Protocol for the evaluation of counterintuitive concepts intervention

Note: This protocol excludes aspects of the evaluation that are the sole responsibility of Birkbeck 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	Class level randomisation within school. Each school will have an intervention class and at least one of two possible control groups
Primary Outcome	Mathematics as measured by Progress Test in Maths or Science as measured by Progress Test in Science. To be decided during development phase and before commencement of trial in October 2017.

Intervention

The counterintuitive concepts learning intervention is being delivered by Birkbeck University in collaboration with IOE (Institute of Education) by principal investigator Professor Denis Mareschal, using funding from the Education Endowment Foundation and the Wellcome Trust. The project runs from January 2016 to December 2018 and is divided into two phases: 18-month development and pilot phase and a randomised controlled trial phase (main trial). 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, will develop 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 by Year 3 and Year 5 pupils receiving three 15-minute sessions a week before a maths or science lesson, where they use a computer 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. The exact nature and length of the

delivery will be determined during the development phase as this could be as a whole-class intervention led by the teacher or on an individual basis via accessing computer games.

It is anticipated that the impact of the study will continue to develop over time, perhaps as a pupils' learning is reinforced. Thie 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, the applicant proposes developing exercises that are more closely related to the attainment measures we are interested in.

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 the inhibition needs to happen in the networks 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/or science achievement?

The secondary analysis question is: what is the impact of the counterintuitive learning intervention on learners' mathematics or science achievement? This will be the dependent on which measure is used for the primary analysis

An additional 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 will start with a development and pilot phase from January 2016 and continue until end of July 2017. It is anticipated that the design for the main trial would be finalised by December 2016 so that recruitment for the main trial can commence from January 2017 with a view to randomisation taking place in summer 2017.Schools would start to implement the intervention in October 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)		
	School recruitment and randomisation for the main trial (January 2017-August 2017)	Intervention introduced (October 2017) and testing take place (Feb/March 2018)

Phase 1-Development phase/ Pilot

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

1. Develop the intervention materials- Birkbeck
2. Determine if whole class or individual implementation will work best for the intervention- Birkbeck
3. Test the materials with the pilot schools- Birkbeck
4. Assess the suitability of three trial groups for implementation and feasibility- intervention, control and control plus groups[1]- Birkbeck
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 and how classes might be split to receive different assessments post intervention- 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

As the development phase is due for completion after school recruitment has started NFER will look to have a meeting with partner organisations in December 2016. NFER will produce a verbal report of any challenges coming out of the pilot and recommendations of changes to the main trial.

Phase 2- Efficacy trial

For the main efficacy trial, a sample of approximately 100 primary schools will be approached who are nationally representative (including areas of particular interest of EEF) from schools with the

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

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 (referred to subsequently as 'intervention')
- 'Business-as-usual' control (referred to subsequently as 'control')
- Social skills learning control in PSHE lessons (referred to subsequently as 'control plus' as this also assesses the impact of the computer programme)

A cluster design is planned, where Years 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 main trial will start in October 2017 and will run until January 2018 and pupils will be tested in February/March 2018. The exact length of the trial will be determined during the development phase and identified in an updated protocol.

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 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 groupshave been allocated to which intervention or control group.

Participants

At this stage we would recommend that there are several recruitment events to explain the evaluation to schools, so they understand the importance of the randomisation methodology before signing up to the trial (NB. there will not be very much information on the intervention at these events to ensure that this does not affect the realistic nature of the trial or to risk contamination in the classes that are allocated to the control group when randomised). Birkbeck 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 to avoid control schools resisting the data requirements after allocation. The main trial will only be looking at Year 3 and 5 learners who are predominantly, but not exclusively, from schools with the above average proportion of students receiving free school meals.

Outcome Measures

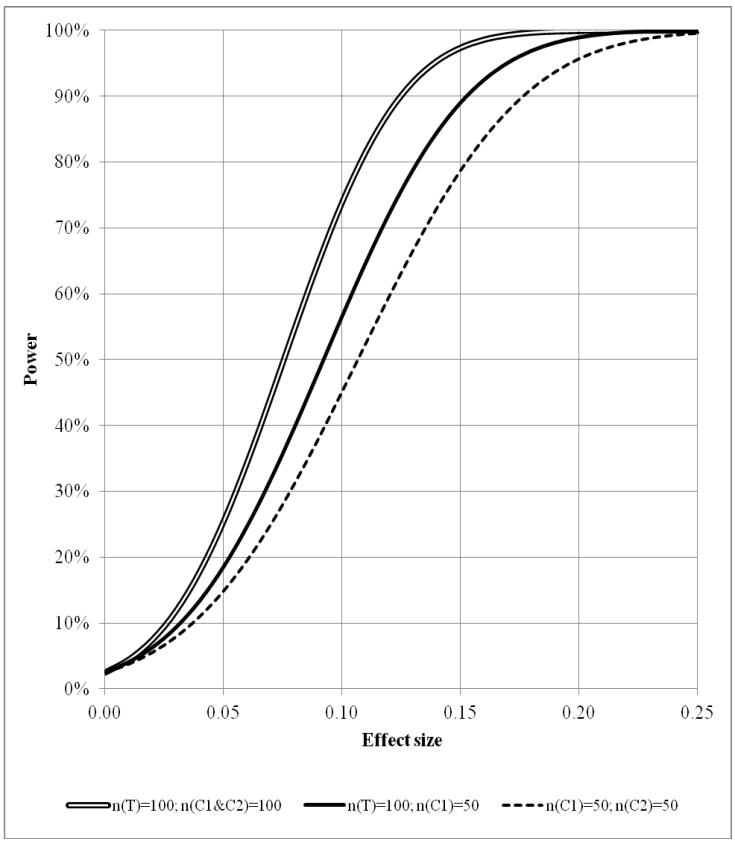
We will measure maths and science outcomes in Year 3 and 5 using the following tests:

