Non-cognitive impacts of Philosophy for Children

Nadia Siddiqui
Stephen Gorard
Beng Huat See

Durham University, School of Education, Leazes Road, DH1 1TA, Durham

Published by

School of Education
Durham University, Leazes Road
Durham
DH1 1TA

ISBN 978-0-907552-06-2

Acknowledgements

The authors fully acknowledge The Nuffield Foundation for providing the main grant support for this project (grant number EDU/42016). The Nuffield Foundation is an endowed charitable trust that aims to improve social well-being in the widest sense. It funds research and innovation in education and social policy and also works to build capacity in education, science and social science research. The Nuffield Foundation has funded this project, but the views expressed are those of the authors and not necessarily those of the Foundation. More information is available at www.nuffieldfoundation.org.

The project is an extension of a study previously funded by Education Endowment Foundation (report available at https://educationendowmentfoundation.org.uk/evaluation/projects/philosophy-for-children/).

The Society for the Advancement of Philosophical and Reflection Enquiry (SAPERE) provided the teacher training necessary. We thank Bob House and Steve Williams from SAPERE for supporting this research project and helping us access the schools.

In particular, we appreciate the participation of the primary schools, teachers and pupils who helped us to complete this study.

Contents

Executive summary

Introduction

Schools are places where children can learn behaviour, skills and attitudes that have lifelong relevance, in addition to the formal curriculum of subjects. In England, despite a continuing emphasis on attainment, there are clear moves to consider also the wider and non-cognitive outcomes of schooling – such as pupils' development of self-confidence, trust, critical thinking and civic-mindedness. However, there is little existing evidence on whether and how such non-cognitive outcomes can be improved through school-based interventions. This is a report of the evaluation of one such intervention - Philosophy for Children (P4C) - in terms of its possible non-cognitive outcomes.

The intervention being evaluated

P4C is a widely implemented teaching approach in UK schools, and previous robust evaluations have provided reasonable evidence that it is linked to improved cognitive outcomes at primary school, especially for disadvantaged pupils (see for example, Gorard, Siddiqui and See 2016). And there are some indications that it is associated with non-cognitive gains as well, but the evidence before the current study was not strong enough to justify a full trial. Therefore, we conducted a quasi-experiment based on schools doing what they were going to do anyway.

P4C is a whole-class intervention which aims to stimulate classroom dialogue in response to children's own philosophical-type questions about shared stories, films and other stimuli. The group sits in a circle facing each other and the teacher introduces a starter activity or game which relates to the theme, skill or disposition that the teacher wants to emphasise. A stimulus such as a video, story, picture or artefact is presented. The group discusses the stimulus, identifies a range of questions prompted by it, and communally chooses one question for discussion. Example questions might be 'What is kindness?', 'Is it OK to deprive someone of their freedom?', and 'Are people's physical looks more important than their actions?'

Following the selection of an appropriate question, the main discussion ensues. Pupils share ideas on the selected question and give their point of views on the raised issue. In a closing stage of the enquiry pupils share their final thoughts and discuss how the enquiry went. In so doing, they can build skills of evaluation and reflection as well as planning for improvement in subsequent sessions.

Summary of research design

This new study is an extension of a large scale randomised control trial (RCT) which demonstrated a positive impact of P4C on children's cognitive skills (Gorard, Siddiqui and See 2016). The control group in that prior study consisted of schools that wanted to use P4C in their classrooms but were randomised to wait 18 months for the necessary training, and start P4C after the trial was over. Of these schools previously on the RCT waiting-list, 16 agreed to become the treatment group for this new evaluation of non-cognitive outcomes.

We have also conducted an even larger RCT of Youth Social Action in primary schools with children in the same year groups as our P4C trial (Years 4 and 5 initially, in Years 5 and 6 by the end). The control group in that study provided responses to the same pupil survey as is used in our new evaluation, and these responses are used as the comparison group for the new evaluation. This group of 26 schools continued with business as usual, undertaking neither the Youth Social Action nor the P4C intervention for the duration of the study. This design was efficient and minimised data collection and intrusion in schools. It was also ethical as all schools continued to do what they were going to do anyway. In all, there were 2,722 pupils in the 42 primary schools. Of these, 1,099 were in the P4C group and 1,623 in the comparison.

The non-cognitive outcomes were assessed before and after the intervention for both groups using a specific survey instrument developed for the purpose and used across a number of previous studies, including in co-operation with the Cabinet Office. The instrument was designed to assess changes in self-reported 'social and communication skills', 'team work and resilience' and 'empathy' and a number of similar constructs. Elements of the instrument, such as vignettes and reverse-coding, were designed to prevent pupils simply giving what might be seen as the most socially desirable answers.

The headline findings are presented in terms of differences between the groups in the post-intervention survey. Both groups were the same average age when they completed this survey. The analysis involved post-intervention comparisons of 'effect' sizes based on differences between means and odds ratios, as appropriate to the data. The same analyses are also conducted for the sub-set of pupils eligible for free school meals (FSM – a measure of family poverty). The key outcomes are also presented as stepped regression models – multiple or logistic as appropriate – using pupil characteristics, their prior responses, and then which group they were in, as predictors.

The use of pupil prior responses to the survey is complicated by the fact that, due to the designs and timing of the two previous studies (above), the P4C group were six months younger, on average, when they did the pre-intervention survey. This matters, as the kinds of outcomes we are trying to assess such as well-being are very sensitive to age (usually worsening over time). This is why the post-intervention results are presented as the headline figures. However, the prior responses and an attempt at adjusted gain scores are presented with every result as a kind of caution that the groups are clearly unbalanced at the outset.

There was also a parallel process evaluation, involving observation of lessons and interviews with trainers, staff and children. This was to assess how well the intervention was implemented, observe any barriers or examples of good practice, and to understand how staff and pupils reacted.

Key findings

The results show that pupils who received the P4C intervention were ahead of their counterparts in the comparison schools in many important respects. For example:

- Children in the P4C group were ahead in self-reported communication skills ('effect' size +0.10), and teamwork and resilience (+0.15), but less so in empathy (+0.01).
- These effect sizes are generally larger for pupils living in relative poverty (FSM-eligible) - communication skills (+0.23), teamwork and resilience (+0.11) and empathy (+0.08).

These and other differences are small, and it is important to note several cautions including the fact that regression analyses suggest perhaps only a small role for the intervention once pupil background characteristics and prior responses have been accounted for.

Teachers mostly reported that positive effects were observed in pupils' confidence in questioning and reasoning, both in P4C sessions and in other lessons. The teachers attributed positive changes in pupils' behaviour to P4C practice. Improvement in pupils' respect and behaviour are related to pupils' social and communication skills, cooperation, teamwork and resilience. In terms of social and emotional literacy, the self-reported benefits suggest that to some extent P4C can be a useful approach in overcoming the challenges of disruption in the classroom, and even bullying and other anti-social behaviours.

Pupils also generally reported that they enjoyed the intervention. Pupil interviews and observations confirmed that pupils enjoyed P4C and actively participated in the sessions. In general, P4C sessions were observed to be very different from ordinary primary classroom settings, and to involve a different communication practice between teachers and pupils. As may be imagined, the topics selected were sometimes difficult to discuss for some pupils.

Strengths and limitations of the study

The design for this study, making use of the waiting-list schools from one randomised study and the control schools from another, was convenient, cost-effective, ethical, and it permitted a larger sample size than otherwise. It is also clearly a better design that a simple cross-section or a before-and-after design. However, it is not ideal for a causal question such as that addressed in this report and this study is intended to be part of working towards a fuller trial.

It is intrinsically hard to collect evidence on the kind of non-cognitive outcomes covered by this study. The instrument used has a clear audit trail to items from standardised questionnaires, and it has been piloted and used extensively in our prior work, internationally, and in collaboration with the Cabinet Office. The use of reverse-coded items was intended to make pupils focus throughout, and the vignette items were developed to prevent pupils simply picking a point towards the most socially desirable end of any scale.

The post-intervention surveys were conducted at the same point in the school year, for children in the same years, in both analytical groups for the evaluation. However, the pre-intervention surveys were completed about six months apart in the school year for the treatment and comparison groups. An attempted adjustment has been made for this, but the most trustworthy results are for the post-intervention scores only. This must reduce our reliance on the results to some extent.

Implications for policy and practice

There are some indications from all elements of the new study described here, that pupils' non-cognitive skills can be altered, at least slightly, through school-based interventions. There is still a gap in the research evidence needed to establish a causal link between non-cognitive skills and academic outcomes. If interventions for the improvement of non-cognitive skills can also yield better academic outcomes, then there is scope for integrating such interventions in the national curriculum and, most importantly, using the pupil premium funds in implementing them. But in their own right they may improve behaviour, co-operation, self-confidence, empathy and tolerance for others. And they may be feasible for practitioners facing the demand that they tackle extremism and radicalisation, and enhance so-called 'British values'.

The overall findings of this research contribute to the ongoing policy concern about the purpose of schools in the society and how far it is possible to make schools meaningful and effective for the most disadvantaged groups. This research suggests the value of developing space in the school curriculum where the focus is the development of pupils' character and values – but through example and process rather than traditional pedagogy. P4C has to be a whole-class approach to be effective, so that all individuals concerned are aware of the objective of cultivating empathy, respect and appropriate or acceptable behaviour. For example, to teach children fairness, teachers themselves have to be seen to be fair. To teach children to be polite, teachers and other (older) pupils have to practise it. And so on.

From successive evaluations of P4C, the results show persistent, small positive links with attainment and non-cognitive outcomes, especially for disadvantaged pupils. It is clear that this approach does not harm children's achievement of academic outcomes or their non-cognitive skills. It seems to reduce the gap between disadvantaged children and the rest, and it is clearly enjoyable for teachers and most pupils. It can also be beneficial for teachers, as the process evaluation results here suggest. It does require training for teachers, especially in how to intervene on the small number of occasions that things go wrong, such as pupils being distressed by something that occurs during the discussions.

There is, therefore, some promise from this study that targeted school-based intervention such as P4C could improve pupils' non-cognitive skills, and there are lessons for how to conduct to such studies, and how to assess the wider outcomes of schooling. This provides a basis for conducting a larger RCT in the future, if deemed appropriate, in which it would also be important to investigate the long term

impact of P4C on pupils' future success, career paths and well-being in adult life. If interventions such as this can improve non-cognitive outcomes as well as yield improved academic outcomes, especially for the poorest children, then there is a case for integrating incorporating thinking skills in the national curriculum, and using the pupil premium funds to implement them.

Chapter 1 – Introduction to the study and non-cognitive outcomes

1.1 Introduction

Schools are where most children and young people first learn how to interact with a wide range of peers and with the formal adult world. They are where pupils begin to see who they can trust, what society is like, and what they can contribute to it. Schools are communities for young people to learn behaviours and develop early perception of the wider social world. School as a community play an important role in the development of skills for social interaction. Policy-makers have traditionally emphasised the economic benefits of education – for society and individuals. But education is also about the happiness of individuals, their preparedness for life other than work, and their general 'flourishing' (Brighouse 2008). Some studies have considered wider outcomes such as enjoyment, self-esteem or determination as merely stepping stones to higher attainment. However, this new study is primarily about taking these kinds of wider outcomes as valuable in their own right whether they also lead to better academic attainment outcomes or not.

In a recent speech Michael Wilshaw, the chief inspector for OFSTED, suggested that (comprehensive) schools in England play a positive and largely unremarked part in developing the integration of immigrant children into society (The Guardian 2016). These schools are largely mixed in terms of the sex, ability, social class, ethnicity and first language of their pupils, while children with special educational needs are now largely included in mainstream school settings. The schools are routinely judged in terms of the academic attainment of their pupils, and perhaps rightly so, but schools do more than that. Schools can provide opportunities for children to develop skills, behaviour and attitudes that have a lifelong impact. A lot of wider knowledge, behaviour, and language can be learned, especially in primary schools, and will play a very important role in life. Growing research evidence is gradually driving the interest of government and policy makers to address the purpose of school in a developed society beyond academic achievement (Common Select Committee: Purpose and Quality of Education in England, 25th January 2016).

This report outlines the kinds of non-cognitive outcomes being considered in our new study, how they might be assessed, the role of schools in developing these outcomes, and how schools might be assisted to improve them. It then describes the intervention, design and methods used in our evaluation of the non-cognitive outcomes of a programme called Philosophy for Children (P4C), before presenting summary results of the impact and process components of our evaluation. The report ends by describing the implications for future research, policy and practice.

1.2 The wider outcomes of schooling

While academic achievements may facilitate access to pathways of success in life, some commentators suggest that there are underlying attitudes and behaviours also influenced by education, which may be important in their own right (Heckman and Rubenstein 2001, Brunello and Schlotter 2011, Gupta and Simonsen 2010). The list of such personal qualities is long and could include social and communication skills, resilience, determination, motivation, confidence, self-esteem, and self-efficacy. Various collective terms are used in the literature such as non-cognitive skills, soft skills, personal characteristics, personality traits, life skills, social and emotional skills. For consistency and clarity we use the term non-cognitive outcomes (of education).

One issue is to investigate through robust research if these skills can be improved through targeted school based interventions, and another is to investigate the effective approaches that are feasible for implementation at national school level. So far there is little evidence regarding the efficacy and effectiveness of interventions and school based programmes that directly target pupils' non-cognitive skills (Siddiqui and Ventista 2017).

Several studies *have* shown that academic performance has a strong correlation with non-cognitive skills but there is not yet enough evidence on any causal relationship between the two. Non-cognitive skills are different from academic abilities because of complexity in the measurement of concepts such as self-esteem, confidence, motivation, social emotional skills, happiness, well-being. There are no age-related standardised criteria for the non-cognitive skills, and the test responses are largely dependent on reporting of attitudes, feelings, choices or experiences based on self-reports or observations recorded by others such as researchers, teachers, parents or peers. There are no 'accurate' levels of these concepts against which a measurement can be judged (Gorard 2010). In other words there are no right or wrong responses when measuring pupils' non-cognitive abilities whilst academic performance is judged upon clear right and wrong criteria. Therefore, assessment of non-cognitive abilities is less reliable when compared with assessment of academic outcomes.

