

Teacher Choices trial: A Winning Start (second run)

Proposed Evaluation Protocol

Evaluator (institution): National Foundation for Educational Research

Principal investigator(s): Dr Ben Styles



The 'A Winning Start' trial was scheduled to run during the spring term of the academic year 2019-20. Due to the Covid-19 pandemic, the UK Government asked all schools to close from 20th March 2020 to all pupils (except vulnerable children and the children of key workers). Consequently, participating science teachers could not fully deliver the intended lesson starters on the second phase of the trial and students did not take the respective end-of-topic tests. As a result, the data collected during the original trial is limited and does not allow us to answer the research questions fully, especially the one that investigates the feasibility of linking attainment data from different end-of-topic tests across a number of topics. Further information on this original trial can be found in the published protocol (EEF, 2020a).

In order to address the research questions fully, EEF and NFER decided to rerun the trial in the Spring and Summer terms of 2021. However, as a result of the third national lockdown in the UK, all schools closed from 6th January 2021 and the trial activities were paused.

In June 2021, it was decided to relaunch the trial that will run in the academic year 2021-22. This protocol relates to the trial activities where the science lessons will be delivered in the autumn term 2021 and spring term 2022. As originally planned, the second run of the trial will allow the full crossover design to be implemented and full assessment data to be collected and analysed. The re-run will build on the learning from the original trial particularly from the implementation and process evaluation (IPE). As such, elements of the IPE will be reduced in the re-run (which will focus on the crossover design). EEF has already developed the Winning Start teacher guidance document; and NFER will run the evaluation to test these choices in a robust way. This second run is important as it enables EEF and NFER to explore the feasibility of running this type of trial and to fully understand more about the best designs and methods to use. Changes in the design and data collection between the first and the second run of 'A Winning Start' trial are summarised in Appendix A.

PROJECT TITLE	A Winning Start trial (second run)
DEVELOPER (INSTITUTION)	Education Endowment Foundation (EEF)
EVALUATOR (INSTITUTION)	National Foundation for Educational Research (NFER)
PRINCIPAL INVESTIGATOR	Dr Ben Styles
PROTOCOL AUTHOR(S)	Palak Roy, Jennie Harland and Maria Galvis
TRIAL DESIGN	Two-armed randomised controlled trial (RCT) with a crossover design; Random allocation of teachers
TRIAL TYPE	Feasibility of RCT
PUPIL AGE RANGE AND KEY STAGE	Ages 12-13 (Year 8, Key Stage 3)
NUMBER OF SCHOOLS	Approximately 40 schools
NUMBER OF PUPILS	80 teachers/classes; approximately 2,400 pupils

PRIMARY OUTCOME MEASURE AND SOURCE	Pupils' improved learning in science as measured by the end-of-topic tests
SECONDARY OUTCOME MEASURE AND SOURCE	None

Version history

VERSION	DATE	REASON FOR REVISION
1.0	30 November 2020	N/A
2.0	December 2021	The trial was scheduled to run in spring 2020. Due to the Covid-19 pandemic related school closures, the data collected during this original trial was limited. Hence, it was jointly decided by EEF and NFER to rerun the trial in 2021. However, it was not possible to progress with this trial as plan due to the third national lockdown in January 2021. The recruitment was launched again in June 2021 with a view to running the trial in 2021-22. This protocol relates to the trial that was successfully implemented in the autumn and spring terms of 2021-22. The changes between the first run and the second run are presented in Appendix A.

Contents

Teacher Choices study rationale and background	3
Teacher Choices trial: 'A winning start (second run)'	7
Evaluation overview	7
Intervention	7
Feasibility of measurement and impact evaluation	13
Implementation and process evaluation	17
Cost information	20
Ethics and registration	20
Data protection	21
Personnel	21
Risks	23
Timeline	24
References	26
Appendix A: Summary changes between the original and second run of 'A Winning Start' trial	27

Teacher Choices study rationale and background

Note: This rationale applies to the original as well as the re-run of 'A Winning Start' Trial.

