Amendments

The following amendments were made to the protocol after the development phase had finished:

- The mode of delivery for the intervention was finalised at the end the development phase to be a whole-class delivery rather than pupils accessing the intervention individually on separate computers. This change was made throughout in the document.
- There were two primary outcomes options in the original protocol. After the development phase was complete, it was decided that both the outcomes will be retained as the primary outcome measures for the trial as suggested in the original protocol.
- Page 4- revised timeline for randomisation to allow more time for recruitment, pupil data collection and aid intervention delivery.
- Page 5- The original protocol mentioned that NFER would find out from the pilot schools how classes might be split to receive different assessments. This did not take place and therefore is removed.
- Page 5- names of the intervention and control plus group updated
- Page 6- in order to aid the training delivery, a decision was made to run the randomisation in two waves
- Page 7- NFER will administer the GL Assessment tests in the schools to ensure that the tests are administered blind to group allocation. This will also help retaining schools in the trial.
- Page 12 – trial registration number added.
- Page 14 (timetable)- revised timeline for randomisation and the intervention delivery

Protocol for the evaluation of counterintuitive concepts intervention

Note: This protocol excludes aspects of the evaluation that are the sole responsibility of Birkbeck College and are not requirements of the EEF or NFER. Following the development phase we anticipate that several aspects of the protocol will need to be redeveloped in an amended protocol.

Evaluation Summary	
Age range	7 to 10
Number of pupils	8100
Number of schools	100
Design	Year group level randomisation within school. Each school will have an intervention class/es and at least one of the two possible control groups
Primary Outcome	Mathematics as measured by Progress Test in Maths and Science as measured by Progress Test in Science.
Date of protocol	04.01.2018
Version	2

Introduction

The Education Endowment Foundation (EEF) and the Wellcome Trust have commissioned Birkbeck College (principal investigator Professor Denis Mareschal) to develop and deliver the counterintuitive concepts learning intervention in collaboration with UCL Institute of Education. The project runs from January 2016 to December 2018 and is divided into two phases: (1) 18-month development and pilot phase and (2) a randomised controlled trial phase (main trial). The sample for the main trial will be year 3 and year 5 pupils from 100 Primary schools, where year groups will be randomised to either receive the intervention or to be part of one of the two control groups- control plus group or business as usual control group.

This protocol was written during the development phase and the design was updated prior to the main trial.

Intervention

When learning new concepts in science and maths, pupils must be able to inhibit prior contradictory knowledge and misconceptions to acquire new knowledge successfully. This skill of "interference control" varies between pupils, with variation evident from an early age. Disadvantaged pupils seem to have weaker control skills than their wealthier peers. The Centre for Educational Neuroscience, a collaboration between Birkbeck College, Institute of Education and University College London, developed a computer game to train a pupils' ability to control such interferences. The programme has the main aim of improving learner's ability to adapt to counterintuitive concepts via training them to inhibit their initial response and instead, give a more delayed and reflective answer to ultimately improve learners' educational outcomes. It seeks to achieve these aims with year 3 and year 5 pupils receiving three 15-minute sessions a week before a maths or science lesson, where they use a teacher-lead computer-based learning activity to practice counterintuitive learning. In the game, a child-friendly character will try to solve problems with help from the player, providing prompts and suggestions. Exercises will relate to specific maths and science content. For example, exercises will help pupils realise that mice and elephants have the same-sized cells, or that the world is round despite seeming flat.

Following the outcome of the development and pilot phase, the intervention will be delivered in a "whole class" mode. The teacher will help the pupils step through the sessions as a whole-class group, all observing and interacting as a class with the session projected on an interactive white board or class wall. The teacher is able to choose which ever method to select the class response (e.g., most popular choice or selecting an individual pupil). The teacher is instructed not to tell the pupils their answer is incorrect or to prompt them towards a correct answer, rather the teacher is instructed to allow children to make errors so they can benefit from the inbuilt support/hierarchy of prompts of the software itself.