There are widespread doubts about the nature and value of some of these reported outcomes, and about how to assess them. For example, is autonomy necessarily a good thing if it means ignoring expert advice? For gratitude to mean something it has to be freely given, in which case how can it be improved via an education intervention (Carr 2015)? Despite its use in policy documents, it is not clear what resilience or 'grit' is. Is it asking others to accept failure (O'Brien 2014)? Does it mean not to change circumstances such as poverty (Ris 2015)? The same queries arise when considering a whole range of mental and personal constructs from self-esteem to aspiration (Gorard et al. 2011). On examination they may be considered by some to be rather sinister, or simply bland and unimpressive (Hyland 2015). In addition, it is hard to envisage how best to assess any of these constructs for an individual. In education, we might approach teachers to ask for their judgements of pupils, we might observe pupils in action and over time, provide them with tasks to perform or games to play, or ask for their self-reports (Duckworth et al. 2015). All of these approaches have some relative advantages and several disadvantages, and none is satisfactory. While accepting these limitations, we have relied largely on repeated self-reports from pupils as being the only feasible way to obtain relevant data at the required scale. However, our instrument contained some vignettes that acted as a kind of practical 'moral' task, in which the socially desirable responses were less obvious, and we also observed and spoke to as many pupils and their teachers as possible (see below).

1.3 The importance of wider outcomes

Many of the wider outcomes of schooling are inter-related. Pupils who struggle to communicate effectively are likely to be at risk of social isolation, rejection and even bullying (Botting and Conti-Ramsden 2000, Clegg et al. 1999, Knox and Conti-Ramsden 2003, Hartshorne 2007). In an international comparison, children in England were the most likely to report being bullied by other pupils (Gorard and Smith 2010). And studies have shown that children bullied in schools have low self-esteem, poor confidence and suicidal/self-harming behaviour (Winsper et al. 2012). Bullying and the inconsiderate behaviour of peers are key issues affecting the subjective well-being of young people (The Good Childhood Report 2015). This group of young people is highly vulnerable to neglect and abuse and most likely to be the victim of sexual exploitation, peer violence, bullying and unfair discrimination in schools (Morgan 2008, Rahilly and Hendry 2014).

Poor communication and social skills can also lead to behavioural and social emotional difficulties. Studies have shown that these children are more likely to be given detention and excluded from school (DfE 2016). Longitudinal studies have shown that children having social emotional and behavioural problems at primary school age are less likely to achieve good results in later schooling (Patalay et al. 2016), more likely to be involved in crime and are at higher risk of poor health, drug addiction, depression and other mental health problems (Carneiro et al. 2012). Children with good social skills, on the other hand, are more engaged in schools and have positive friendship clusters (Gutman and Vorhaus 2012), and these are correlated with later life outcomes such as employment status and hourly wages, and well-being in adult life (Olsson et al. 2013).

Some disadvantaged children appear to have higher resilience to the challenges of poverty and abuse (Schoon 2006, Goodman et al. 2015). Consistent and positive parenting is believed to help develop

resilience among children (Hill et al. 2007). However, not all socially disadvantaged children have permanent homes and positive relationships with parents, foster parents or carers (Farmer and Lutman 2009). Therefore, the role of school could be more important for children who lack such permanence. Schools could provide the care and warmth that might be lacking in the homes of some vulnerable young children.

There is evidence that social skills are malleable at school age (Gutman and Schoon 2013). Supportive peers, school environment and community might develop characteristics that are associated with nurturing good social skills and effective communication behaviours. In particular, active engagement with school (or school connectedness) is thought to be inversely linked with risk-taking behaviours. Schools are a micro-society for children where they learn about trust, mutual respect and expectations from a wider society (Gorard and Smith 2010).

Schools also have a growing concern to tackle the suggested problem of extremism and radicalisation of young people. The common reasons for young people being vulnerable to extremist activities are a sense of isolation, search for answers to question of identity, poor family relationship, experience of bullying and racism, and feeling of failure (Bailey 2015). Schools have been given recommendations to provide support to all children through raising awareness, promoting open and respectful dialogue and using the curriculum to challenge any extremist narrative (DCSF 2008). However, prevention approaches are now also recommended to tackle extremism such as raising young people's resilience so that they can cope with internal and external pressures that push towards violent extremism (Bonnell et al. 2010, Paterson et al. 2014).

Children in any school and in life outside school will encounter situations when their sense of empathy can play a positive role in developing good relationships, social cohesion and overall well-being. One of the challenges for schools is to foster an overall ethos of understanding and valuing individual differences (Humphrey et al. 2010). The existing evidence is very limited to support whether pupils' sense of empathy can be improved through school-based interventions.

1.4 The role of schools

Schools are sites for learning for children where a lot of skills and behaviours are adopted as a consequence of growing up with other children and adults. Schools can help pupil foster skills for effective communication and socially accepted behaviours. OFSTED inspections judge pupils' social skills as one of the school effectiveness criteria. The hallmark of state-funded mainstream schools in the UK is heterogeneity. A comprehensive school, as the name suggests, is where children of different social classes, ethnicities, languages, family backgrounds and religions mix (DEMOS 2015). While heterogeneity in schools can create dissension and conflicts, it more often provides opportunities for children to learn important social skills like empathy, social cohesion, tolerance, acceptance and even critical thinking skills (Braster and Dronkers 2013, Roksa et al. 2016). Following the Special Education Needs and Disability Act 2001, mainstream schools now include children with a range of disabilities. Schools provide an ideal situation for children of different abilities or disabilities and social backgrounds to interact (Ajegbo et al. 2007).

School is also a place where teachers can be role models of acceptable behaviour. Unfortunately, this is often not the case. Less than half of the pupils surveyed in the Good Childhood report fully agreed that their teachers treat them fairly (The Good Childhood Report 2015). We have previously looked at how young people's views on trust, civic participation and their role in society are strongly linked to the way they are treated at school, and to a lesser extent to the kinds of school they attend (Gorard and Smith 2010). The role of the teacher when interacting with young people is therefore paramount in this. Programmes to sensitise children's attitudes towards emotions, feelings and the situation of others might play an important role in overcoming the common challenges of children's disruptive behaviour, bullying, racism, aggression and violence towards peers and teachers. However, teachers' attitudes and sense of empathy towards pupils often gets ignored.

Teacher's fair treatment in the classroom is positively related with children's self-esteem and confidence (De Witt 2000, Burnett 1999). Children's disruptive or aggressive behaviour could be the outcome of unfairness that they encounter, where they are unable to understand and communicate their feelings over it (Hart et al. 2001, Chory-Assad 2002). Children are sensitive to differential treatment on the basis of their social class, ethnicity, and disability (Tenenaum and Ruck 2007). These issues of fairness and justice are also associated with teachers' understanding of individual differences and the ability to deal with them in an effective manner. Children not only demand fair and just treatment from others but they also need to experience a variety of models where fairness and justice are served. This includes school scenarios such as equal opportunities to speak (Corden 2000), teachers' appropriate respect for all children (Gorard and Smith 2010), appropriate feedback and encouragement (Burnett and Mandle 2010), rewards for a wider set of skills and performance (Harlen 2006), freedom to choose tasks and times (Gorard 2011) and required emotional support (Prothero 2007). Pro-social behaviour is encouraged when children see it in adults, and learn to trust adults in schools on the basis of their fair actions and just values. Existing evidence also shows that children's overall well-being is associated with fair and just treatment in schools (Pretsch et al 2015, Paulsel and Chory-Assad 2005).

Schools can make inequalities worse by providing differential opportunities to learn (Schmidt et al. 2015). This might produce differences in political tolerance, volunteering, and intended political participation (Fleming et al. 2014). School systems with tracked or selective pathways or where there is high segregation by SES (social economic status) between schools tend to produce young people who, on average, have weaker civic knowledge (Collado et al. 2014), or poorer social skills (Gottfried 2014). The 2009 International Civic and Citizenship Education Study showed that intended participation in elections and other civic activities was linked to prior experiences at school (Quintelier and Hooghe 2013). In particular, a participatory democratic climate at school was beneficial in this respect.

Enhancing the enjoyment of school for young people, not only for the disadvantaged should also be a part of the policy agenda for education. Poverty is not the only cause of low life satisfaction although children living in relative poverty are likely to report low scores on satisfaction with life in general and specifically life at school (Goodman and Gregg 2010, Chowdry et al. 2010, Statham and Chase 2010, Tomlinson et al. 2008). Enjoyment of school, unlike academic attainment, is much less stratified by young people's background characteristics, and therefore open to improvement for all (Gorard and See 2011).

Schools are spaces where children's experiences shape their perception of the wider social world and therefore the role of school is crucial towards children's readiness to enter into an independent and practical life. Schools have the responsibility to provide each child the best chances of developing skills that a world of success, challenges, opportunities and competition demand from people. If the outcomes of school-based programmes for character building, higher aspirations, social cohesion, well-being and happiness can show a positive impact then there is a need to revise the curriculum by incorporating wider outcomes of schooling. There is a need to see education beyond the limit of academic attainment, where active social participation or citizenship is considered equally important as academic achievement.

But all of this work, including our own up until now, has been passive in design and some or all of the differences could be due to the self-selecting nature of students in some schools and the activities therein. If ideas like trust, citizenship, and self-confidence are malleable (and we believe that they can be) then it should be possible to provide more convincing evidence than above. Can such wider outcomes from schooling be improved through interventions?

Chapter 2 – Philosophy for Children and non-cognitive outcomes in primary schools

2.1 What is P4C?

Philosophy for Children (P4C) is a school-based approach practised in UK schools for over twenty years, developed from the establishment of the Institute for the Advancement of Philosophy for Children (IAPC) in 1970. In 1990 the BBC televised a documentary, 'Socrates for 6 Year olds' which showed P4C being practiced in one of the most challenging schools in Merrick, USA (Lipman 1990 BBC). P4C has since become a worldwide educational approach, and something like it has been adopted by schools in 60 countries across the world, although the nature of the practice varies (Mercer et al. 1999). In the UK, the Society for the Advancement of Philosophical Enquiry and Reflection in Education (SAPERE) was established in 1992 (http://P4C.com/history-P4C) to promote the use of P4C in schools. P4C can be seen as part of a worldwide critical thinking movement in education that has brought changes to teaching approaches and challenged the overall purpose of school education to some extent (Higgins 2015). There are various versions of the P4C approach and different names but all share the core ideas of promoting critical thinking, nurturing young people's curiosity, supporting them in using language of reasoning and argumentation, and sharing views for the better understanding of issues and each other.

P4C has received some criticism for its suggested lack of clarity in the measurement of objectives (Slade 1992), looseness in perceived understanding of P4C (Vansieleghem 2005), adaptability for achieving aims that can go against the fundamentals of free thought and inquiry, and not being a teaching subject where 'real' philosophy is taught (Hayes 2014, 2015, Kitchener 1990).

The P4C intervention for primary schools aims to help pupils to think logically, voice their opinion, use appropriate language in argumentation, and listen to the views and opinions of others. P4C, as promoted by SAPERE, is a template to practice, and to organising a classroom session for philosophical enquiry. It does not have any specified materials or stimuli that must be used; there are only examples and suggestions. It may involve standard material for teaching such as a projector, board, pens or sheets of papers. The steps outlined in training are a guide to organising the classroom dialogue and can be used flexibly as the teacher's expertise grows. For example, the stages do not all need to be completed in one session. Choosing a question in one session and discussing it in another is a popular option. There is also the expectation for teachers to use existing curriculum material in their lessons when they judge it to have the potential to stimulate philosophical discussion and clarify key concepts in subject areas such as democracy, justice, nation, history, truth, cause, evidence, beauty, art, real, belief, knowledge, tolerance, and theory.

Usually, pupils and teacher sit in a circle so everyone can see and hear one another. The teacher negotiates with pupils on guidelines for the conduct of the sessions, and the purpose here is to set some basic rules of communication agreed by all the pupils. The preparation also usually involves the introduction of terminology for supporting and structuring an argument. The teacher then introduces the planned material they have chosen in order to provoke pupils' interest, puzzle them or prompt their sense of what is important. A minute of silence is followed by pupils in pairs sharing interesting issues and themes, or jotting down key words. The teacher often records some of the key words and ideas that emerge. The purpose here is to develop suggested questions for discussion.

Children then present their group's question so that everyone can see and hear it. When all of the questions have been collected children are invited to clarify, link, appreciate or evaluate the questions prior to choosing one for discussion. The selection of one question as a dialogue starter can be made by pupils using one of a range of voting methods. The discussion is then open for all to share their views.

Pupils participate in the discussion, building on other pupils' contributions, clarifying them, questioning them and stating their own opinions. Whether agreeing or disagreeing with others the rule is to justify opinions with reasons. Teachers may prompt pupils to imagine alternatives and consequences, seek

evidence, quantify with expressions like 'all', 'some' or 'most', offer examples and counter examples, and question assumptions.

The session ends with last words from all pupils. Pupils might have the same opinion as in the beginning or it could have changed as a result of dialogue. Pupils are invited to sum up their views concisely and without contradiction from others. This activity could either be a verbal statement or a detailed reflection whereby a teacher could ask pupils to write a summary of their views.

The teacher invites reflective and evaluative comments about the enquiry with reference to broad criteria such as the guidelines the group has adopted. For example, the teacher asks 'What went well?' 'What could we improve on?' or 'What do we need to do next?'. The teacher could point to issues of pupils' behaviour and turn-taking in the session and ask them to reflect on their progress. The review could include suggestions on what needs to be focused on in the next P4C session.

The nature of the intervention in action is discussed further in the process evaluation section (below).

2.2 What is the prior evidence on P4C?

In terms of academic improvement, an initial evaluation of the original scheme reported gains in logical reasoning and reading using a matched comparison design involving only 40 pupils from two schools (Lipman 1980). One of the earliest studies in the UK was conducted by Williams (1993). This small study examined the effects of 27 one-hour P4C lessons (using Lipman's materials) on reading comprehension, reasoning skills and intellectual confidence. Participants were 42 pupils from two Year 7 classes in one school. Results were obtained for only 32 children. Pre- and post-test comparison of reading comprehension using the London Reading Test showed that the P4C group made bigger gains than control pupils. And much of the evidence that has followed has similarly been reported as 'positive' but with weak designs, small scale and often high attrition.