Rationale for the evaluation

In 2018, EEF undertook a review of their grant-making process with the aim of understanding how to make their projects timelier, more relevant for schools and to ensure that the results were as useful as possible to schools and other decision-makers. Through this process, they identified that teachers and headteachers were keen to answer research questions which can more directly feed into existing teaching practice. These questions were often not related to the impact of manualised programmes, which require schools to purchase particular resources or training, but were instead about the everyday choices that teachers have to make when planning their lessons and deciding on the pedagogical approaches they take in their classrooms.

As a result of this review, EEF commissioned NFER to conduct two classroom-based trials to explore the feasibility of running these type of trials and to understand more about the best designs and methods to use, with the intention of EEF then commissioning more of these in the future. These classroom-based trials are called 'Teacher Choices trials'. These trials will explore the feasibility of running robust evaluations of teachers' everyday choices in the classroom.

The key aims of the 'Teacher Choices' trials are to ascertain the feasibility of running such trials in schools and to provide recommendations for future scale-up. The research questions will focus specifically on how the selected everyday choices are implemented in the classroom, how the everyday choices can be defined and described in the guidance, how the trial methodology is followed including if/how the choices can be tested in a robust way using teacher-tests and the most effective way to communicate the results to teachers to support classroom practice/improvement.

The Teacher Choices trials aim to answer the research questions centred on the feasibility of the research method as well as the implementation, process and impact evaluation of the teacher choices. The research questions addressed in the two original trials ('A Winning Start' trial or TC1 and 'The Story Time' Trial or TC2) can be found in the relevant protocols (EEF, 2020a and EEF, 2020b respectively). As this Teacher Choices trial will be a re-run of the curtailed 'A Winning Start' Trial, we have adapted the overall research questions from the previous Teacher Choices studies to focus on the areas we were unable to fully investigate in the original trial. These are:

1. What are the enabling and challenging factors in implementing a 'Teacher Choices' trial?
 - a. How do existing and new schools and teachers respond to recruitment and how well do they engage in the trial?
 - b. Do teachers adhere to the trial design and requirements set out in their information sheet, i.e. randomisation allocation, cross-over and testing?
 - c. Is the design and data collection manageable for teachers?
2. How, and how well, is the choice implemented?
 - a. Is implementation compliant? (see the *Intervention* section for an adjusted approach in this trial)
 - b. How does implementation vary – for example, do teachers adapt the approaches to suit their context? Do they align with usual practice?
 - c. What are teachers' views on the guidance and support provided?
3. Can the end-of-topic tests be used to measure the impact of teacher choices RCTs?
4. Is it feasible to combine or link attainment data from different teacher-made tests across a number of topics?
5. What can we learn from the trials about the expected effect sizes for future teacher choice trials?
6. Does a given teacher choice improve pupil attainment outcomes as measured by teacher-made tests?

Due to the type of research questions and the relative simplicity of the choices, the evaluation designs do not follow standard EEF guidance during the trials. These designs enable the trials to run over shorter timeframes to ensure that the research questions are answered in a timely way.

Teacher Choices

It is envisaged that the everyday teacher choices being investigated as part of these trials and in the future would be of high interest to schools. They should involve a real choice that can be made by classroom teachers, be easy to define so that teachers know what they are expected to do as part of the project and relatively easy to implement without lots of training and resources. Teachers would receive a simple input, which could be easily replicated in the future for teachers wishing to implement the choice outside of the test condition.

Table 1 sets out how the re-run of 'A Winning start' will contribute to the overall aims of the teacher choices feasibility studies to date, and in particular where it will be adapted to complete or complement the first Winning Start trial.

