in assessment in 2016 which resulted in lower correlation values leading to an increased MDES value of 0.15.

Table 11: Minimum detectable effect size at different stages

		Protocol	Randomisation	Analysis
MDES		0.125	0.125	0.15
Pre-test/post-test correlations		0.73	0.73	0.61
Intra-cluster correlations (ICCs)		0.137	0.137	0.17
Alpha		0.05	0.05	0.05
Power		0.8	0.8	0.8
One-sided or two-sided?		2	2	2
Average cluster size		14.38	15.56	18.19
Number of schools	Intervention	75	75	75
	Control	125	123	123
	Total:	200	198	198

Source: NFER sample size calculator

Attrition

In terms of attrition to measurement for the primary outcome, no schools were lost as data was accessed de-identified from NPD for their cohort of pupils. This was matched to a list of schools rather than pupils. Pupils were lost from the analysis as they had either missing baseline or outcome data (n = 243). Table 12 presents pupil-level attrition based on the numbers from the participant flow diagram. On average, 6.75% of pupils were lost from analysis. As for the secondary non-attainment analyses, 12 schools withdrew from the intervention.

Table 12: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised (NPD data)	1342	2259	3601
	Analysed	1248	2110	3358
Pupil attrition (from NPD data to analysis)	Number	94	149	243
	Percentage	7.00%	6.60%	6.75%

Source: NPD data 2019/20

Pupil and school characteristics

In total, 198 schools were involved in the trial. Table 13 below presents key baseline characteristics of the schools that formed the sample for the analysis of the primary outcome. Statistical tests were used to check any imbalance at baseline for analysed cases. As no schools dropped out, results of analysis to check balance at baseline is the same as at analysed for school-level variables.

The national- level mean for pupil-level variables were not calculated as data was obtained for pupils included in the trial, not all pupils nationally. Additionally, as we did not obtain pupil-level data at randomisation, we cannot say that results of analysis to check balance at baseline is the same as at analysed for pupil-level variables. However, if the randomisation is done correctly, there is no reason to suspect imbalance at baseline for randomised cases.

Results from these statistical tests showed that there was not enough evidence to suggest that there was a difference in baseline characteristics between the analysed groups.

In addition to this, we also calculated the baseline effect size using the KS1 data for analysed groups. As seen in the table, the effect size confidence interval straddle zero which suggests no evidence of a difference in KS1 scores of the two randomisation groups.

Table 13: Baseline characteristics of groups as analysed

School-level (categorical)	National-level mean	Intervention group		Control group		p-value[13]	
		n/N[14] (missing)	Count (%)	n/N (missing)	Count (%)		
School Governance Academy or Free School Maintained		21/75 (0) 54/75 (0)	28% 72%	29/123 (0) 94/123 (0)	23.6% 76.4%	0.49	
School type Primary Junior		66/75 (0) 9/75 (0)	88% 12%	106/123 (0) 17/123 (0)	86.2% 13.8%	0.71	
Urban or Rural Urban Rural		67/75 (1) 7/75 (1)	89.3% 9.3%	109/123 (3) 11/123 (3)	88.6% 8.9%	0.86	
Ofsted rating Outstanding Good Requires improvements Inadequate		7/75 (0) 58/75 (0) 9/75 (0) 1/75 (0)	9.3% 77.3% 12.0% 1.3%	11/123 (0) 97/123 (0) 12/123 (0) 3/123 (0)	8.9% 78.9% 9.8% 2.4%	0.91	
School-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)		
% FSM 2016/17	21.27%	75 (0)	36.00% (10.89)	123 (0)	36.19% (10.13)	0.85	
KS2 average scaled score in reading 2016/17	104.37	74 (1)	102.77 (2.81)	120 (3)	102.85 (2.8)	0.90	
Pupil-level (categorical)		n/N (missing)	Count (%)	n/N (missing)	Count (%)		
Eligible for FSM 2016/17[15]		1342/1342 (0)	1342 (100%)	2259/2259 (0)	2259 (100%)		
Pupil-level (continuous)		n/N (missing)	Mean (SD)	n/N (missing)	Mean (SD)		Effect size
KS1 reading point score		1290/1342 (52)	15.61 (3.9)	2185/2259 (74)	15.58 (3.93)	0.87	0.007 (-0.082,0.097)

Source: NFER's Register of Schools and NPD data.

---

[13] p-values were obtained from carrying out chi-sq tests, aside from KS1 which was analysed using a t-test.
[14] n refers to the sample size of a specific group (e.g., intervention schools) while N refers to the total sample size (e.g., all schools in the trial).
[15] This was calculated using the EVERFSM-6 variable obtained from the NPD which are pupils that are known to have been eligible for free school meals (FSM) on any pupil-level census in the last six years.

Outcomes and analysis

Primary analysis

Score distributions are presented for the primary outcome by treatment group and the overall distribution in Appendix I. As the outcome measures are standardised scores, the scores range from 80 to 120. The distributions appear slightly skewed, which can be seen by their bell-like shape[16].

Table 14 presents findings from the primary ITT analysis. As seen in the table, the effect size is 0.02 (-0.09, 0.13). As the confidence interval straddles zero, we cannot reject the null hypothesis. This means that the statistical evidence does not meet the threshold to conclude that the P4C programme has an impact on pupils' attainment in reading. We used ordinal regression to obtain the effect sizes for the non-attainment outcomes. Table 15 presents the parameters that were used in the estimation of the effect size. (The results for outcomes 1A and 1C are presented in Appendix K.)

Table 14: Primary and secondary analyses

| | Unadjusted means [awaiting QA and clearance] | | | | Effect size | | |
| | Intervention group | | Control group | | | | |
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges g (95% CI)	p-value
Year 6 Reading (FSM eligible pupils)	1288 (54)	102.01 (101.56,102.46)	2168 (91)	101.71 (101.35,102.07)	3358 (1248;2110)	0.02 (-0.09,0.13)	0.68
Year 6 Reading (all pupils)	3065 (97)	103.37 (103.08,103.66)	5058 (171)	103.2 (102.97,103.43)	7677 (2878;4799)	0.01 (-0.08,0.11)	0.81
Year 6 Maths (FSM eligible pupils)	1297 (45)	102.85 (102.44,103.26)	2176 (83)	102.36 (102.04,102.68)	3371 (1256;2115)	0.05 (-0.06,0.16)	0.38
Year 6 Maths (all pupils)	3081 (81)	104.49 (104.24,104.75)	5067 (162)	104.12 (103.92,104.33)	7694 (2890;4804)	0.04 (-0.06,0.14)	0.42
Item 1A: 'I am good at explaining my ideas to other people'	2343 (11)		3860 (18)		6137 (2137;3820)	-0.06 (-0.20,0.07)	0.35
Item 1C: 'I can work with someone who has different opinions'	2338 (16)		3842 (36)		6089 (2297;3792)	0.13 (-0.01,0.27)	0.07

Source: NPD data, 2019/20 and NFER pupil data 2019/20

---

[16] The central limit theorem tells us that the sampling distribution tends to be normal if the sample is large enough (n > 30).

Table 15: Effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Total variance from a model without covariates	Population variance (if available)
Year 6 Reading (FSM eligible pupils)	5.09	0.19	70.16	

Source: NPD data, 2019/20

Compliance Average Causal Effect (CACE)

As described in the methods section, SAPERE provided NFER with the award level that each school had achieved after two years of the intervention being underway in intervention schools. As described in the methods section, compliance related to the extent to which schools had engaged with, delivered and embedded P4C across the whole school. As set out in the protocol, the level SAPERE expected intervention schools to have reached by summer 2019 was the bronze award and to have met 50% of the criteria for the silver award[17] - i.e. Level 5 or above. The compliance data showed that 53% (40 of 75) achieved Level 5 or higher; and 47% (35 of 75) of all intervention schools achieved less than what was expected (i.e. Level 4 or lower). The majority of intervention schools received a bronze award (73%, 55 schools); but 27% (20 schools) did not receive an award at all. 12% of intervention schools received a silver award (nine of 75). Further details are included in Appendix J, and in the process evaluation results section.

As for control schools, according to SAPERE monitoring data, three schools received P4C training and participated in the programme at different engagement levels. One of these schoolshad no P4C sessions (level 0), one school was not actively working towards a level (level 1), and one school was working towards bronze (level 3). The process evaluation results section of this report provides further details.

Results from this analysis (shown in the table below) suggested that there was not enough statistical evidence to suggest that the level of engagement was associated with FSM eligible pupils' attainment in reading.

Table 16: Complier Average Causal Effect (CACE) analysis for the primary outcome

Compliance Measure	Total n (intervention; control)	Hedges g (95% CI)	p-value
Engagement level	3358 (1248; 2110)	-0.006 (-0.007, 0.02)	0.661

Secondary analyses

Attainment analyses

Similar to the analysis of the primary outcome, we present the score distributions for the primary outcome by treatment group and the overall distribution in Appendix I. As the outcome measures are standardised scores, the scores range from 80 to 120. The distributions appear normal, which can be seen by their bell-like shape.

---

[17] See Table 5. The section on Compliance in the Implementation and Process evaluation findings provides further details of the award levels in practice (i.e. what a school has to achieve/demonstrate for each of the levels).

Effect sizes and confidence intervals for all these models are also presented in Table 14. As the confidence intervals straddle zero for all models, the statistical evidence does not meet the threshold to conclude that the P4C programme has an impact on pupils' attainment in reading and maths.

Non-attainment analyses

The impact of the P4C programme on pupils' 'social skills and communication' as measured by the survey item 'I am good at explaining my ideas to others', and on 'teamwork and resilience' as measured by the survey item 'I can work with someone who has different opinions', was investigated as outlined in the methods section. Bar charts are presented in Appendix I.

Effect sizes and confidence intervals for these models are also presented in Table 14. As the confidence intervals straddle zero, the statistical evidence does not meet the threshold to conclude that the P4C programme has an impact on pupils' social skills and communication, as measured by item 1A. The effect size for the social skills and communication analysis was -0.06 (p-value= 0.39), and for the teamwork and resilience analysis 0.13 (p-value= 0.07).

Sub-group analyses

Several subgroup analyses to explore the differential impact of the intervention when variables of interest were taken into consideration were conducted, as mentioned in the methods section of the report. Results from the interaction models are summarised in Table 17. In these models, the variables of interest were interacted with the intervention term respectively. We ran four separate models for each variable of interest to explore the differential impact of the intervention.

These results suggest that the intervention did not have a statistically significant differential effect on reading attainment when pupil everFSM status, gender, EAL are taken into account. This means the intervention did not have a differential impact for pupils with everFSM status compared to those who were not everFSM, boys compared to girls, and pupils with EAL status compared to those that are not EAL. The effect of the P4C programme was differential for different levels of prior attainment. There was a greater effect for lower attaining pupils as compared to higher attainers.

Table 17: Results of interaction models

Outcome	Variable of interest	Raw interaction coefficient	Standard error	p-value
Year 6 Reading	Eligible for FSM	0.32	0.29	0.28
Year 6 Reading	Gender	-0.09	0.28	0.76
Year 6 Reading	EAL	-0.64	0.39	0.10
Year 6 Reading	KS1 attainment	-0.012	0.04	0.003

Source: NPD data, 2019/20

Additional exploratory analysis results

Table 18 presents the findings from the exploratory analysis of the two 'factor measure' outcomes and the factor score using results from the follow-up survey adjusting for baseline scores. This analysis was undertaken further to the SAP and with agreement from the EEF and SAPERE, as exploratory analysis. It compares the findings for the intervention group with the control group (Appendix L provides details of the scoring system for each of the outcome measures and the factor score). It presents the results of the analysis from multi-level modelling in terms of an effect size and p-value.

Table 18: Results of multi-level models of factor scores

Outcome	Raw Means				Effect Sizes		
	Intervention		Control		N in model (intervention; Control)	Effect Size (95% Confidence Intervals)	p-value
	N (missing)	Mean	N (missing)	Mean			
Factor 1							
Teachers' approach to caring	2348(5)	2.23	3869 (2)	2.17	6210 (2343, 3867)	-0.03 (-0.149, 0.09)	0.62
Factor 2							
A collaborative and confident approach to learning	2348(0)	6.14	3869(0)	6.11	6217 (2348, 3869)	-0.038 (-0.125,0.049)	0.39

As shown in Table 18, changes in outcomes relating to pupils' self-reported responses that made up teachers' approach to caring and a collaborative and confident approach to learning were not statistically significant, i.e. the differences between the intervention and control group were likely to be due to chance rather than the programme.

Results of the analysis of the distribution scores are presented in Tables 19 and 20.

The distributions of factor scores at baseline and follow up for the control and intervention pupils illustrate the general shift in attitudes for the intervention and control groups. A negative total score indicates that a pupil has an overall negative attitude for a factor. Zero suggests a neutral attitude and a value greater than zero suggest a positive attitude overall. The pattern of attitudes varied for the two measures but both factors represented no effect of the intervention. As shown in Table 19, there was a shift in students' responses to teachers' approach to caring – interestingly a negative shift over time for both the intervention group and control group. For this outcome, 79.5 per cent of the intervention group have positive scores at follow up compared to 84.8 per cent at baseline; meaning that around 5.3 per cent more intervention pupils no longer have a positive rating on teachers' approach to caring. The equivalent change for the control students is 7.9 per cent in the same direction18.

For 'A collaborative and confident approach to learning', there were very minor changes in the distribution of scores amongst the control group for this factor (see Table 20). It is worth noting that there was a very high percentage of pupils with a positive rating at baseline and so there wasn't much room for improvement.

Table 19: Distribution of factor scores: Teachers' approach to caring

% students	Raw Means						
	Intervention				Control		
	Baseline	Follow-up	Change		Baseline	Follow-up	Change
Negative attitude (- 4 to -1)	8.2%	11.40%	3.2%		7.1%	12.00%	4.9%
Indifferent (0)	6.9%	9.10%	2.2%		7.1%	10.10%	3.0%
Positive attitude (1 to 4)	84.8%	79.50%	-5.3%		85.8%	77.90%	-7.9%
Total	100%	100%			100%	100%	
N	2343	2343			3867	3867	

---

[18] This includes pupils that have become less positive in their perception of teachers' approach to caring over time but still remain positive at endpoint.

Table 20: Distribution of factor scores: A collaborative and confident approach to learning

% students	Raw Means					
	Intervention				Control	
	Baseline	Follow-up	Change	Baseline	Follow-up	Change
Negative attitude (-12 to -1)	6.0%	4.60%	-1.4%	5.30%	5.30%	0.0%
Indifferent (0)	2.8%	2.20%	-0.6%	3.00%	2.60%	-0.4%
Positive attitude (1 to 12)	91.2%	93.20%	2.0%	91.70%	92.10%	0.4%
Total	100%	100%		100%	100%	
N	2343	2343		3867	3867	

Implementation and process evaluation results

IPE summary

Just over half of the schools (40 of 75) were implementing P4C at or above the expected level after two years, a substantial minority (35) were working below the expected level (i.e. Level 5, bronze award achieved and working at 50% silver or above) . That said, P4C leads and teaching staff were very positive about the approach, and felt their pupils were enjoying P4C sessions. Almost two thirds (28 of 45) of P4C leads responding to the survey agreed/strongly agreed with the statement 'teachers have fully embraced facilitating P4C sessions'; and around three-quarters of responding P4C leads (34 of 45) agreed/strongly agreed that their teachers valued the P4C approach. The vast majority of teaching staff who facilitated P4C believed that pupils liked the P4C approach (94% of teacher survey respondents agreed/strongly agreed with this; and 88% agreed/strongly agreed with the statement 'pupils are fully engaged with the P4C sessions'.

In the early stages of delivery most interviewees across case studies and telephone interviews were running standalone P4C sessions. This was seen to be vital in order for teachers to gain confidence in delivery of P4C sessions. By 2019, most case-study schools had moved to a more flexible approach to P4C delivery, with standalone P4C sessions running on an ad-hoc basis at least once a month, but more often where possible, and sessions being linked to curriculum topics or themes being taught.

Whilst most intervention schools were applying for their Going for Gold awards, there appeared to be limited awareness amongst classroom teachers of using the P4C indicators to monitor progress and provide feedback to staff and pupils. The survey showed that the P4C leads were aware of the award level indicators. Two of the case study school P4C leads were actively using the awards criteria as progress indicators. However, classroom teachers facilitating P4C were generally not aware of these indicators. This aspect of the programme may require strengthening.

The training and support received from SAPERE was an important facilitating factor for schools. Trainers were reported as being knowledgeable and supportive, often adapting training requirements to suit the needs of teachers and their pupils within the case-study schools. Most interviewees felt that they would not have been able to progress with P4C without the ongoing support of their external trainer.

The case study data suggests that senior leader support is vital for successful implementation of P4C in schools. They need to spearhead the programme to give it urgency and importance. In case study schools, where headteachers valued the P4C model, had attended at least the Level 1 training and kept P4C a priority within school, teachers were seen to value P4C more.

The P4C lead provides valuable support and momentum for the programme within the school. It is particularly worth noting that there were P4C lead changes in 39 of the 75 intervention schools throughout the course of the trial.

Both pupils and teaching staff enjoyed and valued P4C. Pupils particularly enjoyed being able to share and learn from others' opinions in a non-judgemental environment. Teachers felt P4C had positively impacted on a range of pupils' social competencies, but most notably their respect for other pupils' opinions (96% of teachers who responded to this in the teacher survey agreed/strongly agreed with this), their ability to question and reason (91% of teachers agreed/strongly agreed with this) and their ability to express views clearly (93% of teachers agreed/strongly agreed with this).

Teachers felt that P4C had particularly had a positive impact on their professional development (84% of teachers who responded to this in the teacher survey agreed/strongly agreed with this) and their confidence in trying new teaching approaches (84% of teachers agreed/strongly agreed with this).

Overall P4C leads felt that P4C was having a positive impact on their whole school (47 of 53 P4C leads agreed/strongly agreed with this statement). However, there had been very little engagement with parents surrounding the use of P4C (this is an expected part of the programme, particularly in order to gain higher levels of award).

Compliance

Compliance was measured by assessing how far intervention schools had progressed through the 'Going for Gold' programme by the end of Summer term 2019. SAPERE progress indicators are designed to help schools progress through the Bronze, Silver and Gold awards, and cover pupil, teachers and school-level criteria in the following areas:

- Pupils: Building a community of philosophical enquiry; questioning and commenting; development of philosophical thinking; and review and evaluation.

- Teachers: classroom practice; facilitation skills; planning; P4C leadership role; and review and evaluation.

- School: Headteacher and SLT commitment; level of whole school training; involving the whole school; and review and evaluation.

- An example of how this criteria is manifested at the different levels of Bronze, Silver and Gold is given below for teachers' classroom practice[19]:

- Bronze: A minimum of 20% of school staff have facilitated a minimum of six P4C sessions over an academic year. Classroom displays show evidence of P4C.

- Silver: Consistent P4C sessions are planned into the timetable across a key stage or subject area. The P4C methodology and approach is becoming evident in teaching and learning across other areas of the curriculum.

- Gold: Most P4C practitioners show sustained and embedded P4C practice through consistent planning of P4C sessions. The P4C methodology and enquiry approach is explicitly used in teaching and learning in subject activities. P4C practice is communicated to parents and/or the wider school community.

As per the protocol, to ensure all intervention schools had an assessment, those schools who had not yet submitted an application for an award were assessed against the progress indicators by a SAPERE trainer. This approach ensured all schools had a level of progress recorded, even if they had not completed the submission application. SAPERE's trainers work closely with schools so that they know what level each school is working at in terms of progress, even if the school has not completed an award application. The levels of progress used to assess compliance were:

0 - no P4C sessions/withdrew before starting

1 - Not actively working towards a level

2 - Below bronze but some P4C activity and/or P4C activity by a few staff only

3 - Working towards Bronze

4 - Bronze, but not at 50% Silver

5 - Bronze + 50% silver or above

6 - Silver, but not at 50% Gold

7 - Silver +50% Gold, or above.

By the summer term of 2019, SAPERE intervention schools were expected to have done the Bronze award and met 50% of the criteria for the Silver award (specifically 50% of the pupil criteria, 50% of the teacher criteria and 50% of the schools criteria) (i.e. Level 5). By summer 2020 it was expected that schools would have have done the Silver

---

[19] The full criteria is available online at
https://www.sapere.org.uk/Content/Media/P4C%20School%20Award%20Criteria%202018%20for%20website.pdf

award and met 50% of the criteria for the Gold award (i.e. Level 7).Information on compliance up to July 2019 is provided below. Information on compliance for the academic year 2019/20 is provided in a section on Further implementation (2019/20) later in the report.

The analysis revealed that 40 intervention schools had achieved the desired level of progress or above (level 5 or above), whilst 35 schools were working below this level (see Figure 3).

Figure 3: Numbers of intervention schools working at each level of progress.
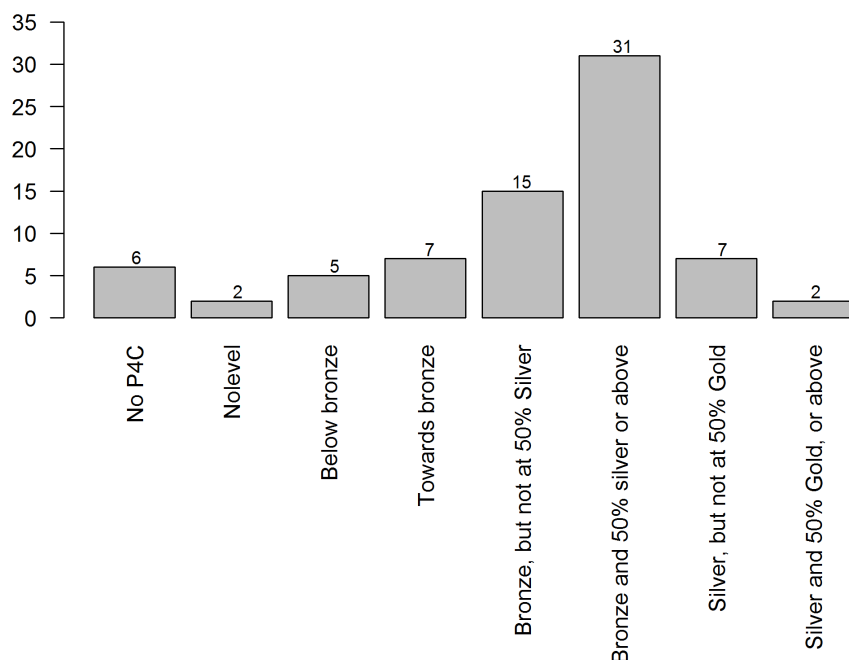
Source: SAPERE award dataset provided to NFER October 2019.

Six of the intervention schools had not delivered any sessions or had withdrawn before any P4C activities had been delivered and a further 2 schools were not working towards a level. Fifteen schools had achieved the Bronze award but had been judged to have not reached the 50% towards Silver criteria.

Self-reported data on progress towards awards was also sought in the teacher survey – the results are discussed in the section on 'How are schools measuring progress?' as part of Fidelity below.

So, whilst the majority of schools were implementing P4C at or above the expected level after two years, a substantial minority were working below the expected level (although 15 schools of these schools had reached bronze). According to case study data, reasons for this would seem to be the need to embed P4C across all year groups, with it taking schools longer than two years to become familiar and confident with, use and embed a new approach across the whole school.

Fidelity

For this study, fidelity focused on issues of training and implementation. At the fundamental level this entailed exploring whether staff were trained as intended and whether the programme was implemented as intended – as a minimum a one-hour session every two weeks to the whole class and gradually rolling out across the whole school. As described in the section on Compliance above, the extent to which the approach is embedded across the whole school is related to practice, rather than a dose per pupil, including areas such as extent of training amongst staff, whether P4C has been planned into the timetable, alignment with the curriculum, and so on. The extent to which the approach is embedded is related to the levels that are awarded.

What is the model of delivery of P4C within schools?

P4C is designed to be a whole-school approach suitable for both KS1 and KS2 children and this was reflected in the survey results. Within the survey we asked schools which year groups had received P4C sessions for each of the

academic years of the trial. Sixty eight per cent (36 of the 53) of P4C leads who responded, stated that they delivered P4C sessions to all year groups in their school in 2018/2019. This increased from 21 per cent (11 of 53) in the first year of the trial, suggesting schools are moving towards embedding P4C across their school. For those respondents who were not delivering P4C across all year groups, there was a tendency for delivery to be focused more in KS2 year groups compared with reception and KS1.

Telephone interviews with senior leaders in 2017, revealed that 11 out of 12 schools had intended to run P4C sessions on a weekly basis with some of their year groups, reflecting the recommendation of SAPERE that P4C be delivered weekly or at least fortnightly. By the second year of the trial, the survey results indicated that this was indeed happening in the majority of schools across KS2 year groups. Indeed, 69 per cent of P4C leads whose school was delivering P4C to KS2 pupils in 2018/19 (33 of 48) reported that they provided P4C sessions at least weekly (29 per cent, 14 of 48) or at least fortnightly (40 per cent, 19 of 48) to KS2 classes. In comparison, those who were delivering to KS1 pupils were generally running sessions less frequently, with these sessions typically being run either a few times a term (39 per cent of those delivering in KS1, or 15 of 39 respondents) or at least fortnightly (36 per cent, 14 of 39 respondents).