Previous reviews have suggested consistent small to medium effects on a range of outcome measures (Trickey and Topping 2004). A recent review identified 16 studies of teaching thinking skills more generally, of at least minimal quality (Gorard et al. 2016). Of these 13 were minimal quality and all reported positive effects on reading, while the two highest quality studies also reported positive effects (Worth et al. 2015, Hanley et al, 2015). One medium quality examination of the impact of dialogic discussion in a P4C tradition on writing and reading comprehension found no benefits in terms of post-intervention assessments (Reznitskaya 2012). A further reasonable quality study found subsequently also suggested gains in terms of CAT scores (Fair et al. 2015). And the results of the largest ever P4C trial, recently conducted in the UK, were promising in terms of Key Stage 2 maths and reading attainment, and slightly less so for CAT scores (Gorard et al. 2017a, 2017b). Disadvantaged pupils showed approximate gains of three to four months of extra progress. The cost per pupil was £16 which is cheap for the possible gains achieved at the end of a two year trial period.

Overall, therefore there is considerable promise from studies using P4C in primary schools and finding better academic outcomes in terms of literacy and sometimes maths. The overall body of work suggests that this kind of critical thinking can be improved in schools and is generally associated with improved outcomes, perhaps especially in science, maths and reading, and for the poorest students.

Some previous studies have also focussed on the non-cognitive benefits of P4C, sometimes in addition to academic attainment or Cognitive Abilities Test (CAT) scores and sometimes on their own (García-Moriyón et al. 2005, Tian and Liao 2016, Fair et al. 2015, Reznitskaya 2012). The Wiser Wales study by the Council for Education in World Citizenship tracked the impact of P4C across seven schools from 2009 to 2012 (Dyfed County Council 1994). The study reported improvements in pupils' social skills, concentration, their ability to question and relationships as a result of increased confidence in self-expression - based on pupils' and teachers' self-reports. Prior research suggests that there may be numerous wider benefits from P4C including self-esteem (Sasseville 1994), self-concept and confidence (Williams and Wegerif 2004), participation in classrooms (Topping and Trickey 2007,

Swain et al. 2013), interpersonal skills (Hedayati et al. 2009), social emotional behaviour (Colom et al. 2014), attendance, self-esteem, respect for others (Meir and McCann 2016), and motivation to learn, perhaps especially so for disadvantaged children (Creative Teaching and Learning 2015).

However, most of these studies are still small-scale with weak designs, unaccounted-for pupil age differences, and poor reporting of attrition. In summary, while there are signs of promise, not that much is really known about the impact of P4C on non-cognitive outcomes and so a much larger pilot trial of the kind we describe below was indicated.

Chapter 3 – Methods of investigation used in our new study

3.1 Design of the new study

The new study is a quasi-experimental evaluation of the non-cognitive outcomes of P4C. A total of 42 schools participated in the study. The schools were from diverse geographical regions of England, and each had a high proportion of disadvantaged pupils. Teachers from 16 schools received P4C training and support for implementing the intervention, following the protocol developed by Society for the Advancement of Philosophical and Reflection Enquiry (SAPERE). These had originally been the waiting-list control schools in our trial of P4C in terms of cognitive outcomes (Gorard et al. 2017). The comparison group was formed of 26 schools, known not to have implemented P4C. This comparison was considered efficient in terms of the cost of the evaluation because the comparison schools were already acting as a clean control as part of another EEF funded Youth Social Action project in which we were using the same non-cognitive survey items as used for P4C groups. It also minimised the extent to which the participating schools were requested to allow access and provide data. The only downside of using these 'naturally' participating groups of schools was that we had no control over the timings (see below). A total of 2,722 pupils provided both pre- and post-intervention survey responses. Of these, 1,099 were in the P4C group and 1,623 in the comparison group. In terms of measures of disadvantaged, 502 pupils were known to be eligible for free school meals, of which 311 were in the P4C group and 191 in the comparison (see below for more on this).

Pupils in the intervention group received P4C for nearly 18 months while those in the comparison schools continued business as usual. A pupil survey on non-cognitive skills was conducted at pre and post intervention stages.

Table 3.1 summarises the timings of when the non-cognitive surveys were conducted in schools. It shows that the P4C group were six months younger, on average, when they did the first survey, although both groups were the same average age for the second survey. This difference in age at the outset makes the use of scores representing change over time more complex than usual, and means that the *headline* findings in this paper are presented in terms of differences between the groups at the post-intervention survey only. However, the pre-intervention scores and the 'gains' from pre- to post are also presented for completeness, and an adjustment is made where possible for the difference of six months in pupils' age between the groups (see analysis).

Table 3.1 – Timing of survey data collection

	June 2014	December 2014	June 2015	June 2016
P4C Year 4		Pre-test Y4 start		Post-test Y5 end
P4C Year 5		Pre-test Y5 start		Post-test Y6 end
Comparator Year 4	Pre-test Y4 end		Post-test Y5 end	
Comparator Year 5	Pre-test Y5 end		Post-test Y6 end	

The timeline of the events can be summarised as:

June 2014
The comparison schools received their pre-test by the end of June 2014, while at the end of Years 4 and 5. Data on the schools and pupils was collected as part of a parallel project.

December 2014
The P4C schools sent a reminder opt-out consent letters to parents (having previously agreed to be part of the attainment trial). The pupil level data from the prior trial was revised for the printing of school and pupil names on individual non-cognitive survey forms. The pre-test was conducted with the Year 4 and Year 5 2014 cohorts. The schools received P4C training from November 2014 to February 2015. The schools began to implement P4C from January 2015 onwards.

June 2015
The comparison schools completed an interim survey of non-cognitive items (as part of a parallel trial) in June 2015. The Year groups were now at the end of Years 5 and 6. The results are treated as the post-intervention findings for this new evaluation. P4C schools were only six months into the intervention at this stage, and continued receiving the intervention during this time and were not given a survey to complete.

June 2016
The P4C schools Years 5 and 6 completed their post-survey.

Pupils in the comparison group continued the usual school activities. There was no chance of diffusion of the intervention because SAPERE confirmed that they had not provided any P4C training or support to the teachers in the list of comparison schools. These schools could have implemented other interventions for the improvement of pupils' non-cognitive skills. However, we had not observed any systematic approach adopted by these schools by June 2016.

There was no school dropout from either group. Pupil drop-out from this study was mainly due to absence on the day of the test and subsequently, a change to home schooling and pupils who changed schools over this time period and could not be traced. We collected information regarding pupils for whom the surveys were returned without completion, and followed them up where possible. There was no case of pupil refusal to complete the survey. The final analysis based on pupils with both pre- and post-intervention scores involved 968 in the treatment group (131 missing data) and 1,469 in the comparator (154 missing). The overall pupil drop-out rate is therefore around 10% of the initial sample. A simple comparison of the pre-tests showed that there were no differences between the mean scores of drop-out students and those who completed the post-test. However, around 50% of the drop-out pupils were FSM-eligible.

Table 3.2 - School characteristics

	No. of P4C schools	No. of comparison schools
Attendance		
Highest Quintile	1	5
2nd Quintile	1	9
3rd Quintile	2	2
4th Quintile	4	5
Lowest Quintile	8	7
FSM eligibility		
Highest Quintile	12	9
2nd Quintile	2	8
3rd Quintile	2	5
4th Quintile	0	3
Lowest Quintile	0	3
SEN		
Highest Quintile	5	4
2nd Quintile	4	8
3rd Quintile	2	6
4th Quintile	5	4
Lowest Quintile	0	6

Schools had volunteered to participate in the P4C intervention and in our prior Youth Social Action study. All of these schools were urban, mixed sex state-maintained primary schools in England. The proportion of disadvantaged pupils was high in most schools – especially for P4C schools. OFSTED has developed a data dashboard for school characteristics which ranks schools into five quintiles where each quintile represents 20% of the data. A school in the highest quintile means it is in the top 20% of schools for that characteristic in comparison to all other schools. Most schools in both groups are in high quintiles for FSM eligibility and SEN and low quintiles for attendance rates (Table 3.2). However,

it is clear that the schools proposing to do P4C are generally more disadvantaged in each category than those randomised to the comparison group in the Youth Social Action trial. This relative disadvantage needs to be taken into account in the analysis when considering the outcomes (see below).

3.2 The non-cognitive outcome measures

The pupil survey used here was developed by the research team at Durham University, in co-operation with researchers from Leicester University and the Cabinet Office. The pupil survey of non-cognitive skills is basically the outcome of a nationwide evaluation of Youth Social Action projects which are funded by the Education Endowment Foundation and the Cabinet Office. All items have a clear audit trail in the published works of Office for National Statistics (2013 and 2014), The Good Childhood Report (2015), Gorard and Smith (2010), and Youth Social Action in the UK (2014). They have been used previously with tens of thousands of children in several research studies in England and as far afield as Japan. The language, sentence structure and relevance to pupils' life, therefore, have been checked a number of times. We conducted a fresh pilot of the survey after selection of items, survey layout and design. We did not identify any problems in pupils' understanding of language, teachers' views on the survey and length of the survey. Using an existing survey means that many of the results will be directly comparable with previous and parallel studies.

The instrument used in the survey included 11 scaled attitude items, representing the best single question available from 11 established tests of psychological constructs such as self-confidence, determination, and well-being. The 11 items represent a wide range of behaviours, attitudes, emotions and feelings (see Appendix B). The questions were developed in association with the Cabinet Office, and represent either the item with the single highest loading on that construct or the item recommended for single use by the test developers. For example, the item 'I feel happy most days' is about pupil's happiness or well-being. Pupils self-report on a scale of 0 to 10 (0 identified as not at all true, and 10 as completely true). The scale is wide to permit variation in responses, especially over time. Two of the items are randomly reverse-coded so that the socially desirable response would be 0 rather than 10. This is intended to encourage a focus on the meaning of each item.

As an additional approach, to try and prevent pupils simply reporting socially desirable answers rather than their own position, we developed three vignettes. These are short stories or scenarios, selected and developed from Gorard and Smith (2010), with three feasible options as responses and no scale. The vignettes are about imaginary characters (children), and the choice of response statements are not right or wrong, but require a judgement based on empathy, or social responsibility respectively.

The instrument contains a small number of further items that are specific to the P4C survey, have no comparison group here, and cannot be compared with other studies. These include a further vignette based on democratic participation - the results of which at pre- and post-intervention appear in Appendix A. The purpose was to begin to calibrate these items for use in future studies on a similar theme.

3.3 Methods of analysis

We pre-selected two of the 11 point scale items as being particularly linked to the activities and aims of P4C. We did this to prevent the developers inadvertently focusing on items with favourable results, once the results were known. We have called these non-cognitive outcomes 'social and communication skills' and 'team work and resilience'. These two items, along with the vignette on empathy, are therefore the headline findings for the study.

All analyses used here are based on 'intention to treat', meaning that all pupils originally identified as eligible were tested and their outcomes analysed, regardless of the time actually spent on the intervention. In order to fulfil the protocol for 'intention to treat' analysis we followed the pupils who were initially pre-tested and later moved schools, where possible.

The non-cognitive measures were assessed by 11 scaled items. Each had the possible responses 0 to 10, which are treated as numbers for analysis, and are reported as 'effect' sizes for pre- and post-intervention. The effect sizes are the differences between the mean scores for the two groups, divided by the overall standard deviation.

The 'effect' sizes are presented primarily in terms of the post-intervention scores because the pre-intervention scores were obtained at different ages for the two groups. Research studies have indicated that in the same cohort pupils' age difference by months is relevant to their overall performance and non-cognitive skills (Bramley et al. 2015, Benton 2014, Gorard 2015). A recent analysis of Longitudinal Study of Young People in England (LSYPE) has shown that pupils born later in the academic year (July to August) are more likely to have lower scholastic competence (Crawford et al. 2013). Therefore, the six month difference in age at pre-test between the P4C and comparison groups is not valid as it stands. For the comparison group with no intervention, the change in attitude pre- to post is an estimate of the impact of one year of aging. Therefore *half* of the change from pre to post for the control group is our best estimate of the changes that would have happened to the P4C group in the extra six months of elapsed time between pre and post intervention surveys even without an intervention (see above). We use this to adjust the gain scores presented for the P4C group.

Although the post intervention results are the headline figures, we also present the gain scores from pre- to post-intervention adjusted in this way. The gain scores for the P4C group are therefore adjusted by half of the gain score for the comparison group, to try and represent the six month age difference at pre-test. However, it must be stressed that this adjustment makes a number of assumptions that cannot be checked. And that is why we made the decision, before analysis, to use the post-intervention differences only for the headline effect sizes. The pre-intervention and adjusted 'gain' scores are included for completeness, and as a caution showing the volatility of small 'effect' sizes under different assumptions.

In addition, we ran multiple regression models with the gain scores of each of the scaled items as the outcome variable, and entering whether the pupil had the following characteristics – FMS-eligible, SEN status, English as an additional language, and sex – as predictors in a first step, and whether the pupil was in the P4C group or not as the second step. This provides a useful illustration of any association with the intervention net of the known differing characteristics of the pupils in the two groups.

The results for the vignettes are reported with effect sizes as odds ratios pre- and post-intervention. The odds ratios (a.d-b.c) are calculated on each occasion for the two groups, comparing the number of cases selecting the option deemed most socially desirable (by the test developers) with the number picking any other option (including none).

In addition, and as a parallel to the regression analysis conducted with the scaled items, we ran a logistic regression model with the empathy vignette post intervention outcome (social desirable response or not) as the outcome, and entering the pre intervention response and whether the pupil had the following characteristics – FMS-eligible, SEN status, English as an additional language, and sex – as predictors in a first step, and whether the pupil was in the P4C group or not as the second step. This provides a useful illustration of any impact from the intervention net of the known differing characteristics of the pupils in the two groups.

These analyses are conducted with all pupils, and then the main analyses are repeated based only on those pupils eligible for free school meals (FSM), in order to estimate the gains, if any, for the most disadvantaged pupils. The number of cases is much smaller in the latter tables (661, with 312 in the treatment and 349 in the comparator group), and so the results are less secure than for the overall figures.

In order to help communicate how secure (or insecure) each effect size or odds ratio is, we also present the *number* of counterfactual cases that would be *needed to disturb* each finding, reported here as NNTD (Gorard and Gorard 2016). This number represents how many cases with outcomes counterfactual to the main finding would need to be added to the existing cases in order for the effect size to disappear. It is computed as the effect size multiplied by the number of cases in the smallest group. The result

(NNTD) can also be compared to the number of missing cases, to estimate how likely the result is to be caused by missing data.