Table 1: Teacher Choices study – research questions, methods and analyses for ‘A Winning Start – second run’

Research Area	Research Question	Study methodology area	Data collection methods	Participants/data sources	Data analysis methods
Feasibility of taking part in the RCT methodology	RQ1: What are the enabling and challenging factors in implementing a teacher choices trial? - How do existing and new schools/teachers respond to and engage in the trial? - Do teachers adhere to the trial design and requirements, in particular randomisation allocation, cross-over and testing? - Is the design and data collection manageable for teachers?	IPE (feasibility)	Recruitment Teacher survey incl. compliance report for each starter Compliance record (in survey) Telephone interviews with heads of dept. and teachers	NFER recruitment data Teachers N = 80-100 for each survey Interviews in 4 schools (2 existing and 2 new) with the head of dept. and up to 2 teachers in each school	Descriptive analysis Formative views
Implementation and process evaluation	RQ2: How, and how well, is the choice implemented? - Is implementation compliant? (see the Intervention section for an adjusted approach in this trial) - How does implementation vary – for example, do teachers adapt the approaches to suit their context? Do they align with usual practice? - What are teachers' views on the guidance and support provided? -	IPE (implementation)	Teacher survey incl. compliance record for each starter Telephone interviews with heads of dept. and teachers	Teachers N = 80-100 As above, interviews in 4 schools (2 existing and 2 new) with the head of dept. and up to 2 teachers in each school	Thematic analysis Perspectives analysis Observational analysis Triangulation
Feasibility of testing / measurement	RQ3: Can the end-of-topic tests be used to measure the impact of teacher choices RCTs?	Testing, measuring impact	Total scores on end-of-topic tests and a standardised test- GL Assessment's Progress Test in Science (PTS13)	Pupil N = approx. 2,400-3,000	Method 1: Meta-analysis of individual test analyses;
Feasibility of testing / measurement	RQ4: Is it feasible to combine or link attainment data from different end-of-topic tests across a number of topics?				

Research Area	Research Question	Study methodology area	Data collection methods	Participants/data sources	Data analysis methods
			End-of-topic tests themselves		Method 2: Calibration of tests across topics within each school; cluster cross-over analysis across whole sample
Trial design	RQ5: What can we learn from the trials about the expected effect sizes for future teacher choice trials?	Measurement, sample sizes			
Impact	RQ6: Does a given teacher choice improve pupil attainment outcomes as measured by end-of-topic tests?	Impact			

Note: We have not included the research question from the original TC1 trial on ‘exploring ways of ensuring accessible and appropriate outputs for teachers’. We intend to conduct focus groups discussions after ‘A Story Time Trial’ (TC2) to explore this, and, if these are conducted in a timely manner, we will include teachers from this trial in the focus group sample.

Teacher Choices trial: 'A winning start (second run)'

Evaluation overview

This protocol relates to the proposed re-run of the first classroom-based Teacher Choices trial: 'A winning start'. It aims to evaluate two different approaches to starting lessons: quizzes and discussions. Quiz starters build on evidence about the impact of retrieval practice to help students remembering key facts and ideas. Discussion starters are a popular way to begin lessons that aim to engage students in the topic they are studying.

This evaluation will run as a two-armed crossover RCT. The intervention delivery will last for two half-terms (Autumn 2 2021 and Spring 1 2022) and split into two phases with the Christmas holidays in the middle, which works as a 'washout' period. Each science teacher will deliver two science topics to their Year 8 class, one in each of the half-terms. For a given topic, teachers will be randomly allocated to start their science lessons with quizzes or discussions. At the end of each topic, teachers will administer the end-of-topic test normally used by their department. These end-of-topic tests could be tests purchased from publishers or tests created by teachers/schools themselves. Outcomes from these tests will be the primary outcome for the trial (within impact analysis).

The process evaluation seeks to understand how the teachers implement the choices, including views on the guidance and support received, and on overall compliance. It also aims to understand how the trial methodology itself is followed by teachers, including adherence to randomisation condition, switching conditions in each half-term, and manageability of the trial requirements such as testing and returning data.