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. The tests are designed to be administered by teachers to a class of pupils, therefore minimising the burden on the participating schools while offering time and monetary savings. 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, Year 3 and 5 pupils will be randomly allocated 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. It is anticipated that the paper versions of these tests will be used. Tests will be administered by schools, with guidance and instructions provided by NFERand NFER statisticians will randomly allocate pupils to sit either the maths or science assessment. Despatch and collection of the tests will be managed by NFER. 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. Prior to the commencement of the trial in autumn 2017 a decision will need to be made as to the primary analysis; maths or science.

Birkbeck will additionally administer an assessment of inhibition control. An appropriate measure will be determined during the development phase and will be administered by Research Assistants appointed by Birkebeck to the project. The administration of the test will be undertaken blind so that RAs are unaware of which group they are administering the assessment to. Analysis of the data will be undertaken by NFER.

Sample size calculations



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) =100schools 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.

Analysis

A full analysis specification will be written and appended to this protocol before outcome data is seen.

The primary outcome will be mathematics and science year 3 and 5 outcomes as measured using the following GL assessments; Progress test in Maths (PTM) and Progress test in Science (PTS). All statistical analysis of the programme's impact will be conducted at pupil level, comparing average pupil maths or 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 gender and age. A three- level model will be created to account for pupils being clustered within classes and within school. The main analysis is to look at 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 a susual 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)[2] will be conducted separately. Prior attainment, gender and age variables will again be included.

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 pupil (depending on the method for the main trial). The main analysis will therefore be followed by an 'on-treatment' analysis where data from the computer system will be used to determine the extent of each pupil's involvement. Creation of the computer programme is due to take place during the development phase of the project. It will be necessary during this time to ensure the computer programme collects information in such a way that high, medium and low fidelity can be reliably identified. These categories will need to be pre-specified and will be created using 'click' data collected by the programme. This analysis will enable us to estimate a 'pure intervention effect' (net of any fidelity issues) that would not necessarily be causal in nature.

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

Implementation and process evaluation methods

The process evaluation will involve a number of methods split into the pilot and main trial phases as follows:

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

- the models of delivering the intervention being used (including any home options)
- 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 schools where interviews will be conducted with a year 3 teacher, a year 5 teacher and a senior leader (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?

We will attend and observe a half-day intervention-related training workshop for teachers. 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 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.

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

Finally, we recommend running 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 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 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 interview 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 (particularly in the case of a pupil level intervention on individual computers). We will reduce the burden on schools by gaining as much data as we can from Birkbeck.
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)

Opt-out parental consent is appropriate for this study and this will be collected by Birkbeck 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 will need to collect opt-in parental consent forms. This testing

is not part of this design and will be undertaken solely by Birkbeck. The sample of pupils will come from the intervention and control plus groups and testing will not occur during 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 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 Mehta – trial management, randomisation and analysis.

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

Dr Gareth Hathway, Associate Professor at the University of Nottingham -will 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:** 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. Process evaluation will monitor this.
Schools do not have enough computers to run the trial	**Likelihood:** low **Impact:** high	According to recent research[3] the average number of desktop computers in primary schools in 2012 was 28.9 and the number of laptop computers was 20.2. The trends over time suggested that number of desktops would slightly decrease over time and that laptops would increase. As an average primary class for Years 3 and 5 is around 30 pupils, this suggests that most schools should have adequate provision to supply computers for a whole class for this intervention if the development phase decides this is the best implementation of the intervention. However, we could add in that this was a requirement for

[3] 6 Information and Communication Technology in UK State Schools Full report 2012 Edition Volume II Provision and Spending

		taking part in the trial as long as this did not discriminate against more disadvantaged schools. This will need to be discussed further with EEF and Birkbeck
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. 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, write and register protocol
Jan 2016-Jul 2017:	Development phase

Jan 2017- Jul 2017:	Pilot of intervention including process interviews
Jan 2017- Jul 2017:	Recruit and consent schools and learners for main trial (including recruitment events)
Aug 2017:	Randomisation of schools
Sep- Oct 2017:	Training of teachers and attending a training session
Oct 2017 - Jan 2018:	Implementation of intervention programme and case studies
Feb/March 2018:	Testing
Apr 2018:	Process phone interviews
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, we will provide a short summary report containing recommendations of any methodological changes for the main trial (December 2016). 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.