It was evident from all five case-study schools that P4C sessions tended to last for up to an hour in KS2 year groups, reflecting the recommendation from SAPERE. However, for KS1 pupils, most teachers interviewed were running shorter sessions of up to 30 minutes. These teachers explained that they felt the shorter 30-minute sessions were a suitable length of time for the younger children to engage with the P4C process. In comparison, most teachers interviewed explained that the older children were comfortably able to engage with a P4C enquiry for an hour. Indeed, case-study revisits in 2019 revealed that KS2 year groups can benefit from running an enquiry over two separate sessions as this provided them with more time to explore the topic. P4C leads explained that this had helped allow time to effectively progress through the 10-step process of an enquiry.

All case-study schools were delivering P4C as a whole-class activity. In a minority of instances, case-study interviewees identified occasions when it was considered more beneficial to divide the class into two groups, for example groups who struggle with language at sentence level (some younger children, EAL or SEN pupils). However, this was not regular practice.

In the early stages of delivery most interviewees across case studies and telephone interviews were running standalone P4C sessions. This was seen to be vital in order to assign time for teachers to engage with the programme, the 10-step process and for teachers to gain confidence in delivery of P4C sessions. However, the case-study visits in 2019 found that, while all of the P4C leads were keen to run weekly P4C sessions, they generally felt that this weekly commitment to standalone sessions was unsustainable in the longer term due to other curriculum demands on the timetable.

By 2019, most case-study schools had moved to a more flexible approach to P4C delivery, with standalone P4C sessions running on an ad-hoc basis at least once a month, but more often where possible. Most teachers interviewed explained that, over the trial period, they had become more comfortable with the processes, had come to value the programme and the benefits it brought to pupils' level of engagement with a P4C topic and as such were incorporating P4C practice into curriculum time, such as using P4C sessions as a means of introducing a new topic in art, for example. This flexibility helped overcome the challenge of timetable demands for standalone P4C sessions. This was considered by the majority of teachers interviewed in case-study schools to be pivotal to the programmes' success and to encourage sustainability and teacher commitment. This fits well with the theory of change that assumes P4C success can be achieved if senior leaders and teachers are committed to its delivery and see it as part of their pedagogy.

What do P4C lessons look like?

While a designated P4C area in school was considered beneficial, interviewees across all case-study visits felt that this was not essential to success. The majority of case-study schools delivered P4C sessions within the teachers' own classrooms. Most teachers interviewed during case-study visits explained that it was imperative to pupils' understanding of the parameters of discussion to make it explicit to pupils when a P4C session was taking place.

In case-study schools, P4C was predominantly delivered by class teachers with the option of Teaching Assistants (TAs) supporting the session. However, there were examples reported by interviewees of Higher-Level Teaching Assistants (HLTAs) delivering P4C sessions to support the class teacher, for example where a class teacher required

time away from class for planning or to deliver other senior-level responsibilities. Note that HLTAs could receive training at INSET days.

P4C sessions observed as part of the case studies typically began with pupils moving tables to the outside edge of the classroom. This was seen to be a quick process. Chairs were then typically arranged in a circle in the centre of the room. The session begins with an introductory game. Teachers interviewed as part of the case studies explained that the game inspiration came from the training that they had attended, or ideas shared by their trainer or with colleagues. After the introductory game, the teacher highlighted the P4C rules (often already displayed on a classroom wall) and then used a stimulus to introduce a topic. The general topic area was typically decided by the teacher. Most teachers interviewed in the first year of case studies explained that they chose topic areas that reflected current affairs, topics from PSHE sessions or ideas collated from their P4C advisors or teaching colleagues. In the second year of case studies, most schools had incorporated P4C into curriculum time and as such the P4C enquiry was dictated by the subject or curriculum topic area. One example of how this is done is given below by a teacher who embedded P4C into their Art topic area:

> 'I've used it really well in art – we would look at an artist and a painting they have done. The idea is that they will create something in the style of that artist. We will look at one or two of their paintings first and run a P4C [session] about that painting – what do you think they are trying to say? What's the mood? What does it make you feel? They've then been more interested in the art, the style of art, and what they've produced in the end has been more in that style because they've understood what the artist was trying to say.'

Below is another example of how P4C has been used in science:

> 'In a recent Science lesson on evolution focussing on how intelligence has developed over time, the class ended up having a conversation about whether humans are born with intelligence or whether it is dependent on environment (nature versus nurture). From that the children developed their own P4C question: "What's the best quality to have – intelligence or wisdom?" This then became the question for the class's next P4C enquiry.'

Children then developed their own questions based on the stimulus. The children then discussd which of their questions they wanted to take forward to discuss more fully. The teacher's role was to facilitate this discussion. This is different to normal teaching practice in which the teacher would normally pose the question to the children to discuss and the teacher would lead this. During discussion, teachers used various ways to encourage every child to talk, for example, only pupils holding a particular item such as a ball were allowed to talk. Pupils also used a variety of hand signals to show agreement or disagreement with a statement. This was seen to be a way of including those pupils who might find verbal agreement or disagreement uncomfortable. Observed P4C sessions ended with a review of the P4C rules, reflecting on how well they worked during the enquiry.

In case-study schools where P4C sessions were observed, the teacher facilitating the discussion played an important role in encouraging pupils to follow the P4C rules, facilitating pupil thought, encouraging pupils to explain their thinking and encouraging pupils to consider their own opinions. Teachers appeared confident in facilitating the discussion. When interviewed, these teachers expressed gratitude to their trainers and the support that they had received over the three-year trial period in order to increase their confidence with P4C delivery. They perceived the training and support from their trainer to be imperative to this confidence.

How are schools measuring progress?

As part of the trial, intervention schools were meant to start moving through the progress indicators/awards from Bronze through to Gold, as written in the memorandum of understanding. Within the survey, P4C leads showed awareness of the progress awards: when asked what awards they were working towards, ten of 53 respondents said they were working towards Bronze; almost three-quarters of respondents (38 of 53) said that they had achieved Bronze award, with a further 19 per cent (10 of 53) stating that they were working towards this level. Two-thirds (35 of 53) said they were working towards Silver; with 1 P4C lead stating they had achieved this award. Four of the 53 responding P4C leads said they were working towards Gold.

When asked directly about implementation fidelity, eight per cent of teachers facilitating P4C sessions (17 of 213) strongly agreed that they had delivered P4C exactly as intended while a further 57 per cent (122 of 213) agreed with this. However 29% neither agreed nor disagreed with this (61 of 213) and six per cent actively disagreed (12 of 213).

However, while P4C leads were aware of these awards, it was evident from the case-study schools that the wider teaching staff had limited awareness and use of the progress indicators. Only two of the case-study school P4C leads were actively using the awards criteria as progress indicators. In these two cases they explained that these indicators had helped identify gaps, had been useful in identifying the next steps required in order to prevent stagnation and, more importantly to them, in order to achieve the Gold, Silver or Bronze status awards. One P4C lead commented: '*P4C is reasonably well established, so what do we do next to make it have a bigger impact? Otherwise, it would stagnate slightly.*'

As described above, over the trial period, most case-study schools had moved from standalone P4C sessions to a more flexible approach delivering P4C within curriculum time. However, this made it more difficult to monitor how often teachers were delivering P4C on a regular basis. As such, formal monitoring of P4C delivery was very limited in the case-study schools. In these schools, where monitoring was taking place, it was on an ad-hoc basis and usually took the form of verbal feedback and observations from learning walks by the P4C lead.

What are the barriers to delivery?

Respondents involved in all aspects of data collection activities were largely extremely positive about P4C and when asked were unable to identify many barriers to delivery. Of those who were able to identify barriers, the most common barrier was the initial lack of confidence of some teaching staff in delivering P4C. In particular, these teachers were seen to struggle to adapt to the child-led features of the delivery model. However, these interviewees explained that they had supported their staff to develop their skills and confidence in P4C through session observations and feedback and talking through techniques (this was often done with the remote support of the external trainer, through their allocation of additional support days).This highlights the important role of both the external trainer and internal support from the P4C lead to support staff through the early stages of P4C implementation. This also supports the theory of change that states the importance of appointing a P4C lead to champion P4C in the school.

During initial telephone interviews with senior staff members at the beginning of the trial period in 2017, most interviewees noted some level of scepticism from their staff about the thought of a new initiative. It was made clear by all senior staff interviewed, however, that these barriers were reportedly eradicated early on, through informative Level 1 training that was said to inspire and motivate staff and increase enthusiasm for the P4C model.

The other barrier mentioned by some interviewees in the case-study schools related to apprehension of sustainability. In particular, a few interviewees in a minority of case-study schools were concerned about how to incorporate P4C sessions into curriculum areas, in order to relieve timetabling pressures caused by running P4C as standalone sessions.

Implementation factors: What are the necessary conditions for success of P4C?

Training

The feedback from case-study interviewees aligned well with SAPERE's training strategy. All case-study school teaching staff had received Level 1 training at the time of the first case-study visits. Training was typically delivered during whole INSET days by an external trainer. Follow-up twilight sessions were provided where needed or requested by the school. Without exception, case-study school interviewees reported training as valuable and practical and often reported it as being inspiring.

Similarly, all P4C lead survey respondents reported that 25 staff members had accessed Level 1 training as well as the Tools for Thinking Together Training. Of the P4C leads surveyed, 79 per cent (42 of 53) had completed training up to Advanced Level 2A. However, the survey also revealed that none of the responding P4C leads had completed Advanced Level 3 and Qualified P4C trainer levels.

Survey respondents were very positive about the training they had received. Nearly all (92 per cent) P4C leads either agreed (47 per cent) or strongly agreed (45 per cent) that the P4C training had enabled them to effectively fulfil the role of P4C lead teacher. While of the teachers surveyed, 92 per cent agreed (61 per cent) or strongly agreed (31 per cent) that the P4C training had enabled them to facilitate P4C sessions effectively.

The theory of change assumes that approved trainers deliver an on-going programme of support to schools. Interviews with teaching staff, P4C Leads and senior staff identified the important role of good-quality training and support from external trainers. Case-study interviewees continually praised the training and support that they had received. Trainers

were reported as being knowledgeable and supportive, often adapting training requirements to suit the needs of teachers and their pupils within the case-study schools. Indeed, most interviewees (both P4C leads and other teachers facilitating sessions) explained that they would not have been able to progress with P4C without the ongoing support of their external trainer.

Support from trainers can ensure practical advice is given to improve the way teachers facilitate discussions. Clear from the survey data, respondents reported confidence with the P4C model, as highlighted in figure 4 below:

Figure 4: Teachers' confidence with the P4C model
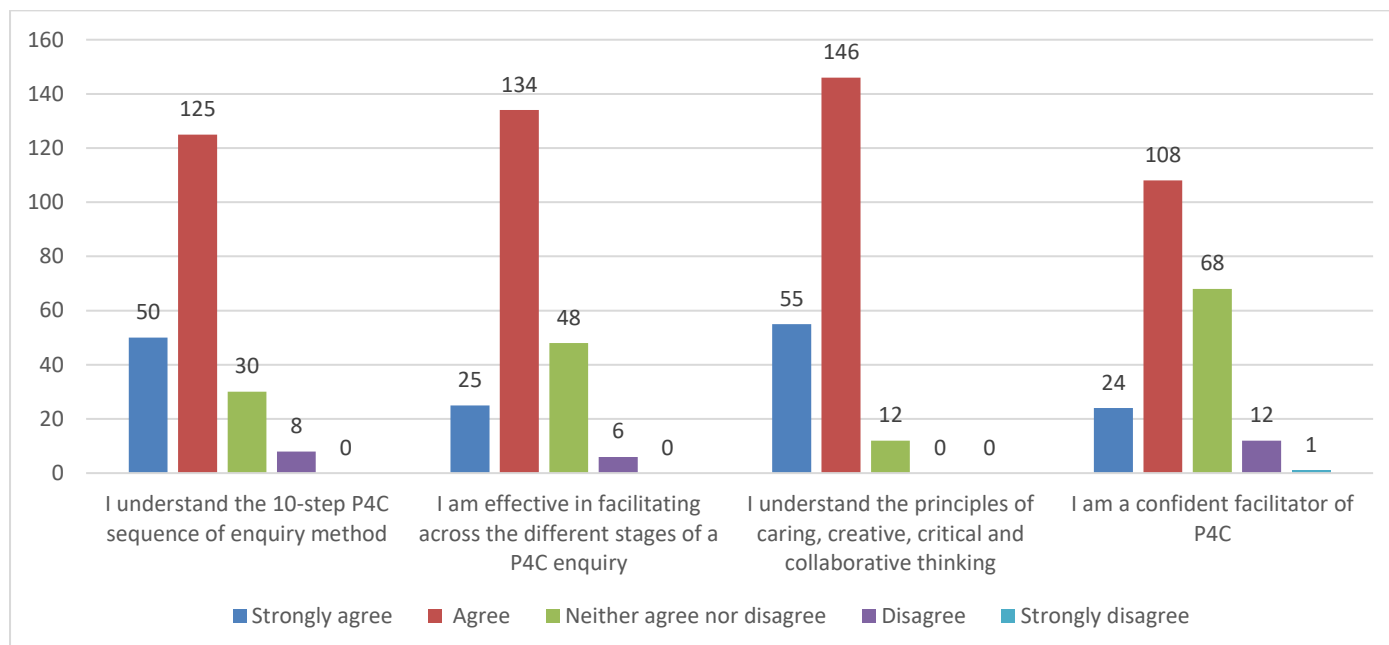
Source: NFER P4C intervention staff survey, 2019.

The majority of those surveyed (82 per cent, 175 of 213) agreed or strongly agreed that they understood the 10-step P4C sequence of enquiry method, whilst 75 per cent (159 of 213) reported that they felt effective in facilitating across the different stages of a P4C enquiry and a positive 94 per cent (201 of 213) stating that they understood the principles of the 4Cs. In addition, a total of 62 per cent of those surveyed (132 of 213) felt that they were a confident facilitator of P4C. Conversely, the importance of access to training was very evident in one of the case-study schools in which staff turnover was high, including senior staff. In this school it was reported by the P4C lead that a lack of training for new staff had meant that P4C understanding was low, which had resulted in the enthusiasm for the programme being depleted over time.

Case-study schools were positive about the training they had received, for example one P4C lead commented: '*The training has been very useful. […] Staff have said that it's been one of the better CPD sessions they've had.*'. During discussions with case-study staff, interviewees reflected on the training that they had received and what worked well for them. They explained that training worked well when trainers understand the curriculum, when the aims of the training were clear, when trainers included practical examples in their training and when there were opportunities for teachers to observe a trainer leading an enquiry. Teachers stated that they preferred the practical aspects of training to the theoretical content.

Staff often felt that their success with P4C was related to the supportive training they had received, which they could revisit through the ongoing support of the trainer and the P4C lead. It was most helpful to have Level 1 training in the same term in which P4C was going to be delivered to ensure that staff enthusiasm was maintained and then a spot-check later on in the term by the trainer to reaffirm the delivery processes are in place. This fits well with the SAPERE assumption that an ongoing programme of support is necessary to achieve P4C success.

Ongoing support and resources

Survey respondents reported that they had received between three and eight days' support from their external trainer. Of the 50 P4C leads surveyed that had received in-school coaching and support from a trainer, 72 per cent (36 of 50)

found it very useful with a further 26 per cent (13 respondents) finding it quite useful. In total, 64 per cent (34 of 53) of P4C leads surveyed had received remote P4C administration and planning support and of these 47 per cent (16 of 34) had found this very useful while a further 44 per cent (15 respondents) had found this quite useful.

Similarly, all P4C leads in the case-study schools explained that they had valued the ongoing support from their trainer. Interviewees particularly valued the trainers providing useful resources and signposting teaching staff and P4C leads to useful stimuli and resources online; modelling P4C sessions in school for teaching staff; and supporting facilitators in incorporating P4C into curriculum areas. Examples of comments included: '*I'm in contact with* [trainer] *whenever I need resources or help or ideas. She's given us wads and wads of lesson plans, pictures, book titles, training pamphlets, and questions prompts. We've had all the resources we need to do it and I can call her whenever I want.'.* It was explained that whilst it was immediately apparent to case-study school teachers how to link P4C to guided reading and PSHE, support from trainers exposed teaching staff to how it can be applied in other topic areas, such as maths and history. Interviewees in case-study schools explained how this ensured that P4C became part of curriculum areas and would help P4C become embedded in practice long term.

All P4C trial schools had unlimited access to SAPERE's online P4C resources and practice guides. Case-study interviewees frequently referred to the SAPERE's website as a useful resource of support, especially during the early stages of implementation. In total, 91 per cent of P4C leads surveyed (48 of 53) said that they had accessed the online P4C resources and guides. Of these respondents, 46 per cent (22 of 48) reported them to have been very useful, whilst 54 per cent (26 of 48) found them quite useful.

P4C trial schools also received two reference copies of SAPERE's Level 1 and Level 2 Handbooks. The Level 1 Handbook received positive feedback. Of the 92 per cent of P4C leads (49 of 53) responding to the survey that had received the Handbook, 59 per cent (29 of 49) reported it to have been very useful with 41 per cent (20 of 49) finding it quite useful. Forty five P4C leads stated that they had received the Level 2 handbook (85 per cent). Of these, 60 per cent (27 of 45) found it very useful and 40 per cent (18 of 45) found it quite useful.

All case-study schools reported having a central P4C resource folder on their school's shared computer drive. This central resource was an important aspect of the support provided to help teachers deliver P4C. Teachers explained how useful this was as a reference point. It was often stated by case-study interviewees that P4C does not require lengthy preparation time as long as there is a well-stocked bank of resources and stimuli for teachers to refer to and incorporate into their class topics. The P4C leads often explained how they would cascade information to colleagues via this shared folder too. In addition, the shared folder was useful for new staff to identify resources to support their P4C development. It helped to have resources in this folder filed by year group in order to identify appropriate P4C resources for different year groups.

The value of P4C within the school

Without exception, case-study interviewees explained the importance of senior staff understanding and valuing P4C in order for it to become embedded within school. One P4C lead commented: '*It's support from the leadership. It's got to come from the top down that it's important.'.* What was clear during case-study visits was that senior leaders are needed to spearhead the programme to give it urgency and importance. Case-study visits revealed that, where headteachers valued the P4C model, had attended at least the Level 1 training and kept P4C a priority within school, teachers interviewed in these schools valued P4C more. However, a revisit to one case-study school showed that without this support, P4C is unable to flourish. Indeed, P4C has been discontinued in this school as the appointment of a new headteacher had meant that P4C was not considered a priority, P4C training was not approved and P4C had been side-lined.

In contrast, senior management in the other case-study schools were prioritising P4C in practical ways that supported teachers in their delivery of P4C. This included making it part of the school development plan, displaying P4C visuals in the library and classrooms, keeping P4C on the agenda at staff meetings and tweeting a P4C question of the week. Such examples ensured that P4C remained visible within the school, highlighting its importance.

The findings above support the theory of change that assumes committed senior leaders, as well as stability of the leadership team and the whole-school culture reflecting the P4C's ethos to support teachers to develop the 4Cs. It is worth noting that 10 of the 75 intervention schools had at least one headteacher change over the course of the trial (4 of these headteachers were also P4C leads) (according to NFER contact details log), a point we will return to in the

interpretation section of the report. (For context, 20 of the 123 control group schools had at least one headteacher change during the same time period.)

Role of P4C Lead

All schools involved in the telephone interviews or case studies had a named P4C Lead, reflecting the approach outlined in the theory of change. Typically, this role was assigned to a class teacher. Of those P4C Leads surveyed, the majority (70 per cent, 37 of 53) reported that they were not part of the senior leadership team; the largest group of P4C leads were classroom teachers (47 per cent, 25 of 53 respondents).

Indeed, interviews with headteachers during telephone interviews at the start of the trial reported that they had spent little time on P4C themselves, instead they saw their role as keeping P4C a priority in school. As such, the main responsibility for coordination lay with the P4C lead. The P4C lead role involved liaising with external trainers to organise training, as well as supporting colleagues with P4C delivery.

All P4C leads interviewed during the initial telephone interviews at the start of the trial period and during the case study visits, explained that their coordinating role was manageable in the early stages of implementation and became increasingly easier as teachers became more confident with the programme. Typically, P4C leads within case-study schools reported spending around two hours per term undertaking P4C tasks (in addition to training sessions and planning their own P4C sessions). A few P4C leads explained the benefits of having a second staff member who had attended external training or P4C conferences in order to have someone to consult with in school.

P4C leads within case-study schools often described how, while time-consuming, it was important to run P4C distinctly in the early stages, (e.g. the first few terms) in order to build staff confidence to deliver the programme, develop ideas and overcome practical issues. They reported that teachers needed to have a P4C lead to share concerns with and seek support from and that teachers needed to disseminate their successes with each other in order to see the benefits of the programme.

Interviews with senior leaders and teaching staff certainly valued the role of the P4C Lead. Most interviewees across all case studies explained that one of the most important roles of the lead was having in-house support and maintaining momentum for P4C. In a case-study school where P4C had dissipated, the senior team was not involved at all. For this reason, the role of the P4C lead was of upmost importance in championing P4C in the school so that teachers who wanted to continue with it were supported to do so. This is in line with the theory of change which emphasises this important role of a P4C lead. It is particularly worth noting that there were P4C lead changes in 39 of the 75 intervention schools (4 of these were also headteachers) over the course of the trial (according to NFER contact details log), a point that will be revisited in the interpretation section of the report. For comparison, 38 of the 123 control group schools had at least one key day-to-day contact change during the same time period. The research team did not collect data on whether P4C staffing changes were related to staff turnover or internal role changes within schools. However, the case study and interview data highlighted mainly strong engagement with the programme from staff, and did not raise internal role changes as an issue. We can therefore assume it is likely that the P4C lead changes and key contact changes noted on the trial records are associated with staff turnover.

Other facilitating factors

Incorporating P4C into an already stretched timetable was a concern for teaching staff in case-study schools in the initial stages of the trial period. However, in most schools by the time of second visits, schools were confidently linking the P4C to the curriculum and delivering sessions as part of other subject areas and topics and were seeing the benefits of this. For example, one teacher commented: *'I think you get much more P4C discussed and used if it's in our curriculum rather than an add-on'*. Most P4C Leads explained that in order to facilitate this move to embedding within the curriculum, support from the external trainer on how to forge these links is vital. A small proportion of interviewees across the case-study schools reported how useful it had been to liaise with the curriculum leads when lesson planning in order to incorporate a P4C link where there is an opportunity to do so. In one school the school's curriculum lead had built P4C into the school's curriculum rationale document as a key element, ensuring that P4C was being used by all teachers across the school. Literacy coordinators explained that P4C was used widely in literacy or English lessons. For example one literacy coordinator explained that speaking and listening activities were often set out as a P4C enquiry. However, numeracy coordinators felt that P4C was less easily linked with numeracy and therefore was not being embedded to the same extent as it was in other subjects.

Other facilitating factors raised by both teachers and pupils in case-study schools related to the running of the sessions to ensure they were successful. These were:

Facilitators of P4C interviewed in case-study schools frequently explained the importance of making it explicit to pupils when a P4C session is running in order for them to understand that the parameters of discussion have broadened.

Facilitators and pupils both agreed that the circular seating formation of sessions was important. Once in the circle, teachers often reported how surprised they were at how the dynamic between pupils changed positively. One pupil commented: *'It's nice to see everyone when we're sat in a circle. You can see everyone else's face.'*

P4C displays were reported by case-study interviewees to work well and were described as being a visual reminder to pupils, highlighting P4C rules of engagement and helping pupils become acquainted with P4C rules. Displays might also include P4C sentence stems. This was particularly important where pupils were becoming challenging during discussions and helped teachers have a visual reference for pupils.

The regular repetition of language was often reported by case-study interviewees as helpful in encouraging the less articulate pupils to contribute.

Children interviewed within focus-group discussions often referred to the importance of hand signals as a means of communicating their opinions. They also valued having the P4C sentence stems (such as 'I agree with…', 'I'd like to question…') and having the normal parameters of discussion extended during P4C enquiries.

Adapting the model for different groups of pupils

During the initial case-study visits and exploratory interviews with senior leaders, there was some concern that P4C would be too complex for younger children. However, the second-year case-study schools revealed that teaching staff were becoming more confident in delivering P4C to younger year groups. Modelling from trainers and observations were ways in which teachers reported becoming more confident facilitating sessions with younger pupils.

If younger children or certain groups of pupils struggled with the 10-step rule, there was enough flexibility in place to adapt the programme to suit the pupils' needs. For example, splitting a class into smaller groups was a worthwhile way of introducing language that would prepare the group for larger group discussions. Case-study interviewees often explained that the programme has enough scaffolds in place for schools to build the programme around their own learning needs.

Some case-study schools explained their early struggles with encouraging EAL, SEN and young children to fully engage with the P4C 10-step process. Through on-going support from trainers, teachers reported how their confidence increased. In turn, interviewees explained how over the course of the trial period, they had noticed how well these learning groups responded to the P4C process; the slower pace of the discussion allowed these children the time to explore their own thinking rather than mirror the answers of others. This in turn was felt to increase pupils' self-esteem and confidence to participate. One P4C lead commented: *'Having that silence has allowed some students who would never talk normally in a session to talk.'*

Splitting the 10-step enquiry model so that pupils progress more slowly through an enquiry over a number of sessions were seen as beneficial to young year groups. For example, initially, younger children were not asked to devise their own questions or build on others' opinions, however they benefited from learning P4C language which would put them in good stead for when they participate in P4C further up the school.

