

Evaluation Summary	
Age range	Primary (Years 3 to 6)
Number of pupils	c. 3,000
Number of schools	50
Design	Whole school randomised controlled trial
Primary Outcome	Literacy and Maths

Evaluation of 'SAPERE: philosophy for children'

Introduction

'Philosophy for Children' (P4C) aims to improve children's abilities and dispositions to question, reason, construct arguments and collaborate with others. Through the training and development of teachers, the initiative is intended to foster cognitive improvement and greater self-confidence in the children they teach, leading to higher levels of recorded attainment.

In this study, P4C will be delivered by The Society for the Advancement of Philosophical Enquiry and Reflection in Education (SAPERE). SAPERE have a record of developing teachers for this purpose. P4C has been previously tested on a small-scale (e.g. N=177, Topping and Trickey 2007), or with incomplete randomisation, and without follow through to standard attainment results. There are also some unsystematic observations of beneficial impact from OFSTED reports. There is equipoise at present and so a trial is appropriate.

Sampling and recruitment

The evaluation will be conducted with a total of 50 primary schools, involving years 3 to 6 over two academic years – 2012/13 and 2013/14. The schools will be based in five areas of England (to be determined) representing a range of geography, economy, local political control, population density, and levels of disadvantage. There will be approximately 10 schools in each area. All schools must be at least single form entry. All schools will have, or recently had, at least 25% of their pupils known to be eligible for free school meals. No school will have previously formally implemented P4C. At least 10 of the schools will have fewer than 60% of pupils achieving level 4+ in English and Maths, and with pupils making below-average progress in English and Maths, in 2012 (or 2011).

The initial year 3 pupils can be assessed for long-term progress and their eventual KS2 results and subsequent attainment. Their involvement will be part of the process evaluation. However, the evaluation will end before these pupils reach the end of KS2. Year 3 will therefore not be assessed using Cognitive Abilities Tests (CATs), as part of the evaluation.

Due to the difficulty of following the initial year 6 pupils, the evaluation will not be able to assess their progress after at least one full year of progress or experience of the intervention. Their KS2 results will be the first available for comparison between the two

groups, but this is not deemed a fair test of the impact of the intervention. Year 6 will therefore also not be assessed using CATs, as part of the evaluation. Their involvement will be part of the process evaluation.

The impact evaluation, and the resources for administering CATs, will focus on years 4 and 5. Assuming an average of 30 pupils per year group, this means that the first phase of 20 schools will have 1,200 pupils and the second phase of 30 schools will have 1,800 pupils. The research proposal cites an effect size of 0.4 (Cohen's d) on the basis of a review of a range of attainment scores. If the intervention did indeed have an average effect size of 0.4, and assuming an intra-cluster correlation of 0.2 for the outcomes scores, then Lehr's formula suggests a minimum sample size of 480 cases per arm (for 80% power to detect a difference with alpha of 5%). This suggests that the sample provides sufficient power to detect an effect in terms of CATs and in terms of KS2 outcomes.

Schools will be approached both formally and informally, using existing contacts and local networks. Each possible sample school will be provided with a pack describing the project, the data requirements, and information about the evaluation. Where possible, recruitment will include local events, at which project members and evaluation team members will be present to answer questions. It is essential for all participating schools to agree to be in either phase, and to be ready for January 2013 start.

The resultant sample may or may not be representative of England or the areas in which schools were recruited. However, the constraints on size, location, attainment, prior experience of P4C, and interest in P4C, all mean that the sample is not likely to be easily generalizable to a wider population, using sampling theory techniques. Once the sample has been agreed, it will become the population for the study, and the groups/phases will be selected randomly from that population. However, the sample will be substantial enough for any substantive findings to be worthy of note.

Allocation to groups

All schools must agree to the data requirements, including any testing in December 2012, before allocation to groups. Therefore all schools must be prepared to begin training in January 2013, and must be aware that their training may not start until September 2014. Allocation to groups will take place in early December 2012, but will not be revealed to schools until the day after the first testing. This maximises the time schools will have available to prepare for January training, but allows the pre-test to be conducted blind of the results of the allocation.

The first phase treatment group (starting P4C in January 2013) will consist of 20 schools. The second phase group (starting P4C in September 2014) will consist of 30 schools. All 50 schools will be listed in descending order of their average 2012 (or 2011) KS2 results. The evaluators will create syntax to generate 10 pairs of pseudo-random integers in the range 1 to 5. These will then be applied to the sorted list in groups of 5, starting with the school with the highest 2012 (or 2011) KS2 results. This procedure will create a list of 20 schools for the first phase, with the remaining 30 schools in the second phase. The two groups will be stratified to some extent in terms of prior attainment. Because attainment is often linked to geography and the relative disadvantage of school intakes, this procedure should also ensure a reasonable range of areas and school intakes in both groups. However, the precise nature of the stratification may be subject to change depending upon the number of regions and number of schools per region once the sample has been enlisted.