Intervention

Quiz starters build on evidence about the impact of retrieval practice, which has been identified as one of the best-evidenced approaches from cognitive science (Dunlosky *et al.*, 2013). Evidence suggests that repeatedly getting students to try to remember what they have learned is a good strategy, especially if they get feedback on the correct answer (Roediger and Butler, 2011). However, there is some debate about whether this applies to complex learning, which is more like the kind of learning we would typically expect in Key Stage 3 science (see Van Gog and Gog and Sweller 2015; Rohrer *et al.*, 2019). To date, most of the existing studies have been done in laboratories with relatively trivial kinds of learning; some are school-based, but generally have other limitations that make it hard to know whether the ways that typical teachers might use retrieval practice are in fact beneficial for complex learning.

Discussion starters have been identified as an active control group but it does not have any specific supporting evidence. It is believed to be a widely adopted practice with a plausible rationale that allows us to make a comparator more specific and homogeneous than 'business as usual'.

The following framework outlines what is involved in 'A Winning Start'. In addition, a TIDIER framework for each starter has been updated and agreed between EEF and NFER. These will be included in the final report.

Brief name

A Winning Start – Quiz and Discussion lesson starters

Why

The study aims to investigate the impact of two different approaches of starting lessons (quiz and discussion). Retrieval practice has been identified as one of the best-evidenced approaches from cognitive science (Dunlosky *et al.* 2013). It supports pupils to recall/remember, and embed in longer-term memory. Discussion starters are believed to be common practice in foregrounding new learning.

Who

Year 8 teachers in the science departments are the recipients of the guidance, and Year 8 pupils from the participating science classes are the recipients of the lesson starter approach.

What

Materials

Year 8 teachers from the participating schools will be given a short teacher guide to enable them to implement both approaches of lesson starters (i.e. retrieval practices or discussions). The teacher guide was developed by the EEF and used by the teachers in the original trial (EEF, 2020a). NFER and EEF have jointly amended the guidance document, based on feedback from these teachers and taking into account that some teachers may have taken part in the original trial, whilst some will be new to the trial.

As in the original trial of 'A Winning Start', teachers will be provided with the guidance document with their first allocated starter appearing first (i.e. two versions of the document will be printed and provided according to randomisation). A cover sticker or a letter will be attached with details of their allocation and topic. In addition to the guidance on how to use the approaches, EEF will provide heads of department with an email helpline for teachers and ad hoc support via telephone, where requested.

Teachers will need to find or develop their lesson starters (questions, prompts etc.) appropriate to their topic and class. The teacher guide mentioned above contains links to resource information. Note, teachers who have taken part in the original trial of 'A Winning Start' may wish to use the starter resources they developed for that trial. We will include a survey question to establish how often this is the case.

Procedures

The teachers will experience delivering both types of lesson starters – one in the first half-term of the trial and then the other in the second half-term. The type of lesson starter for each topic will be randomly allocated to teachers by NFER.

Who

EEF has developed the teacher guidance. Year 8 science teachers are the implementers of the lessons starter approach. Teachers taking part can include experienced teachers, Newly Qualified Teachers (NQTs), science specialists or non-specialists. Teachers may or may not have taken part in 'A Winning Start' trial before. Teachers may or may not have used the starter approach before; and they may or may not have taught that half term's topic before.

At the end of the trial, schools will receive an incentive payment of £150 per participating teacher.

How

After randomisation, NFER will give Year 8 teachers the teacher guide (see above *Materials*) to enable them to implement both approaches (quiz starters and discussion starters). The teachers will then need to plan and develop their lesson starter questions and discussion prompts, appropriate to their topic and class. They will experience both types of lesson starters: one in the first half-term of the trial and the other in the second half-term. The type of lesson starter for each topic will be randomly allocated to teachers by researchers at NFER.

Where

The starters will be used in Year 8 science classes in participating state-funded secondary schools. Schools can nominate as many teachers as they wish but at the minimum two teachers with two classes are required (each participating teacher must teach only one class). The topics will be selected and communicated to NFER by participating teachers prior to randomisation. Classes using any type of student grouping approach in science are eligible i.e., streaming, setting or mixed ability

grouping. Joint-teaching (where one class is taught the same topic by two teachers) will not be eligible for the trial.