Attractiveness of intervention to stakeholders

Schools and teachers

The survey responses from P4C leads suggest that they believed their teaching staff had generally responded positively to P4C. Almost two thirds (62 per cent, or 28 of 45) of the P4C leads agreed with the statement 'teachers have fully embraced facilitating P4C sessions' with six respondents disagreeing with this statement. Around three-quarters of responding P4C leads (76 per cent, or 34 of 45) agreed that their teachers valued the P4C approach, with just three P4C leads disagreeing with this. The responsiveness of the teaching staff to P4C was explored in more depth through the telephone interviews and case studies, discussed below.

In the early stages of implementation, most senior leaders interviewed by telephone reported that their staff were enthusiastic about the programme and saw the delivery as innovative. Where there had been some scepticism initially by teaching staff, interviewees reported that staff had been pleasantly surprised by how much impact the programme was perceived to be having on the children. The training in particular was seen to have been a factor in alleviating teachers' concerns or apprehensions over delivery and fostering a motivation and enthusiasm for P4C.

The case studies revealed a similar picture with the majority of staff being seen to be responding positively to P4C, and enthusiastic to try something new. However, in nearly all case-study schools, P4C leads and senior leaders reported that a very small number of teachers in each of their schools were a little sceptical of P4C. This was always reported to be in relation to the practicalities of delivering P4C rather than in relation to the ideology of it. For example, in nearly all cases this circumspection was due to staff being concerned over what they viewed as an already full timetable. These staff raised concerns about how to fit additional tasks into the curriculum. One headteacher commented:

> *'Timetabling has been a huge challenge. Some people see it as an additional task. We explained that it is an additional tool. This was tricky as first – to get all teachers on board.'* (Headteacher, year one case study)

As with the telephone interviews, those who took part in the case studies felt that those staff who had been concerned initially, became enthusiastic after they had delivered P4C and perceived some positive impact, particularly in relation to improvements in the pupils' communication skills.

Pupils

The vast majority of teaching staff who facilitated P4C believed that pupils liked the P4C approach (94 per cent of survey respondents (201 of 213) agreed with the statement pupils like the 4C approach). Furthermore 88 per cent (187 of 213) agreed with the statement 'pupils are fully engaged with the P4C sessions'. Together these findings suggest that teaching staff believe their pupils are reacting positively and enjoying P4C sessions.

Teachers in all case-study schools explained that pupils enjoyed P4C lessons and were engaged in sessions. They also believed that pupils enjoyed the freedom of expression that came with the sessions. Some of the staff interviewed felt that certain groups of pupils were particularly engaged in P4C sessions, including those with SEND, those with EAL and those who perhaps do not engage as well in a traditional classroom setting. One teacher commented: '*It engages children who don't necessarily engage with standard teaching practice and thinking processes*'.

There were mixed views as to whether the younger pupils were as engaged as the older pupils in schools, with a small number of staff questioning whether the younger pupils were able to understand and therefore engage fully in the sessions. However, other staff interviewed explained that they had been surprised at the high level of engagement from the younger pupils in the schools.

A minority of P4C leads explained that some pupils had been initially reluctant to take part as they did not want to 'open up'. However, in all instances of this the interviewees said that the pupils now enjoyed sessions and were excited to take part.

When interviewed as part of the case studies, pupils were overwhelmingly positive about P4C. This positivity spanned the two case-study time points, highlighting the extent to which pupils enjoy P4C. One pupil commented: *'When the teacher says that it's P4C time, we all cheer, YES!'*

Pupils enjoyed P4C for a number of reasons, however the most commonly cited reasons referred to the pupils being able to express themselves and their own opinions in a non-judgemental arena. For example, one pupil commented: *'I like how everyone respects each other's opinion. It's quite fair and everyone is allowed their opinion,'* while another pupil in a different case-study school commented: '*You can express how you feel because there's no right or wrong answer.'* Pupils in one school took this further and explained how P4C sessions helped them discuss their problems: '*We get to express our feelings, you're not shy. This is very different to other lessons, it's more peaceful. You don't have to worry about anything and you can talk about your problems.'*

A large proportion of the pupils across the case-study schools commented that they enjoyed the more practical mechanics of a session, particularly sitting in a circle rather than at a desk. Pupils explained that this made them feel more relaxed.

When pupils in case studies were asked if there was anything they did not enjoy, most pupils found it difficult to think of anything. However, where pupils were able to think of elements they dislike of P4C, the most commonly reported element was the use of negative vocabulary when pupils were sharing their opinions. In two of the schools pupils said that the phrase 'you're wrong' was used a lot by pupils in sessions and this made some pupils feel embarrassed about sharing their views, which appears to be at odds with the principles of P4C.

A notable minority of pupils felt that sometimes, they do not enjoy sessions due to the question that had been chosen for that particular session. This could be either because the question was not of interest to them or, in some instances, it was because they did not like the topic or found it upsetting, for example one pupil commented: '*I don't like questions about evil people.*'.

Other elements the pupils did not like related to the skill of the facilitator, for example, a minority of pupils said that sometimes certain pupils were chosen to talk multiple times while other pupils did not get a turn, or there were instances of children talking over each other and shouting when they disagreed.

Parents

Overall, staff interviewed across the case studies explained that parents had not yet been engaged in P4C to any great extent and did not understand it. For example one P4C lead commented: '*I don't think they* [parents] *understand the benefits of it. They say it's just a talking session'.* In some schools this was because they had not tried to engage parents, while in others this was because parents were not responsive to attempts to engage them with P4C. Schools which had attempted to engage parents were doing activities such as tweeting about P4C topics, adding a question of the week to the newsletter and putting up displays around the school. However, these schools noted that they had very little evidence that parents are engaging with any of these activities.

Note that by 2019, schools were expected to be working at Level 5 (i.e. bronze award and to have met 50% silver or above). Parental engagement is an area that schools work on towards their gold award. In the phase 2 case studies, P4C leads saw the lack of parental engagement as an area they needed to address, particularly if they wanted to work towards their Gold Award, for which evidence of parental engagement was required to demonstrate involving the wider school community (for example, evidence of communicating P4C practice to parents). However, schools were planning to do this in the next academic year, rather than addressing it immediately. P4C leads had ideas such as running a P4C information session at a Parents' Evening or adding a P4C question to discuss with parents at home to pupils' homework books to raise P4C's profile amongst parents.

Perceived outcomes

What is the perceived impact on pupils?

Teachers facilitating P4C sessions who responded to the survey were asked whether they perceived P4C had positively impacted on a range of pupils' social competencies. As shown in figure 5, the areas in which survey respondents felt P4C was having the most positive impact on pupils were:

- Respect for other pupils' opinions (96 per cent, 204 of 213, agreed that P4C had positively impacted on this)

- Ability to question and reason (91 per cent, 196 of 213, agreed that P4C had positively impacted on this)

- Ability to express views clearly (93 per cent, 193 of 213, agreed that P4C had positively impacted on this)

- Listening skills (89 per cent, 188 of 213, agreed that P4C had positively impacted on this).

Figure 5: Extent to which P4C has impacted on pupils
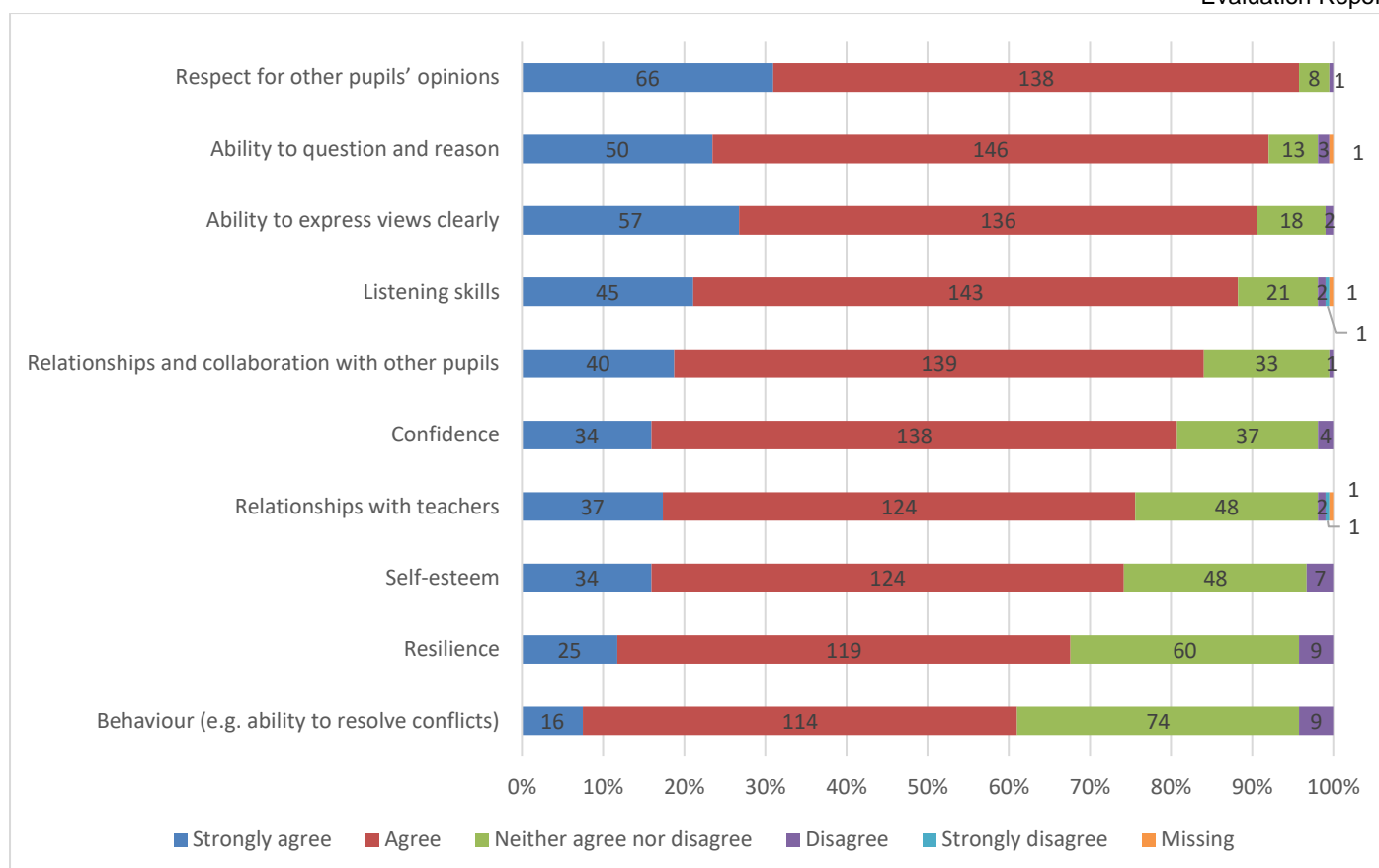
Source: NFER P4C intervention staff survey, 2019.

The majority of staff thought P4C had positively impacted behaviour (61 per cent, 130 of 213, agreed with this) and resilience (67 per cent, 144 of 213, agreed with this).

These findings were reflected in the views of those staff involved in the case-study discussions. By the end of the first year of delivery, the majority of P4C leads and teachers described how they believed P4C had positively impacted pupils' listening skills, their ability to respect other's opinions, their critical thinking and their confidence, although most of the improvements were limited to changes in children within the P4C lessons. By the end of the second year of delivery, the schools were reporting that these impacts were now being seen more widely across the curriculum. For example, interviewees in most schools reported that pupils were now applying their questioning and reasoning skills in other curriculum areas such as English and History. Interviewees also commonly noted that the turn taking modelled in P4C was now being seen to trickle through into other areas of the school, including the playground.

While teachers in the first year of the case studies generally felt it was too early to comment on any perceived impact of P4C in literacy and mathematics, by the second visit most schools were reporting seeing some benefits in literacy. For example, some schools reported improvements in reading comprehension due to greater discussion generation, which helps to support and deepen their understanding. Others noted pupils had a much wider vocabulary and had seen improved sentence structures in pupils' writing. Two schools felt that P4C linked very well with their reading programme, and as such, these two programmes combined had resulted in strong improvements in reading.

As reflected in the survey responses, there was less agreement across interviewees in case-study schools as to whether P4C had positively impacted pupils' behaviour. Generally, staff believed that pupils' behaviour had improved but often this improvement was limited to the P4C sessions themselves and was not translated into better behaviour in other lessons.

Pupils were also asked about how if at all P4C had helped them. By the end of the first year of delivery, most pupils felt that they had become more confident, that their listening skills had improved and their critical thinking. One pupil commented: *'Since we've been doing P4C I've started to think more and I'm thinking in, like, a different way.'* Another pupil commented: *'It has given me more confidence. When we do P4C lessons, everyone's well behaved but in normal lessons sometimes they can be a bit naughty and loud. People listen more.'*

By the second case-study visits, pupils across three of the schools were able to explicitly see how the skills they had developed in P4C were linking to other lessons, and how they were using the language from P4C (for example, 'I challenge you…') to other lessons. One pupil explained: '*P4C comes in handy in English when you're writing a diary entry or a news report and you have to give an opinion about something you use the language that we use in P4C, or you need to give a rhetorical question.*'

Similar to the staff views, the pupils taking part in case-study discussions felt that P4C has impacted positively on behaviour of pupils but that this was limited to behaviour within the P4C sessions. One pupil commented: '*If the teacher leaves the classroom during another lesson, things can turn pretty silly, but during P4C it doesn't because everyone is so caught up in the discussion.*'

Which pupils are perceived to benefit most from P4C?

Staff across all case studies generally agreed that those pupils who struggle to 'have a voice' in normal lessons have benefited the most from P4C. Often, staff reflected that these were EAL pupils, those who lacked confidence or SEN pupils. Staff expressed their surprise at how vocal these pupils can be in P4C sessions and how their confidence had improved. One teacher commented: '*Nearly every member of staff has said that children who you wouldn't normally expect to speak in Literacy lessons are.*'

Some schools reported that pupils from disadvantaged backgrounds or those with EAL had benefited the most in terms of their language development. EAL pupils in particular were seen to have benefited considerably in terms of their sentence structures. However, one school did note that with SEN pupils, lessons had to be very carefully planned in order that these pupils could understand the language. This school explained that they had received a lot of support from their trainer in order to enable their sessions to be accessible for this group of pupils. One school felt that their disadvantaged pupils had benefited the most from P4C sessions as the question areas discussed in sessions, as chosen by the children themselves, tended to be areas that affected them directly, such as poverty. The P4C lead explained:

> '*We have a huge number of disadvantaged children in this school, we are way above the national average. It's quite interesting how quite a lot of the P4C sessions have ended up going towards big questions in terms of things like poverty, how people are treated, all sorts of things like that. So I think for them, personally, it has been quite useful because they've been able to talk about massive issues that affect them as a community and in their personal lives as well. […] It's giving them a voice to talk about things that they probably wouldn't have had a voice to talk about before.*'

While one P4C lead believed that only older pupils would benefit from P4C, generally staff in case-study schools thought all age groups had benefited from P4C, although schools did believe they benefited in different ways. For example, older pupils benefited more from the development of critical thinking and reasoning skills, while younger pupils benefited more from improvements in confidence and developing an understanding that it is acceptable to have your own opinions and answers.

What were the perceived outcomes on teachers and the wider school?

Teaching staff who responded to the survey were generally positive about the perceived impact of P4C on them and their practice. Overall, the majority of teachers (84 per cent, 185 of 221) agreed with the statement 'P4C has had a positive impact on my professional development'. Almost two-thirds (65 per cent, 144 of 221) agreed that P4C had positively impacted on their overall effectiveness as a teacher, while just over half (56 per cent, 123 of 221) agreed that P4C had positively impacted on their self-esteem as a teacher. There was generally very little disagreement with these statements.

Confidence in trying new teaching approaches was reported most commonly as an area of impact on teaching and pedagogy by respondents to the survey. As the chart below indicates, 84 per cent (178 of 213 respondents) agreed that 'P4C had impacted positively on your confidence in trying out new teaching approaches'. Over three quarters of respondents (78 per cent, 166 of 213 respondents) agreed that P4C had a positive impact on their relationships with pupils, while a similar proportion (77 per cent, 165 of 213) agreed that P4C had positively impacted their all-round practice.

Figure 6: Perceived impact of P4C on teachers' pedagogy and teaching practice
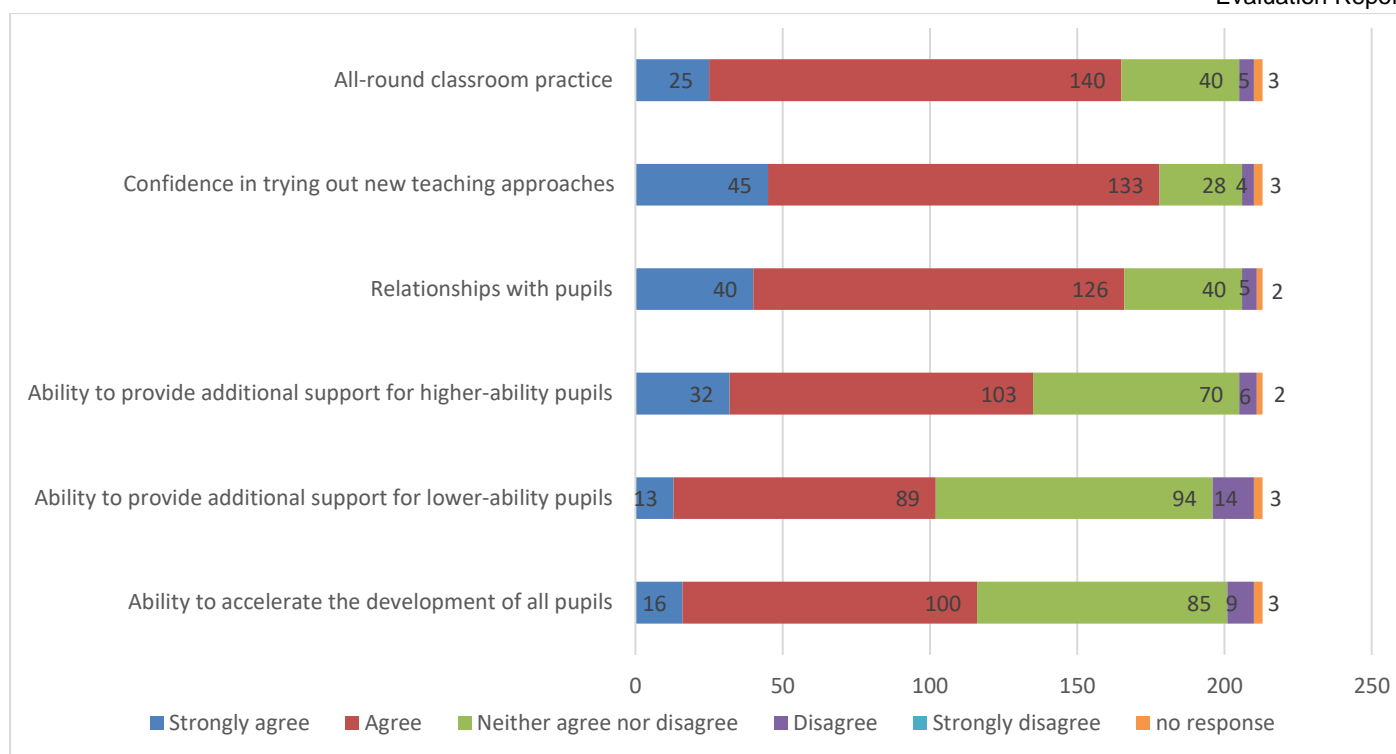
Source: NFER P4C intervention staff survey, 2019.

Teachers were less likely to believe that P4C had positively impacted on their ability to provide additional support for pupils. This was particularly the case in relation to supporting those pupils with lower ability (48 per cent, or 102 of 213, agreed with the statement 'P4C has positively impacted on my ability to provide additional support for lower-ability pupils' while 44 per cent (94 of 213) neither agreed nor disagreed with this statement). It is important to note that only a very small proportion of teachers actively disagreed with any of the statements, with none strongly disagreeing with any of these statements, reflecting a generally positive view of the impact of P4C on them and their practice.

Four of the five P4C leads in case-study schools explicitly spoke about the perceived impact of P4C on staff. Two considered that teachers' pedagogy had been developed. One P4C lead explained that their teachers further up the school had developed better questioning skills which then translated into their questioning in guided reading. One P4C lead said of their own teaching practice that *'learning to let go and let it be student led has been quite helpful because then that's led into other academic subjects.'*

All four of the P4C leads that mentioned impact on staff felt that TAs and HLTAs who were facilitating or supporting P4C sessions had benefited enormously from P4C, both in terms of their own development, and also in terms of their status within the school. One P4C lead commented: *'The TAs that have done some sessions have really enjoyed them and they feel it's helped improve their relationship with the children and level of respect with the children.'*

P4C leads were asked within the survey to what extent do you agree that P4C has had a positive impact on your school. The vast majority of P4C Leads (89 per cent, or 47 of 53) agreed with this statement, with 15 leads strongly agreeing and 32 agreeing with this statement. Just one respondent disagreed with this statement, indicating that P4C leads felt that P4C was a positive initiative for their school as a whole.

Is the theory of change fit for purpose?

The evidence collected and analysed in this process evaluation indicates that parts of the theory of change are fit for purpose; whilst others may need greater specificity. The findings corroborate the assumptions outlined in the theory of change. In particular, there was evidence of the importance of the senior leaders and teachers showing commitment, and that the role of the P4C lead and the stability of the school leadership seemed key: where there was leadership staff turnover this had the potential to change school priorities and prevent further embedding of P4C, a challenge experienced in one of the case study schools. The turnover in headteachers, and particularly in P4C leads, logged during the course of the trial is particularly worth noting, and it may be that this is an area where greater support both within schools and from SAPERE to ensure handover and continuity of P4C practice and ethos. Indeed, in addition to priorities and commitment, school leaders' understanding of the P4C approach seemed paramount – pointing to senior

leader support not only as a moderating or contextual factor (i.e. affecting the strength of outcomes), but also as a mediator (a mechanism in itself for ensuring that the P4C ethos pervades the whole school).

Reflecting the theory of change, findings from the evaluation highlight the importance of good-quality training and the ongoing support that trainers provide. The model of training and support provided by P4C was welcomed by schools and seen as vital in ensuring schools are able to embed P4C across the curriculum.

The findings also highlight some elements of metacognition in both the approach used and the outcomes (in particular, respecting others' opinions, questioning and reasoning skills and expressing views clearly), which suggest that the specific contribution of metacognition within the theory of change perhaps needs to be strengthened. Most schools were indeed exposing pupils regularly to P4C, through standalone sessions at the beginning of the trial or through embedding P4C within the curriculum over time. Both teachers and pupils could see the impact of P4C on pupils' critical thinking and ability to question outside of P4C sessions. Whilst there were some comments about how P4C linked well with reading comprehension, there was very little comment from participants in the process evaluation that P4C was having a wider impact on pupils' attainment, or on the mechanisms by which that would occur.

The perceptions of the interviewees in the process evaluation appear to support the theory of change's description of the impact P4C should have on teachers in relation to the development of the 4Cs. However, there was very little evidence from the process evaluation data that P4C was having a wider impact yet in relation to parents – an area that schools would need to work on for the gold stages of their Going for Gold awards (note, by 2019 schools were expected to be working at the bronze level with 50% silver, which does not focus on parental engagement). It may be that engaging with parents through and with P4C approaches takes more time to implement; and perhaps the planning stages for this need to be strengthened.

Further discussion on the theory of change, integrating the process findings with the findings on impact, can be found in the Interpretation section of this report.

Sustainability and scalability of P4C

Four of the five schools involved in the case studies were planning on continuing with P4C in the future. The P4C lead in the one school which was not continuing stated that this was because a new headteacher had joined the school since the trial began who did not support their involvement with P4C. As such, it was no longer timetabled and high staff turnover had meant that most teachers in the school had not received training.

The importance of having the support of the headteacher was seen as crucial to the sustainability of P4C in schools by all involved in the case studies. Without this support, P4C would not be timetabled, staff would not receive the level of training required to be confident in delivery, and importantly, P4C would not be kept on the school agenda and would not be discussed in staff meetings and ideas and resources would not be shared.

All case study schools felt that standalone sessions were required within the first year of delivery. This was deemed essential to help both teachers and pupils gain confidence and skills in the P4C approach. However all schools felt that in the longer term, in order to continue delivering P4C they needed to move from standalone sessions to sessions embedded within the curriculum. This was due to the time pressures of an already busy curriculum. Teachers gave examples of P4C being embedded successfully in a range of subjects, most notably English, History, Geography and Art.

With an embedded model, teachers explained that achieving 60 minutes of P4C a week would be sustainable, although perhaps not delivered in one session, but broken up and split over short sessions. This was seen as particularly important for younger pupils in KS1 and lower KS2 whereby it was felt that a 60 minute session on any subject was too long.

Teachers in the case study schools were asked what advice they would give other schools wanting to participate in P4C. Their summarised advice was to:

- Ensure schools receive good quality training which include lots of practical examples (not just covering the theory of P4C).

- Pair with a similar school who is already experienced in delivering P4C to share best practice, give support and advice.

- Start early with pupils as the earlier you start the more comfortable the pupils become with it.

- Keep P4C on the agenda at staff meetings to keep the discussion going.

- Set up a resource bank to share ideas for sessions.

- Ensure the headteacher fully supports P4C.