Design

The design is a stratified randomised controlled trial with two outcome measures, a treatment group of 20 schools, and a waiting list control of 30 schools. The second phase schools will wait approximately 18 months for training.

Treatment

P4C is an initiative that aspires to achieve high-quality classroom dialogue in response to children's own questions about shared stories, films and other stimuli. With guidance from the teacher, the dialogue focuses not only on the chosen questions but also on the assumptions that lie behind the answers and the criteria used to make judgements. The main aim of P4C is to achieve a community of enquiry involving all pupils, guided by the teacher. A community of enquiry is a group of people able to delve critically and rigorously into their own questions and understandings while at the same time feeling safe and being open to the views of others. Some of the routines of P4C, such as questioning, logical thinking, evaluating the grounds of arguments, and using precise language, might transfer across the curriculum. The goal of the intervention is to raise educational attainment in primary schools, particularly those with a preponderance of disadvantaged children.

The treatment is in two main stages. First, SAPERE will train and support teachers of years 3 to 6 in the first phase primary schools. Then the intervention itself will be delivered by trained teachers for at least one hour per week. The initial training will take place over 2 days or the equivalent time in shorter sessions, and involve teachers, assistants and senior management. Ongoing support will involve advanced training, and on-line help.

Prior measures

All of the prior background and contextual data will be generated automatically by schools. No further data collection is necessary.

Most of the prior background and contextual data will come from the individual pupil NPD records for pupils in both phases. This will include KS1 results (levels and points), sex, month of birth, FSM status, SEN status, ethnicity, and first language.

In addition, it would be useful to have individual attendance records, date of leaving (if during the project), and any suspensions or exclusions (where applicable). These would come from existing school records.

The CAT (below) will be used as a pre-test for years 4 and 5.

Outcome measures

There will be two key outcomes measures for the impact evaluation. The first is the pupil KS2 scores by subject and overall, used as the terminal measure of school attainment. The second is the pre and post difference in CAT scores (assuming cost permits), used as a measure of progress in pupil reasoning, both for its own sake and because it may lead to enhanced measures of school attainment at KS2 or later.

Therefore, the primary evidence used to determine the effectiveness or otherwise of this intervention will be:

The effect size of the difference between groups for initial years 4, 5 and 6 in their combined KS2 scores;

And

The effect size of the difference in gain scores between groups for initial years 4, 5 in their combined CAT scores.

All other measures and outcomes will be secondary. These include the results of analyses by year groups separately, for different subjects at KS2, and for components within the CAT (such as verbal reasoning).

The KS2 scores will be available in December 2013, for the original year 6, December 2014, for the original year 5, and December 2015 for the original year 4.

CATs will be administered by schools, under the supervision of the evaluating team, for all years 4 and 5, in December 2012, December 2013, and July 2014 (when the original year 5 are reaching the end of KS2). The CAT can be administered during a convenient lesson time during normal schooling. The first measurement will be before schools know which phase/group they have been allocated to.

Longer term, there will be KS2 scores for other cohorts and KS3 and KS4 scores for all cohorts. But these do not form part of the evaluation at this stage.

Analysis of outcome measures

Comparison between the available first and second phase pupil outcome measures will take place at the end of the first school year, after one calendar year, and again at the end of the second school year. The same cohorts can also be followed into further years, using future qualifications. However, this is beyond the scope of the current evaluation.

Every attempt must be made to get complete test scores for all pupils even where they are initially absent or where they leave the schools during the trial. Where dropout, turnover or exchange between schools occurs, the results will be analysed both in terms of the original phase for each pupil (intention to treat) and in terms of the eventual groups. Where scores for pupils are missing, the results will be subject to a sensitivity analysis estimating how different the missing data would have to be in order for any difference between the groups to disappear (Gorard 2013). Differences will be calculated for the post-test scores alone, and for the gain scores from pre-test to post-test. And differences will be analysed in terms of pupil prior attainment (at KS1) and background (from NPD).

Differences that are robust enough to appear under all of these conditions will be considered substantial. Differences will be presented as raw-score and in standardised form such as Cohen's d effect size. Where possible, differences will also be calculated for appropriate sub-sets of potentially disadvantaged pupils, such as those known to be eligible for FSM.