It is anticipated that the impact of the study will continue to develop over time, perhaps as a pupils' learning is reinforced. The expectation is that this longer-term impact would be looked at separately to this evaluation by the EEF's overarching evaluator.

Significance

Educational Evidence

Studies of interventions designed to improve such "executive function" skills have shown improvements on skills like working memory, but have often failed to show an impact on related attainment measures. As a result, Birkbeck College proposed developing exercises that were more closely related to the attainment measures of interest.

Neuroscience Evidence

Evidence from neuroscience research supports the hypothesis that inhibition control is necessary to develop the reasoning skills required in maths and science. The intervention draws on work which suggests that when being trained in inhibition control, participants begin to engage more of the parts of the brain required for logical thinking. Emerging neuroscience research suggests that inhibition needs to happen in the networks which are specific to the skills being developed, thus the need for exercises to be related to specific subject knowledge.

When learning new concepts in science and maths, pupils must be able to inhibit prior contradictory knowledge to successfully acquire new knowledge. This skill of "interference control" varies between pupils, with variation evident from an early age.

Methods

Research questions

The primary research question is: does the use of the counterintuitive learning intervention impact on learners' mathematics and science achievement?

The secondary research question is: what is the impact of the counterintuitive learning intervention on learners' inhibition control? An additional secondary research question will explore if there is an impact of the social skills computer programme, used by the control plus group, on learners' mathematics and science achievement in comparison to the counterintuitive concepts programme. This will determine if any identifiable effect is due to using a computer programme rather than any specific content.

Trial design

The evaluation started with a development and pilot phase in January 2016 and finished in the end of July 2017. The design for the main trial was finalised in May 2017 so that recruitment for the main trial could commence from January 2017 with a view to randomisation taking place in the autumn 2017. Schools would start to implement the intervention in November 2017 (see Table 1 below).

Table 1 Overview

Academic year 2015/16	Academic year 2016/17	Academic year 2017/18
Development and pilot phase (January 2016-July 2017)		
	Main trial- School recruitment from January 2017 and randomisation in October 2017	Intervention introduced (November 2017) and testing of intervention takes place (Feb/March 2018)

Phase 1-Development phase/ Pilot

The research design was piloted in eight schools by Birkbeck College. The purpose of the pilot was to:

1. develop the intervention materials - Birkbeck College
2. determine if whole class or individual implementation will work best for the intervention- Birkbeck College
3. test the materials with the pilot schools - Birkbeck College
4. assess the suitability of three trial groups for implementation and feasibility- intervention, control and control plus groups[1] - Birkbeck College
5. carry out a small scale process evaluation involving case studies with three of the schools involving interviews with senior leaders and those taking part in the pilot to check the feasibility and scalability of the intervention, including gathering feedback on whether the three groups are practical - NFER
6. develop a theory of change model for the intervention to be tested in the main trial through process and impact evaluation - NFER and Birkbeck College.

NFER produced a verbal and written summary report of findings coming out of the pilot and recommendations of changes to the main trial. Findings from this report and the feedback from Birkbeck College's pilot study recommended that the schools' preference was to run the intervention as a whole-class programme.

Phase 2- Efficacy trial

For the main efficacy trial, a sample of approximately 100 primary schools will be approached that are nationally representative (including areas of particular interest of EEF) from schools with the highest percentage of pupils on free school meals (FSM). The 100 schools (50 being one form entry and 50 being greater than one form entry) recruited will then have year 3 and year 5 classes randomly assigned to one of three groups:

- Counterintuitive concepts in mathematics/ science lessons through a computer-based learning activity called '**Stop and Think**' (referred to subsequently as 'intervention')
- 'Business-as-usual' control (referred to subsequently as 'control')
- Social skills learning control in PSHE lessons through a computer-based learning activity called **See+**[2] (referred to subsequently as 'control plus' as this also assesses the impact of the computer programme) which captures the content of the age-appropriate PSHE and SEAL curricula[3] . During this activity, children observe and reflect upon social interactions and engage in social-emotional learning through a series of computerised animated stories with virtual characters engaging in social scenarios.