3.4 Methods used in the process evaluation

In our prior evaluation study of P4C (Gorard, Siddiqui and See 2016), we had previously looked at what the comparison schools were doing, and making sure that it did not involve P4C. We now collected in-depth information on the more recent implementation of P4C in these other schools. We collected formative evidence on all phases and aspects of the intervention from the selection and retention of schools, through the training of teachers, to testing the eventual outcomes. This was used to assess fidelity to treatment, challenges of implementation, and the perceptions of teachers and pupils, including any resentment or resistance. We attended P4C teachers' training sessions as participant observers noting the process of implementing P4C, the methods of delivery and also teachers' responses to the training

We then observed the delivery of the programme in the schools to assess the implementation challenges, and to get feedback from teachers and pupils on the perceived outcomes of P4C. Eight schools were selected for close observation. These schools agreed to our request to support us in collecting formative evidence at least two times in a year. We observed 16 P4C sessions and interviewed their teachers. The purpose of re-visiting the same schools at the end of the project was to observe the changes in the quality of P4C sessions, and to see whether the regularity of implementation had been maintained. We also looked for changes in teachers' behaviour (their confidence and competence in the delivery) as well as changes in pupils' behaviour and attitude.

We observed P4C sessions as participants of the circle of enquiry and sometimes just sitting outside the circle. Following P4C session in their schools, we asked pupils for feedback regarding P4C, the topics chosen for sessions, their life in school and relationship with teachers and peers. Detailed field notes of these observations and interviews were taken and analysed. The analysis involved synthesising recurring themes and reporting the results in the light of non-cognitive survey outcomes.

This information was collected to report in-depth details about the programme as practised in real classroom condition. P4C is a whole-group approach which can be incorporated in classes for other subjects or conducted as separate sessions. The effects of P4C are expected to be seen in other school activities and areas of learning. Therefore, observations and interviews with teachers and pupils were deemed the appropriate approaches to gather information on the wider-effects of P4C.

The observations of P4C in action were non-intrusive, with the evaluator sitting either inconspicuously at the back of the classroom or more usually as part of a circle but not taking full part in the dialogue unless directly addressed. Interviews with teachers and pupils were also conducted as convenient during these visits. The interviews were relatively informal conversations with those involved in the intervention. On each visit a prior meeting was set up between the P4C lead and the teaching staff to discuss the lesson to be taught that day. In addition to observations and interviews we used teachers' feedback collected by SAPERE through an online survey. Teachers' feedback survey provided data on the regularity of P4C sessions implemented, the challenges reported in the implementation and information on the SAPERE resources utilised by the teachers for P4C sessions. The teachers also reported their experiences of using P4C in class through the online feedback survey.

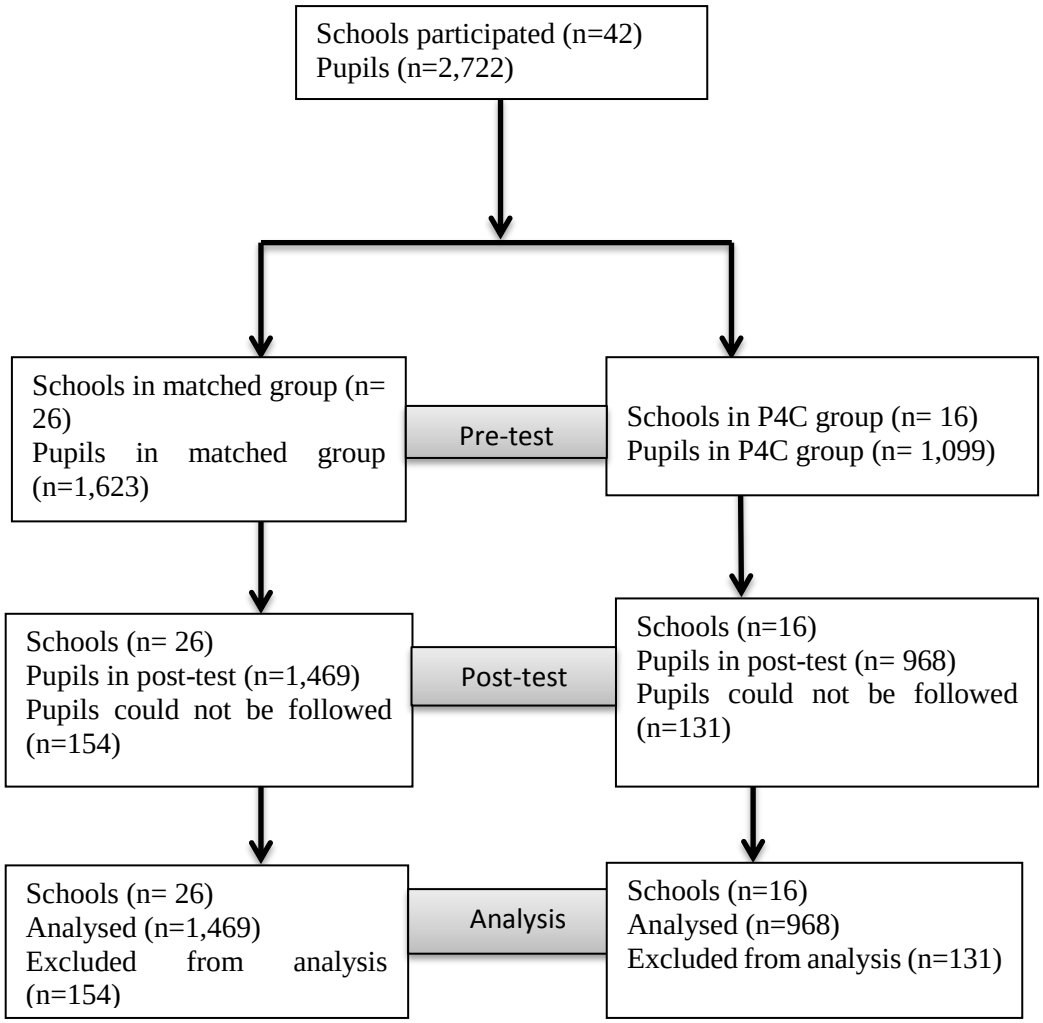
4.1 Pupil dropout

This project lasted nearly 19 months in total. There was no school dropout. Changes in school leadership, pupil absenteeism and pupil mobility between schools were all part of this real-life experiment. According to existing evidence, pupil school mobility in England is around 3% per annum. This is higher among pupils in primary schools than secondary schools. During Key Stage 1 6% of pupils change schools, and nearly 5% change school during Key Stage 2 years (Machin et al. 2006). On this basis, the approximately 10% of initial pupils with missing post-intervention scores is understandable (Table 4.1 and Figure 4.1).

Table 4.1 - Pupil dropout

	Pupil changed schools	Pupil absent	Total
P4C schools	45	86	131
Comparison schools	50	104	154
Total	95	190	285

Figure 4.1 Flowchart of school and pupil numbers in the evaluation



Student drop-out in this study is mainly due to pupil absence, home schooling and pupils who have changed schools and area over this time period. There was no case of a pupil refusing to complete the survey. As far as we could track pupils who did not complete the post-tests we found that 33% had changed schools and 66% were absent on the day when the post-test was conducted. A simple mean

comparison of the pre-tests showed that there were no differences between the drop-out students and those who completed the post-test. However, while 24% of the initial sample had been FSM-eligible, 50% of the drop-out pupils were FSM eligible perhaps reflecting the generally higher absence rate for this group.

4.2 Findings for the attitude scales

On the first headline indicator of impact – the self-reported response on communication with others – the P4C group are slightly ahead of the comparison group but were also ahead from the start (Table 4.2). The number of standard counterfactual cases needed to disturb this finding (NNTD) would be 97 (the smallest group times the effect size), which is smaller than pre- to post- intervention attrition. This must lead to further caution about the result. However, it is unclear if there is indeed any attrition where the effect size is based on post-test scores only, as it is here.

Table 4.2 - Differences in social and communication skills, all pupils

I am good at explaining my ideas to other people	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	6.42	2.81	6.25	2.58	-0.16	3.19	-
Comparison	6.03	2.64	6.00	2.29	-0.03	2.92	-
Total	6.19	2.72	6.10	2.41	-0.08	3.03	0.10

Note: the gain score for the P4C group is adjusted by half of the gain score for the untreated comparison group

Using the same headline indicator, the P4C group are considerably further ahead when considering only the disadvantaged pupils known to be eligible for FSM, and here have also shown greater gains over time (Table 4.3). The number of standard counterfactual cases needed to disturb this finding would be 44, which is larger than the pre- to post- intervention attrition for FSM-eligible pupils suggesting a strong result.

Table 4.3 - Differences in social and communication skills, FSM-eligible pupils only

I am good at explaining my ideas to other people	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	6.37	2.84	6.22	2.49	-0.13	3.07	
Comparison	5.92	2.69	5.65	2.43	-0.27	3.06	
Total	6.13	2.77	5.92	2.47	-0.20	3.07	0.23

A similar picture appears for the second headline indicator of impact – the self-reported response on ability to work with others (Tables 4.4 and 4.5). On post-test scores, the P4C group are ahead, whether for all pupils (NNTD=155) or FSM-eligible pupils only (NNTD=21). But again, the P4C were ahead at the outset anyway and this weakens our trust in the scale of the post-intervention differences.

Table 4.4 - Differences in co-operation, teamwork and resilience, all pupils

I can work with someone who has different opinions to me	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.26	3.03	7.16	2.77	+0.09	3.64	-
Comparison	6.51	3.12	6.75	2.76	+0.23	3.68	-
Total	6.81	3.11	6.91	2.77	+0.17	3.67	0.15

Table 4.5 - Overall post-test differences on cooperation, teamwork and resilience: FSM eligible pupils only

I can work with someone who has different opinions to me	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.18	3.19	6.92	2.88	-0.08	4.07	-
Comparison	6.25	3.48	6.61	2.75	+0.36	3.88	-
Total	6.69	3.38	6.75	2.81	+0.16	3.97	0.11

Table 4.6 shows the overall results in terms of 'effect' sizes for all 11 scaled attitude items. The P4C group are most obviously ahead in terms of three items - representing communication, team work and resilience, and social responsibility. They are behind in terms of one – happiness. And the differences negligible or zero on all other items. Considering only pupils known to be eligible for free school meals, the P4C group of disadvantaged pupils is most obviously ahead in terms of items representing communication, team work, self-confidence, empathy, and seeking help. Further details of the underlying appear in Appendix A.

Table 4.6 - Overall post-test differences on all 11 items, all pupils and FSM-eligible pupils

	Post-intervention 'effect' size, all pupils	Post-intervention 'effect' size, FSM-eligible pupils
'Communication skills' I am good at explaining my ideas to other people	+0.10	+0.23
'Sociability' I like meeting new people	+0.05	+0.06
'Cooperation and teamwork' I can work with someone who has different opinions to me	+0.15	+0.11
'Self-confidence' I can do most things if I try	+0.04	+0.10
'Determination' Once I have started a task I like to finish it	-0.02	+0.02
'Social responsibility' I want to try and make my local area a better place	+0.08	-0.03
I like to be told exactly what to do	-0.04	-0.03
I am often afraid to try to new things	-0.02	0
'Well-being' I feel happy most days	-0.09	-0.01
'Empathy' I try to understand other people's problems	+0.01	+0.08
I know where to go for help with a problem	-0.02	+0.08

Note: The highlighted items were independently selected as the most likely outcomes relevant to P4C, before the pre and post test results were known.

The regression models for gain scores in the two headline indicators are both very weak in the sense that the predictors explain very little (R of around 0.06 in Table 4.7). This could be because the gain scores are not valid, due to the different ages of the two groups at pre-test (see methods section).

Nevertheless, in both models, knowledge of which treatment group a pupil is in does not improve the prediction by much at all ('effect' sizes of near zero in Table 4.8). This is not a test of the intervention – the design and headline findings do that – but it is a way of presenting the results that suggest caution.

Table 4.7 – R from multi-stage regression models predicting total responses to two headline attitude items, all pupils

	Communication	Teamwork
Background predictors	0.057	0.070
Whether in P4C intervention	0.058	0.070

Table 4.8 – Standardised coefficients from multi-stage regression models predicting total responses to two headline attitude items, all pupils

	Communication	Teamwork
Gender	+0.01	+0.02
Ethnicity	-0.03	-0.07
FSM-eligibility	-0.01	-0
Special educational needs	-0.03	+0.04
English as an additional language	-0.18	+0.04
Whether in intervention group	+0.01	+0

4.3 Findings for the vignettes

The survey included three vignettes as measures of change in pupils' sense of empathy and generosity, social responsibility and understanding of democracy. The outcomes for the two main vignette results are presented in terms of odds ratios at each time. Table 4.9 shows that the responses of the two groups to the vignette on empathy and generosity are very similar at post-intervention, with the comparison group slightly ahead. At post-intervention, around 40% of all pupils showed a sense of generosity/empathy for a struggling pupil (..it is fair that the teacher should spend more time helping Jacintha, even if the other pupils have to wait) as compared to the other two options (2. Jacintha should work harder and 3. Jacintha should be taught in a separate class). This is in contrast to the pre-test where the comparison group reported similar results, but the younger P4C group had less empathy at that stage. Therefore, there may be some slight evidence of an improvement for the P4C group, and this is assessed further in the regression model below.

Table 4.9 - Vignette on empathy/generosity: Percentage agreeing and pre and post-intervention odds ratios, all pupils

	Pre-'empathy'	Pre-'Not-empathy'	Pre-Odds ratio	Post-'empathy'	Post-'Not-empathy'	Post-Odds ratio
P4C	31%	69%	0.67	39%	61%	0.92
Comparison	40%	60%	-	41%	59%	-

At post-intervention, the results for FSM-eligible pupils participating in P4C show them more empathetic than the control. They also show an increase in the selection of the 'empathy' option over time (Table 4.10). Although both groups increased in terms of this response, the difference of six months between the groups at the pre-intervention survey cannot account for the difference in favour of the P4C group. These are therefore reasonably strong results.