KS2 scores will be available for all pupils reaching the end of year 6, and CAT scores will be available for all pupils originally in years 4 and 5.

Process evaluation and fidelity to treatment

A substantial part of the evaluation fieldwork will be conducted with the aims of assessing how closely schools adhere to the intended intervention, and what the short term or intermediate impacts are (such as changes in classroom interaction). In co-operation with the teachers and trainers, it will address questions such as:

1. Is P4C being done regularly?
2. Are children sharing their ideas more with each other in a critical but friendly way.
3. Are questioning and reasoning being prompted and demonstrated in lessons?
4. Are instances of questioning and reasoning increasing?
5. Is there less dominance by the teacher in discussions?
6. Are children taking more responsibility for the questioning and reasoning?
7. Are teachers and children talking about significant concepts?
8. Are teachers' perceptions of children changing?
9. Are teachers' perceptions of their own work changing?
10. Are children's perceptions of themselves and school changing?

The process evaluation will provide some formative evidence on all phases and aspects of the intervention from the selection and retention of schools, through the training of teachers to evaluating the outcomes. It will involve the perceptions of participants including any resentment or resistance, and lead to advice on improvements and issues for subsequent scaling up.

The evaluators will make about 25 person trips to the research sites per year. This will necessitate the generation of some additional data from observation and interviews with staff, pupils, and parents as well - also observation of training, delivery and testing. These will all be as simple and integrated as possible. The areas of interest include but are not limited to:

- the contents and use of training materials
- the reaction to training
- the fidelity of subsequent implementation
- assessments
- how missing data is handled
- changes in teacher behaviour
- how pupils with additional learning needs react
- and whether there appears to be an impact on how children are thinking and constructing arguments in other areas of schooling.

The evaluators will produce a brief report each year.

Evaluation timeline (steps with direct involvement of evaluators in *italics*)

2012

September- Project starts
Initial discussions with project team, EEF, and evaluators
School recruitment continues
Information-giving meetings for clusters of schools
Observation of school recruitment process
Contracts prepared for project schools and trainers
Design and production of special project materials

November- *Resolve any outstanding decisions on design, measures and sample size*
Trainers selected for each cluster
Pre-test (CATs) takes place in all schools
Allocation of schools to first or second phase
Dates agreed for training days and support

2013

January- Training begins in first phase schools
Observation of training for intervention
Observation of intervention taking place in schools
December- *Post-test (CATs) takes place in all schools*
Analyse progress in CATs
Analyse KS2 results for initial year 6
Write end of year brief report
First phase schools apply for SAPERE Bronze Award

2014

January- Selected staff from first phase schools trained to Level 2
Training and support for first phase schools continues
First phase schools apply for SAPERE Silver Award when ready
Observation of in-fill teacher training for intervention
Observation of intervention taking place in schools
July- *Delayed post-test (CATs) takes place in all schools*
Analyse progress in CATs
September - Training begins for waiting list schools)
December – *Analyse KS2 results for initial year 5*
Write end of year brief report

2015

January- Training continues for all schools
Second phase schools apply for SAPERE Bronze/Silver Awards when ready
Analyse KS2 results for initial year 4 (if available in time)
Complete full EEF report
Advise on the future of the intervention and its evaluation

Ethical issues

This evaluation of the intervention, as distinct from the intervention itself, raises few additional ethical issues. The CAT is a standardised test in frequent use in schools, and the KS2 tests happen as a matter of course. The remainder of the outcome and contextual data is generated by schools as a matter of course. SAPERE will arrange for agreement from the schools, and from parents where necessary, to taking part in the evaluation as part of the access agreement.

The evaluators will need to match data anonymously at an individual pupil level, and this requires use of the unique pupil id, as it appears in the National Pupil Database (NDP). The link between the pupil id and any identifying information will be held separately from all other data, and will be destroyed immediately after use. The evaluators do not require to know, and do not wish to know, who each pupil is.

All participants in interviews and observations will be informed that participation is voluntary, and that they can withdraw at any stage. The evaluation will receive ethical approval from the University of Birmingham's Ethical Review Board, and will be conducted in accordance with BERA's professional Code of Practice. No individual or school will be identifiable in any report or publication arising from this evaluation.

References

Gorard, S. (2013) *Research design: robust approaches for the social sciences*, London: Sage

Topping, K. and Trickey, S (2007) Collaborative philosophical enquiry for school children: Cognitive effects at 10-12 Years, *British Journal of Educational Psychology*, 77, 2, 271-288