- Run standalone sessions to begin with but then think about the cross-curriculum potential and embed P4C across other subjects, to ensure sustainability in the longer term.

The P4C team at SAPERE explained that, as P4C is a well-established programme, they were continuing to roll it out across the country and could not foresee any problems should numbers of schools signing up increase substantially. Their model of using freelance trainers based across the country works well and provides flexibility to schools and the central team. Similarly their model of training courses; providing a mix of whole-school training based at the school, alongside a limited number of days adhoc support from their trainer, and then higher level and top up training courses at centralised locations allows the team to adapt to changes as and when required.

Usual practice

Through the survey, intervention schools were asked whether they had partaken in any other enquiry-based philosophy-based education interventions, training or support while involved in the trial. In total 30 P4C Leads mentioned other sources of support they had received in addition to the support they had received through the intervention. Seventeen of these had received other support through SAPERE, while the other most commonly cited support was through Jason Buckley (15 respondents), an independent provider of P4C training and resources.

Schools allocated to the control group were required to continue 'business as usual' and commit to not participating in P4C until September 2019 (from when they would be able to take part in P4C with any year group other that Year 6 as this cohort is involved in a further year of KS2 analyses). As an incentive to take part in this trial, control schools each received £5,700 to ensure their commitment to the research project while not receiving the intervention. In order to assess 'business as usual', control group schools were sent a pro-forma in Summer term 2019 to complete with information aboutany other philosophy-based or enquiry-based programmes or training (in addition to P4C) that the school may have accessed during the trial period. They were also asked to indicate if they had accessed P4C during this time (i.e. a contamination check).

Of the 105 control schools that returned the proforma, six reported having run P4C sessions with pupils between the academic years 2016/17 and 2018/19 (note, from SAPERE's monitoring data, three control schools accessed some P4C via SAPERE during the trial period). Pupils who received sessions tended to be in Years 4, 5 or 6. Most of these schools (five of six) reported that these pupils had received ten or fewer sessions of P4C across this time. Four of the six had not accessed any P4C training in order to do this. However, one school had received some training from SAPERE and had delivered 15 sessions across in 2018. This had occurred as the school was part of a federation; the lead school in the federation (not part of the trial) had booked onto P4C training and had invited members of staff from across the federation to the training. At the time, SAPERE was unaware that staff members from a control group school had attended training (as they had attended under the booked school's name). As soon as this was discovered in early 2019, the school agreed to stop any delivery of P4C sessions to Years 5 and 6 (the measurement years). Compliance data provided in July 2019 (i.e. SAPERE's monitoring data) revealed that this control school was working towards bronze but had not yet achieved any award level. Compliance data also revealed that another control school had taken part in P4C during 2016/17 but as soon as they were allocated to the control group they ceased P4C activity (they were not on the exclusions list provided by SAPERE prior to NFER starting recruitment); and another control school was recorded as having withdrawn from the trial and taken up P4C (this school was not working towards any award level and did take part in the follow-up pupil survey or complete the control online proforma).

Overall, exposure to P4C was minimal across the control schools; and any progress recorded towards award levesl (i.e. by the one school noted above) was taken into account in the on-treatment analysis.

Seven control schools had run other philosophy-based or enquiry-based programmes between the academic years 2016/17 and 2018/19 (such as an oracy project, Bloom's Taxonomy, Zippy's Friends and Growth Mindsets). Most of these were linked to research programmes. Schools all reported different programmes with no two schools mentioning the same programme and most were running the programmes with Year 6 pupils in the academic year 2018/19. While

nearly all schools were delivering just one programme, one school was delivering three different programmes. The number of control schools reporting other enquiry-based programmes being used was fairly small (seven of the 105 returns for control schools), so this would not unduly affect the interpretation of the findings. However, the broader policy backdrop of encouraging character education (Ofsted, 2019b) and metacognition (EEF, 2016) may have percolated both intervention and control schools during the trial period (although our study did not collect data on this).

Further implementation (academic year 2019/20)

After the completion of the main trial data collection (pupil survey and pupils' sitting their end of KS2 assessments), control schools were able to take part in induction sessions and book training for P4C, with the proviso that should they run P4C in their school they would not deliver it to any Year 6 children (in 2019/20). This year group would continue to act as a control for the planned addendum cohort analysis (using their KS2 assessments from the end of 2019/20). However, due to school closures in England response to Covid-19, the 2019/20 KS2 tests were cancelled. No assessment data was collected, but SAPERE were still able to supply NFER with implementation data for intervention and control schools for this academic year.

Schools were able to implement P4C and apply for awards in the usual way, up until schools closed in mid-March 2020. Between March and July 2020, SAPERE were able to contact trainers and schools to ensure the latest award level data was captured. This was particularly important for the levels where schools are judged as 'working towards' or 'working above a certain level but not yet at the next level'. Trainers views here were important to collect up until the end of term, rather than provide the incomplete dataset as at mid-March.

The same seven award categories were used in 2020 as in 2019. As can be seen in Figure 7, more intervention schools had achieved the highest award levels (silver, levels 6 and 7) in 2020 than in 2019. However, there had also been some drop off in engagement with the programme; the number of schools not taking part in any P4C sessions during the year or not actively working towards a level was higher in 2020 than in 2019. Notably, the proportion working at 'compliant' levels (i.e. level 5 or above) remained the same (40 schools in 2019, 39 in 2020).

Figure 7: Intervention schools awards 2019 and 2020
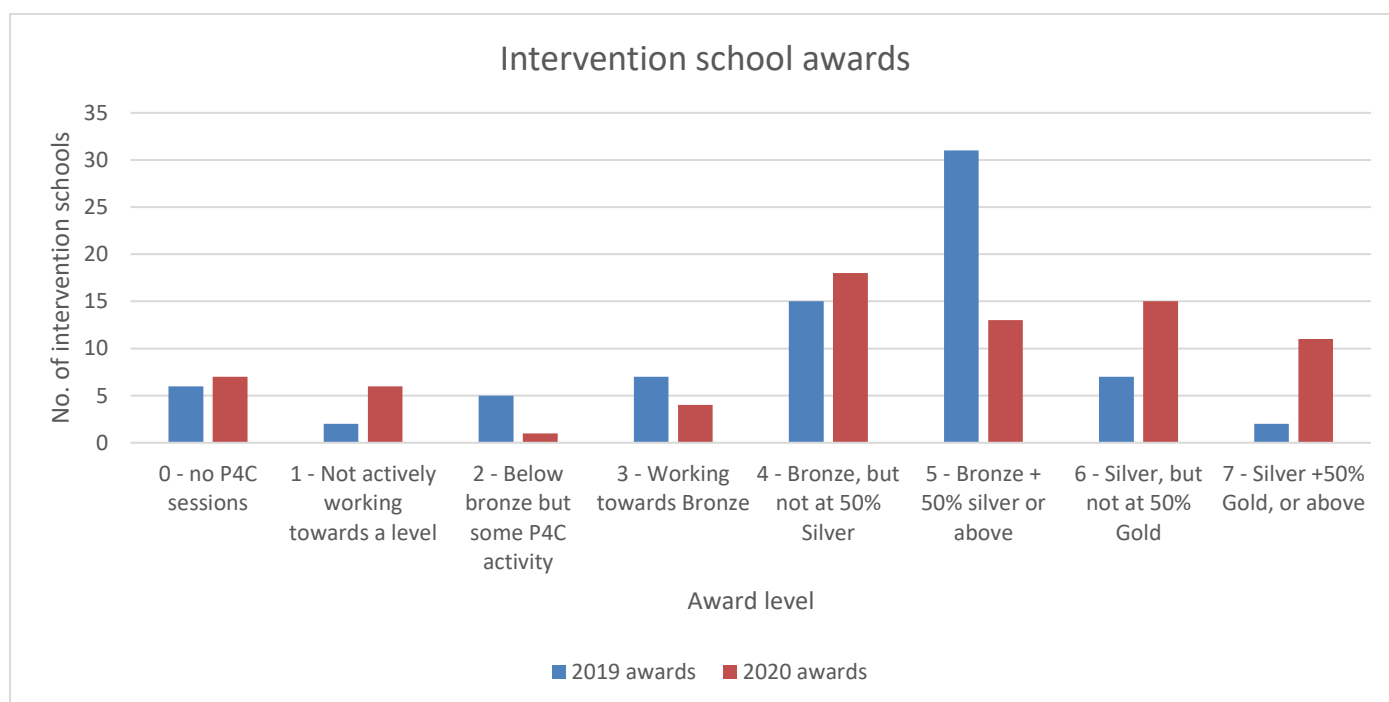
Source: SAPERE P4C award data, July 2019 and July 2020

It is helpful to explore the change in award levels to understand school engagement further. As shown in Figure 8, almost half of the intervention schools (36 out of 75) remained at the same award level in 2020 as they had achieved a year earlier. About a quarter of the schools (19 out of 75) were awarded the next level up in the subsequent year, indicating continued progress with implementing P4C. A small handful of schools (7) raised their awards by two or three levels. These were schools that had previously been awarded level 4 or 5, but had embedded P4C to the extent of

achieving level 7 (i.e. silver and on the way to 50+ of gold) in 2020; i.e. the greatest increases are seen amongst schools that were already engaged, and were able to embed the work during the next academic year (indeed up until March 2020).

Interestingly, a number of schools (13) were recorded with reduced award levels in 2020. For some, this was related to no engagement due to other priorities in 2020; for others the decreased engagement was in terms of no longer 'working towards' the next level.

Figure 8: change in award level from 2019 to 2020
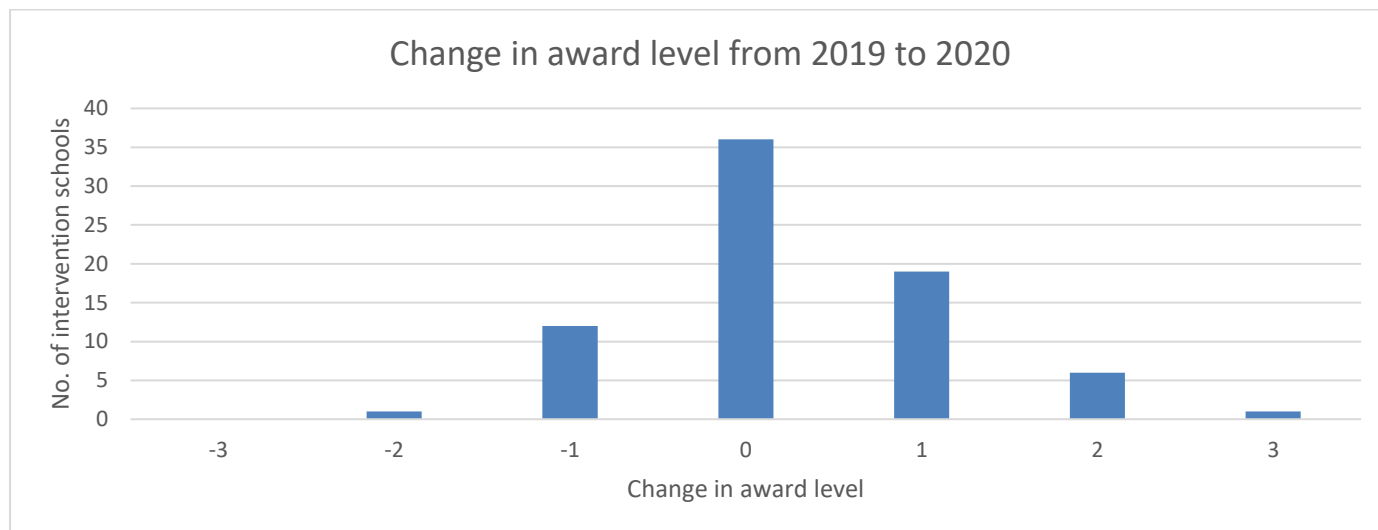
Source: SAPERE P4C award data, July 2019 and July 2020

As noted above, control schools could take part in P4C in 2019/20 (but not with Year 6) as part of a waitlist design. SAPERE collected award level data for control schools, and reported that 9 control schools had taken part in P4C in 2019/20, with one of these schools achieving a Bronze award (level 4) (see Fig 9). SAPERE confirmed that none of these 9 schools had delivered P4C to Year 6 in 2019/20. Take up was lower than expected. SAPERE felt some of this was due to the waitlist design necessitating waiting until Key Stage assessments were complete (24th May 2019 was the last date for schools to submit any special considerations about pupils' KS2 assessments), before they could approach control schools to invite them to introductory sessions on P4C and to book training. This timing may not have aligned with schools' planning cycles, thus reducing take up of P4C.

Figure 9: Waitlist control awards 2019/20
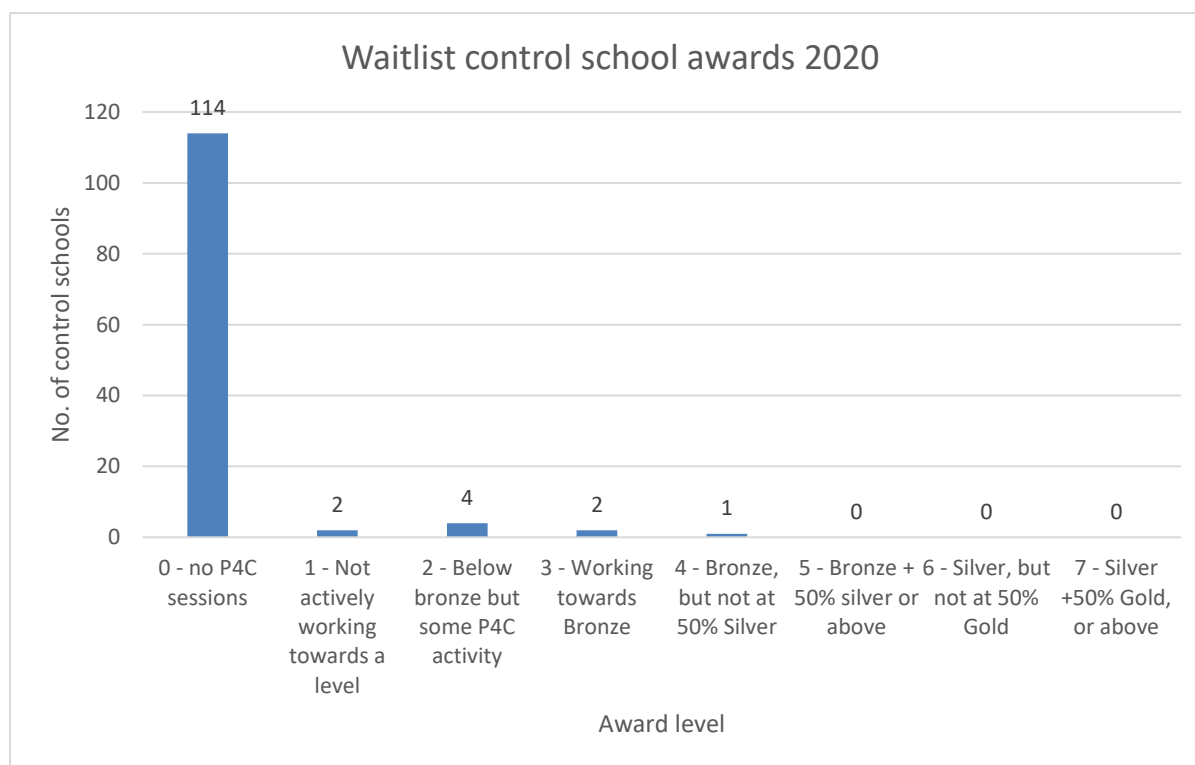
Source: SAPERE P4C award data, July 2019 and July 2020

Cost

The average cost of SAPERE's Going for Gold programme is £13.50 per pupil per year when averaged over three years. This estimate is based on the delivery of the intervention across all pupils in an average sized primary school of 282 pupils (Department for Education, 2019)[20], with schools signing up for the full package of training and support from SAPERE.

We collected direct costs of the Going for Gold programme from SAPERE. The costs presented in Table 21 are based on a comprehensive support and training package provided to schools over three years. While schools can, to some extent, choose the elements they want from the package in order to suits their needs, the assumption for the costs calculations is that schools sign up for the full level of support from the Going for Gold programme, as was received by the schools in the intervention group.

---

[20]
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/812539/Schools_Pupils_and_their_Characteristics_2019_Main_Text.pdf

Table 21: Cost of delivering Going for Gold

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
Level 1 training	INSET training. Cost for year 1 only	£2,200	£2,200	£7.80
Tailored support	7 days of support across the three years (amount varies per year)	£500 per day	£3,500	£12.41
Trainer planning and preparation days	4 days of support across the three years (amount varies per year)	£250 per day	£1,000	£3.55
Tools for Thinking training	INSET training. Cost for year 2 only	£1,100	£1,100	£3.90
Level 2A training	Training course. Cost for year 2 only	£1,200	£1,200	£4.26
Office support from SAPERE	1 day per year of office support	£100 per day	£300	£1.06
Level 2B training	Cost for 2 places on course. Cost for year 3 only	£600	£600	£2.13
Level 1 top-up places	Cost for 5 places on course. Cost for year 3 only	£1,000	£1,000	£3.55
Award visit	Gold award visit in year 3 only	£300	£300	£1.06
SAPERE membership	Yearly membership fee	£75	£225	£0.80
Total			£11,425	£40.51

The Going for Gold package cost for three years is £11,425. However, the level of support varies per year, with more support being provided in the first year compared with the following years; as such the costs per year are not equal and break down as follows:

- Year 1 costs a school £4,375. It includes Level 1 INSET training for 22 participants over 2 days, 3 days of support, 1 day of administrative time and a SAPERE yearly membership fee.

- Year 2 costs a school £3,725. It includes Tools for Thinking Together INSET training day for 22 participants, 4 places on the Level 2A course, 1 day of support, 1 days of administrative time and SAPERE yearly membership fee.

- Year 3 costs a school £3,325. It includes 2 places on the 2-day Level 2B course, 5 places at a Level 1 top-up course, 1 day of support, 1 days of administrative time, Gold award visit cost and SAPERE yearly membership fee.

Further details of costs for Going for Gold are presented in Table 22 below.

Table 22: Cumulative costs of Going for Gold (assuming delivery over three years)

	Year 1	Year 2	Year 3
Cost per school for Going for Gold	£4,375	£3,725	£3,325
Cost per pupil	£15.51	£13.21	£11.79

Within the effectiveness trial, intervention schools paid a discounted rate of £1,600 per year (£4,800 over the three years) for the Going for Gold programme. This equates to £5.67 per pupil per year.

Indirect costs were collected from interviews with case-study schools. Two of the five schools reported indirect costs of £200 to purchase sets of books. However, one of these schools considered that this was not necessary as online resources such as film and music clips were more useful and potentially less leading than books, as they felt books often led the reader to a pretermined answer. In both these instances these were one-off costs at the outset of the intervention. The other schools stated that there were no indirect costs associated and that all resources were either supplied or they could access free of charge.

The majority of training for staff is delivered through INSET days, and as such, does not require staff to be taken out of school for training. The exception to this is the Level 2 courses or Level 1 top up courses. This was mentioned as an additional cost and time resource by one school which incurred costs associated with travel, subsistence and overnight accommodation as these sessions were at a central location. The school was unable to provide the costs of these.

As P4C lessons were seen as part of the curriculum or embedded into other subjects, the schools felt that planning for P4C was subsumed in teachers' planning time. However, schools did report that additional planning time was needed by staff at the outset of P4C when they were unfamiliar with the teaching methods. Interviewees delivering P4C in case-study schools explained that, in the early stages of implementation, standalone P4C sessions required an hour of preparation time to find a suitable stimulus, review the 10-step model and draft a plan. This planning time was said to decrease over the trial period as teachers became more confident in their delivery. Planning time also decreased as P4C became more aligned with the curriculum. In addition to planning time for delivery, P4C lead typically reported spending around two hours per term undertaking P4C tasks.

Conclusion

Key Conclusions

1. There is no evidence that P4C had an impact on reading outcomes on average for KS2 pupils from disadvantaged backgrounds (i.e. FSM eligible pupils). This result has a high security rating.

2. Similarly, there is no evidence that P4C had an impact on reading attainment at KS2 for the whole cohort of Year 6 pupils. There is also no evidence that P4C had an impact on attainment in maths for KS2 pupils – either for the whole cohort, or for pupils from disadvantaged backgrounds.

3. Whilst teacher feedback on P4C was positive - 96% of intervention teachers felt that pupils had improved their level of respect for others' opinions, and 93% felt that pupils had improved their ability to express their views clearly; there was no evidence of impact on children's social and communication skills, as measured by the pupil survey.

4. Of the 75 intervention schools, after two years from commencement, a substantial minority (35 of 75 schools) were not implementing P4C at the expected level. Of these, six did not implement P4C at all due to other priorities and/or senior leader turnover. The evaluation suggests that it takes time for teachers to become confident with, use and embed the P4C approach and this could have impacted the outcomes.

5. Where schools were implementing P4C, teachers and pupils found it enjoyable, engaging and that it encouraged pupils to share opinions in a non-judgmental way, finding it particularly beneficial for EAL pupils, those who lacked confidence or SEN pupils. Teachers and P4C leads felt that the training and ongoing support was high-quality and that it had enabled them to facilitate P4C sessions effectively in their school. Important factors for sustaining and embedding implementation included: starting with sessions based on standalone topics before incorporating cross-curricular work into sessions; and senior staff support, particularly around understanding and valuing the P4C approach.

Impact evaluation and IPE integration

Evidence to support the logic model

A theory of change, rather than logic model per se, was agreed in collaboration with SAPERE at the start of this study (see Appendix D). The results of the impact and process evaluation support parts of the theory of change; but further development of the model may need to establish the particular specificity of metacognition (an underlying assumption) to P4C and greater connectedness across the steps of the theory of change.

Starting with the overall purpose of the evaluation and of P4C, whilst the study clearly set out to explore the 4Cs (thinking in caring, collaborative, creative and critical way) by incorporating them into process instruments and the teacher survey; the link from these 4Cs to the expected pupil outcomes in the theory of change (attainment and social skills in terms of explaining ideas to others and working with others who have different opinions) is perhaps not explicit and needs strengthening. Alternatively, other outcome measures more closely related to the 4Cs could be valuable to explore. The theory of change was drawn up to be relevant in terms of disadvantaged pupils. Again, greater specificity about how P4C might benefit disadvantaged pupils moreso than their peers would be useful to strengthen – or, indeed, to refute. Certainly, the process evaluation highlighted P4C as enjoyable, engaging and beneficial to all pupils, especially those who were less confident.

Moving on to underlying assumptions, the findings do corroborate some of those outlined in the theory of change. In particular, having committed senior leaders and teachers, appointing a P4C lead to champion P4C in the school, and the stability of the school and its leadership team seemed key. These areas were strongly emphasised as key factors in the process evaluation, and indeed, where leadership changes had occurred this had the potential to change school priorities and prevent further embedding of P4C, a challenge experienced in one of the case study schools. Moreover, in addition to the assumptions of leadership commitment and stability, the data we collected suggested that school leaders' understanding and valuing of the P4C approach seemed paramount – i.e. senior leader support not only as a moderating or contextual factor (i.e. affecting the strength of outcomes), but also as a mediator (a mechanism in itself for ensuring that the P4C ethos pervades the whole school). It is therefore of particular note that during the course of this trial, 10 of the 75 interventions schools experienced changes in headteacher, and 39 of the 75 intervention schools had P4C lead changes – meaning this part of the theory of change was substantially weakened in practice. Whilst external organisations cannot do anything about staff turnover, SAPERE could strengthen this mechanism in the theory of change through specific input to support handover and continuity of understanding and ethos during periods of senior staff/P4C lead change.

The findings highlight some elements of the assumptions about metacognition in the approach used (in particular, using critical thinking and reasoning skills), but the specific contribution of metacognition to P4C within the theory of change needs strengthening. Most schools were indeed exposing pupils regularly to P4C, whether that be through standalone sessions at the beginning of the trial or more recently through embedding P4C within the curriculum. That said, there was some variation in delivery (for example in terms of the extent to which it was embedded across the curriculum) and this may be contributing to the lack of effects found.

Reflecting the strategies set out in the theory of change, findings from the evaluation highlight the importance of good-quality training and the ongoing support that trainers provide. The model of training and support provided by P4C was welcomed by schools and seen as vital in ensuring schools were able to embed P4C across the curriculum. It is worth noting that the Going for Gold progress indicators were not mentioned in the agreed theory of change. Our findings suggest that whilst P4C leads may have been aware of them, other teachers were not, and certainly in case study schools the indicators were not being widely used to monitor pupil progress and provide feedback (reasons being that it could be relatively easy to monitor a weekly standalone subject but less so when the approach is 'up to teachers' and embedded into the curriculum). That said, where they were being used for monitoring they were felt to be highly valuable. It is unclear how vital these indicators are to the theory of change (for example, are they a key mediator/mechanism for change?). Further work is needed on the theory of change in this regard, and in terms of supporting schools to engage with and use the indicators, if they are indeed important.