Teachers can choose when to teach each topic across the trial period. However, they will decide and communicate this order with NFER prior to randomisation. Random assignment of topic and approach must be adhered to.

When and how much

The teacher should ideally use the randomly allocated class starter approach in every lesson of the topic during the relevant half-term (four or five weeks approximately) with their trial class. Topics should be selected so that the allocated starter can be used in no fewer than six lessons. Teachers are expected to use the allocated lesson starter approach for no more than ten minutes at the start of each lesson in their topic (unless unforeseen circumstances occur such as a fire alarm).

Tailoring

The process evaluation will explore the level and amount of tailoring by teachers. This will include information on how teachers developed their starters, number of lessons within their topic, and examples of adaptation to suit their context.

How well

How well teachers were able to deliver the starters was explored in the original trial of 'A Winning Start' through logs, exploration of materials and case study observations and interviews. In the second run of the trial, fidelity will be explored through survey questions in each half term on usual practice and delivery, and through telephone interviews with teachers and heads of department (see IPE methods).

Costs

Costs in terms of planning time will be explored.

Implementation fidelity

We will explore the extent and nature of fidelity at three levels: i) compliant starter usage (compliance being using the correct starter for no more than ten minutes in all lessons with the exception of unforeseen circumstances such as fire alarms); ii) implementation fidelity (i.e. teachers' reports of implementation of the approach itself); iii) implementation support (i.e. the provision of the guidance materials and the offer of a support call to heads of department. These are outlined further below.

- i) **Starter usage (compliance):** We will explore teachers' reported **starter usage** (using compliance questions in a survey at the end of each half term) as a way of understanding teachers' usage of the starters and students' exposure to the approach. Teachers are expected to use their allocated approach in all their lessons (excepting unforeseen issues such as fire alarms¹). As the number of lessons per topic will be different across the teachers in the trial (as they are teaching to their own/department's timetable), we will consider asking teachers to indicate in the survey the number of lessons in their topic and the number in which they used the starter for no more than ten minutes. However, we would discuss the level of detail required here with EEF, and whether self-report using a measure of 'always, mostly, sometimes, rarely, never' would be sufficient.
- ii) We will explore the extent of **implementation fidelity** in terms of: whether teachers report reading the guidance document, whether teachers have planned how they are going to deliver their lesson starters, and how teachers report using their allocated approach. In addition, as we know from TC1 that there was variation in the nature of the

¹ We will also be mindful of local lockdowns or teachers' self-isolating etc. when we report compliance, including a 'Not Applicable' category should a local lockdown prevent many lessons from being delivered.

starters being delivered, we will explore further examples of implementation variation through telephone interviews.

- iii) We will also explore **other aspects of implementation support and characteristics including**: how the teachers went about planning their starters (including exploration of planning approach/re-use of materials amongst any teachers who were part of TC1), the nature of the starters, and whether their school took up a support call with EEF, and the kinds of support issues/questions that teachers raised on emails with EEF and NFER. As a minimum expected level of 'fidelity' we would expect teachers to have read the guidance and planned their starters.