[1] Control and control plus are described in more detail within the design section

[2] See+ stands for social-emotional engagement

[3] See https://www.pshe-association.org.uk/ and http://webarchive.nationalarchives.gov.uk/20110812101121/http://nsonline.org.uk/node/87009 for PSHE and SEAL curriculum contents respectively.

A cluster design is planned, where year 3 and year 5 year groups will be randomly allocated to either the intervention or control/control plus with 1:1 ratio. This will be an unbalanced design with a ratio of 2:1:1 for the intervention versus the control and control plus. This will mean that no school is solely a control school which will help with recruitment and reduce attrition. In all schools, the year group that is not allocated to intervention will be randomly allocated to control or control plus.

The intervention delivery for the main trial will start in November 2017 and will run until February 2018 and pupils will be tested in February/March 2018.

The trial will be designed, conducted and reported to CONSORT standards (http://www.consort-statement.org/consort-statement/) and registered on http://www.controlled-trials.com/.

Randomisation

Randomisation on all levels will be carried out blind by a statistician at NFER. As discussed above, schools will have year 3 and year 5 randomised equally to treatment or control/control plus. Therefore every school will have at least one intervention class and one control/control plus class. Within the 50 one-form entry schools there will be 50 intervention classes, 25 control classes and 25 control plus classes. Within the 50 larger schools there will be 100 intervention classes, 50 control classes and 50 control plus classes. In the larger schools whole year groups will be randomly assigned to the same group. Therefore in a single school all of year 3 might be in the intervention group and all of year 5 will be in the control group. In another school all of year 3 will be in control plus and all of year 5 will be in the intervention group. The process will result in a ratio of 2:1:1.

Randomisation will be conducted using a full syntax audit trail and records kept by NFER of which year groups have been allocated to which intervention or control group. There will be two waves of randomisation to aid the training delivery. This means schools within each wave will start delivering the intervention at different time-points.

Participants

Several recruitment events took place to explain the evaluation to schools, so they understand the importance of the randomisation methodology before signing up to the trial. The school representatives were able to access all the information that was also available on the trial website[4].

Birkbeck College will be responsible for the recruitment of schools. To be included in the trial, schools will need to provide a list of pupils UPNs, names and date of births and class lists to Birkbeck College to avoid schools resisting the data requirements after allocation. The main trial will only be looking at year 3 and year 5 learners who are predominantly, but not exclusively, from schools with the above average proportion of students receiving free school meals.

Outcome Measures

We will measure maths and science outcomes in year 3 and 5 using the following tests:

[4] http://unlocke.org/schools.html

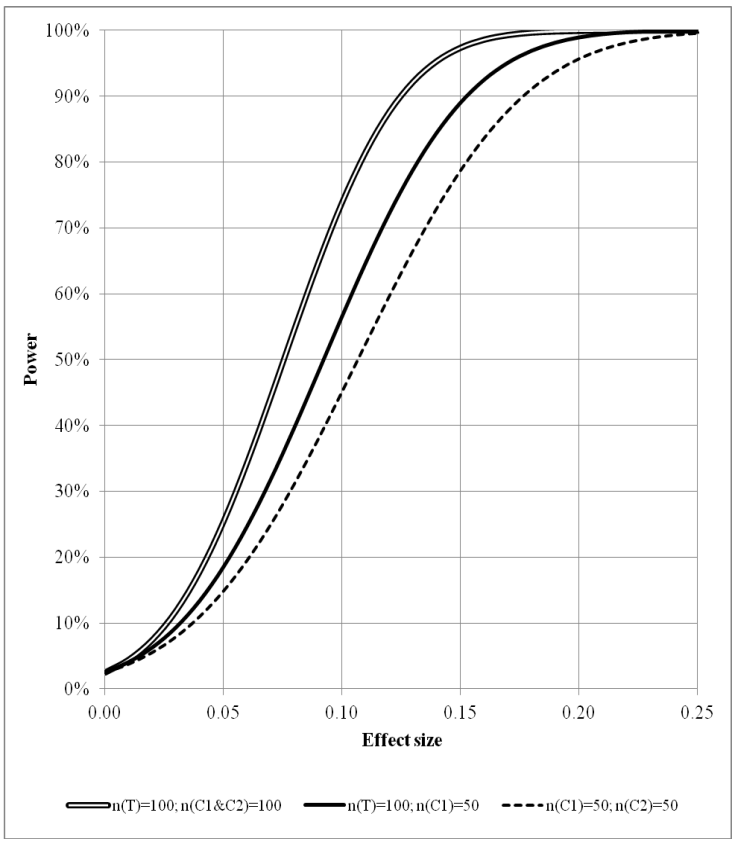
Progress test in Maths (PTM) and Progress tests in Science (PTS) produced by GL Assessment will be used to measure mathematics and science outcomes in both year 3 and year 5. PTM8 and PTS8 will be used for year 3 pupils and PTM10 and PTS10 will be used for year 5 pupils. The tests are designed to be administered to a class of pupils. The availability of a standardised test for each year group would allow a different test to be administered either side of a two term intervention. However, to reduce costs we will use Key Stage 1 assessment data as a pre- test measure (and therefore a covariate in any analysis). We will use a combined KS1 assessment measure that would include task/test assessments in reading, writing and maths as the correlation with science outcomes at KS2 is still strong. In addition, test outcomes can be easily related to performance at Key Stage 2.

There is considerable cost and burden associated with developing, printing, administering and marking the year 3 and 5 tests. For this reason, NFER will randomly allocate year 3 and 5 pupils to sit either the maths or science test in such a way that within each class, half the students will sit a maths test and half will sit a science test. The paper versions of these tests will be used. NFER will manage the test administration by sending test administrators to schools to reduce burden placed on schools. This will also help with the response rate. Schools will receive pupil level test results by directly accessing the GL Assessment's results portal. This will act as an incentive for taking part in the trial. The raw scores from the maths and science tests will be used as the primary outcome measures.

Birkbeck College will additionally administer an assessment of inhibition control using an adaptation of the chimeric animal-stroop task[5]. This will be administered by the Research Assistants (RAs) appointed by Birkbeck College to the project. The administration of the test will be undertaken after the schools have completed the GL Assessment tests at post-test (February/ March). The administration will be blind so that the RAs are unaware of which group they are administering the assessment to. Analysis of the data will be undertaken by NFER.

Sample size calculations

[5] Wright, Waterman, Prescott, and Murdoch-Eaton (2003). A new Stroop-like measure of inhibitory function development: typical developmental trends. *Journal of Child Psychology and Psychiatry, 44*, 561-575.



Power calculations use the following two assumptions which were obtained from EEF's paper on pre-test effects. The trial design is using KS1 as a covariate with Progress in Maths and Progress in Science being used to measure post test outcomes. The correlation between KS1 and these assessments in year 3 and year 5 is assumed to be 0.75. The intra-class correlation is assumed to be 0.126. These figures are used in the calculation of optimum sample sizes for desired levels of power. These assumptions allow for the following comparisons:

- n (intervention) =100 schools and 150 classes; n (control and control plus) =100 schools and 150 classes represents the comparison between intervention classes and both control and control plus classes grouped together and assumes an average cluster size of 27 (average cohort size for eligible primary schools class in England. Power calculations are based on half of theses pupils taking a maths test and half taking a science test. Calculations are based on an effect size for either of these tests).