Table 4.10 - Vignette on empathy/generosity: Percentage agreeing and pre and post-intervention odds ratios, FSM-eligible pupils only

	Pre-'empathy'	Pre-'Not-empathy'	Pre-Odds ratio	Post-'empathy'	Post-'Not-empathy'	Post-Odds ratio
P4C	28%	72%	1.05	36%	64%	1.31
Comparison	27%	73%		30%	70%	-

The second main vignette on social responsibility also shows the comparison group to be ahead, with a slight relative increase over time for P4C pupils compared to others, both when considering all pupils and for FSM-eligible pupils only (Tables 4.11 to 4.12). And again this cannot be explained in terms of the difference of six months at the outset. However, our prior pilot study did not pick up how skewed the responses in this vignette would be, with well over 80% of pupils selecting only one of the three options (taking responsibility). This threshold makes it harder to see any changes over time or differences between the groups, in contrast to the first vignette. Therefore, this vignette is not considered any further.

Table 4.11 - Vignette on social responsibility: Percentage agreeing and pre and post-intervention odds ratios, all pupils

	Pre-'responsibility'	Pre-Not 'responsibility'	Pre-Odds ratio	Post-'responsibility'	Post-Not 'responsibility'	Post-Odds ratio
P4C	81%	19%	0.37	84%	16%	0.40
Comparison	92%	8%	-	93%	7%	-

Table 4.12 - Vignette on social responsibility: Percentage agreeing and pre and post-intervention odds ratios, FSM-eligible pupils only

	Pre-'responsibility'	Pre-Not 'responsibility'	Pre-Odds ratio	Post-'responsibility'	Post-Not 'responsibility'	Post-Odds ratio
P4C	81%	19%	0.27	82%	18%	0.29
Comparison	94%	6%	-	94%	6%	-

The third vignette on democracy is an additional item which cannot be compared to the results for the comparison group. Our main purpose was to pilot this item and get a sense of the likely responses, for use in future studies. The vignette is based on whether a leader for a group task should be picked blindly, selected by an adult or voted on by all pupils. For what they are worth, the results can only be reported here in terms of before and after (P4C) figures.

The percentage of children picking the voting response increased noticeably after P4C (Table 4.13). Because the comparison group did not answer this question, it is not possible to conclude whether the result is a consequence of exposure to P4C or maturation in pupils' age or experience. However, P4C sessions involve pupils 'voting' to select questions for discussions so it is possible that by doing P4C regularly pupils come to understand the process of democracy and the value of voting better.

Table 4.13 - 'Vote for a leader' vignette on democracy: Percentage pre and post-tests odds ratio, all pupils, and FSM eligible pupils

	'Vote for a leader'	'Not vote for a leader'	Odds ratio
P4C pre-test	45%	55%	1.2
P4C post-test	50%	50%	-
P4C pre-test FSM only	40%	60%	1.4
P4C post-test FSM only	49%	51%	-

The binary regression model for the empathy vignette is as weak as the model for the attitude items (Table 4.14). The pre-intervention score is the best single predictor of the later score (Table 4.18). Those in the P4C group have only slightly higher improvement (odds of 1.05). As above, this is not a test of the intervention – the design and headline findings do that – but it is a way of presenting the results that suggest caution.

Table 4.14 – Accuracy of logistic regression model predicting responses to empathy vignette, all pupils

	Percentage of case predicted
Baseline model	60.0
Background predictors	61.3
Whether in P4C intervention	61.4

Table 4.15 – Odds ratios from logistic regression model predicting responses to empathy vignette, all pupils

Female pupil	1.03
Minority ethnic group	0.93
FSM-eligible	0.94
Having special educational needs	0.79
English as an additional language	1.01
Prior response to same item	2.13
In intervention group	1.05

Summary

As a feasibility study for assessing non-cognitive outcomes, the study shows some promise from P4C for areas like empathy, co-operation, and communication, and few if any downsides. This is perhaps especially the case for disadvantaged pupils as indicated by FSM-eligibility. What did the participants say?

Chapter 5 – Findings from the formative evaluation

5.1 The implementation of P4C

The schools were supported by SAPERE staff trainers who provided a P4C demonstration in each school. The recommended minimum 'dosage' for P4C implementation is one session per week which teachers can either incorporate in the time slots for other subjects or conduct separately. Teachers reported that finding regular gaps in school schedules to be devoted to P4C was the biggest challenge, because the main priority of subject teaching is to focus on the content-based knowledge. Many of them integrated P4C in subjects like History, English and PSHE. The trainers followed-up each of the schools by providing plans and worksheet ideas to the teachers in order to embed P4C in regular teaching. However, due to other curriculum activities and events such as OFSTED inspections, it was not always possible to maintain the regularity recommended by SAPERE. A senior management team leader reported that the national assessment system requires schools to achieve literacy and numeracy targets and unless those targets are directly addressed through teaching then school resources and staff time cannot be devoted to other aims or approaches. A few school leaders also perceived the P4C approaches as a waste of time with a risk of diverting pupils' attention from focus on learning. There was sometimes a tension between the school research lead and others in this respect.

Two of the 16 schools could not continue P4C on a full and regular basis because of school conversion to academy status leading to changes in school staff and senior leadership. This also highlighted the role of management in supporting teachers when embedding any new practice in teaching and school culture. In schools where P4C was observed to be fully embedded, the senior leaders were found actively engaged with teachers and P4C trainers. As far as we can tell, the other 14 schools implemented the intervention fully.

P4C is a practice of dialogic teaching. There is no syllabus or necessary methodology for the SAPERE approach to P4C. There is a danger that this approach may be open to the influence of teachers' biases, beliefs and ideologies, and examples of this were noted in our fieldwork both in this study and our previous work in the area.

A few pupils in some of the necessarily large enquiry groups were sometimes neglected by the teachers and their peers. It was observed in the sessions and was also reported by the pupils that they wanted to contribute at certain points and put their hand forward but teachers just moved on or gave the opportunity to another pupil. Where, as is desirable, the speaker decides who speaks next there is a fine line between a genuine back and forth between two pupils necessary for sustained argument, and abuse of the system by groups of friends.

It was observed by the evaluators that a complete P4C session should usually cover a complete circle of enquiry otherwise pupils would not gain the sense and purpose of the whole activity. For example, in one of the sessions the discussions initiated were not summed up and sufficiently reviewed. The session was rushed to the end as the time for the session was passing quickly. As observers we felt that the pupils had not really understood the sense and purpose of the discussion because there was no proper conclusion. Sometimes pupils said that the questions were not selected fairly and that other pupils cheated and voted for their friends' questions. In one of the sessions the pupils were not given enough thinking time, and this was possibly the reason that they could not reflect on the issues to develop interesting questions for later discussion.

5.2 Examples of topics and questions

Some of the examples of questions discussed in P4C observed sessions were as follows:

- Is it acceptable for people to wear their religious symbols at work places?
- Are people's physical looks more important than their actions?

- What is kindness?
- Can you and should you stop free thought?
- Is it OK to deprive someone of their freedom?

This list of questions, and many others, was created by pupils themselves from a given stimulus such as a story or short video, using a blind voting system. The substance of these questions is clearly relevant to the broader purpose of schools.

One of the sessions was aimed at the creative thinking of pupils. The stimulus was to trigger imagination and allow pupils to think beyond the abilities of human senses. The teacher presented a stimulus in the form of a real object, a replica of the object, a picture of the object and an imaginary object. The pupils were then asked to say a few sentences individually about the forms of object presented and then asked to make questions in pairs. The top three voted questions were:

- If we can't see many things does it mean that they don't exist?
- Why are different things named the same?
- Is it important to name things?

The follow up discussion was based on the question that the pupils voted for. The teacher encouraged pupils to debate, using justification for their views. As pupils knew that there are often no right or wrong answers in P4C session this seemed to encourage them to take part in the discussion.

Another P4C session was observed in which the teacher shared the stimulus of a picture based story about a bird that could not fit into any other families of birds, due its peculiar looks. The following questions were suggested by pupils:

- Why one needs to have a family?
- Is it important to know who you are?
- Is it important to belong to a family and look like the family members?
- Why did the bird look different from the other birds?
- If you look different does it mean you are also different from inside?
- Why are unique things not easily accepted?

The pupils shared their views very openly about what they thought a family meant to them. Pupils appeared to be used to the rules of communication in P4C. No one talked over others, and despite contrasting opinions on the topic the pupils listened to each other openly and carefully. The teacher reported at the end of the session that it took a few sessions to embed these rules of communication. She said that P4C had a great benefit for pupils in learning communication skills, and very swiftly it became a norm to listen and wait for a turn to speak. She said that P4C sessions are an opportunity for her to do things that otherwise gets ignored, such as listening to individual pupils' points of view and sharing good practice in communication skills.

In another school, the teacher presented a stimulus based on an animated story of a Roman soldier. The soldier was the shortest in height and shown being bullied and laughed at by others in the army. Despite his short height he never gave up on any task and mostly won over other soldiers by using his clever tricks. The questions raised by pupils for discussion were:

- What is bravery?
- Is being clever more important than being physically strong?
- Why people bully others?
- Is it fair to win by cheating?

One session involved a stimulus of a documentary on baby whales nurtured in a zoo, separated from their mothers and from an open sea environment. The pupils seemed interested in the topic of animal

welfare, and the topic was discussed with several differing arguments for and against zoos. The popular questions from pupils were as follows:

- Should baby whales be kept in captivity?
- Should animals be paid to perform for entertainment?
- What if endangered species are not preserved?

The teacher encouraged pupils to debate, using justification for their views. As pupils knew that there are often no right or wrong answers in P4C session this seemed to encourage them to take part in the discussion. As a result of discussion some pupils changed their views regarding keeping animals in a zoo. After the session ended the pupils resumed their chairs and places in the class. The teacher showed us her P4C planner and evaluation sheets. Detailed records were kept for each session, and individual pupil participation was closely monitored. The teacher reported that she found P4C a very valuable approach in the vocabulary development of pupils. She also reported that the benefits are much more profound for the disadvantaged pupils as they gained a lot of confidence and motivation through participating in the sessions.

In a school where the class was made up of pupils from many different ethnic origins, the teacher used an Irish song about migrating to America in the hope of a better life as the stimulus. It was a large class, and there was a class teacher and a teaching assistant who conducted the session together. The teaching assistant's role was observed to be very helpful in managing a large group and specifically in supporting pupils who needed assistance in writing questions. After a few starter activities the teacher read the poem and showed a video clip based on the song. The difficult words and meaning of lines were discussed before the pupils were asked to make questions. The following questions were proposed by pupils after discussion within groups of three to four:

- Is the grass always green on the other side?
- Does migration solve people's problem?
- Why is hope so important?
- If you are wealthy, will you always have a good life?

The pupils actively participated in the discussion. The teacher discussed words from the song such as feat, feast, homeless and encouraged pupils to share their views in the context of the song. There were several debating points on the theme of migration and pupils shared their views openly. The teacher reported that she had conducted P4C quite regularly and it helped children to integrate with the school culture where there are pupils and teachers from different ethnic backgrounds and religions. She found that P4C gave pupils a voice of their own and confidence by encouraging them to justify their views. Although this was not their purpose, many of the sessions had themes about integration and tolerance that could form a useful and non-didactic element of teaching concerns about radicalisation.

In general the success of P4C sessions were found to rely on the teacher's preparation of the session, enthusiasm to conduct the enquiry regularly, willingness to accept challenging arguments from pupils, and being aware of personal bias and readiness to accept justifications against personal beliefs and choices. The perceived benefits of P4C, according to leads, teachers, and pupils, can be envisaged as improved respect for others, social and communication skills, language and logic, and enjoyment.

5.3 Respect and behaviour

The P4C sessions we observed in eight schools covered a wide range of skills and concepts. The intervention is apparently appealing to many schools as a way of raising and debating pupil-school discipline problems in an enquiry group. The school leads reported that they discussed the concepts of bullying, racism, lying and cheating, equality and fairness which are core issues of school discipline and ethos. Teachers reported a wide range of positive behavioural outcomes as an impact of P4C. These were noticed within P4C settings during the circle of enquiry, in more general classroom settings as well as during play time.

> We have noticed a fall in the number of behavioural incident forms being submitted, suggesting a fall in the number of incidences of poor behaviour. We think this is because the children are now more able to communicate with each other.
>
> Children seem less likely to squabble at play times. The occasions of telling tales has decreased a lot.
>
> Children deal with conflict resolution in a mature and grown up way and think about actions after an argument.
>
> [They are] more respectful of each other now.
>
> The children are able to articulate more about how they feel and feel confident to disagree without it being seen as a cause for quarrelling.

P4C promotes the practice of valuing differing views and opinions. The initial stages of setting P4C rules are mainly about embedding the concept of effective communication practice in the circle of enquiry. This includes setting the codes of behaviour such as waiting for a turn to speak, valuing an individual's opinion by listening carefully, and the use of language for disagreement. The teachers attributed positive changes in pupils' behaviour to P4C practice. Improvement in pupils' respect and behaviour are related to pupils' social and communication skills, cooperation, team work and resilience. In terms of social and emotional literacy, the self-reported benefits suggest that to some extent P4C can be a useful approach in overcoming the challenges of disruption in the classroom, and even bullying and other anti-social behaviours.

The teachers also reported an important change in pupils' behaviour in response to arguments. Pupils were thought to be becoming more respectful of different opinions:

> Children aren't afraid to question each other, or refer back to someone else's point from earlier on in any lesson. Children are confident to challenge me, the teacher, without being disrespectful or rude.
>
> Children have become more accepting of the different opinions within the group. One SEN pupil, who is a very good speaker, seems to have gained more respect from his peers as he is often vocal and persuasive in our debates.
>
> The children in class are more willing to reason with each other and listen carefully before jumping in and talking over each other.

5.4 Social and communication skills

Related to this apparent increase in respect for others is better communication more generally, as commented on by several P4C teachers. A teacher reported in the interview that P4C sessions are the opportunity for her and the class to learn and practice listening skills which otherwise gets neglected. In another interview a teacher said that listening to each other in P4C has helped pupils learning to build their own arguments. Other teachers commented:

> I think it had helped lots of children with listening to each other and the relationship with any learning partner.
>
> Children given more time to talk without me talking as much, and more time to listen to each other.