Moving on to the outputs expected in the theory of change, training take up was high and well received, and only three intervention schools withdrew from P4C implementation during the course of the trial. In terms of outcomes, this evaluation explored the number of schools achieving bronze, silver and gold awards. 40 of the 75 schools reached the expected level after two years, meaning a substantial minority did not (reasons related to school priorities, as well as to it requiring time to become familiar and confident with a new approach, in order to embed it across a school). The theory of change could benefit from being more explicit about the requirements for each level in schools' Going for Gold journeys, including having a measure of the frequency/usage of P4C, as well as the timescales for these achievements. It could be that two years was too short for schools to reach the expected level. The following year, about one in three intervention schools raised their award level, suggesting that it may take time to embed P4C across the school. This is also reflected in the teachers' reports of the benefits to pupils – where the benefits were only just being seen outside P4C lessons; so it may have been too early for these to be translated into measurable results.

The expected perceived outcomes for pupils, in terms of improvements in metacognition including reasoning skills and oracy, self-esteem, resilience and confidence, and behaviour (including tolerance and relationships) – were reported in varying degrees in the process evaluation. In particular, according to the teacher survey, teachers could see the impact of P4C on pupils in terms of their respect for other pupils' opinions (96% of teachers responding to the survey agreed or strongly agreed with this), their ability to question and reason (91% of teachers reported this) including outside of P4C sessions and their ability to express views clearly (93% of teachers reported this). The first of these is particularly interesting, given its similarity to the non-attainment item 'I can work with someone who has different opinions' where there was no evidence of an effect. Therefore, given the positive quantitative teacher-reported views on these outcomes, as well pupils' own qualitative views relating to these benefits, the theory of change for P4C needs greater specificity in terms of the non-cognitive impacts expected – so that a standardised instrument can be used to measure appropriate outcomes.

Whilst there were some comments about how P4C linked well with reading comprehension, there was very little comment from participants in the process evaluation that P4C was having a wider impact on pupils' attainment, or on the mechanisms by which that would occur. Implementation of P4C does not specifically cover reading and maths. Mechanisms in the theory of change appear limited in terms of attainment outcomes. These may need defining more precisely if implementation is to more directly affect attainment outcomes – perhaps with more emphasis on the pedagogical changes expected by P4C and how these might transfer from sessions to wider curriculum areas – including literacy and maths. Furthermore, if P4C is to affect attainment, SAPERE may wish to consider focusing P4C lesson stimuli more directly on specific subjects including reading and numeracy.

The theory of change highlights that teachers' practice should improve in relation to the 4Cs. The survey data and perceptions from interviewees support this. According to the survey, the majority of teachers understood the 4Cs, and felt that their teaching practice had improved particularly around confidence to try out new ideas and their relationships with pupils (arguably related to thinking in a creative, collaborative and caring way). However, there was little evidence from the process evaluation data that P4C was having the wider impacts set out in relation to parents – an area that

schools would need to work on for the gold stages of their Going for Gold awards. It may be that engaging with parents through and with P4C approaches takes more time to implement; and perhaps the planning stages for this need to be strengthened.

Interpretation

This project aimed to measure the educational and social skills improvement of pupils in primary schools after two years of P4C delivery. This evaluation, however, found no evidence that P4C had an impact on reading or maths for pupils in the whole cohort (Y6) or for those from disadvantaged backgrounds (Y6 FSM). There was also no differential impact on reading attainment for children with English as an additional language, and no difference in the reading outcomes of boys compared with girls. There was a greater but not significant effect for lower attaining pupils. The previous trial reported impacts on reading and maths attainment at Key Stage 2, including for children from disadvantaged backgrounds (Gorard et al., 2015), but this study, at a larger scale has not found evidence of impact on attainment. As discussed above, the mechanisms to support attainment outcomes were not strong in the theory of change, and, apart from some comments on aiding reading comprehension, teachers and pupils did not emphasise attainment outcomes in the process evaluation. Although Trickey and Topping's (2004) systematic review did find evidence of positive outcomes for reading, the UK programme is different to the original US programme, for example in its use of alternatives to philosophical novels. Furthermore, we note that had the data from Gorard et al. (2015) been analysed using a conventional ANCOVA or regression approach, such as that used in this study and recommended by EEF, it is unlikely to have demonstrated such a positive effect.

The P4C programme aims for children to develop social and communication skills relating to thinking in a more caring, collaborative, creative and critical way (the '4Cs'). Arguably, the secondary non-attainment outcomes for this trial were of greater relevance to the theory of change for P4C than the attainment outcomes, and indeed, of greater proximal relevance to children's development within the steps of the theory of change. The theory of change posits that through the programme's enquiry- and philosophy-based approaches to topics such as fairness, truth and bullying, children are supported to become more willing and able to ask questions, construct arguments, and engage in reasoned discussion. We suggested some validated scales to explore these areas; however, for continuity a survey instrument from a previous quasi-experimental study of non-cognitive impacts of P4C was used (Siddiqui et al., 2017). That study reported that the children in the P4C group were 'ahead' of those in the comparison group on two items 'I am good at explaining ideas to other people' and 'I can work with someone who has different opinions' (ibid). It also reported limitations, including imbalance between the intervention and comparison groups. However, when using the same items on a large scale and with randomised groups, we have not found evidence of an effect.

It could be that these single items were not sufficiently sensitive to capture what are quite complex character and metacognitive outcomes highlighted in the theory of change. However, the exploratory analysis of pupil survey data also did not yield any evidence of impact.

It is also possible that the backdrop within schools may have changed over the course of the trial, with an increasing focus on metacognitive approaches (EEF, 2016) as well as more emphasis on character education (Ofsted, 2019b) as part of usual classroom practice. Certainly a small number of the control schools reported using other enquiry-based programmes during the trial period, although the small number (seven) reporting this would not unduly affect the interpretation of the results. The lack of impact could also be related to maturation and age-sensitivity: authors of the previous evaluation using this instrument commented that 'the kinds of outcomes we [were] trying to assess such as wellbeing are very sensitive to age (usually worsening over time)' (Siddiqui et al., 2017). It could also be that implementation was not yet sufficiently embedded to influence these non-attainment outcomes (a substantial minority of schools had not implemented P4C to the level expected). However, given the positive perceptions from the process evaluation, we believe there may still be value in evaluating the effects of P4C using other, validated, instruments.

That said, the evidence from our process evaluation highlights that teachers and pupils found P4C to be suitable and engaging for all pupils, and indeed, particularly helpful at an individual level for children who were less self-confident, who were shy, or who had not previously experienced this kind of opportunity to share and discuss ideas. The majority of teaching staff who facilitated P4C believed that pupils liked the P4C approach (94% of surveyed teacher respondents agreed/strongly agreed with this; and 88% agreed/strongly agreed with the statement 'pupils are fully engaged with the P4C sessions'. Teachers particularly felt that P4C helped pupils to respect other pupils' opinions (96% of the 213 teachers surveyed agreed or strongly agreed with this), and pupils themselves emphasised that they enjoyed being able to express their own opinions in a non-judgemental way. These findings, particularly those of the teacher survey on

perceptions of outcomes for pupils, reflect the outcomes of P4C described in existing literature (e.g. Trickey and Topping, 2004).

In terms of sustaining the approach and embedding P4C into the life of the school, the theory of change highlights the importance of a whole school approach and an embedded ethos reflecting P4C values – in addition to ongoing weekly 60-minute sessions. It is therefore noteworthy, that only just over half of the intervention schools had reached the level of engagement expected in the P4C programme after two years of delivery (40 of the 75 schools), and indeed 11% (8 of the 75 schools) had not done any P4C/were not working towards any level/or had withdrawn from delivery early on. Our evaluation highlights that it takes time to develop and embed a new approach – for both teachers and pupils to become familiar and confident with P4C, in order to move from sessions on standalone topics to sessions linking to the curriculum. To support this, greater attention to frequency of sessions (akin to exposure to the approach or dose) could be considered in the award critiera. It could be that a longer time period is needed for implementation and embedding to reach 'Level 5' on the Going for Gold journey. That said, level of engagement was not associated with impact on reading attainment – the primary outcome for the trial. Unfortunately, it will no longer be possible to measure attainment in a cohort with greater experience of P4C due to the cancellation of Key Stage 2 assessments in 2020.

Furthermore, the theory of change and the process evaluation both emphasised the importance of senior leader support, not only as a moderating or contextual factor, but also in fact as a mediator – senior leaders' understanding and valuing of the P4C ethos was felt to be crucial for maintaining and embedding the P4C approach in the school. In addition, having a trained P4C lead to champion the approach was key to supporting and modelling the approaches (in both the theory of change and in the process evaluation data). It is therefore noteworthy that over half of the schools had a change in P4C lead during the course of the trial, and 1 in 7 had a change of headteacher. Given that senior leader/P4C stability is so essential to the theory of change, this must have had some bearing on the ability of schools to embed P4C sufficiently, and in turn, to achieve expected impact.

In terms of implementation of P4C, teachers and P4C leads particularly praised the training and support available through SAPERE for P4C. The P4C model used by SAPERE is readily scalable. Freelance trainers based across the country provide a mix of whole-school training based at the school, alongside a limited number of days ad hoc support. Higher level and top up training courses at centralised locations allow the team to adapt to changes as and when required. Our findings suggest that it may be important to strengthen any training on the use of the indicators in the Going for Gold programme – particularly amongst classroom teachers attending training, but also by ensuring P4C leads are trained to cascade these to their colleagues. Key advice from teachers for other schools wanting to participate in P4C included: ensuring schools receive good quality training with lots of practice examples, pairing with a similar school already experienced in delivery P4C to share good practice, starting early with pupils so that they become comfortable with the approach, running standalone sessions to begin with before using cross-curricular themes for P4C sessions, and ensuring the headteacher fully supports P4C. As above, it would seem that the stability and continuity of senior leadership and the P4C lead is paramount, to establishing and maintaining the P4C ethos in a school. Greater curriculum innovations may also be required in order to establish P4C within cross-curricular themes.

The **EEF Toolkit** highlights the benefits of programmes and approaches that support metacognition and character education. The aims of P4C appear to align well to these areas of the toolkit – particularly character education (which emphasises social interactions akin to working well with others' with different opinions). Indeed, the evidence from our process evaluation highlights the value of P4C to pupils' critical thinking, to their ability to question, and to being able to express their opinions in a non-judgmental way. However, whilst other metacognitive approaches might be quite specific to a particular skill (e.g. self-regulation), or to say a particular practice (e.g. retrieval practice), P4C requires a change in whole school ethos and curriculum innovation – a larger undertaking for schools (see for example, **EEF's guidance on managing whole school change**, which again highlights the importance of leadership and continued nurturing of the approach).

Limitations and lessons learned

This was an effectiveness trial, which examined whether the P4C intervention worked under everyday conditions in a large number of schools. The trial was a well-designed school-level randomised controlled trial and was well-powered to detect an effect on attainment for FSM children. Pupils in the intervention group were similar to those in the control group. The primary and secondary attainment outcomes used data accessed from the National Pupil Database (NPD). All schools' data was used in the attainment analysis. Only 6.75% of pupils did not have test scores for both Key Stage

1 reading and Key Stage 2 reading. As this data is from statutory age-related assessments, there is no reason to suspect any biased reason for missing test scores so no missing data analysis was necessary.

It is worth noting that the compliance measure used for this trial was a measure of the extent to which schools engaged with, delivered and embedded P4C across the whole school. This reflected the nature of the intervention and SAPERE's progress criteria that their trainers use with schools. No pupil level engagement data was collected, as this was an intervention about whole-school practice/embedding. The design was a cohort analysis, of FSM-eligible pupils in Year 6 in trial schools in 2019, to explore learning outcomes in schools after two year of teachers' implementing and emedding the P4C approach in the school. A limitation in this design is that pupils joining the school since 2017 (Year 4) may not have experience as much P4C as other children in the school, and joining date was not recorded. Should SAPERE need to understand class level or even pupil-level exposure to P4C then future research could look at class level or pupil-level dosage, for example the number of hours per week that P4C is delivered to the class and the number of hours in total.

As noted earlier, the award level used in the analysis included both awards achieved and trainers' assessments of level of progress. Whilst there may be some element of trainer influence on assessed levels, as SAPERE's trainers work closely with schools they are in a position to assess the level each school is working at even if the school has not completed an award application. This approach ensured all schools had a level of progress recorded, even if they had not completed the submission application. That said, SAPERE may want to consider developing their award criteria further by outlining further quality assurance and monitoring procedures (to minimise any trainer bias in awarding levels), and setting out measurable criteria for each level alongside the existing descriptive indicators.

The non-attainment secondary outcomes for the study were measured using the same single self-reported survey items from a previous quasi-experimental study of P4C ('I am good at explaining my ideas to other people' and 'I can work with someone who has different opinions') (Siddiqui et al., 2017), and had shown positive results. Whilst using the same instrument in this effectiveness trial had the potential to measure non-cognitive at a greater scale (over 6,000 pupils), using only single items to reflect what are quite complex metacognitive and character-related outcomes is a weakness in the design. At the time of designing the trial, other scales were considered (i.e. SSIS-RS multi-rater tool21, Harter22 with sub-scales relating to social competence and behavioural competence, and the Chi instrument23; but as none of these specifically captured what the P4C theory of change proposed, it was decided by SAPERE, EEF and the evaluation team to use the instrument from the previous study of P4C in order to test at greater scale. This evaluation has shown that P4C is a rich and complex approach, and finding a scale that can better capture the elements of the 4Cs is important. Similarly, the theory of change needs to be more specific about the outcomes to capture.

The process evaluation was designed to include a focus on the progress indicators and their use. It transpired that whilst P4C leads were aware of them, teachers were less aware, and schools were not widely using them to monitor and review progress or provide feedback to pupils. As such, data on the second aim for the process evaluation was limited.

Future research and publications

Future work needs to develop the theory of change to focus on the outcomes teachers reported in the process evaluation (such as communication, reasoning, relationships), as well as the 4Cs, rather than on literacy and numeracy. Indeed, it could be argued that reading and numeracy outcomes are only peripheral to P4C, and future trials may wish to consider more suitable outcome measures. In addition, differentiating expected P4C outcomes from those of other enquiry-based or metacognitive approaches could be important in further developing the theory of change.

Future evaluation should seek to use a standardised non-cognitive instrument or scale to measure such outcomes. In order to select such a scale, greater clarity is needed around the mechanisms of change for P4C and the specific contribution of P4C to metacognitive outcomes, and nuances in children's development associated with P4C (particularly around the 4Cs). Future research should also explore in greater detail the role of the headteacher and P4C lead, given how crucial they appear to have been in this research (both in terms of their presence, but also given there were so many P4C lead changes over the course of the trial); and P4C support be developed to ensure continuity when

---

[21] **http://www.pearsonclinical.com/education/products/100000322/social-skills-improvement-system-ssis-rating-scales.html**
[22] **https://portfolio.du.edu/downloadItem/221383**
[23] **http://digitalcommons.unomaha.edu/cgi/viewcontent.cgi?article=1009&context=slceciviceng**

staff changes do occur. Future evaluation should also consider the timelines for implementation and embedding P4C in schools, and seek to align exploration of outcomes at the most relevant time in schools' implementation journey (two years may be too short). The P4C Going for Gold programme might also need to re-adjust expectations, set out straightforward and measurable criteria for each award level, and ensure these are cascaded to all teachers in P4C schools.

References

Angrist, J.D. and Imbens, G.W. (1995). 'Two-stage least squares estimation of average causal effects in models with variable treatment intensity', *Journal of the American Statistical Association,* **90**, 430, 431-442. [online]. Available: http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1995.10476535   [19 May, 2017].

Christensen, R. H. B. (2019). R*egression Models for Ordinal Data.* Vienna: The R Foundation [online]. Available: https://cran.r-project.org/web/packages/ordinal/index.html [28 April, 2020].

Education Endowment Foundation (2013). *Pre-Testing in EEF Evaluations.* London: EEF Available: https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/Pre-testing_paper.pdf   [28 April, 2020].

Education Endowment Foundation (2016). *Meta-Cognition and Self-Regulation.* London: EEF [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Publications/Metacognition/EEF_Metacognition_and_self-regulated_learning.pdf [28 April, 2020].

Education Endowment Foundation (2018). *Statistical Analysis Guidance for EEF Evaluations*. London: EEF [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Grantee_guide_and_EEF_policies/E valuation/Writing_a_Protocol_or_SAP/EEF_statistical_analysis_guidance_2018.pdf [27 March, 2020].

Gorard, S., Siddiqui, N. and See, B.H. (2015). *Philosophy for Children Evaluation Report and Executive Summary*. London: EEF [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/EEF_Project_Report_PhilosophyForChildren.pdf [27 March, 2020].

Jiang, Y. and Small, D. (2014). *Ivpack: Instrumental Variable Estimation'. R package version 1.2.* Vienna: The R Foundation [online]. Available: https://CRAN.R-project.org/package=ivpack [27 March, 2020].

Lipman, M. and Bierman, J. (1970). 'Appendix B: Experimental research in philosophy for children.' In: Lipman, M., Sharp, A.M. and Oscanyo, F. (Eds) *Philosophy in the Classroom.* Philadelphia, PA: Temple University Press.

Lord, P. and Styles, B. (2019). *Philosophy for Children- Evaluation Protocol (amended).* London: EEF [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Philosophy_for_Children_-_Evaluation_protocol_(amended).pdf [28 April, 2020].

Education Endowment Foundation (2016). *Technical Appendix: Meta-Cognition and Self-Regulation.* London: EEF [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Toolkit/Technical_Appendix/EEF_Technical_Appendix_Meta_Cognition_and_Self_Regulation.pdf [9 December, 2016].

The Office for Standards in Education, Children's Services and Skills (2019a). *The Education Inspection Framework*. London: OFSTED [online]. Available: https://www.gov.uk/government/publications/education-inspection-framework [26 February, 2020].

The Office for Standards in Education, Children's Services and Skills (2019b). *Character Education Framework.* London: OFSTED [online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/849654/Character_Education_Framework_Guidance.pdf [28 April, 2020].

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D. and R Core Team (2019). *'nlme: Linear and Nonlinear Mixed Effects Models', R Package Version 3.1-142.* Vienna: The R Foundation [online]. Available: https://CRAN.R-project.org/package=nlme [27 March, 2020].

The R Foundation (2019). *R: A Language and Environment for Statistical Computing' R Foundation for Statistical Computing.* Vienna: The R Foundation [online]. Available: https://www.R-project.org/ [27 March, 2020].

Roy, P., Rennie, C., Dirie, A. (2019). *Statistical Analysis Plan for Philosophy for Children.* London: EEF [online]. Available: https://educationendowmentfoundation.org.uk/public/files/Projects/EEPS_Statistical_Analysis_Plan_AMENDED_2019 0509.pdf [27 March, 2020].

Sasseville, M. (1994). 'Self-esteem, logical skills and philosophy for children', *Thinking*, **11**, 2, 30 – 32 [online]. Available: https://www.pdcnet.org/thinking/content/thinking_1994_0011_0002_0030_0032 [27 March, 2020].

Society for the Advancement of Philosophical Enquiry and Reflection in Education. (2020). *P4C and the Curriculum* [online]. Available: https://www.sapere.org.uk/about-us/p4c-and-the-curriculum.aspx [28 April, 2020].

Siddiqui, N., Gorard, S. and See, B.H. (2019). 'Can programmes like Philosophy for Children help schools to look beyond academic attainment?'. *Educational Review,* **71**, 1, 146-165 [online]. DOI 10.1080/00131911.2017.1400948.

Siddiqui, N., Gorard, S. and See, B.H. (2017). *Non-Cognitive Impacts of Philosophy for Children, Project Report*. Durham: School of Education, Durham University [online]. Available: http://dro.dur.ac.uk/20880/1/20880.pdf?DDD34+DDD29+czwc58+d700tmt [27 March, 2020].

Trickey. S. and Topping, K. J. (2004). 'Philosophy for Children: a Systematic Review', *Research Papers in Education.* **19**, 3, 365-380 [online]. DOI 10.1080/0267152042000248016.

Appendix A: EEF cost rating

Appendix Figure 1: Cost Rating

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

OUTCOME: *Reading KS2, FSM-eligible pupils*

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	MDES	Attrition					
5 🔒	Randomised design	<= 0.2	0-10%					5 🔒
4 🔒	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%					
3 🔒	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **[0]**		
2 🔒	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1 🔒	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0 🔒	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	Large-scale randomised trial, little evidence of baseline imbalance.
Threat 2: Concurrent Interventions	Low	Some very limited evidence of control schools delivering similar interventions. However, the mechanism of pupil talk and metacognition might be more prevalent in the English school system and therefore part of the 'background' control condition. This was considered in the interpretation of the results.
Threat 3: Experimental effects	Low	High-stakes tests used for the main analysis, so no reason to think that this would be affected by taking part in the trial
Threat 4: Implementation fidelity	Moderate	Fidelity was mixed, but this was an effectiveness trial; no evidence that fidelity was less than could possibly be expected if the programme was rolled out; the CACE results found no evidence that better implemented programmes had different effects. More detail on fidelity/ compliance measures would have been helpful in the report.
Threat 5: Missing Data	Low	Very little missing data.
Threat 6: Measurement of Outcomes	Moderate	The main outcomes have no threat, but the secondary 'process' outcomes were based on single items and not validated. A concern is that many pupils said 'I don't know' in response to those single items, which contains almost no information. There is also some concern over alignment of outcomes to the logic model. However, these risks do not concern the primary analysis.
Threat 7: Selective reporting	Low	Highly adherent to the protocol and SAP.

- **Initial padlock score:** 5 padlocks (well designed trial, powered to detect change in FSM cohort; low attrition)

- **Reason for adjustment for threats to validity**: N/A (two moderate threats to validity; direction of likely biases unknown – no adjustment required)
- **Final padlock score:** initial score adjusted for threats to validity = 5 padlocks

Appendix C: Changes since the previous evaluation[24]

Appendix table 1: Changes since the previous evaluation

	Feature	Previous trial to effectiveness re-grant stage
Intervention	Intervention content	In the previous trial schools committed to facilitating at least one hour per week of P4C for all children at Key Stage 2. In addition to this commitment, in the effectiveness trial intervention schools worked through the Going for Gold awards scheme.
	Delivery model	Schools in both trials received the same initial training of 2 days for up to 25 staff members and 4 days of additional training for 2 members of staff. In the previous trial schools received 9 days of support across 2 years. In the effectiveness trial schools received 12 days of support across 3 years. Schools in the effectiveness trial also received 1 day of Thinking Together training for up to 25 staff and 4 top up places for level 1 training for new teachers.
	Intervention duration	The previous trial lasted two years while the effectiveness trial lasted three years.
Evaluation	Eligibility criteria	Geographic location of eligible schools extended from 5 regions (Northeast, Northwest, Southeast, Southwest, and the Midlands) to the whole of England.
	Level of randomisation	No changes
	Outcomes and baseline	**Primary outcome** narrowed down from KS2 Reading, Writing and Maths results for all pupils in Y5 at baseline (previous trial) to KS2 Reading score for FSM-eligible pupils in Y6 at endpoint (re-grant effectiveness stage). **Baseline for Primary outcome** narrowed down from KS1 Reading, Writing and Maths results for all pupils in Y5 at baseline (previous trial) to KS1 Reading score for FSM-eligible pupils in Y6 at endpoint (re-grant effectiveness stage). **Secondary outcomes** changed from cognitive abilities test scores for all pupils in Y5 and Y6 at endpoint, measured through CAT4 test, (previous trial) to the following for the re-grant effectiveness stage:

[24] Please delete this section if it is not applicable.

		- KS2 Reading score for all pupils in endpoint cohort - KS2 Maths score for FSM-eligible pupils in endpoint cohort KS2 Maths score for all pupils in endpoint cohort - Two items for all pupils in endpoint cohort, measured at endpoint through a bespoke survey ('I am good at explaining my ideas to other people' and 'I can work with someone who has different opinions'). **Baseline for Secondary outcomes** changed from cognitive abilities test scores for all pupils in Y4 and Y5 at baseline, measured through CAT4 test, (previous trial) to the following for the re-grant effectiveness stage: - KS1 Reading score for all pupils in endpoint cohort - KS1 Maths score for FSM-eligible pupils in endpoint cohort - KS1 Maths score for all pupils in endpoint cohort - Two items for all pupils in baseline cohort, measured at baseline through a bespoke survey ('I am good at explaining my ideas to other people'; and 'I can work with someone who has different opinions') ○
	Control condition	No changes - "Business as usual" control condition in both the previous trial and in the re-grant effectiveness stage. The incentive structure changed from pure waitlist design (control schools receive P4C after the end intervention - no cost nor incentive) in the previous trial, to a design at re-grant effectiveness stage where control schools received a monetary incentive equivalent to the discount guaranteed to intervention schools, which could be used to implement P4C in the school after the end of the intervention (except with Y6 pupils in 2019/20).

Appendix D: Theory of Change

Overall purpose of the evaluation

To evaluate the impact of Philosophy of Children (P4C) on KS2 reading and maths scores, and social competencies for EverFSM pupils.

Purpose of the P4C intervention

To improve pupils' and teachers' capability to think in a caring, collaborative, creative and critical way ('the 4Cs') in order to support their personal, social and educational development.

Assumptions

Extensive research (EEF, 2016) shows a link between meta-cognition and improved attainment, particularly in maths and English but also in other areas of the curriculum. This assumption underpins the evaluation.

For P4C to achieve success certain in-school conditions must be met, including:
- committed senior leaders and teachers
- appointing a P4C lead to champion P4C in the school
- stability of the school and its leadership team
- pupils being exposed to regular P4C sessions (e.g. weekly)
- school culture reflect P4C's ethos to support pupils and teachers to develop the 4Cs.

Approved trainers must deliver an ongoing programme of support to schools.