- n (intervention) =100 schools and 150 classes; n (control) =50 schools and 75 classes represents the comparison between the intervention classes and the control plus group. This again assumes an average cluster size of 27 (average cohort size for eligible primary schools in England).

- n (control) =50 schools and 75 classes; n (control plus) =50 schools and 75 classes represents the comparison between the control and control plus groups. This assumes an average cluster size of 27 for the size of each class.

From the power curves, it can be seen the main trial is well powered with a minimum detectable effect size (MDES, at 80% power) of less than 0.2 for all three types of analyses. The intervention and control/control plus comparison has an MDES of around 0.11 with the other two comparisons, intervention to control plus and control plus to control, between 0.13 and 0.15. Given that all schools will contain treatment and control group pupils we are not anticipating high rates of attrition. The design allows for five schools to leave the evaluation without power being unduly affected. Assuming that there are 22.5%[6] pupils who are EVERFSM (pupils eligible for FSM at any time during the past six years), the MDES will be 0.17 at 80% power.

Analysis

A full statistical analysis plan (SAP) will be written and will be published in addition to this protocol before any outcome data is seen. This will be written within three months of randomisation.

The primary outcomes will be mathematics and science year 3 and 5 outcomes as measured using the following GL Assessments: PTM and PTS. All statistical analysis of the intervention impact will be conducted at pupil level, comparing average pupil maths and science scores in the intervention group with average scores in the control groups. The average difference will be measured in a multilevel linear regression model that takes account of pre-test score as a covariate, along with the stratification variable used at randomisation(whether the school is a one form entry or two form entry) . A three-level model will be created to account for pupils being clustered within classes and within schools. The main analysis will examine/assess differences between the treatment group and a combination of both control groups. Additional analysis will look at differences between the intervention and the control plus group and a minor analysis will look at differences between the control plus group and the business as usual control group.

The main analysis will be 'intention to treat', reflecting the reality of how interventions are delivered in practice and avoiding attrition bias. The main analysis will include prior attainment as measured by KS1 along with group identifiers. Sub-group analysis will include the fixed effects and interactions for gender and age. Analysis on a cohort of pupils eligible for free school meals (FSM)[7] will be conducted separately. Prior attainment, gender and age variables will again be included.

Secondary outcomes analysis will use the pupil scores on the chimeric animal stroop task and will be analysed in the same way as the primary outcome models.

It is likely that the number of interventions in each class will not be exactly as planned. Due to the method of the intervention (computers), it should be possible to discern the exact number of interventions carried out by the teacher or completed by the class. The main analysis will therefore be followed by a CACE analysis (Complier Average Causal Effect) in order to assess the effect of non-compliance on outcome measures where data from the computer system will be used to determine the extent of each class's involvement. The developer will provide data on whole class activities such as number of completed sessions, average amount of time per session, spacing of sessions, and average number of structured practice activities completed. This analysis will enable us to estimate

[6] Average proportion of EVERFSM pupils for state-funded Primary schools in school census 2015

[7] The everFSM variable from the National Pupil Database will be used to identify these pupils.

a ‘pure intervention effect’ (net of any fidelity issues) that would not necessarily be causal in nature. We will use the number of session completed to evaluate compliance. Classes completing 1 to 10 sessions will be classified as having “low” compliance, those completing 11-20 sessions will be classified as having “medium” compliance and those classes completing 21-30 sessions will be classified as having “high” compliance.

Implementation and process evaluation methods

The process evaluation will involve a number of methods split into the pilot and main trial phases as outlined below.

The process evaluation for the Phase 1 (the pilot) will collect information on:

- the models of delivering the intervention being used
- the feasibility of delivering the intervention
- the teacher training related to the intervention
- any other mathematics/science interventions taking place in the schools
- any other neuroscience based interventions taking place in the schools.

At Phase 1, the process evaluation will comprise case study visits to three pilot schools where interviews will be conducted with a year 3 teacher, a year 5 teacher and a senior leader (up to nine interviews).