Figure 1: Logic model for the quiz starters

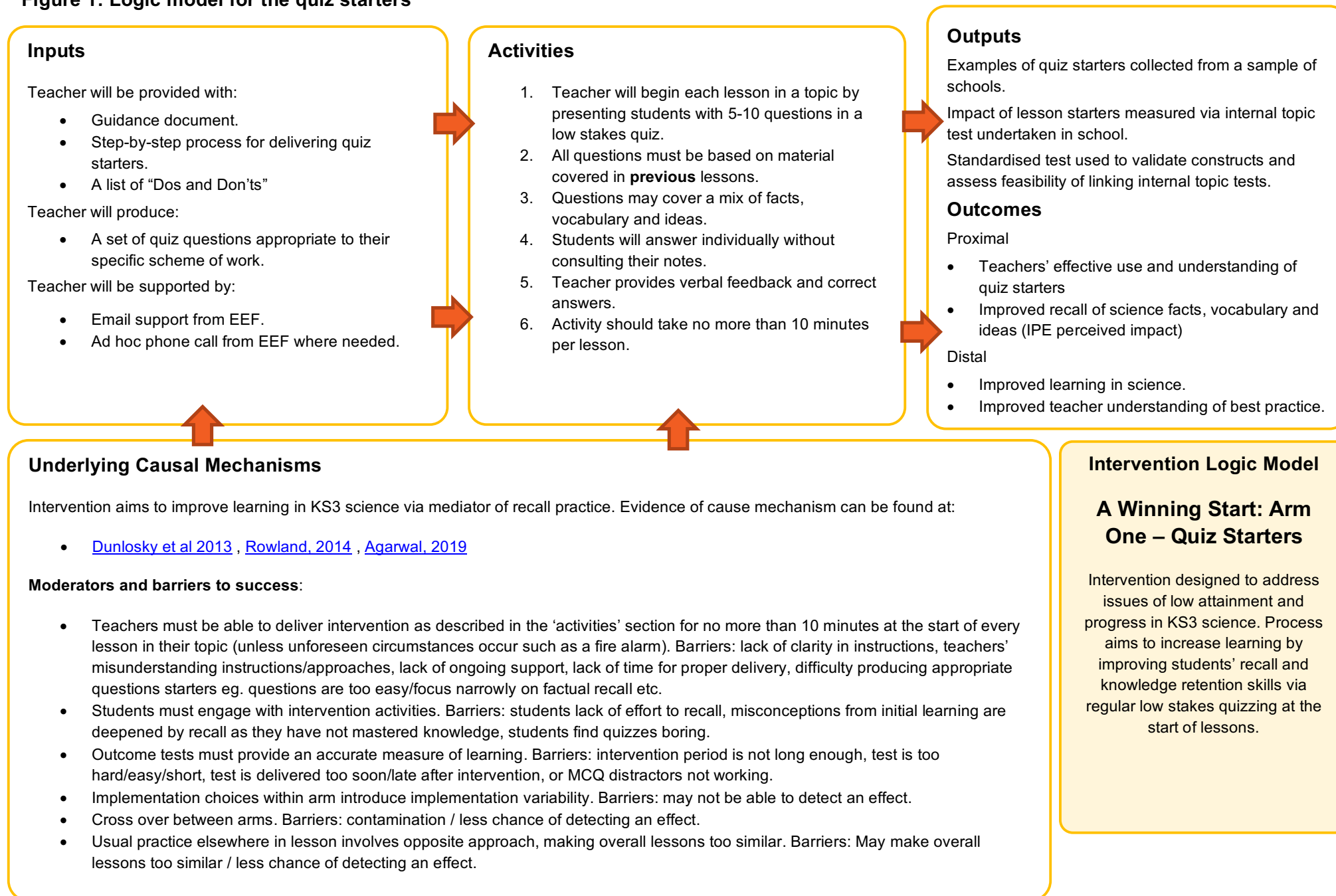
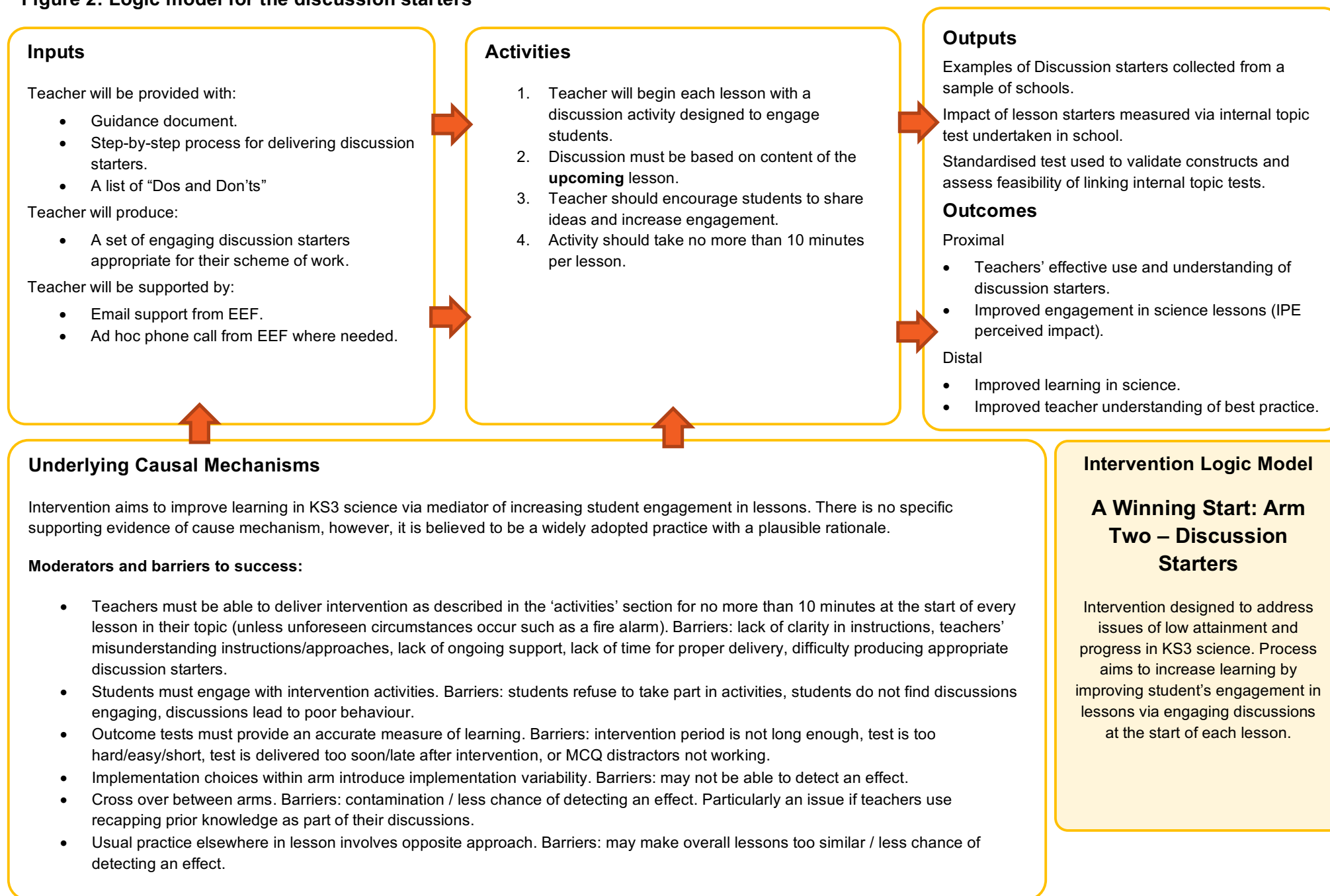