Target Groups

- Schools: with 25 per cent of EverFSM pupils and who have not previously implemented P4C
- Pupils: KS2 EverFSM
- Teachers: other beneficiaries of P4C.

Impact, e.g. I would expect P4C to make the following difference(s):

Outputs:
- Number of schools that received training
- Number of training and coaching days accessed by schools
- Number of teachers trained at each level (Level 1 and 2)
- Number of sessions delivered to pupils
- P4C leader in place who has attended advanced training and offers support to colleagues in each school.

Outcomes:
- Number of teachers achieving practitioner certificate and self-reported changes as a result of practising P4C on teaching practices and classroom management
- Number of schools achieving bronze, silver and gold awards
- Improvement in pupils' meta-cognition including reasoning skills and oracy. Improvement in self-esteem, resilience and confidence, and behaviour (including tolerance and relationships).

Impact:
- Attainment measured through higher KS2 scores in reading and maths when compared to a control group.
- Pupils' social competencies measured through higher scores on the social skills questionnaire when compared with a control group.
- Teachers' perceived impact of the 4Cs on themselves and their teaching practices.

Wider impact:
- Impact on the whole-school community, e.g. teachers, senior leaders, non-teaching staff, parents/carers.

Strategies

What is the P4C approach?

Programme of support over three years consisting of:
- 2 days of Foundation 1 Training (Level 1) for up to 25 staff
- 1 day of P4C Tools for Thinking Together Training for up to 25 staff
- 4 days of Advanced P4C Training (Level 2A and 2B) for two staff
- 7 days of in-school P4C coaching and support (approx. 3 days year 1, 3 days in year 2 and 1 day in year 3)
- 5 days of remote P4C accredited trainer administration and planning support
- Up to 4 top up places for schools on open Level 1 courses for new teachers joining the schools during the programme.

Resources:
- Unlimited access to SAPERE's online P4C resources and practice guides
- Two reference copies of SAPERE's Level 1 and Level 2 handbooks
- Other resources available through individual

Education Endowment Foundation (2016). *Technical Appendix: Meta-Cognition and Self-Regulation.* London: EEF [online]. Available: **https://educationendowmentfoundation.org.uk/public/files/Toolkit/Technical_Appendix/EEF_Technical_Appendix_Meta_Cognition _and_Self_Regulation.pdf** [9 December, 2016].

Appendix E: MoUs, letters and privacy notices

Agreement to participate in the Evaluation of the Philosophy for Children Programme

Memorandum of Understanding (MoU)/Information sheet

Please sign both copies of this MoU, retaining one and returning the second copy to NFER in the pre-paid envelope provided. Alternatively, please scan and email this agreement to P4C@nfer.ac.uk

School name: «Description»

Aims of the evaluation

The Education Endowment Foundation (EEF) has commissioned the National Foundation for Educational Research (NFER) to evaluate the effectiveness of the Philosophy for Children (P4C) programme.

The evaluation will assess:

- whether P4C is effective in improving pupil's reading and maths performance and social skills

- how P4C has been implemented in schools.

What is Philosophy for Children and how does it work?

P4C is a teaching method which centres on philosophical enquiry and dialogue. It was introduced in the UK in the 1990s, and has been adopted by hundreds of schools. P4C seeks to develop pupils' independent thinking, reasoning, communication and collaboration skills. It aims to develop intellectual dispositions, such as curiosity and clarity of expression, as well as emotional intelligence. The impact of P4C on attainment and social skills is perceived by SAPERE as a valuable by-product rather than being seen as the programme's primary aim. Please see the P4C website for further information: http://www.sapere.org.uk/

Information for all schools

All schools that agree to participate in the evaluation will be randomly assigned to one of two groups: the intervention group or the control group. Schools will find out which group they are in on Monday 30th January 2017. Schools allocated to the intervention group will receive P4C training and support over three years, with the intention of reaching SAPERE's Going for Gold Award. Between March and September 2017, SAPERE will deliver initial training during INSET days to up to 25 teaching staff on dates of your school's choice. Schools allocated to the control group will continue 'business as usual' and will be required to commit to not participating in P4C until September 2019. Control schools will receive a financial incentive of £5,700 for participating in the research.

An important aspect of the evaluation is to measure pupils' performance in reading and maths in both the control and intervention groups. For this evaluation we will use the pupils' key stage 2 results in reading and maths as the main outcome measure. With your permission we would like to access the pupils' end of key stage 2 national curriculum assessment results when they become available from the National Pupil Database in summer 2019 and 2020.

In order for us to do this, you will need to provide NFER with a list of the names, dates of birth and Unique Pupil Numbers (UPNs) of the pupils currently in year 4. We will provide you with a template for completion which you will need to share with us using NFER's secure portal. Please be assured that this process is very easy to complete and NFER staff will be happy to help you with any queries during the process.

We will also measure the impact of P4C on pupils' social skills and require you to administer a short questionnaire to year 4 pupils by 20th January 2017. In June 2019, an NFER test administrator will arrange to visit your school to administer the follow-up questionnaire.

For schools randomly assigned to the intervention group

SAPERE will provide training and assistance to help schools implement the programme. Intervention schools must commit to:

- releasing up to 25 teachers for one day's training during the 2016/17 academic year and for two further INSET days training in 2017/8
- all Key Stage 1 and 2 teachers to be trained in P4C
- introducing weekly P4C sessions of around 45 minutes from September 2017 onwards, at least for Year 4-6 classes
- designating one or more teachers a 'P4C leader' and releasing them for Advanced P4C training at various points in the programme
- providing pupil data for year 4 pupils (name, date of birth and Unique Pupil Number)
- administering a social skills questionnaire to Year 4 pupils by 20th January 2017. The questionnaire will be repeated in June 2019.
- contributing towards the annual cost of P4C: £1,600 per year for three years (a total of £4,800).
- writing P4C Going for Gold programme into the school's strategic plan.

An information sheet is enclosed which provides more details about the activities schools will be asked to undertake. A video providing an overview of P4C and its benefits can be accessed online at
https://vimeo.com/121408660

A second video showing P4C in action with year 6 pupils can be seen at
https://vimeo.com/113524022

For schools randomly assigned to the control group

Control schools must commit to:

- maintaining a 'business as usual' approach: teaching pupils as normal and not using P4C materials until September 2019
- designating one teacher to be NFER's key contact throughout the duration of the trial
- providing pupil data for year 4 pupils (name, date of birth and Unique Pupil Number)
- administering a social skills questionnaire to Year 4 pupils by 20th January 2017. The questionnaire will be repeated in June 2019.
- receiving a financial incentive of £5,700 at the end of the 2019 summer term.

Teacher questionnaires

We would like to ask teachers to take part in a short questionnaire in summer term 2019.

How your data will be used

The pupil information you provide, including questionnaire responses, will be linked with the National Pupil Database (held by the Department for Education) and other official records. Named data will be shared with SAPERE, the Department for Education, Education Endowment Foundation (EEF), EEF's data contractor FFT Education and stored anonymously on the UK Data Archive. Please note that no individual schools or pupils will be identified in any reports arising from the research.

We will send all participating schools an information letter for the parents/carers of the pupils involved in the evaluation. The letter will provide the option for parents/carers to withdraw their child/ren from the evaluation should they wish to do so.

Is your school eligible to participate in the evaluation?

Your school is one of a randomly selected sample of primary schools from across England invited to take part in this evaluation. We are asking all schools in the sample if they would like to take part, although any school which has already completed the P4C programme with their whole school would not be eligible.

How to take part in the evaluation

Schools that would like to take part in the evaluation should complete the enclosed reply form and this MoU and return it to NFER in the pre-paid envelope provided (or scan and email it to P4C@nfer.ac.uk). On the reply form please nominate a member of staff with whom NFER and SAPERE can liaise with about the evaluation. The reply form also asks you to indicate in which half term you would prefer for your initial full day

training to take place during the 2016/17 academic year. Should your school have any queries about the evaluation, please contact our helpdesk on 01753 637096 or by email to **P4C@nfer.ac.uk**

Key dates for the evaluation

Key Dates	Activities for intervention schools	Activities for control schools
October 2016	NFER invites a random sample of schools to participate in the trial.	
October – December 2016	**Schools agree to participate in the trial by returning their reply form and signed MoU** Once a school has returned their reply form and signed MoU, NFER will contact the person named on the reply form via phone to confirm their participation in the trial. At this point NFER will answer any questions and explain the next steps relating to the sharing of pupil data and the administration of the social skills questionnaire.	
November – December 2016	After speaking with the named teachers, NFER will write to each participating school providing them with instructions for sharing pupil data for their current year 4 classes via NFER's secure portal.	
Beginning of January 2017	NFER will send schools the social skills questionnaire which all year 4 pupils should complete. Schools must return all surveys to NFER via pre-paid secure delivery by 20th January.	
End of January 2017	Schools that have returned a signed reply form and MoU, along with pupils' data and completed social skills questionnaire will be eligible to participate in the evaluation. Schools are randomly assigned to the intervention group or control group.	
30th January 2017	NFER writes to all intervention schools letting them know they will receive the P4C programme and that a SAPERE representative will contact the schools to set up the programme.	NFER writes to all control schools letting them know that they have been allocated to the control group and that they should maintain a 'business as usual' approach.
February – September 2017	SAPERE contact schools to book the training sessions based on school's preferred INSET dates (as supplied on the reply form) SAPERE provides initial full day of P4C Foundation training for all participating staff.	Schools maintain a 'business as usual' approach.
September 2017 – July 2019	Schools implement weekly P4C sessions; SAPERE provide further training and support schools to reach to Bronze and Silver Awards.	Schools maintain a 'business as usual' approach.
June 2019	NFER analyses Key Stage 2 reading and maths results for intervention and control schools using data from the National Pupil Database. NFER test administrator visits schools to administer the social skills questionnaire to pupils (this is the same questionnaire as the pupils completed in autumn 2016). Teachers will be asked to complete a short questionnaire.	
July 2019		Receive financial incentive.

Autumn 2019		For schools that choose to implement P4C, SAPERE start to deliver training with all teachers **except** year 6.
Autumn 2019 – Summer 2020	Schools follow Going for Gold programme up to Gold level.	For schools that choose to implement P4C, P4C weekly session are introduced to all year groups, **except** year 6.
June 2020	NFER analyses the Key Stage 2 reading and maths results for year 6 pupils in the intervention and control schools using data from the National Pupil Database.	

We commit to the Evaluation of Philosophy for Children as detailed above.

Signed:..   Date:..................................

Name:..

Parent letter

December 2016

Dear Parent /Guardian

Research into the Philosophy for Children (P4C) Programme

We are writing to you to let you know that your school has asked to be part of our research into the Philosophy for Children (P4C) programme. P4C is way for children to learn about new and different things. P4C aims to help children talk and listen to others in a kind and caring way. It shows children how to work as part of a team as well as being able to think about and share their own views with others. It enables children to make choices, be creative and think about their own questions to talk to others about.

As part of the research, we are trying to find out whether P4C makes a difference to children's maths and reading. The only way to find out if P4C really makes a difference is for some schools to use P4C and for other schools not to use it. All schools that have asked to be part of the research will be randomly chosen (e.g. names picked out of a hat) to either take part in the P4C programme or not. Schools that will be doing P4C will teach one P4C lesson a week.

NFER (the organisation that has been asked to carry out the research) is only interested in pupils who will be in Year 5 in September 2017. We will be asking these pupils who are in Year 4 now to complete a short questionnaire survey in January 2017. We will ask these pupils to do the questionnaire survey again in June 2019. As part of the research, schools will also give us the name, date of birth and 'unique pupil number' (UPN) for each pupil taking the questionnaire survey and the P4C programme.

If you would prefer for your child **not** to take part in the research, please inform their teacher or complete the form below. If you are happy for your child to take part in the research, you **do not** need to return the reply slip. Your child may withdraw from the research at any time. Please inform their teacher if you or child would like to withdraw from P4C or the research at a later stage.

Pupils' questionnaire survey responses and any other information collected as part of the research will be treated with the strictest confidence. NFER will match pupil names, dates of birth and UPNs to the National Pupil Database[25] to collect other information about the pupil such as whether English is the main language used at home. At the end of the research, this information will also be shared with SAPERE (the company that runs P4C), the Department for Education, the Education Endowment Foundation (EEF), who are paying for the research, and EEF's data contractor, FFT Education, where the data will be held anonymously in the UK Data Archive. Please note that no individual children will be identified in any reports arising from the research.

If you would like more information about the research, please contact me or my colleague, Michael Neaves, on 01753 637014 or email **P4C@nfer.ac.uk**.

---

[25] Every child is given a UPN when they first start school. This is a number which stays with the child until they leave school at age 16/18.

Yours sincerely

Kathryn Hurd

Head of Survey Operations

Research and Product Operations

Research into the Philosophy for Children (P4C) Programme

I would not like my child to take part in this research.

Print child's full name: _____

Parent's signature: _____   Date: _____

NFER code: EEPS

We would be grateful if you would briefly outline your reasons for not wishing your child to participate in this study:

Privacy Notice for Philosophy for Children

1. Why are we collecting this data?

The Education Trials Unit at the National Foundation for Educational Research (NFER) is collecting personal data to enable the evaluation of Philosophy for Children (P4C) using a Randomised Controlled Trial (RCT). Philosophy for Children is an approach to teaching in which students participate in group dialogues focused on philosophical issues (e.g. 'truth', 'fairness' or 'bullying'). The main aim of P4C is to improve pupils' and teachers' capabilities to think in a caring, collaborative, creative and critical way ('the 4Cs') in order to support their personal, social and educational development.

The RCT aims to ascertain the impact of P4C on disadvantaged children's reading and maths at Key Stage 2, and on their social competencies/skills. The disadvantaged children involved are those who were in Year 4 in 2016/17 who have ever been in receipt of free school meals (known as EverFSM). The RCT will also explore the impact on reading and maths at Key Stage 2 for two cohorts of children – those who were in Year 4 in 2016/17 and those who were in Year 3 in 2016/17.

The Education Endowment Foundation (EEF) has commissioned the Society for the Advancement of Philosophical Enquiry and Reflection in Education (SAPERE) to develop and deliver the P4C programme. The EEF has commissioned and funded the NFER to undertake the independent evaluation of P4C. NFER is the Data Controller for this evaluation.

2. What is the legal basis for processing activities?

The legal basis for processing personal data is covered by:

- GDPR Article 6 (1) (f) which states that 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

Our legitimate interest for processing personal data for this trial is to administer the RCT and analyse all data in this RCT.

3. How will personal data be obtained?

Personal data will be obtained by NFER in the following ways:

- from the Department for Education's (DfE) National Pupil Database (NPD) using DfE's secure data exchange portal and the Office for National Statistics secure access procedures
- on the school memorandum of understanding (MoU) for the trial
- directly from schools, via NFER's secure school portal
- through pupil paper surveys, administered by schools at baseline and by NFER test administrators at follow-up26)
- directly from teachers, via an online survey
- through telephone interviews and case study visits with school senior leaders, teachers, pupils and SAPERE staff

---

[26] NFER will ensure that appropriate contracts are in place for NFER test administrators, and that they obtain personal data in accordance with the GDPR and other applicable legislation.

- through updates of contact details provided by the P4C lead and other teachers, and shared securely between NFER and SAPERE using NFER's secure portal.

Note, NFER is also collecting information from SAPERE about programme delivery. This data will capture school names/school IDs only – it will not include personal data.

4. What personal data is being collected by this project?

Personal data for this RCT includes data about pupils and teachers from participating schools as described below:

- NFER is collecting the name, job title and contact details about the nominated P4C lead teacher within participating schools so that we can liaise with them about the evaluation.

- The personal data we will collect about pupils includes pupil name, date of birth, unique pupil number (UPN) and school name for all pupils in Year 4 (in 2016/17). This will be matched to data from the DfE National Pupil Database (NPD), including pupils' free schools meals status (EverFSM).

- Attitudinal survey responses will be collected on a pupil baseline and follow-up pupil survey. The personal information that pupils provide to us in response to the survey will be used in the RCT.

In addition, NFER will also access and process the following **anonymised pupil data** for this RCT:

- Key Stage 1 and Key Stage 2 results (Reading and Maths scores) for the 2019 Year 6 cohort (i.e. those who were in Y4 in 2016/17), and for the 2020 Year 6 cohort (i.e. those who were in Y3 in 2016/17) – from the DfE NPD.

- Background data such as EverFSM status, month of birth/age in months, gender, and EAL status, where needed for the analysis.

Other personal data being collected includes staff attitudes and views as follows:

- Attitudinal survey responses will be collected on a teacher survey. The personal information that teachers provide to us in response to the survey will be used in the RCT.

- Attitudes and views of school staff and of SAPERE staff will be collected through telephone interviews and case studies; contact details will be collected to allow us to organise these.

5. Who will personal data be shared with?

NFER will share the names, job titles and contact details of the P4C leads in each school with SAPERE, so that they can be contacted about the programme. If your P4C contact details change during the evaluation, SAPERE and NFER will share and update your contact information so that we can continue to contact you about the programme and the RCT.

For the purposes of the research, pupils' names, dates of birth, UPNs and questionnaire responses will be linked with information about the pupils from the National Pupil Database (NPD). In addition, NFER will request the following pupil data from the DfE: anonymised Key Stage 2 Reading and Maths assessment results, KS1 attainment data, and anonymised background data including EverFSM status, month of birth and EAL status where needed for the analysis. Anonymised data for all pupils involved in the trial will also be shared with the Office for National Statistics and potentially other research teams, and stored in the UK Data Archive. Further matching to NPD and other administrative data may take place during subsequent research.

We will not share personal data collected through telephone interviews, case studies or in the teacher survey with other organisations.

All data will be treated with the strictest confidence in line with the GDPR 2016/679 and the Data Protection Act 2018. **No individuals or schools will be named in any report arising from this research.**

6. Is personal data being transferred outside of the European Economic Areas (EEA)?

No personal data for this RCT is being stored or transferred outside of the EEA.

7. How long will personal data be retained?

NFER will delete any personal data within one year from the completion of the project

NFER will send all the pupil data it has collected for this RCT to the Fisher Family Trust for archiving within three months of project completion, at which point EEF will take responsibility for data protection compliance.

8. Can I stop my personal data being used?

NFER handles your personal data in accordance with the rights given to individuals under data protection legislation. If at any time you wish us to withdraw your data or correct errors in it, please contact Guido Miani at P4C@nfer.ac.uk

In certain circumstances, data subjects have the right to restrict or object to processing. They also have the right to make a subject access request to see all the information held about them. To exercise any of these rights, please contact our Compliance Officer.

9. Who can I contact about this project?

NFER is responsible for the day-to-day management of this project. Contact Guido Miani at P4C@nfer.ac.uk at NFER with any queries.

If you have a concern about the way this project processes personal data, we request that you raise your concern with us in the first instance (see the details above). Alternatively, you can contact the Information Commissioner's Office, the body responsible for enforcing data protection legislation in the UK, at https://ico.org.uk/concerns/.

10. Last updated

We may need to update this privacy notice periodically so we recommend that you revisit this information from time to time. This version was last updated on 4th April 2019.

Privacy notice for parents

1. Why are we collecting your child's data?

Philosophy for Children (P4C) is an approach to teaching in which pupils take part in group discussions focused on philosophical issues (e.g. 'truth', 'fairness', 'bullying'). P4C aims to help pupils and teachers to think in a caring, collaborative, creative and critical way. It aims to support pupils' personal, social and educational development.

Your child/ren's school is involved in a Randomised Controlled Trial (RCT) to evaluate P4C. The evaluation is exploring what difference P4C makes to children's reading and maths at Key Stage 2, and to their social skills. The RCT involves children who are currently in Year 6 (in the school year 2018/19), and those who will be in Year 6 next year (school year 2019/20). The RCT includes a focus on disadvantaged children.

The Education Endowment Foundation (EEF) has commissioned the study. The Society for the Advancement of Philosophical Enquiry and Reflection in Education (SAPERE) is overseeing the P4C programme. The NFER is carrying out the RCT evaluation of P4C. NFER is the Data Controller – this means they are in charge of what happens to your child's data during the evaluation.

2. What personal data is being collected by this project?

The RCT includes children's Key Stage 1 and Key Stage 2 results. It also involves short questionnaires which children will fill in at the beginning and end of the trial. Children may also be invited to take part in interviews, to tell researchers about their experiences of taking part in P4C.

NFER will collect the following **personal data** about children who were in Year 4 at the start of the trial and are in Year 6 this year (2018/19), for a pupil questionnaire:

- Your child's name, date of birth, unique pupil number (UPN) and school name.

- This will be matched to background data from the DfE National Pupil Database (NPD) such as whether they have Free Schools Meals (EverFSM).

- Pupils' responses to a social skills questionnaire – at baseline 2016/17 and follow-up (March 2019).

NFER will also collect the following **anonymous (i.e. unnamed) data** about children who are in Year 6 this year (2018/19) and children who will be in Year 6 (2019/20):

- Unnamed Key Stage 1 and Key Stage 2 results in Reading and Maths for each year group.

- Unnamed background data held on the DfE NPD including: whether they have Free School Meals (EverFSM), month of birth, and whether English is an Additional Language (EAL) for them.

Our purpose for processing your child's personal and anonymised data for this trial is to assess children's reading, maths and social skills outcomes. We will also see if the programme is beneficial for those from disadvantaged backgrounds by looking at the results for children who have free school meals, and for those with English as an additional language.

3. What is the legal basis for processing activities?

The legal basis for processing your child's personal data is covered by:

- GDPR Article 6 (1) (f) which states that 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interest are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of the personal data'.

Our legitimate interest for processing personal data for this trial is to carry out the evaluation and to find out if P4C makes a difference to children's learning and social skills.

4. How will personal data be obtained?

Your child's personal data will be obtained by NFER in the following ways:

- from the Department for Education's (DfE) National Pupil Database (NPD)
- on pupil lists provided by schools
- on questionnaires completed by pupils at baseline and at follow-up, with an NFER survey administrator
- during interviews with pupils and NFER researchers.

5. Who will personal data be shared with?

All personal data provided electronically will be shared using NFER's secure school portal. All personal data provided in hard copy will be transferred by secure courier. All NPD data will be accessed using the Office for National Statistics secure access procedures.

For the purposes of the research, pupils' questionnaire responses will be linked with information about the pupils from the National Pupil Database (NPD). NFER will share pupil data with the DfE, and, in an anonymised form, with the Office for National Statistics, the UK Data Archive and potentially with other research teams. Further matching to NPD and other administrative data may take place during subsequent research.

We will not share personal data collected through interviews with children with other organisations.

Your child's data will be treated with the strictest confidence in line with the GDPR 2016/679 and the Data Protection Act 2018. **No child will be named in any report for this project.**

6. Is personal data being transferred outside of the European Economic Areas (EEA)?

No personal data for this RCT is being stored or transferred outside of the EEA.

7. How long will personal data be retained?

NFER will delete any personal data within one year from the completion of the project.

NFER will send all the pupil data it has collected for this RCT to the Fisher Family Trust for archiving within three months of project completion, at which point EEF will take responsibility for data protection compliance.

8. Can I stop my personal data being used?

NFER handles personal data in accordance with the rights given to individuals under data protection legislation. If at any time you wish us to withdraw data or correct errors in it, please contact Guido Miani at P4C@nfer.ac.uk

In certain circumstances, data subjects have the right to restrict or object to processing. You also have the right to make a subject access request to see all the information held about your child(ren). To exercise any of these rights, please contact our Compliance Officer.

9. Who can I contact about this project?

NFER is responsible for the day-to-day management of this RCT evaluation. Contact Guido Miani at P4C@nfer.ac.uk at NFER with any queries.

If you have a concern about the way this project processes personal data, we request that you raise your concern with us in the first instance (see the details above). Alternatively, you can contact the Information Commissioner's Office, the body responsible for enforcing data protection legislation in the UK, at https://ico.org.uk/concerns/.

10. Last updated

We may need to update this privacy notice periodically so we recommend that you revisit this information from time to time. This version was last updated on 4th April 2019.

Appendix F: Baseline survey

NFER
National Foundation for
Educational Research

| Bar code | 123456 |

Please help us with our research!

We are doing some research about schools teaching maths and reading. We'd really like it if you could help us by answering our questions. We will ask you questions about things you do and don't like and what you think about school.

You do not have to answer the questions unless you want to. This is **not** a test and there are no right or wrong answers. Please tell the truth when you answer the questions.

Your name and answers will be seen by the people in charge of the research, this includes the Department for Education. Your teachers, parents/carers and friends will **not** know what you say.

If you need help reading some words or if you are not sure what a question means, please put up your hand. We would like you to answer all the questions but if you don't want to answer a question, you don't have to, please move to the next question.

We have sent some information for parents/carers to your school, explaining how your data will be used. You can ask your parent/carer or a teacher from your school to tell you more about this. The information is in a Privacy Notice, which you can find here:
https://www.nfer.ac.uk/media/3367/philosophy_for_children_parent_privacy_notice.pdf

Please use a pen to fill in this questionnaire if you have one. If you do not have a pen, please use a pencil.

Name

Date of
birth

School
name

NFER
number

1 **How much do you agree with the following sentences?**

Please tick only one box on each line.