The process evaluation for Phase 2 will aim to investigate the following research questions:

- Was the theory of change model identified in the pilot an accurate representation of the intervention and its outcomes?
- Have schools implemented the intervention in the way it was intended? If not, why not.
- Is the intervention appropriate for pupils of this age and in these lessons?
- Can programme materials and delivery be improved for the future?
- Is the roll-out of the intervention feasible for schools?

During the spring term, we will then collect evidence from schools through six school case studies which will examine:

- Teachers’ views on the adequacy of the intervention-specific training they have received.
- How effectively the delivery model has worked including what factors have facilitated or acted as a barrier to delivery.
- The perceived impact of the intervention on pupils, particularly on their ‘interference control’ skills and logical thinking.
- The impact on teachers’ professional development, particularly on their expertise in using reasoning techniques in maths and science lessons and how they assess the impact on pupils’ learning.
- Whether there have been any unintended consequences and outcomes.
- Cost information (see below).

The school visits will include observations of the intervention and control plus groups and in-depth interviews with senior leaders and interviews and/or focus groups with year 3 and 5 teachers and subject specialists (if different) in mathematics and science. The case studies are a cost-effective way of gaining a range of practitioner experience and perspectives on the intervention and its impact. We will also collect relevant documentation from the schools. At the end of the trial (i.e. at the beginning of the summer term) we will arrange telephone interviews (30 minutes) with a representative (either year 3 or year 5 teacher) of the school to seek follow-up feedback on how the intervention went

We will investigate fidelity to the intervention from Birkbeck College's records of teachers' accessing the training materials. This will be analysed to see if the programme is being used in the way it was expected.

Finally, we will administer an online survey of teachers at the end of the intervention to get a broad overview of the implementation of the intervention, any barriers they have experienced and any outcomes observed by teachers on pupils and their teaching. This will capture any changes to delivery and perceived impact as well as asking teachers some basic questions about cost (see below).

This combination of methods will give us a full understanding of how and why the intervention has/ has not worked including implementation challenges and adaptations, any unexpected outcomes and perceived impacts, perceived importance of, benefits from and commitment to the programme and participants' views on its sustainability and suitability for national roll out.

A final report in November 2018 will include the findings of the process evaluation and make recommendations to ensure the sustainability and replicability of counterintuitive concepts learning if it were scaled up.

Costs

We will collect information from Birkbeck College on teacher time required to complete the intervention (time to deliver the intervention, time for preparation, time for training etc) and any resources needed to implement the intervention (number of computers, internet access etc., photocopying needed).

We will also collect schools cost information through the case study interviews and also the teacher survey. We will collect information on supply cover needed to run the intervention and if any digital technology has had to be purchased. We will reduce the burden on schools by gaining as much data as we can from Birkbeck College.

Cost data will be calculated at a per pupil level for each intervention arm. The cost data, combined with the outcome measures would provide an indication of the relative cost effectiveness of the trial arms.

Ethics and registration

The final trial design will be reviewed by NFERs' ethics and code of conduct committee. The trial will be registered by NFER at www.controlled-trials.com to obtain the International Standard Randomised Controlled Trial Number (ISRCTN). (Trial registration number: ISRCTN20284041.)

Opt-out parental consent is appropriate for this study and this will be collected by Birkbeck College through signed memorandums of understanding from headteachers and then letters home to parents. For additional neurocognitive testing of a smaller sample of 180 pupils and neuroscience testing of a smaller sample of 80 pupils, Birkbeck College will need to collect opt-in parental consent forms. This testing is not part of the NFER design and will be undertaken solely by Birkbeck College. The sample of pupils will come from all three groups and the post-testing will occur after the main trial period.

Personnel

The project will be led by Professor Denis Mareschal. The evaluation will be overseen by Simon Rutt from NFER. The impact evaluation will be led by Palak **Roy (née** Mehta) and the process evaluation will be led by David Sims. Anneka Dawson will have overview of the evaluation at EEF and Eleanor Stringer will oversee the grant.