> I think P4C has really helped the children to be aware of each other and the views that other people have that may conflict with their own. Their listening and reflecting skills have improved as a whole.
>
> Children ask deeper questions and are beginning to listen more attentively.
>
> Children seem to listen to one another's points of view after a disagreement.
>
> The children use the words I agree and I disagree, and listen carefully to each other. They often follow on from each other and ask each other questions.
>
> There is a real emphasis that children need to learn to listen to each other more, asking more questions and starting to ask why things occur etc... Communication is definitely the key to understanding each other which might improve relationships in a forever changing world that we live in.

5.5 Language and logic

Most teachers reported that they found P4C effective in helping pupils to develop their voice and giving them confidence to speak:

> Children have become more aware of other children's opinions in the classroom and understand that they can have different opinions without it arising to being an issue.
>
> Noticed improved confidence in one child who is very able but quietly spoken, he seems to have found his 'voice' through P4C and this is now impacting on all curriculum areas.
>
> I've noticed children have been more considerate towards other's opinions and views. There is an appreciation of different opinions, which I hope children will be able to transfer to other social situations rather than just in P4C sessions.
>
> It can be linked to other core areas such as reading and if children are encouraged to support their answer and give reasons for their opinion.
>
> I am enjoying listening to the conversations and reasoning for when the children give their opinions and are trying to convince others to follow their view. I can see a lot of benefits of doing this regularly in class.
>
> Unprompted feedback from staff who are involved in a long term Animation/art project who visit the school every six months have stated "that the children have become far more capable at explaining their ideas and concepts in lesson times".
>
> The children are now more willing to challenge each other's opinions. They are also more willing to suggest reasons for their thinking. As a teacher I am more confident with accepting differing answers to questions.
>
> Giving the child the chance to be heard and have their ideas listened to has vastly improved their self-esteem. They understand it's ok to be different and have different ideas. This makes them unique and their learning and questioning is now tailored. I feel its impact has been a success. We are educating children not robots. Individuality is not seen as weird amongst their peers but more accepting. It's been great.

Some teachers reported that they experienced an improvement in their own understanding of pupils, and their perspectives on pupils' world views – and that this was very valuable. Teachers' personal values, beliefs, actions and attitudes towards others can play an important role in pupils' understanding

and attitudes to life outside and after school. Therefore, P4C teachers need to be aware of their position, bias, actions and attitudes in their classroom and be able to balance their influence on pupils' perception and participation during P4C sessions.

The teachers reported benefits of P4C in the form of improvement in pupils' ability to use the language of reasoning, and becoming confident in using these language skills in other lessons and subject areas:

> Children are using the language used in P4C sessions in other subjects.
>
> Pupils occasionally use language of inquiry during other subjects and conversations.
>
> Encouraging children to use P4C language in other curriculum areas
>
> More children answering questions using appropriate language, sentence starters.
>
> The children use the language of disagreement and agreement far more than they used to. They also now link their ideas more succinctly and refer back to previous points. It has had a really good effect on the less able children who have grown in confidence in sharing their ideas.

The teachers also believed that P4C provided an opportunity for discussions that further supported pupils to think independently and talk freely:

> The formation of questions has developed well during an enquiry with more open-ended questions linked to the big ideas discussed. I feel it is an important time for children to think about things in the world; with the opportunity for them to make their own mind up about an issue rather than being told.
>
> Made me more aware of allowing discussion in classroom.
>
> Better, clearer explanations whilst pupils are more willing to discuss topics in greater detail.
>
> Allows the children to more openly discuss their ideas and feelings in a safe environment and gives them the skills to do this outside the classroom.
>
> Children have been using agree, disagree and extend questions and have been bouncing questions and answers between one another in ways they previously were not.

Many teachers reported improvements in pupils' thinking skills. One stated that P4C makes pupils think out of the box and see different perspectives of one thing. Another reported that allowing pupils some time to reflect is often neglected, and P4C is the time where she thought pupils get time to practice reflection. Teachers gave the following comments on the association between P4C and thinking skills:

> It has given the children room to investigate different situations in a different way.
>
> Children ask questions that involve reasoning in more of a logical way.
>
> It has helped children appreciate other may think differently from themselves - and that this is ok!
>
> Children are able to develop their thinking further, sometimes allowing themselves to change their minds.
>
> Better learning community. Greater thinking skill and ability to understand that not all questions have a right answer.
>
> Children ask deeper questions and are beginning to listen more attentively.

> Children are now more able to accept others' alternate viewpoints, without viewing these as criticism of themselves.
>
> It gives children an opportunity to see teacher in a different light.

5.6 Enjoyment

The upshot of all of these other factors was that pupils were generally seen to be enjoying the sessions. The teachers reported the following changes in pupils' responses:

> Children look forward to P4C sessions and enjoy the active 'starter games' that we have termed 'PE-4-C'. Children who are perhaps more reluctant to volunteer information in Literacy and Numeracy, feel much more confident to take an active part in discussions.
>
> Certain lessons have become much more pupil led.
>
> Some children enjoy the idea of exploring ideas but others find it difficult as they prefer concrete answers.
>
> The children contribute more to lessons and this has made lessons more interactive.
>
> The children respond to pupil comments with more questions.
>
> The children in my class had already established friendship groups and there are now core values of respect between the children. It is therefore difficult to know whether this has improved further due to P4C or because of maturity levels.

Pupils generally enjoyed P4C sessions, perhaps because they are interactive, pupil-centred and liberating. Pupil interviews and observations confirmed that pupils enjoyed P4C and actively participated in the sessions. In general, P4C sessions were observed to be very different from ordinary primary classroom settings, and to involve a different communication practice between teachers and pupils. Pupils were observed in the P4C sessions to have a certain level of excitement in discussion and voting for the questions. Many pupils commented on this in our interviews with them:

> I really enjoy P4C because I love to share my feelings in class because I feel my classmates and teacher will listen to me. I also enjoy looking at videos because we get to talk about things together.
>
> Sometimes you talk about things you had always wanted to let out.
>
> You can always express your feeling and you can never get judged or bullied for what you think.
>
> I like to hear people's opinions and problems.
>
> I like sitting in a circle and hearing different ideas and other things about other people's life and what they get up to on school holidays, I also love talking about things that you can have a debate and discuss it.
>
> P4C is a great opportunity to share our feelings and share the things that are troubling us. I also enjoy talking about things happening around the world.

The pupils also reported that sharing opinions helped them in building ideas:

> P4C is a good lesson because you get to share your feelings and what could happen in life and what can't. I like P4C because you can talk about what is true or false.
>
> We can say our own opinions on what we think and the teacher helps us we also build on each other's opinions. And after school I go home and speak to my parents about the things that we done in P4C
>
> I like P4C because you can say any think [*sic*] you like.
>
> I enjoy doing P4C because I like to do different stuff on each day, and I like to talk in different groups and I like to share my ideas to the class.

5.7 Possible barriers and downsides

A few (the five reported here) pupils described their reluctance to share feelings publicly, or that they were sometimes upset by the topics:

> I don't really like P4C that much because I don't really like talking so I get really bored when other people talk. But I like it when people talk about nice things.
>
> I do not like listening to other peoples' ideas for a long time!
>
> I don't like sharing my ideas in case people don't agree. In a way I do like P4C because I like listening to others points of review [sic].
>
> It is boring things to talk about and we sometimes say personal things.
>
> It upsets me sometimes and I like it sometimes because I can learn about different things.
>
> I get to share my opinion with other people and I like doing it with Miss*** and sometimes I get a little sad.

The pupils' views show that they are not all the same in terms of sharing ideas, talking and willingness to participate in the discussions. Some pupils are naturally quiet but keen on listening to others, whilst some get bored easily when they have to listen to others. It is perhaps part of the process of learning through dialogue that pupils would sometimes become anxious about conflicting perspectives. A crucial role for the teachers is to ensure that challenging issues are covered, but that they do not lead to distress.

Chapter 6 - Conclusions

6.1 Strengths and limitations of the study

This was a cost-effective independent evaluation, using existing groups of schools for a comparison that involved no artificial allocation to treatment or not. All of the schools did what they would have been doing anyway. We are certain that the comparison schools have not undertaken the P4C intervention. It was a large study involving 42 schools and 2,437pupils, with no school dropout and relatively low loss of pupil data.

The design for this study, making use of the waiting-list schools from one randomised study and the control schools from another, was convenient, cost-effective and permitted a larger sample size than otherwise. It is also clearly a better design that a simple cross-section or a before and after design. However, it is not ideal for a causal question such as that addressed in this paper.

It is intrinsically hard to collect evidence on the kind of non-cognitive outcomes covered by this study. The instrument used has a clear audit trail to items from standardised questionnaires, and it has been piloted and used extensively in prior work in collaboration with the Cabinet Office. The use of reverse-coded items was intended to make pupils focus, and the vignette items were developed to prevent pupils simply picking a point towards the most socially desirable end of any scale.

The post-intervention surveys were conducted at the same point in the school year, for children of equivalent year groups, in both analytical groups for the evaluation. The pre-intervention surveys were completed about six months apart in the school year for the treatment and comparison groups. An adjustment has been made for this, but the most trustworthy results are for the post-intervention scores only. All of these factors must reduce our reliance on the results to some extent.

There was an integrated process evaluation, involving consideration of observations, interviews and documents. We found no other threats to the security of the study.

Using the approach to judging the trustworthiness of research findings proposed by Gorard (2014), this study is judged to be 2★ out of a possible 4 (see Appendix C). The cases are at school-level, not randomised, there is some initial imbalance, some attrition, and the outcome measures are inevitably weaker than an examination or standard test of attainment. This means that the effect sizes need to be interpreted with considerable caution. The results are deemed promising both substantively and for the assessment of non-cognitive outcomes, but for more secure findings a true randomised control trial of similar or even larger scale is advised.

6.2 Implications for future research

The use of vignettes in this study, providing stories with multiple responses that overcome the social demands of a scale with a clearly desirable end point, has been promising. The vignettes were first used by and reported in Gorard and Smith (2010), and have now been completed by over 50,000 pupils in 10 countries around the world. We have developed further vignettes on democracy, trust in others, and religious diversity, as part of this project. They were piloted and we will be using these in future projects. We recommend other researchers consider this approach either instead of or in addition to the more usual psychometric scaling.

However, the vignettes still rely on self-report. In future studies, we would like to go beyond this approach, and are currently considering two ways forward which would unfortunately have additional cost implications because of their time-intensive demands. The first is to create real-life scenarios rather than vignettes, in which pupils have to respond through behaviour other than self-report. An example of the kind was used by (Kirkman et al. 2016). In one task, pupils were given four 50 pence pieces to

decide whether they would keep the money and how much to donate to charity. In another, pupils were interviewed for a post, and their performance was assessed by experienced job hirers.

The second realistic approach to assessing impact was suggested by the process evaluation for this study. Teachers reported that children became more responsive to each other, and that some individuals who had previously kept quiet in class became more communicative. Some teachers reported that pupils were using the language and structures of argumentation, as learnt via P4C, in other subjects at schools. These and other claims could be assessed more rigorously in a randomised control trial. A schedule of observation would be developed for researchers to record these kinds of pupil behavior for control and intervention group, at the outset and once the trial was ended. These and other approaches would give more secure knowledge of the extent to which non-cognitive or at least behavioural outcomes at school had really changed.

It is important to investigate the long term impact of P4C or similar approaches on pupils' future success, career paths and well-being in adult life. This would be possible by tracking the children over time to secondary school and comparing them with the rest of their cohort. A particular focus on the life outcomes of disadvantaged pupils would give further insight into the impact of P4C on this group.

6.3 Implications for policy and practice

There are some indications from all elements of the new study described here, and from other work such as our true trials or randomised control trials of Youth Social Action (Gorard et al. 2017c), that pupils' non-cognitive skills can be altered at least slightly through school-based interventions. There is still a gap in research evidence needed to establish a causal link between non-cognitive skills and academic outcomes. If interventions for the improvement of non-cognitive skills can also yield better academic outcomes, then there is scope for integrating these interventions in the national curriculum and most importantly using the pupil premium funds in implementing these interventions. But in their own right they may improve behaviour, co-operation, self-confidence, empathy and tolerance for others. And they may be feasible for practitioners facing the demand that they tackle extremism, radicalisation and so-called 'British values'.

From repeated evaluations of P4C we have gathered results showing persistent, small and positive effects of this kind of intervention. It is clear that this approach does not harm children's achievement of academic outcomes or their non-cognitive skills. It seems to reduce the gap between disadvantaged children and the rest, and it is clearly enjoyable for teachers and most pupils. It can also be beneficial for teachers, as the process evaluation results show. It does require training for teachers, especially in how to intervene on the small number of occasions that things go wrong, such as pupils being distressed.

The overall findings of this research contribute to the ongoing policy concern about the purpose of schools in the society and how far it is possible to make schools meaningful and effective for the most disadvantaged groups. This research suggests the value of developing space in the school curriculum where the focus is the development of pupils' character and values – but not through dirigiste pedagogy. It is important that for P4C to be effective it has to be a whole-class approach so that all individuals concerned are aware of the aims and objectives of P4C in cultivating empathy, respect and appropriate or acceptable behaviour. For example, to teach children fairness, teachers themselves have to be seen to be fair. To teach children to be polite, teachers and other (older) pupils have to practise it.