55030/EEPS

	Not at all true	Not sure			Completely true
	1	**2**	**3**	**4**	**5**
I am good at explaining my ideas to other people.	☐	☐	☐	☐	☐
I like meeting new people.	☐	☐	☐	☐	☐
I can work with someone who has different opinions.	☐	☐	☐	☐	☐
I can do most things if I try.	☐	☐	☐	☐	☐
Once I have started a task I like to finish it.	☐	☐	☐	☐	☐
I want to try and make my local area a better place.	☐	☐	☐	☐	☐
I like to be told exactly what to do.	☐	☐	☐	☐	☐
I am often afraid to try new things.	☐	☐	☐	☐	☐
I try to understand other people's problems.	☐	☐	☐	☐	☐
I know where to go for help with a problem.	☐	☐	☐	☐	☐
Teachers treat children fairly at my school.	☐	☐	☐	☐	☐
Teachers and other grown-ups at school care about me.	☐	☐	☐	☐	☐

PSSQF

Please read the sentence below and tick the box next to the sentence you agree with the most.

2 **The teacher helps Jo a lot with her reading and to keep up in class.**

Tick one box only.

Jo needs extra help so it is fair that the teacher should spend more time helping her, even if the other pupils have to wait. ☐

Jo should work harder to keep up with the rest of the class. ☐

Jo should be taught in a separate class. ☐

Thank you for answering the questions.
Please sit quietly until the survey administrator collects your paper.

Appendix G: Randomisation code January block (SPSS syntax)

```
GET DATA /TYPE=XLSX

 /FILE='K: \EEPS\RPO\Data\55002 schools for first randomisation.xlsx'

 /SHEET=name '55002 3a 26-01-17'

 /CELLRANGE=full

 /READNAMES=on

 /ASSUMEDSTRWIDTH=32767.



* Restricted *.



*Check for duplicates.

sort cases by nfer_no.

match files file=*/first=f/last=l/by nfer_no.

cross f by l.

temp.

select if any(0, f, l).

list vars=nfer_no description post_code.



*Randomise schools.

set rng=mt, mtindex=26012017.

compute random=rv.uniform(0,1).

sort cases by random.

*Allocate schools to intervention and control according to the overall ratio of intended numbers
(75:125).

* This translates to 41.25 intervention and 68.75 to control (total 110). Round to 41 and 69.

* Review overall balance at second randomisation to achieve (close to) ratio.

if $casenum le 41 Group=1.

if $casenum gt 41 Group=2.
```

ADD VALUE LABELS Group 1 'Intervention group' 2 'Control group'.

frequencies Group.

sort cases by Group NFER_no.

SAVE TRANSLATE OUTFILE='K:\EEPS\CfS\Randomisation\Block 1 randomisation.xls'

 /TYPE=XLS

 /VERSION=8

 /MAP

 /REPLACE

 /FIELDNAMES

 /CELLS=LABELS

 /DROP=random f l.

save outfile='K:\EEPS\CfS\Randomisation\Block 1 randomisation.sav' /drop=random f l.

output save outfile='K:\EEPS\CfS\Randomisation\Block 1 randomisation.spv'.

Appendix H: Randomisation code March block (SPSS syntax)

```
GET DATA /TYPE=XLSX

 /FILE='\\milesan1\projects\EEPS\RPO\Data\EEPS dates for second randomisation.xlsx'

 /SHEET=name 'Sheet1'

 /CELLRANGE=full

 /READNAMES=on

 /ASSUMEDSTRWIDTH=32767.



* Restricted *.



*Check for duplicates.

sort cases by nfer_no.

match files file=*/first=f/last=l/by nfer_no.

cross f by l.

temp.

select if any(0, f, l).

list vars=nfer_no description post_code.



*Randomise schools.

set rng=mt, mtindex=24032017.

compute random=rv.uniform(0,1).

sort cases by random.

* Allocate schools to intervention and control according to the overall ratio of intended numbers (75:125).

* Overall number is 198, so make 75 intervention schools and the remainder (123) control schools.

* Randomisation 1 allocated 41 to intervention and 69 to control.

* Therefore, allocate 34 to intervention and the remainder (54) to control.

if $casenum le 34 Group=1.
```

```
if $casenum gt 34 Group=2.

ADD VALUE LABELS Group 1 'Intervention group' 2 'Control group'.

frequencies Group.

sort cases by NFER_no.


SAVE TRANSLATE OUTFILE='K:\EEPS\CfS\Randomisation\Block 2 randomisation.xls'

 /TYPE=XLS

 /VERSION=8

 /MAP

 /REPLACE

 /FIELDNAMES

 /CELLS=LABELS

 /DROP=random f l.


save outfile='K:\EEPS\CfS\Randomisation\Block 2 randomisation.sav' /drop=random f l.

* Check that there are no duplicates across the two randomisations.

get file = "K:\EEPS\CfS\Randomisation\Block 1 randomisation.sav".

alter type Post_Code (A8).

add files

  /file=*

  /file = "K:\EEPS\CfS\Randomisation\Block 2 randomisation.sav"

  /in = RandBlock2.

frequencies RandBlock2.

*Check for duplicates.

sort cases by nfer_no.

match files file=*/first=f/last=l/by nfer_no.

cross f by l.

output save outfile='K:\EEPS\CfS\Randomisation\Block 2 randomisation.spv'.
```

Appendix I: Histograms of outcome measures[27]

Overall (Intervention and Control)
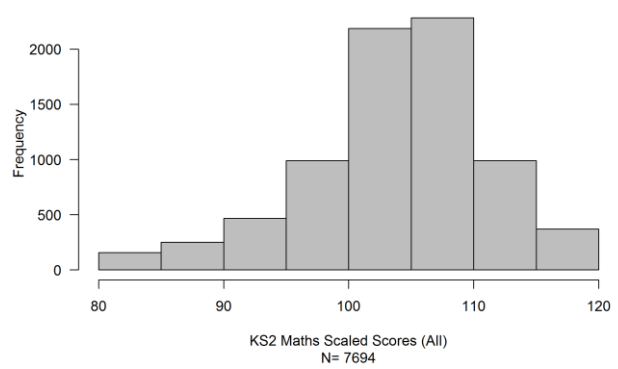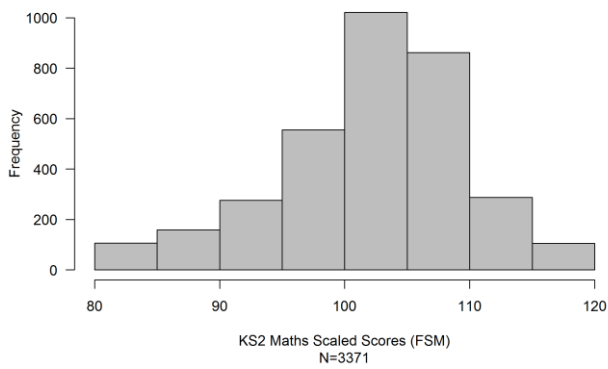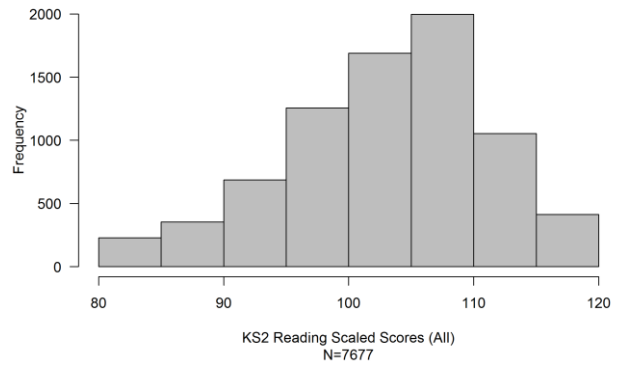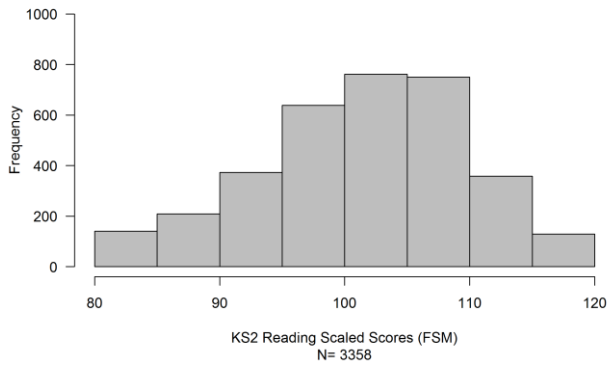
27

For non-attainment outcomes, barcharts on the left represent pupils' responses at baseline and barcharts on the right represents pupils' responses at follow up.

**Social and communication skills -
I am good at explaining my ideas to other people**
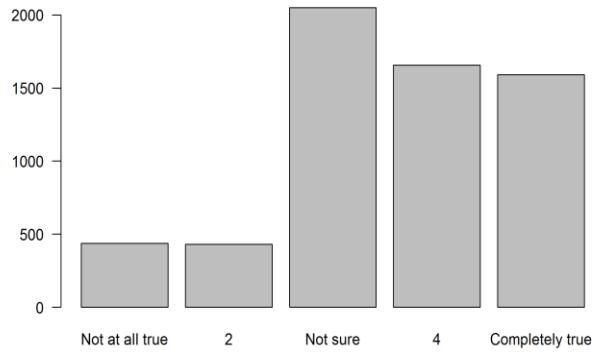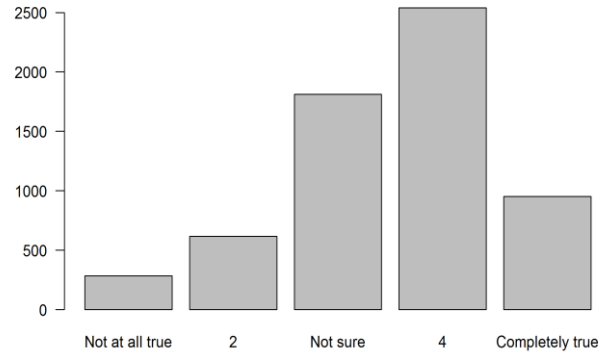
**Social and communication skills -
I am good at explaining my ideas to other people**

**Team work and resilience -
I can work with someone who has different opinions**

**Team work and resilience -
I can work with someone who has different opinions**

Intervention

KS2 Reading Scaled Scores (FSM)
N=1248 Intervention

KS2 Reading Scaled Scores (All)
N=2878 Intervention

KS2 Maths Scaled Scores (FSM)
N=1256 Intervention

KS2 Maths Scaled Scores (All)
N=2890 Intervention

Social and communication skills - I am good at explaining my ideas to other people

Intervention Group

Social and communication skills - I am good at explaining my ideas to other people

Intervention Group

Team work and resilience - I can work with someone who has different opinions

Intervention Group

Team work and resilience - I can work with someone who has different opinions

Intervention Group

Control

KS2 Reading Scaled Scores (FSM)
N=2110 Control

KS2 Reading Scaled Scores (All)
N=4799 Control

KS2 Maths Scaled Scores (FSM)
N=2115 Control

KS2 Maths Scaled Scores (All)
N=1256 Intervention

Social and communication skills -
I am good at explaining my ideas to other people

Control Group

Social and communication skills -
I am good at explaining my ideas to other people

Control Group

Team work and resilience -
I can work with someone who has different opinions

Control Group

Team work and resilience -
I can work with someone who has different opinions

Control Group

Appendix J: School-level engagement data (compliance)
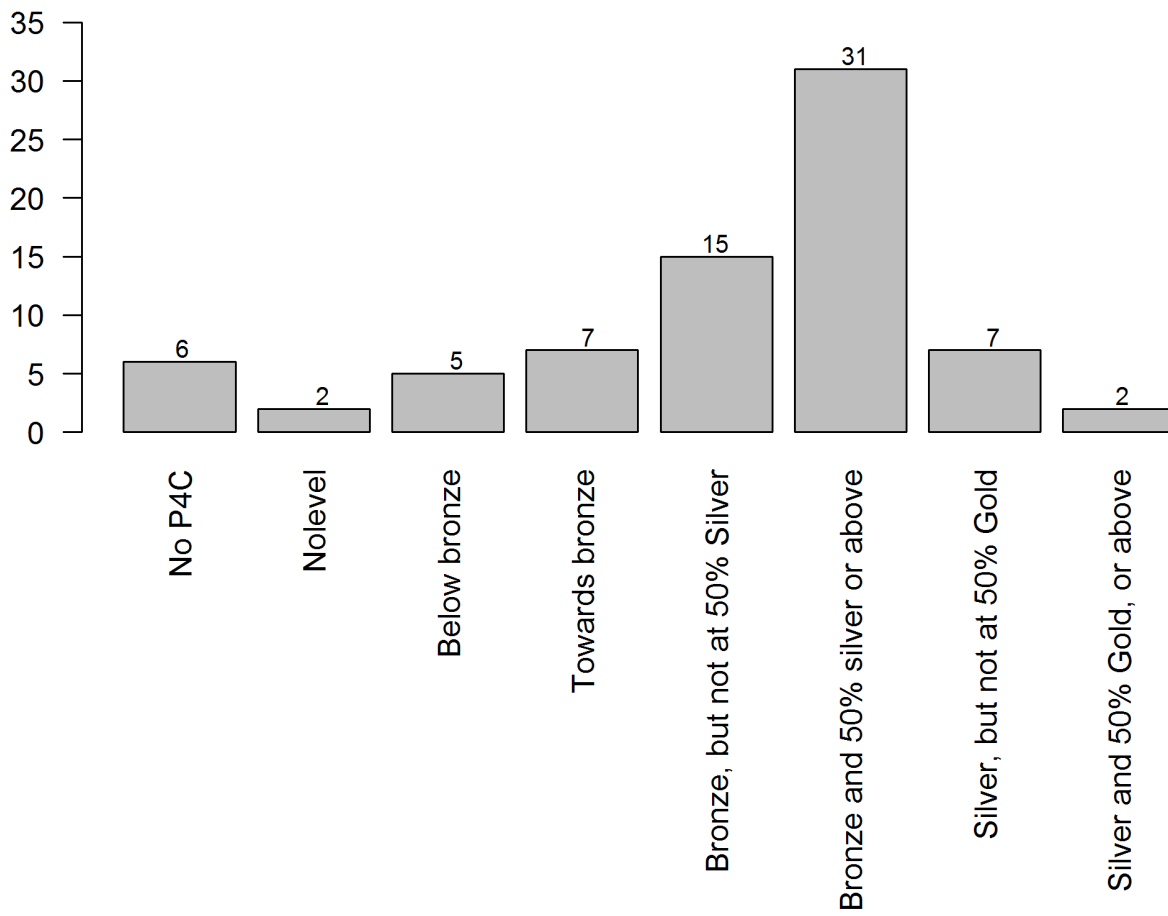
Appendix K: Cross tabulations of items in social skills questionnaire

To encourage a focus on the meaning of each item, Items 1G (I like to be told exactly what to do) and 1H (I am often afraid to try to new things) are worded negatively (as per the survey used by Siddiqui *et al.*, 2017) so that the socially desirable response would be 'not at all true' rather than 'completely true'.

| | | Randomisation group | | |
		Intervention group	Control group	Total
Social and communication skills - I am good at explaining my ideas to other people (baseline)	Not at all true	157	279	436
	2	159	271	430
	Not sure	811	1245	2056
	4	622	1026	1648
	Completely true	577	1014	1591
Total		2326	3835	6161

| | | Randomisation group | | |
		Intervention group	Control group	Total
Social and communication skills - I am good at explaining my ideas to other people (follow up)	Not at all true	99	181	280
	2	235	376	611
	Not sure	713	1096	1809
	4	965	1563	2528
	Completely true	314	619	933
Total		2326	3835	6161

| | | Randomisation group | | |
		Intervention group	Control group	Total
Team work and resilience - I can work with someone who has different opinions (baseline)	Not at all true	214	340	554
	2	115	227	342
	Not sure	555	944	1499
	4	442	675	1117
	Completely true	1000	1649	2649
Total		2326	3835	6161

| | | Randomisation group | | |
		Intervention group	Control group	Total
Team work and resilience - I can work with someone who has different opinions (follow up)	Not at all true	88	175	263
	2	132	274	406
	Not sure	447	748	1195
	4	663	1070	1733
	Completely true	996	1568	2564
Total		2326	3835	6161

| | | Randomisation group | | |
		Intervention group	Control group	Total
I like meeting new people (baseline)	Not at all true	87	117	204
	2	75	135	210
	Not sure	271	385	656
	4	292	493	785
	Completely true	1577	2683	4260
Total		2302	3813	6115

| | | Randomisation group | | |
		Intervention group	Control group	Total
I like meeting new people (follow up)	Not at all true	86	162	248
	2	172	291	463
	Not sure	352	615	967
	4	654	1034	1688
	Completely true	1038	1711	2749
Total		2302	3813	6115

| | | Randomisation group | | |
		Intervention group	Control group	Total
I can do most things if I try (baseline)	Not at all true	76	98	174
	2	57	102	159
	Not sure	311	533	844
	4	418	689	1107
	Completely true	1451	2388	3839
Total		2313	3810	6123

| | | Randomisation group | | |
		Intervention group	Control group	Total
I can do most things if I try (follow up)	Not at all true	44	81	125
	2	82	156	238
	Not sure	356	560	916
	4	747	1185	1932
	Completely true	1084	1828	2912
Total		2313	3810	6123

| | | Randomisation group | | |
		Intervention group	Control group	Total
Once I have started a task I like to finish it (baseline)	Not at all true	147	231	378
	2	106	169	275
	Not sure	373	664	1037
	4	358	588	946
	Completely true	1334	2176	3510
Total		2318	3828	6146

| | | Randomisation group | | |
		Intervention group	Control group	Total
Once I have started a task I like to finish it (follow up)	Not at all true	88	168	256
	2	153	274	427
	Not sure	411	705	1116
	4	669	1002	1671
	Completely true	997	1679	2676
Total		2318	3828	6146

| | | Randomisation group | | |
		Intervention group	Control group	Total
I want to try and make my local area a better place (baseline)	Not at all true	142	210	352
	2	67	101	168
	Not sure	329	631	960
	4	298	483	781
	Completely true	1457	2374	3831
Total		2293	3799	6092

| | | Randomisation group | | |
		Intervention group	Control group	Total
I want to try and make my local area a better place (follow up)	Not at all true	89	128	217
	2	106	175	281
	Not sure	399	680	1079
	4	527	854	1381
	Completely true	1172	1962	3134
Total		2293	3799	6092

| | | Randomisation group | | |
		Intervention group	Control group	Total
I like to be told exactly what to do (baseline)	Not at all true	476	685	1161
	2	175	258	433
	Not sure	477	762	1239
	4	313	576	889
	Completely true	855	1512	2367
Total		2296	3793	6089

| | | Randomisation group | | |
		Intervention group	Control group	Total
I like to be told exactly what to do (follow up)	Not at all true	381	709	1090
	2	425	600	1025
	Not sure	529	883	1412
	4	430	708	1138
	Completely true	531	893	1424
Total		2296	3793	6089

		Randomisation group	Total

		Randomisation group	Total

		Intervention group	Control group	
I am often afraid to try new things (baseline)	Not at all true	790	1325	2115
	2	237	384	621
	Not sure	409	684	1093
	4	272	427	699
	Completely true	580	983	1563
Total		2288	3803	6091

		Intervention group	Control group	
I am often afraid to try new things (follow up)	Not at all true	735	1243	1978
	2	449	727	1176
	Not sure	470	754	1224
	4	350	564	914
	Completely true	284	515	799
Total		2288	3803	6091

		Randomisation group		
		Intervention group	Control group	Total
I try to understand other people's problems (basleine)	Not at all true	149	227	376
	2	88	147	235
	Not sure	424	607	1031
	4	396	697	1093
	Completely true	1247	2124	3371
Total		2304	3802	6106

		Randomisation group		
		Intervention group	Control group	Total
I try to understand other people's problems (follow up)	Not at all true	79	122	201
	2	90	170	260
	Not sure	366	577	943
	4	772	1200	1972
	Completely true	997	1733	2730
Total		2304	3802	6106

		Randomisation group		
		Intervention group	Control group	Total
I know where to go for help with a problem (baseline)	Not at all true	110	159	269
	2	61	88	149
	Not sure	248	426	674
	4	258	391	649
	Completely true	1614	2734	4348
Total		2291	3798	6089

		Randomisation group		
		Intervention group	Control group	Total
I know where to go for help with a problem (follow up)	Not at all true	64	123	187
	2	95	121	216
	Not sure	240	385	625
	4	403	689	1092
	Completely true	1489	2480	3969
Total		2291	3798	6089

		Randomisation group		
		Intervention group	Control group	Total
Teachers treat children fairly at my school (baseline)	Not at all true	107	193	300
	2	108	131	239
	Not sure	268	461	729
	4	293	451	744
	Completely true	1534	2583	4117
Total		2310	3819	6129

		Randomisation group		
		Intervention group	Control group	Total
Teachers treat children fairly at my school (follow up)	Not at all true	108	211	319
	2	223	329	552
	Not sure	355	626	981
	4	483	792	1275
	Completely true	1141	1861	3002
Total		2310	3819	6129

		Randomisation group		
		Intervention group	Control group	Total
Teachers and other grown-ups at school care about me (baseline)	Not at all true	86	152	238
	2	79	96	175
	Not sure	313	493	806
	4	241	399	640
	Completely true	1601	2688	4289
Total		2320	3828	6148

		Randomisation group		
		Intervention group	Control group	Total
Teachers and other grown-ups at school care about me (follow up)	Not at all true	52	91	143
	2	83	188	271
	Not sure	433	704	1137
	4	445	712	1157
	Completely true	1307	2133	3440
Total		2320	3828	6148

		Randomisation group		
		Intervention group	Control group	Total
The teacher helps Jo a lot with her reading and to keep up in class (baseline)	Socially desirable option	804	1229	2033
	Other options	1451	2447	3898
Total		2255	3676	5931

		Randomisation group		
		Intervention group	Control group	Total
The teacher helps Jo a lot with her reading and to keep up in class (follow up)	Socially desirable option	933	1423	2356
	Other options	1322	2253	3575
Total		2255	3676	5931

Appendix L: Score distributions of factor scores[28]

Overall (Intervention and Control)

Factor 1 Baseline — Score (-4 to +4), Mean = 2.7

Factor 1 Endpoint — Score (-4 to +4), Mean = 2.2

Factor 2 Baseline — Score (-12 to +12), Mean = 6.33

Factor 2 Endpoint — Score (-12 to +12), Mean = 6.12
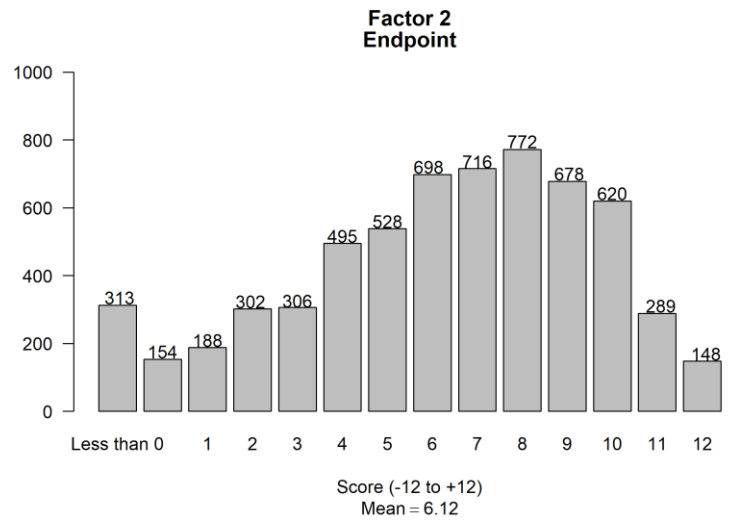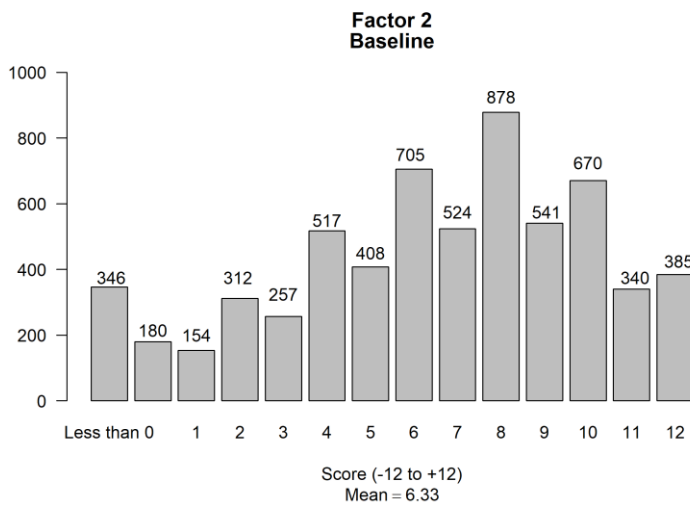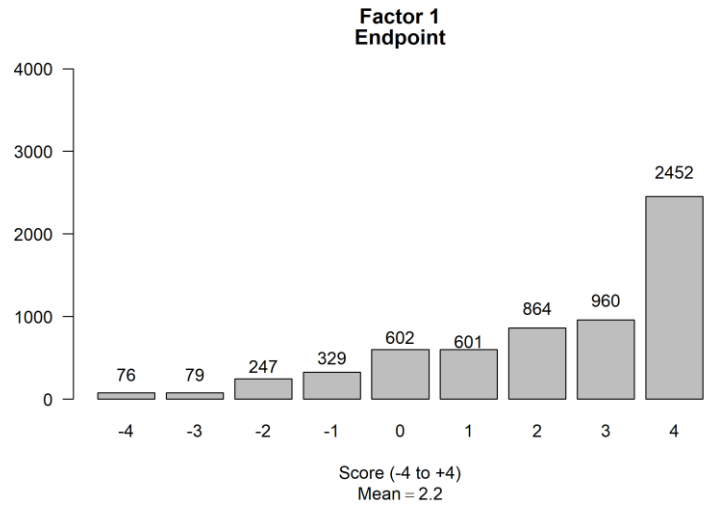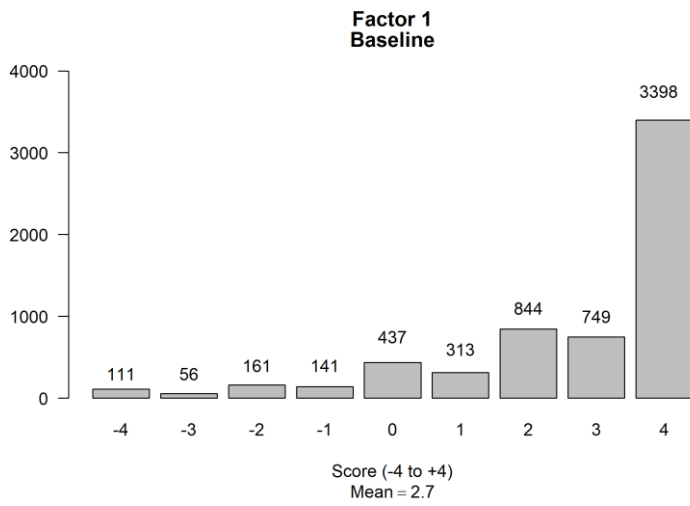
[28] Factor1: **Teachers' approach to caring**
Factor 2: A collaborative and confident approach to learning