Each person will carry out their duties with the assistance of teams at their respective institutions:

Denis Mareschal – development phase, recruitment and retention of schools, training and delivery of intervention, supply of list of eligible schools for randomisation

Simon Rutt/ Palak Roy (née Mehta) – trial management, randomisation and analysis.

David Sims and Claire Easton – process evaluation telephone interviews, visits, observations and on line survey.

Dr Gareth Hathway, Associate Professor at the University of Nottingham -will work with NFER and act as a neuroscience consultant for the evaluation.

Risks

Risk	Assessment	Countermeasures and contingencies
School attrition	**Likelihood:** moderate **Impact:** moderate	Clear information/recruitment events and/or initial meeting with schools explaining the principles of the trial and expectations. Both 'intention to treat' and 'on-treatment' analysis will be used. Attrition will be monitored and reported according to CONSORT guidelines.
Interventions are not implemented well	**Likelihood:** low **Impact:**	Clear information/initial meeting with schools explaining the principles of the trial and expectations. Both 'intention to treat' and 'on-treatment' analysis will be used.

	moderate	Process evaluation will monitor this.
Control classes also implement the intervention	**Likelihood:** moderate **Impact:** moderate	Clear information/ initial meeting with schools explaining the principles of the trial and expectations. Both 'intention to treat' and 'on-treatment' analysis will be used.
Delays in training of teachers and commencing interventions	**Likelihood:** moderate **Impact:** low	Agree a clear timetable with project teams up-front.
Failure in recruiting schools	**Likelihood:** low **Impact:** high	Project teams could make use of NFER's Research Operations Department to recruit more schools (at additional cost). Timescale could be revised.
Administrative data required not available or supplied in incorrect format	**Likelihood:** moderate **Impact:** moderate	Data sharing procedures will be agreed in advance with Birkbeck College. Limited impact on primary analysis.
Intervention is not properly defined by the time of recruitment and therefore recruitment is challenging	**Likelihood:** moderate **Impact:** moderate	Recruitment could be postponed until intervention is determined. NFER will oversee recruitment materials to ensure schools are properly informed.
Researchers lost to project due to sickness or absence	**Likelihood:** moderate **Impact:** low	NFER has a large research department with numerous researchers experienced in evaluation who could be redeployed. Senior staff can stand in if necessary.
Project teams do not follow correct trial protocols	**Likelihood:** moderate **Impact:** high	Meetings with project teams at start of project. Provision of clear guidance describing protocols for distribution to all schools.

Data protection statement

NFER's data protection policy is available at:

http://www.nfer.ac.uk/nfer/about-nfer/code-of-practice/nfercop.pdf

Timeline

Date	*Activity*
Sep- Dec 2015:	Meeting with partner organisations, draft the protocol
Jan 2016-Jul 2017:	Development phase
Jan 2017- Jul	Pilot of intervention including pilot process interviews

2017:	
Jan 2017- Jul 2017:	Recruit and consent schools and learners for main trial (including recruitment events)
Oct 2017:	Randomisation of schools
Oct 2017:	Training of teachers and attending a training session
Nov 2017 - Feb 2018:	Implementation of intervention programme and case studies
Feb/March 2018:	Process phone interviews Testing Teacher survey
May- Sep 2018:	Review of usage data and reporting
Nov 2018:	Final report- including results from process and impact evaluations
Nov- Dec 2018:	Post analysis conference

Reporting

NFER will provide six-monthly progress reports to EEF. After the development phase/pilot, NFER provided a short summary report containing recommendations of any methodological changes for the main trial in May 2017. We will then provide a report of the evaluation findings to CONSORT standards for the EEF Board at the end of the evaluation (November 2018). This will include the findings of the impact evaluation together with the findings from the process evaluation. We will be pleased to discuss dissemination options, such as an article in a refereed or practitioner journal.