References

Ajegbo, K., Kiwan, D. and Sharma**,** S**.** (2007) *Curriculum review: Diversity and Citizenship*, London: DfES

Bailey, P. (2015) *Guidance for Working with Children and Young People who are vulnerable to the messages of Radicalisation and Extremism*, London Borough of Merton: Merton Safeguarding Children Board, http://www.merton.gov.uk/mscb_prevent_guidance_final.pdf

Benton, T. (2014) Should we age-standardise GCSEs? Cambridge Assessment Research Report, http://www.cambridgeassessment.org.uk/Images/285155-should-we-age-standardise-gcses-.pdf

Bonnell, J., Copestake, P., Kerr, D., Passy, R., Reed, C., Salter, R., Sarwar, S. and Sheikh, S. (2011) Teaching approaches that help to build resilience to extremism among young people. Department for Education (DFE): Research Report DFE-RR119. (Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/182675/DFE-RR119.pdf) (Last seen on 2nd September 2016)

Botting, N. and Conti-Ramsden, G. (2000) Social and Behavioural Difficulties in Children with Language Impairment Child Language, *Teaching and Therapy*, 16, 105–120

Bramley, T., Rodeiro, C. and Vitello, S. (2015) Gender differences at GCSE. Cambridge Assessment Research Report, http://www.cambridgeassessment.org.uk/Images/gender-differences-at-gcse-report.pdf)

Braster, S. and Dronkers, J. (2013) *The positive effects of ethnic diversity in class on the educational performance of pupils in a multiethnic European metropole*, London: Centre for Research and Analysis of Migration

Brighouse, H. (2008) Education for a Flourishing Life, Yearbook of the National Society for the Study of Education, 107, 1, 58–71

Brunello, G. and Schlotter, M. (2011**)** Non-Cognitive Skills and Personality Traits: Labour Market Relevance and Their Development in Education and Training Systems. IZA Discussion Paper No. 5743, http://ssrn.com/abstract=1858066

Burnett, P. (1999) Children's Self-Talk and Academic Self-Concepts: The impact of teachers' statements. *Educational Psychology in Practice*,*15*(3), 195-200

Burnett, P. and Mandel, V. (2010) Praise and Feedback in the Primary Classroom: Teachers' and Students' Perspectives, *Australian Journal of Educational & Developmental Psychology*, 10, 145-154

Carneiro, P., Crawford, C. and Goodman, A. (2012) The impact of early cognitive and non-cognitive skills on later outcomes, IFS, https://www.ifs.org.uk/publications/5986

Carr, D. (2015) The paradox of gratitude, British Journal of Educational Studies, 10.1080/00071005.2015.1011077

Chory-Assad, R. (2002) Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and aggression. *Communication Quarterly*, 50(1), 58-77

Chowdry, H., Crawford, C., Dearden, L., Joyce, R., Sibieta, L., Sylva, K. and Washbrook, E. (2010) Poorer children's educational attainment: how important are attitudes and behaviour?, *Joseph Rowntree Foundation*

Clegg, J., Hollis, C. and Rutter, M. (1999) A Survey of the Nature and Extent of Bully / Victim Problems in Junior / Middle and Secondary Schools, *Educational Research*, 35, 3–25

Collado, D., Lomos, C. and Nicaise, I. (2014) The effects of classroom socioeconomic composition in student's civic knowledge in Chile, *School Effectiveness and School Improvement*, 26, 3, 415-440

Colom, R., Moriyon, F. G., Magro, C., and Morilla, E. (2014). The long-term impact of Philosophy for Children: A longitudinal study (preliminary results). *Analytic Teaching and Philosophical Praxis*, 35(1), 50-56.

Common Select Committee: Purpose and Quality of Education in England, 25th January 2016 https://www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/purpose-quality-education-forum-15-16/ (Last seen on 31st July 2016)

Corden, R. (2000) *Literacy and learning through talk: Strategies for the primary classroom*. McGraw-Hill Education (UK)

Crawford, C., Dearden, L. and Greaves, E. (2013) *When you are born matters: evidence for England*. London: Institute for Fiscal Studies

Creative Teaching and Learning (2016) Closing the achievement gap with P4C, http://www.jamesnottingham.co.uk/media/2344/closing-the-achievement-gap-with-P4C.pdf

DEMOS (2015) *61% of ethnic minority kids in England – and 90% in London – begin Year 1 in schools where ethnic minorities are the majority of the student body*, http://www.demos.co.uk/press-release/61-of-ethnic-minority-kids-in-england-and-90-in-london-begin-year-1-in-schools-where-ethnic-minorities-are-the-majority-of-the-student-body/

Department for Children, School and Families (DCSF) (2008) *Learning together to be safe: A toolkit to help schools contribute to the prevention of violent extremism*, http://dera.ioe.ac.uk/8396/1/DCSF-Learning%20Together_bkmk.pdf

Department for Education (2016) Behaviour and discipline in schools Advice for headteachers and school staff, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/488034/Behaviour_and_Discipline_in_Schools_-_A_guide_for_headteachers_and_School_Staff.pdf

DeWitt, D., Offord, D., Sanford, M., Rye, B., Shain, M. and Wright, R. (2000) The effect of school culture on adolescent behavioural problems: selfesteem, attachment to learning, and peer approval of deviance as mediating mechanisms. *Canadian Journal of School Psychology*, 16(1), 15-38

Duckworth, A., Shulman, E., Mastronarde, A., Patrick, S., Zhang, J. and Druckman, J. (2015) Will not want: Self-control rather than motivation explains the female advantage in report card grades, *Learning and individual differences*, 39, 13-23

Dyfed County Council (1994) *Improving Reading Standards in Primary Schools Project*, Dyfed County Council, Wales

Fair, F., Haas, L., Gardosik, C., Johnson, D., Price, D. and Leipnik, O. (2015) Socrates in the schools: Gains at three-year follow-up, *Journal of Philosophy in Schools*, 2(2), 5-16

Farmer, E., and Lutman, E. (2009). Case management and outcomes for neglected children returned to their parents: A five year follow-up study. *Report to the Department for Children, Schools and Families: University of Bristol*.

Fleming, J., Mitchell, W. and McNally, M. (2014) Can markets make citizens?, Journal of School Choice, 8, 2, 213-236

García-Moriyón, F., Rebollo, I. and Colom, R. (2005) Evaluating Philosophy for Children.*Thinking: The Journal of Philosophy for Children*,17(4), 14-22

Goodman, A. and Gregg, P. (2010) *Poorer children's educational attainment: how important are attitudes and behaviour?*, York: Joseph Rowntree Foundation

Goodman, A., Joshi, H., Nasim, B., and Tyler, C. (2015). Social and emotional skills in childhood and their long-term effects on adult life. *Early Intervention Foundation*. Cabinet Office. (Avilable at: http://www.eif.org.uk/wp-content/uploads/2015/03/EIF-Strand-1-Report-FINAL1.pdf) (Last seen on 7th July 2016)

Gorard, S. (2010) Measuring is more than assigning numbers. Sage *Handbook of Measurement*, Los Angeles: SAGE, 389-408

Gorard, S. (2011) The potential determinants of young people's sense of justice: an international study, *British Journal of Sociology of Education*, 32(1), 35-52

Gorard, S. (2014) A proposal for judging the trustworthiness of research findings. *Radical Statistics* 110: 47-59

Gorard, S. (2015) Helping August's child. *New Scientist*, 228(3043), 28-29

Gorard, S. and Gorard, J. (2016) What to do instead of significance testing? Calculating the 'number of counterfactual cases needed to disturb a finding', *International Journal of Social Research Methodology*, 19, 4, 481-489

Gorard, S. and See, BH (2011) How can we enhance enjoyment of secondary school? The student view. *British Educational Research Journal*, 37(4), 671-690

Gorard, S. and See, BH and Morris, R. (2016) *The most effective approaches to teaching in primary schools*, Saarbrucken: Lambert Academic Publishing

Gorard, S. and Smith, E. (2010) *Equity in Education: an international comparison of pupil perspectives*. London: Palgrave Macmillan

Gorard, S., See, BH and Davies, P. (2011) *Do attitudes and aspirations matter in education?: A review of the research evidence*, Saarbrucken: Lambert Academic Publishing

Gorard, S., Siddiqui, N. and See, BH (2017a) Can 'Philosophy for Children' improve primary school attainment, *Journal of Philosophy of Education*, http://onlinelibrary.wiley.com/doi/10.1111/1467-9752.12227/abstract

Gorard, S., See, BH and Siddiqui, N. (2017b) What works and what fails? Evidence from seven popular literacy catch-up schemes for the transition to secondary school in England, *Research Papers in Education*, http://dx.doi.org/10.1080/02671522.2016.1225811

Gorard, S., See, BH and Siddiqui, N. (2017c) *The trials of evidence-based education, London*: Routledge

Gottfried, M. (2014) Does classmate ability influence students' social skills?, School Effectiveness and School Improvement, doi: 10.1080/09243453.2014.988731

Gupta, N., and Simonsen, M. (2010) Non-cognitive child outcomes and universal high quality child care, *Journal of Public Economics*, 94(1), 30-43

Gutman, L. and Schoon, I. (2013) The impact of non-cognitive skills on outcomes for young people. Education Endowment Foundation. (Available at: h ttp://educationendowmentfoundation. org. uk/uploads/pdf/Non-cognitive_skills_ literature_review. pdf.) (Last seen on 4th September 2015)

Gutman, L., and Vorhaus, J. (2012) The impact of pupil behaviour and wellbeing on educational outcomes. Department for Education. DFE- RR253. (Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/219638/DFE-RR253.pdf) (Last seen on 14th December 2015)

Hanley, P., Slavin, R. and Elliott, L. (2015) *Thinking, doing and talking science. Evaluation report and summary*. London: Education Endowment Foundation.

Harlen, W. (2006) The role of teachers in the assessment of learning. *Pamphlet produced by the Assessment Systems for the Future projects (ASF) Assessment Reform group, UK*, 347-364

Hart, S. (2001) *Children's rights in education* (Vol. 11), Jessica Kingsley Publishers

Hartshorne, M. (2007). *The Cost to the Nation of Children's Poor Communication: Scotland Edition*. I CAN(2). (Available at: http://www.ican.org.uk/~/media/Ican2/Whats%20the%20Issue/Evidence/2%20The%20Cost%20to%20the%20Nation%20of%20Children%20s%20Poor%20Communication%20pdf.ashx) (Last seen on 9th September 2016).

Hayes, D. (2014) Can kids do Kant? The Conversation. (Available at https://theconversation.com/can-kids-do-kant-22623) (Last seen on 12th August 2016).

Hayes, D. (2015) Philosophy for Children isn't real philosophy. Available at: http://www.spiked-online.com/newsite/article/philosophy-for-children-isnt-real-philosophy/17193#.V63HQ1srJpg (Last seen on 12th August 2016)

Heckman, J., and Rubinstein, Y. (2001) The importance of non-cognitive skills: Lessons from the GED testing program. *American Economic Review*, 145-149

Hedayati, M. and Ghaedi, Y. (2009) Effects of the Philosophy for Children Program through the Community of Inquiry method on the improvement of interpersonal relationship skills in primary school students, *Childhood and Philosophy*, 5, 9, 199-217

Higgins, S. (2015). A recent history of teaching thinking. *The Routledge international handbook of research on teaching thinking, Routledge, New York, NY*, 19-28.

Hill, M., Stafford, A., Seaman, P., Ross, N. and Daniel, B. (2007) Parenting and resilience. *York: Joseph Rowntree Foundation*

Humphrey, N., Lendrum, A. and Wigelsworth, M. (2010) Social and emotional aspects of learning (SEAL) programme in secondary schools: National Evaluation, London: DfE

Hyland, T. (2015) The limits of mindfulness, *British Journal of Educational Studies*, 10.1080/00071005.2015-1051946

Kirkman, E., Sanders, M., Emanuel, N. and Larkin, C. (2016) *Does participating in social action boost the skills young people need to succeed in adult life? Evaluating Youth Social Action: Final report*, London: Behavioural Insights Team

Kitchener, R. F. (1990). Do children think philosophically? *Metaphilosophy*, 21(4), 416-431.

Knox, E. and Conti-Ramsden, G. (2003), Bullying risks of 11-year-old children with specific language impairment (SLI): does school placement matter? *International Journal of Language and Communication Disorders*, 38(1), 1-12.

Lipman, M. (1980) Philosophy for Children by Mathew Lipman, (https://youtu.be/fp5lB3YVnlE)

Lipman, M. (1990) Socrates for six years old, (https://www.youtube.com/watch?v=gP1NycQaEyI)

Machin, S., Telhaj, S., and Wilson, J. (2006). The mobility of English school children. *Fiscal Studies*,27(3), 253-280.
Meir, S. and McCann, J. (2016) An evaluation of P4C, in Anderson, B. (ed) *Philosophy for Children: Theories and praxis in teacher education*, Routledge
Mercer, N., Wegerif R. and Dawes, L. (1999) Children's talk and the development of reasoning in the classroom, *British Educational Research Journal*, 25, 1, 95-111
Morgan, R. (2008) *Children on Bullying: A report by the Children's Rights Director of England*. London: Ofsted
O'Brien, S. (2014) 'Graceful failure': the privatization of resilience, *International Studies in Sociology of Education*, 24, 3, 260-273
Office for National Statistics (2013). *Review of available sources and measures for children and young people's well-being.* (Available at: http://sourcesmeasureschildrenswellbeing_tcm77-294952.pdf) (Last seen on 23rd July 2015).
Office for National Statistics. (2014). *Measuring National Well-being - Exploring the Well-being of Children in the UK*. (Available at: http://webarchive.nationalarchives.gov.uk/20160105160709/http://www.ons.gov.uk/ons/dcp171776_379712.pdf) (Last seen on 23rd July 2015).
Olsson, C. , McGee, R., Nada-Raja, S. and Williams, S. (2013). A 32-year longitudinal study of child and adolescent pathways to well-being in adulthood, *Journal of Happiness Studies, 14*(3), 1069-1083
Patalay, P., Fink, E., Fonagy, P. and Deighton, J. (2016) Unpacking the associations between heterogeneous externalising symptom development and academic attainment in middle childhood. *European child and adolescent psychiatry*, 25(5), 493-500
Paterson, C., Tyler, C. and Lexmond, J. (2014) *Character and resilience manifesto. All party parliamentary group on Social Mobility*, (Available at: http://www. educationengland. org. uk/documents/pdfs/2014-appg-social-mobility. pdf.) (Last seen on 18th January 2016)
Paulsel, M. and Chory-Assad, R. (2005) Perceptions of instructor interactional justice as a predictor of student resistance, *Communication Research Reports*, 22(4), 283–291
Pretsch, J., Ehrhardt, N., Engl, L., Risch, B., Roth, J., Schumacher, S. and Schmitt, M. (2016) Injustice in school and students' emotions, well-being, and behavior: A longitudinal study, *Social Justice Research*, *29*(1), 119-138
Protheroe, N. (2007) Emotional Support and Student Learning, *Principal*, 86(4), 50-54
Quintelier, E. and Hooghe, M. (2013) The relationship between political participation intentions of adolescents and a participatory democratic climate at school in 25 countries, *Oxford Review of Education*, 10.1080/03054985.2013.830097
Rahilly, T. and Hendry, E. (2014) Promoting the wellbeing of children in care. London: NSPCC
Reznitskaya, A. (2012) Dialogic teaching: Rethinking language use during literature discussions. *The Reading Teacher*, 65(7), 446-456
Ris, E.W. (2015) Grit: A Short History of a Useful Concept, *Journal of Educational Controversy*: (10) 3, 1-11
Roksa, J., Trolian, T., Pascarella, E., Kilgo, C., Blaich, C. and Wise, K. (2016) Racial Inequality in Critical Thinking Skills: The Role of Academic and Diversity Experiences, *Research in Higher Education*, 1-22
Sasseville, M. (1994) Self-esteem, logical skills and philosophy for children. *Thinking*, 4 (2) (1994), pp. 30–32
Schmidt, W., Burroughs, N., Zoido, P. and Houang, R. (2015) The role of schooling in perpetuating educational inequality: an international perspective, *Educational Researcher*, 44, 7, 371-386
Schoon, I. (2006) *Risk and resilience: Adaptations in changing times*, Cambridge: Cambridge University Press
Siddiqui, N. and Ventista, O. (2017) A Review of School-Based Interventions for the Improvement of Non-Cognitive Skills (Under review)
Slade, C. (1992). Creative and Critical Thinking: An Evaluation of Philosophy for Children. *Analytic Teaching*, *13*(1), 25-36.
Statham, J. and Chase, E. (2010) Childhood Wellbeing: A brief overview, *Loughborough: Childhood Wellbeing Research Centre*,