Intervention

Factor 1 (Baseline)
Intervention group

Score (-4 to +4)
Mean = 2.67
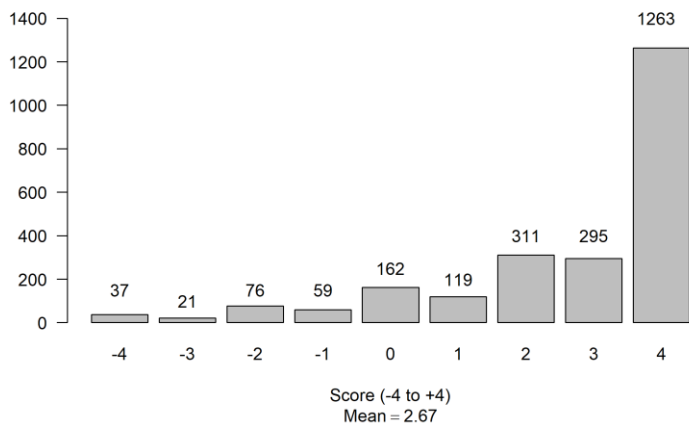
Factor 1 (Endpoint)
Intervention group

Score (-4 to +4)
Mean = 2.24
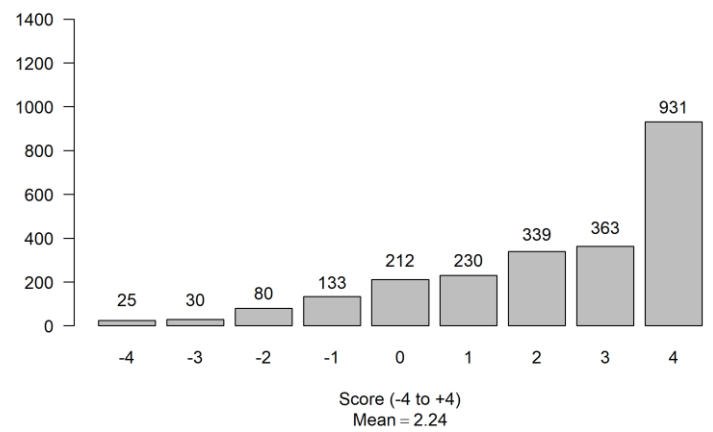
Factor 2 (Baseline)
Intervention group

Score (-12 to +12)
Mean = 6.26
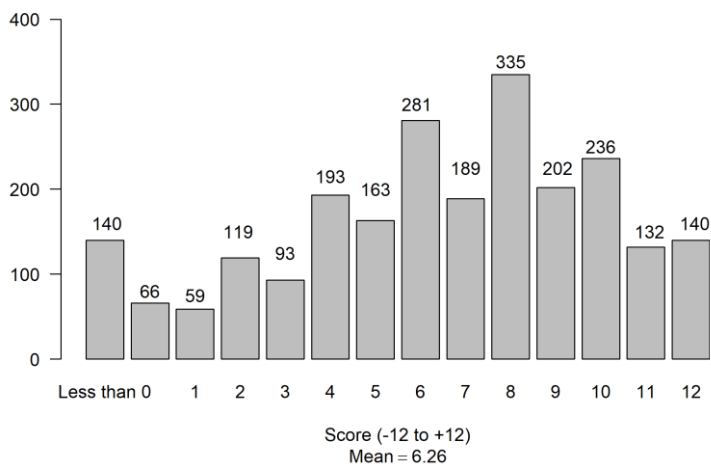
Factor 2 (Endpoint)
Intervention group

Score (-12 to +12)
Mean = 6.14
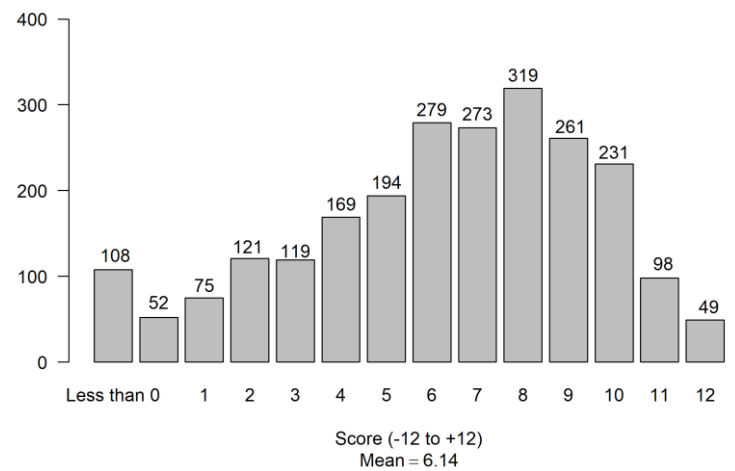
Control

Factor 1 (Baseline)
Control group

Score (-4 to +4)
Mean = 2.72
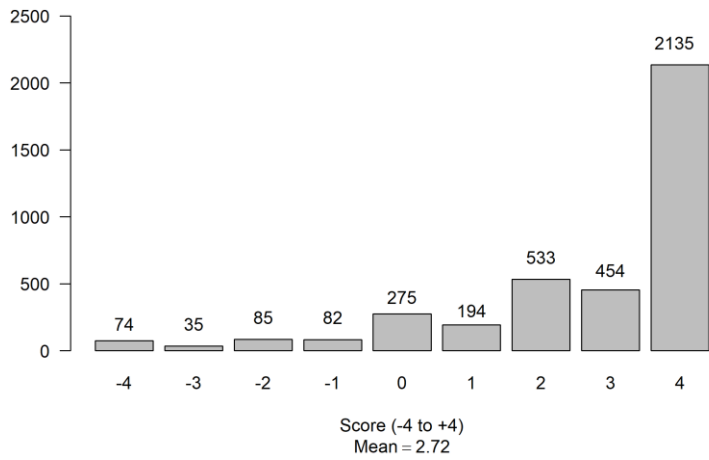
Factor 1 (Endpoint)
Control group

Score (-4 to +4)
Mean = 2.18
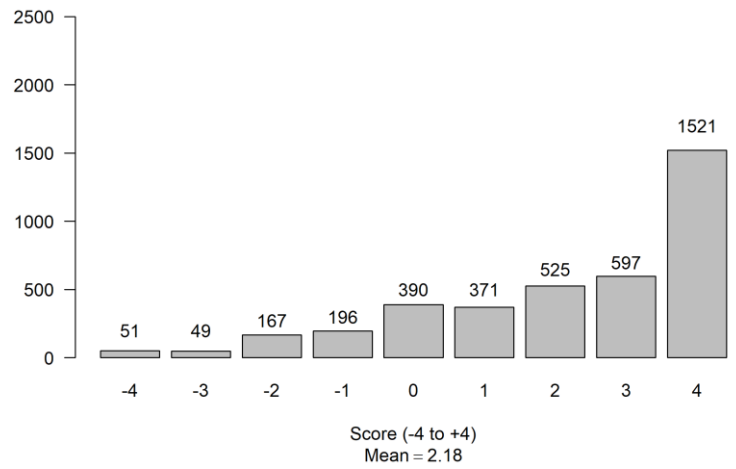
Factor 2 (Baseline)
Control group

Score (-12 to +12)
Mean = 6.38
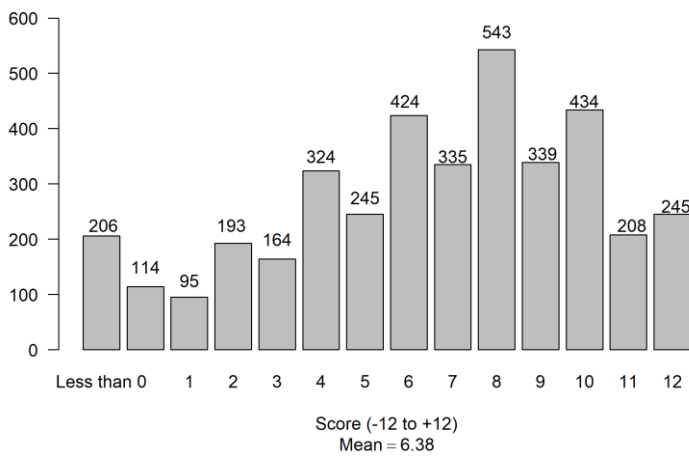
Factor 2 (Endpoint)
Control group
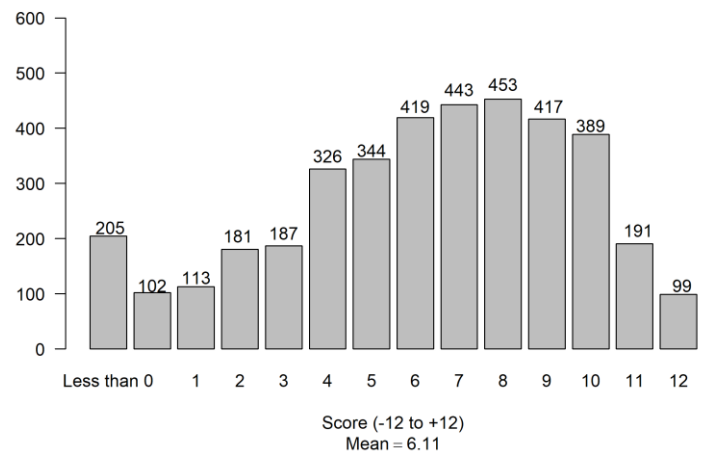
Score (-12 to +12)
Mean = 6.11

Appendix: M Intervention schools: teacher survey

NFER
National Foundation for
Educational Research

Education
Endowment
Foundation

Appendix M: Intervention schools: teacher survey

Evaluation of Philosophy for Children (P4C) Questionnaire

The Education Endowment Foundation (EEF) has commissioned NFER to undertake a survey of staff who are delivering the Philosophy for Children (P4C) programme in their schools. The P4C programme was developed by SAPERE.

The purpose of this survey is to explore how P4C has been implemented and whether it is meeting its aims. The survey findings will inform our overall assessment of the impact of the P4C programme.

Your views are invaluable to us so please take the time to complete this survey. All responses will be treated in confidence and reported only in aggregated or anonymised form. If you exit the survey before the end, your partial answers (i.e. any answers that you have given before exiting the survey) may still be analysed. The information collected will be used for research purposes only and will not be shared with EEF or SAPERE.

This survey will take between 10 and 20 minutes to complete, depending on your involvement with P4C.

If you have any queries, please contact NFER on 01753 637252 or P4C@nfer.ac.uk

First name: _____

Last name: _____

Your role within school

1. **What is your main role within the school?** *Please select one option.*

Headteacher	☐₁
Deputy/Assistant Headteacher	☐₂
Phase, team or year group leader	☐₃
Classroom teacher	☐₄
SENCO	☐₅
Other (please specify)	☐₆

_____

2. **What is your P4C role in the school?** Please select one option.

P4C Lead and I also facilitate P4C sessions [Route to Q3] ☐₁

P4C Lead but I do not facilitate P4C sessions [Route to Q3] ☐₂

I facilitate P4C sessions, but I'm not the P4C Lead [Route to Q14] ☐₃

About you and your school

3. How long have you been the P4C Lead teacher in your school? *Please select one option.*

Less than one academic year ☐₁

One academic year ☐₂

Two academic years or longer ☐₃

4. What P4C training levels have you completed? *Please select all that apply*

Foundation Level 1 ☐₁

Advanced Level 2A ☐₂

Advanced Level 2B ☐₃

Advanced Level 3 ☐₄

Qualified P4C trainer ☐₅

5. To what extent do you agree that the P4C training you have received enables you to fulfil the role of P4C Lead teacher in your school effectively? *Please select one option.*

Strongly agree ☐₁

Agree ☐₂

Neither agree nor disagree ☐₃

Disagree ☐₄

Strongly disagree ☐₅

6. How many teachers in your school (including yourself) have completed the following levels of P4C training? *Please type in the number* [input number 0-25]

Foundation Level 1 ☐₁

Tools for thinking together training ☐₂

Advanced Level 2A ☐₃

Advanced Level 2B ☐₄

Qualified P4C trainer ☐₅

7. **In addition to the training courses, what other support or resources have you received from P4C, and how useful have they been?** *Please select one box per row.*

	Very useful 1	Quite useful 2	Not very useful 3	Not at all useful 4	Did not receive this 5
In-school coaching and support from a trainer	☐	☐	☐	☐	☐
Remote (over telephone or email) administration and planning support	☐	☐	☐	☐	☐
Online P4C resources and guides	☐	☐	☐	☐	☐
Copies of Level 1 Handbook	☐	☐	☐	☐	☐
Copies of Level 2 Handbook	☐	☐	☐	☐	☐

[If have received in-school coaching and support from a trainer ask Q8 all else to 9]

8. **How many days of in-school support has your school received from P4C since becoming involved with this project?** [Drop down box of 0-10]

Days ☐

9. **Which P4C awards, if any, has your school achieved?** *Please select one box per row.*

	Yes, have achieved 1	Working towards but not yet achieved 2	Not working towards 3
Gold	☐	☐	☐
Silver	☐	☐	☐
Bronze	☐	☐	☐

10 **Over the past 3 academic years, which year group(s) in your school have participated in P4C sessions?** *Please select all that apply.*

	2016/17	2017/18	2018/19
All year groups	☐ a 1	☐ b 1	☐ c 1
Reception	☐ 2	☐ 2	☐ 2
Year 1	☐ 3	☐ 3	☐ 3
Year 2	☐ 4	☐ 4	☐ 4
Year 3	☐ 5	☐ 5	☐ 5
Year 4	☐ 6	☐ 6	☐ 6
Year 5	☐ 7	☐ 7	☐ 7
Year 6	☐ 8	☐ 8	☐ 8

11. **In this academic year, how often does your school provide P4C sessions for your pupils?** *Please select one box per row.* [Year groups routed from Q10C]

	At least weekly	At least fortnightly	A few times a term	Rarely
	1	2	3	4
Reception	☐	☐	☐	☐
Year 1	☐	☐	☐	☐
Year 2	☐	☐	☐	☐
Year 3	☐	☐	☐	☐
Year 4	☐	☐	☐	☐
Year 5	☐	☐	☐	☐
Year 6	☐	☐	☐	☐

12 How many pupils has your school been teaching P4C to during this academic year?
[Year groups routed from Q10c (2018/19)]

Reception		1
Year 1		2
Year 2		3
Year 3		4
Year 4		5
Year 5		6
Year 6		7

13 Have you received any additional support or resources, related to using philosophy in schools, from any of the following sources: *Please select all that apply.*

Philosophy Foundation	☐	1
Jason Buckley	☐	2
Dot Lenton or Barbara Videon (Philosophy4Children)	☐	3
Philosophy trainers independent of P4C or Sapere	☐	4
Other philosophy support or resources (please list below)	☐	5

[Respondents who ticked option 2 on question 2 (do not facilitate P4C) route to Q23]

Facilitating P4C

14. How long have you <u>facilitated</u> P4C sessions in your school? *Please select one option.*

Less than one academic year	☐₁
One academic year	☐₂
Two academic years or longer	☐₃

15. Which levels of P4C training have you completed? *Please tick all that apply.*

Foundation Level 1	☐₁
Advanced Level 2A	☐₂
Advanced Level 2B	☐₃
Advanced Level 3	☐₄
Qualified P4C trainer	☐₅
None of the above [go to Q17]	☐₆

16. To what extent do you agree that the P4C training you have received enables you to facilitate P4C sessions effectively? *Please select one option.*

Strongly agree	☐₁
Agree	☐₂
Neither agree nor disagree	☐₃
Disagree	☐₄
Strongly disagree	☐₅

17. For which year group(s) do you facilitate P4C sessions? *Please select all that apply.*

Reception	☐₁
Year 1	☐₂
Year 2	☐₃
Year 3	☐₄
Year 4	☐₅
Year 5	☐₆
Year 6	☐₇
Not applicable	☐₈

NFER
National Foundation for
Educational Research

Education
Endowment
Foundation

Delivery of P4C in your school

18. To what extent do you agree with the following statements about facilitating P4C in your school: *Please select one box per row.*

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	1	2	3	4	5
I understand the 10-step P4C sequence of enquiry method	☐	☐	☐	☐	☐
I am effective in facilitating across the different stages of a P4C enquiry	☐	☐	☐	☐	☐
I understand the principles of caring, creative, critical and collaborative thinking	☐	☐	☐	☐	☐
I am a confident facilitator of P4C	☐	☐	☐	☐	☐

19. To what extent to you agree with this statement: 'I have delivered P4C in our school exactly as intended'. *Please select one option.*

Strongly agree	☐₁
Agree	☐₂
Neither agree nor disagree	☐₃
Disagree	☐₄
Strongly disagree	☐₅

20. To what extent do you agree with the following statements about the engagement with P4C in your school? *Please select one box per row.*

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	1	2	3	4	5
Pupils like the P4C approach	☐	☐	☐	☐	☐
Pupils are fully engaged with P4C sessions	☐	☐	☐	☐	☐
[P4C Lead only (Q2 option1)] Teachers value the P4C approach	☐	☐	☐	☐	☐
[P4C Lead only (Q2 option 1)] Teachers have fully embraced facilitating P4C sessions	☐	☐	☐	☐	☐

Perceived impact of P4C

21. This question focuses on whether participating in P4C sessions has had an effect on pupils in your school. To what extent do you agree that participating in P4C sessions has impacted positively on pupils' … (please select one box per row).

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	1	2	3	4	5
Ability to question and reason	☐	☐	☐	☐	☐
Ability to express views clearly	☐	☐	☐	☐	☐
Listening skills	☐	☐	☐	☐	☐
Respect for other pupils' opinions	☐	☐	☐	☐	☐
Self-esteem	☐	☐	☐	☐	☐
Resilience	☐	☐	☐	☐	☐
Confidence	☐	☐	☐	☐	☐
Relationships with teachers	☐	☐	☐	☐	☐
Relationships and collaboration with other pupils	☐	☐	☐	☐	☐
Behaviour (e.g. ability to resolve conflicts)	☐	☐	☐	☐	☐

22. This question focuses on whether facilitating P4C sessions has had an effect on you as a teacher. To what extent do you agree that facilitating P4C sessions has impacted positively on your: (please select one box per row)

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	1	2	3	4	5
Ability to accelerate the development of all pupils	☐	☐	☐	☐	☐
Ability to provide additional support for lower-ability pupils	☐	☐	☐	☐	☐

NFER
National Foundation for
Educational Research

Education
Endowment
Foundation

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
Ability to provide additional support for higher-ability pupils	☐	☐	☐	☐	☐
Relationships with pupils	☐	☐	☐	☐	☐
Confidence in trying out new teaching approaches	☐	☐	☐	☐	☐
All-round classroom practice	☐	☐	☐	☐	☐

23. **To what extent do you agree that P4C has impacted positively on your:** *(please select one box per row)*

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree
	1	2	3	4	5
Professional development	☐	☐	☐	☐	☐
Self esteem as a teacher	☐	☐	☐	☐	☐
Overall effectiveness as a teacher	☐	☐	☐	☐	☐

[Those who answered Q2 Option 3 route to end]

[P4C lead only (Q2 options 1 and 2)]

24. **Overall to what extent do you agree that P4C has had a positive impact on your school.** *Please select one option*

Strongly agree	☐	1
Agree	☐	2
Neither agree nor disagree	☐	3
Disagree	☐	4
Strongly disagree	☐	5

THANK YOU FOR COMPLETING THIS SURVEY

Appendix N: Control school survey

∴∴NFER
National Foundation for
Educational Research

Appendix N: Control school survey
INTRODUCTION PAGE 1 – SHOW TO ALL

NFER Evaluation of Philosophy for Children (P4C)

As part of the data collection for the trial of Philosophy for Children (P4C) in primary schools, please could you complete the following information about any Philosophy programmes or support you have been involved with during 2016-2019.

This research is being conducted by the National Foundation for Educational Research (NFER) and has been commissioned by the Education Endowment Foundation. Your information will be held in accordance with the Data Protection Act and will be treated in the strictest confidence by the NFER Research Team. If you have any queries, please contact XXX on 01753 63XXX or P4C@nfer.ac.uk.

For more information about how we will use you and your school's data for this project, please see the project Privacy Notice:
https://www.nfer.ac.uk/media/2763/philosophy_for_children_privacy_notice.pdf

ASK ALL, SR, MANDATORY

1. **Since September 2016, has your school run any P4C sessions with pupils?**
   (Please tick one option).

   Yes ☐ ₁ *(please provide details below)*   No ☐ ₂ *(please go to question 3)*

ASK IF Q1 = YES (1), NOT MANDATORY

2. **Which year groups have you run P4C sessions with since September 2016?**
   2.1)
   a) **Which year groups were involved in 2016/17?** Please select all that apply
      *Reception, Year 1, Year 2, Year 3, Year 4, Year 5, Year 6
   b) **Which year groups were involved in 2017/18?** Please select all that apply
      *Reception, Year 1, Year 2, Year 3, Year 4, Year 5, Year 6
   c) **Which year groups were involved in 2018/19?** Please select all that apply
      *Reception, Year 1, Year 2, Year 3, Year 4, Year 5, Year 6

   ASK IF Q2.1 is Y4 in 2016/17, and/or Y5 in 2017/8 and or Y6 in 2018/19:
   2.2)
   You indicated that your school has run P4C sessions with the current Year 6 pupils at some point since September 2016. For these pupils:
   a) **How many pupils were involved in these sessions?** *whole year group / whole class / targeted pupils

b) When did you start these P4C sessions? *month/year

c) How many sessions did these pupils attend? *number

ASK ALL, SR, MANDATORY

3. Since September 2016, has any member of staff at your school undertaken any P4C training? *(Please tick one option).*

Yes ☐ 1 *(please provide details below)*   No ☐ 2 *(please go to question 6)*

ASK IF Q3 = YES (1), OE, NOT MANDATORY, MAX CHAR 50

4. You mentioned that a member of staff at your school has undertaken training since 2016, please could you provide the name of the trainer/organisation who provided this training:

```
┌─────────────────────────────────────────────────────────────┐
│                                                             │
│                                                             │
│                                                             │
│                                                             │
└─────────────────────────────────────────────────────────────┘
```

ASK IF Q3 = 1, MC, NOT MANDATORY

5. What level of P4C training have you/ members of staff at your school completed? *(Please tick all that apply).*

Foundation Level 1 ☐ 1

Tools for thinking together training ☐ 2

Advanced Level 2A ☐ 3

Advanced Level 2B ☐ 4

Qualified P4C trainer 5

ASK ALL, MC, GRID, NOT MANDATORY

6. Please tell us about any other philosophy- or enquiry-based programmes your school has been involved with since September 2016.

Programme name (OE TEXT BOX)	Start date (month/year)	End date (month/year)	Year groups involved (e.g. Y1, Y2)	Was this part of a research programme? (Yes/No)

Thank you for your support.

Appendix O: Award Level Data Collection

NFER no.	School Name	Town	Postcode	Group	Received P4C Training as at July 2019	Control group: Received P4C Training (June 2019 - July 2020) only L1	Control group 2019: Did Year 6 pupils receive any P4C this year (September 2019 - July 2020)?	0 - no P4C sessions/ withdrew before P4C activity started	1 - Not actively working towards a level (please use the notes column to indicate what level the school might be at should they apply for an award)	2 - Below bronze but some P4C activity and/or P4C activity by a few staff only	3 - Working towards Bronze	4 - Bronze, but not at 50% Silver	5 - Bronze + 50% silver or above	6 - Silver, but not at 50% Gold	7 - Silver +50% Gold, or above	Notes - please use this column to indicate what level a school that has not applied for an award might be at, should they apply for one (see category 1)
ID no.	XXX school	XXX town	Postcode	Intervention	Yes/No	Yes/No	Yes/No									