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/183197/Child-Wellbeing-Brief.pdf

Swain, J., Cara, O. and Litster, J. (2013) Doing philosophy in schools: An evaluation report prepared by the Institute of Education, University of London. (Available at: https://www.philosophy-foundation.org/asset/download/416.) (Last seen 20th August 2016)

Tenenbaum, H. and Ruck, M. (2007) Are teachers' expectations different for racial minority than for European American students? A meta-analysis. *Journal of educational psychology*, 99(2), 253

The Good Childhood Report (2015) *The subjective well-being of children in the UK*, The Children's Society, London. (Available at: https://www.childrenssociety.org.uk/sites/default/files/TheGoodChildhoodReport2015.pdf) (Last seen on 22nd July 2016)

The Guardian (2016) OFSTED head praises England's schools for immigrant integration, *The Guardian*, 2nd November 2016, https://www.theguardian.com/education/2016/nov/02/ofsted-head-michael-wilshaw-praises-england-schools-for-immigrant-integration

Tian, S. and Liao, P. (2016) Philosophy for children with learners of English as a foreign language. *Journal of Philosophy in Schools*, *3*(1)

Tomlinson, M., Walker, R. and Williams, G. (2008) *The relationship between poverty and childhood well-Being in Great Britain*, Barnet Papers in Social Research, Oxford, Department of Social Policy and Social Work

Trickey, S. and Topping K. (2004) 'Philosophy for Children': A Systematic review. *Research Papers in Education.* 19(3), 363-378

Trickey, S., and Topping, K. (2007) Collaborative Philosophical Enquiry for School Children, *Thinking: The Journal of Philosophy for Children*, 18(3), 25-36

Vansieleghem, N. (2005) Philosophy for Children as the Wind of Thinking, *Journal of Philosophy of Education*, 39, 1, 19-35

Williams, S. (1993) *Evaluating the Effects of Philosophical Enquiry in a Secondary School.* Derbyshire, England: Derbyshire County Council

Williams, S. and Wegerif, R. (2004) *Radical Encouragement: Creating Cultures for Learning*, Imaginative Minds: Birmingham

Winsper, C., Lereya, T., Zanarini, M., and Wolke, D. (2012) Involvement in bullying and suicide-related behavior at 11 years: a prospective birth cohort study, *Journal of the American Academy of Child and Adolescent Psychiatry*, 51, 3, 271-282

Worth, J., Sizmur, J., Ager, R. and Styles, B. (2015) *Improving numeracy and literacy, Evaluation report and summary*, London: Education Endowment Foundation

Youth Social Action in the UK (2014) *A face-to-face survey of 10–20 year olds in the UK*, IPSOS MORI, https://www.ipsos-mori.com/ researchpublications/publications/1711/Youth-social-actionin-the-UK.aspx

Appendix A - Detailed impact evaluation results

These are the detailed results for all 11 scaled attitude items in the survey, based on all pupil respondents. Each table shows the results for one item, including the pre-intervention means gathered six months apart for the two groups, the post-intervention means gathered at the same stage for both groups, a gain score of the difference between the pre- and post- means adjusted by six months for the P4C group, and an 'effect' size based on the post-intervention means. The information in the tables can be used to compute a pre-intervention 'effect' size or one based on the gain scores.

Table A1

I am good at explaining my ideas to other people	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	6.42	2.81	6.25	2.58	-0.16	3.19	-
Comparison	6.03	2.64	6.00	2.29	-0.03	2.92	-
Total	6.19	2.72	6.10	2.41	-0.08	3.03	0.10

Note: the gain score for the P4C group is adjusted by half of the gain score for the untreated comparison group

Table A2

I like meeting new people	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	8.27	2.62	7.79	2.62	-0.34	3.21	-
Comparison	7.97	2.67	7.67	2.55	-0.30	3.04	-
Total	8.09	2.65	7.71	2.58	-0.31	3.11	0.05

Table A3

I can work with someone who has different opinions to me	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.26	3.03	7.16	2.77	+0.09	3.64	-
Comparison	6.51	3.12	6.75	2.76	+0.23	3.68	-
Total	6.81	3.11	6.91	2.77	+0.17	3.67	0.15

Table A4

I can do most things if I try	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	8.15	2.41	8.13	3.90	+0.07	4.28	-
Comparison	8.16	2.20	8.00	2.14	-0.17	2.66	-
Total	8.16	2.29	8.05	2.96	-0.07	3.40	0.04

Table A5

Once I have started a task I like to finish it	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.84	2.78	7.50	2.81	-0.25	3.42	-
Comparison	7.73	2.70	7.55	2.59	-0.18	3.36	-
Total	7.78	2.73	7.53	2.68	-0.21	3.38	-0.02

Table A6

I want to try and make my local area a better place	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.88	2.93	7.73	2.78	+0.08	3.58	-
Comparison	7.99	2.83	7.52	2.75	-0.48	3.54	-
Total	7.95	2.87	7.60	2.76	-0.26	3.56	0.08

Table A7

I like to be told exactly what to do	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	6.11	3.75	5.54	3.53	-0.46	4.22	-
Comparison	5.64	3.61	5.39	3.29	-0.26	4.21	-
Total	5.83	3.67	5.45	3.39	-0.34	4.21	+0.04

Table A8

I am often afraid to try to new things	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	4.45	3.89	3.83	3.54	-0.48	4.83	-
Comparison	4.17	3.66	3.89	3.34	-0.28	4.48	-
Total	4.28	3.75	3.86	3.42	-0.36	4.63	-0.02

Table A9

I feel happy most days	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.45	3.08	7.22	2.94	-0.18	3.81	-
Comparison	7.56	2.73	7.46	2.59	-0.09	3.23	-
Total	7.51	2.87	7.37	2.74	-0.13	3.47	-0.09

Table A10

I try to understand other people's problems	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.59	2.93	7.59	2.63	-0.02	3.48	-
Comparison	7.51	2.77	7.56	2.40	+0.05	3.15	-
Total	7.54	2.84	7.57	2.49	+0.02	3.28	0.01

Table A11

I know where to go for help with a problem	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	8.51	2.64	8.36	2.53	-0.04	3.34	-
Comparison	8.54	2.41	8.40	2.41	-0.15	3.01	-
Total	8.53	2.50	8.39	2.45	-0.11	3.14	-0.02

What follows are the detailed results for all 11 scaled attitude items in the survey, based only on pupils known to be eligible for free school meals. As above, they portray the pre-, post- and adjusted gain scores, with an estimated post-intervention 'effect' size.

Table A12

I am good at explaining my ideas to other people	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	6.37	2.84	6.22	2.49	-0.13	3.07	
Comparison	5.92	2.69	5.65	2.43	-0.27	3.06	
Total	6.13	2.77	5.92	2.47	-0.20	3.07	0.23

Table A13

I like meeting new people	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	8.43	2.61	7.82	2.61	-0.48	3.20	-
Comparison	7.72	2.91	7.66	2.63	-0.06	3.18	-
Total	8.06	2.79	7.73	2.62	-0.26	3.19	0.06

Table A14

I can work with someone who has different opinions to me	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.18	3.19	6.92	2.88	-0.08	4.07	-
Comparison	6.25	3.48	6.61	2.75	+0.36	3.88	-
Total	6.69	3.38	6.75	2.81	+0.16	3.97	0.11

Table A15

I can do most things if I try	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	8.05	2.55	8.25	6.05	+0.30	6.18	-
Comparison	8.02	2.30	7.81	2.36	-0.23	2.88	-
Total	8.03	2.42	8.01	4.48	+0.02	4.73	0.10

Table A16

Once I have started a task I like to finish it	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.91	2.81	7.43	2.98	-0.42	3.44	-
Comparison	7.92	2.66	7.38	2.63	-0.55	3.35	-
Total	7.92	2.73	7.40	2.80	-0.49	3.39	0.02

Table A17

I want to try and make my local area a better place	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.97	2.87	7.67	2.97	-0.06	3.74	-
Comparison	7.76	3.11	7.77	2.74	-0.01	3.67	-
Total	7.86	3.00	7.72	2.85	-0.03	3.70	-0.04

Table A18

I like to be told exactly what to do	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	6.24	3.81	5.30	3.60	-0.84	4.07	-
Comparison	5.58	3.83	5.39	3.29	-0.19	4.30	-
Total	5.89	3.83	5.35	3.44	-0.50	4.20	-0.03

Table A19

I am often afraid to try to new things	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	4.98	4.04	4.11	3.68	-0.69	5.17	
Comparison	4.57	3.80	4.08	3.47	-0.50	4.64	
Total	4.76	3.91	4.10	3.57	-0.59	4.89	0.01

Table A20

I feel happy most days	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.60	2.96	7.25	3.03	-0.31	3.83	
Comparison	7.24	2.91	7.29	2.72	+0.03	3.52	
Total	7.41	2.94	7.27	2.86	-0.13	3.67	-0.01

Table A21

I try to understand other people's problems	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	7.63	3.07	7.60	2.73	-0.06	3.69	
Comparison	7.19	3.10	7.40	2.40	+0.19	3.64	
Total	7.40	3.10	7.349	2.60	+0.07	3.67	0.08

Table A22

I know where to go for help with a problem	Pre-intervention mean	Standard deviation	Post-intervention mean	Standard deviation	Adjusted gain score	Standard deviation	'Effect' size
P4C	8.68	2.51	8.61	2.45	+0.01	3.24	
Comparison	8.58	2.47	8.42	2.46	-0.16	3.00	
Total	8.62	2.49	8.51	2.45	-0.08	3.11	0.08

Tables A23 and A24 show the regression models for all 11 attitude items. Only the social responsibility model has a noticeable difference for P4C pupils.

Table A23 – R from multi-stage regression models predicting total responses to all attitude items, all pupils

	Communication	Social skills	Teamwork	Confidence
Background predictors	0.057	0.083	0.070	0.053
Whether in P4C intervention	0.058	0.087	0.070	0.054

	Determination	Responsibility	Be told	Afraid
Background predictors	0.076	0.110	0.053	0.079
Whether in P4C intervention	0.080	0.135	0.053	0.081

	Happy	Empathy	Knowledge
Background predictors	0.066	0.032	0.049
Whether in P4C intervention	0.067	0.032	0.051

Table A24 – Standardised coefficients from multi-stage regression models predicting total responses to all attitude items, all pupils

	Communication	Social skills	Teamwork	Confidence
Gender	+0.01	+0.05	+0.02	-0.02
Ethnicity	-0.03	-0.06	-0.07	+0.04
FSM-eligibility	-0.01	+0.01	0	+0.01
Special educational needs	-0.03	+0.02	+0.04	-0.01
English as an additional language	-0.18	+0.08	+0.04	-0.03
Whether in intervention group	-0.01	+0.03	0	+0.02

	Determination	Responsibility	Be told	Afraid
Gender	-0.01	-0.07	+0.03	-0.04
Ethnicity	-0.06	+0.03	0	-0.02
FSM-eligibility	-0.04	+0.02	-0.02	-0.02
Special educational needs	-0.04	-0.04	-0.04	-0.02
English as an additional language	+0.03	+0.03	-0.01	-0.05
Whether in intervention group	+0.03	+0.10	0	-0.02

	Happy	Empathy	Knowledge
Gender	-0.02	-0.01	0
Ethnicity	-0.07	-0.02	-0.03
FSM-eligibility	+0.01	+0.02	+0.02
Special educational needs	+0.01	+0.01	+0.01
English as an additional language	+0.02	+0.02	-0.02
Whether in intervention group	_0.01	0	+0.02

Appendix B – The survey items

The instrument could not be published here because it is still being used in other studies that are ongoing. Please contact nadia.siddiqui@durham.ac.uk for further information on this.

Appendix C: Gorard's 'sieve' to assist in the estimation of trustworthiness of descriptive work

The shaded areas suggest our judgements on the factors affecting the security of this new study, leading to an overall rating of 2*.

Design	Scale	Dropout	Data quality	Threats	Rating
Strong design for research question (RQ)	Large number of cases (per comparison group)	Minimal attrition, no evidence of impact on findings	Standardised, pre-specified, independent	No evidence of diffusion, demand, or other threat	4★
Good design for RQ	Medium number of cases (per comparison group)	Some attrition (or initial imbalance)	Pre-specified, not standardised or not independent	Little evidence of diffusion, demand or other threat	3★
Weak design for RQ	Small number of cases (per comparison group)	Moderate attrition (or initial imbalance)	Not pre-specified but valid in context	Evidence of diffusion, demand or other threat	2★
Very weak design for RQ	Very small number of cases (per comparison group)	High attrition (or initial imbalance)	Issues of validity or appropriateness	Strong indication of diffusion, demand or other threat	1★
No consideration of design	A trivial scale of study, or N unclear	Attrition huge or not reported	Poor reliability, too many outcomes, weak measures	No consideration of threats to validity	0