Figure 2: Logic model for the discussion starters



Feasibility of measurement and impact evaluation

Research questions

This trial will investigate the overall research aims of the Teacher Choices study as per Table 1. The primary aim of the trial is to explore the feasibility of the Teacher Choices approach. In order to do this, we will first explore the feasibility of the tests and subsequently conduct an impact evaluation which would enable us to estimate the impact of the choice (See Table 1, Research Questions 3 to 6).

Within the impact evaluation, the primary research question is: What is the difference in pupils' science attainment measured by end-of-topic tests for pupils who received science lessons with quiz-starters compared to those pupils who received science lessons with discussion-starters?

Design

When schools sign-up to the trial, they are asked to nominate as many Year 8 science teachers as they wish to but a minimum of two teachers are required. Only one class per participating teacher can take part in the trial. This is different from the original trial and the change is intended to reduce overall data collection burden for teachers where they will be asked to provide trial data only for one class. We will select all nominated classes irrespective to their science grouping practices. Further details on eligibility and selection are included in the section '*participants*'.

The trial follows a crossover design using two separate science topics in each school. The delivery will last for two half terms with a Christmas holiday in the middle, which works as an interval to 'washout' the effects of the treatment, received in the first phase. Each science teacher will deliver two science topics (topic A & topic B) to their Year 8 class. Randomisation will determine whether the science lessons will be started with quizzes or discussions. For a given teacher, the lesson starter approach will be randomly allocated to each topic. The order of the topics taught can be different across teachers within a school. Each teacher will switch the topics and lesson starter approaches for the second half-term. Prior to randomisation, each participating school determines the two science topics and the order of the topics that each teacher will teach so these decisions are not based on the randomisation.

Table 2: Trial design

Trial design, including number of arms		Two-arm crossover RCT
Unit of randomisation		Science teachers (or classes)
Stratification variables (if applicable)		School
Primary outcome	variable measure (instrument, scale, source)	End-of-topic scores in science
		School selected end-of-topic tests
Baseline for primary outcome	variable measure (instrument, scale, source)	N/A
		N/A

Randomisation

For a given topic within a school, science teachers will be randomly allocated to quiz or discussion starters. This means, within a school, teachers will be randomised to each lesson starter for a given topic. Unlike the original TC1 trial, we will not collect student grouping practices and therefore the randomisation will not be stratified by grouping practices.

Figure 3 illustrates a simple example of a trial design for a secondary school with four participating teachers. The school takes part in the trial with topics A and B where the order in which teachers teach the topic are pre-specified prior to randomisation and may not occur in the same half-terms across the school. However, within school randomisation of teachers/classes will mean that two

teachers will be randomly allocated to start Topic A lessons with the quizzes and after the Christmas holidays, they will teach Topic B lessons with the discussions.

Figure 3: An example of a crossover design for ‘A Winning Start’ trial (rerun)

	Topic A		Topic B	
First intervention period (second half of the Autumn term)	Teacher 1	Teacher 2	Teacher 3	Teacher 4
	Quiz	Discussion	Quiz	Discussion
	End-of-topic test – Topic A		End-of-topic test – Topic B	
Christmas holidays ('washout' period)				
	Topic B		Topic A	
Second intervention period (first half of the Spring term)	Teacher 1	Teacher 2	Teacher 3	Teacher 4
	Discussion	Quiz	Discussion	Quiz
	End-of-topic test – Topic B		End-of-topic test – Topic A	

An NFER statistician will undertake randomisation in October 2021. They will not be blinded to group allocation. This means the statistician will be able to see which teachers are assigned to which lesson starter and their topics. Due to the staggered recruitment and data collection, it may be possible to randomise in more than one wave. Further details will be included in the Statistical Analysis Plan.

Recruitment

NFER will be responsible for school recruitment for this trial and EEF. In summer term 2021 we wrote to schools that expressed interest in this trial previously, to ask whether they wish to take part in the rerun of the trial. The trial will also be promoted via EEF channels, e.g. promoting it on their website and via a research blog.

The intended trial sample in TC1 was 80 teachers although due to the popularity of the trial, we recruited beyond the target sample. As a result, 99 teachers with 107 classes from 33 schools took part in the trial. During the course of the trial, some schools and teachers withdrew from the trial. As a result, we were able to collect end-of-topic test results from 70 classes, 66 teachers from 28 schools. This lower follow-up rate was partly due to the national lockdown in March 2020.

We recognise that academic year 2021-22 poses many challenges and uncertainty to the schools. Given the current situation and the response rate in TC1, we propose to reduce the data we collect from schools during the rerun of the trial. This will not only assist in achieving the recruitment target, reduced burden placed on participating teachers and schools will also help with higher percentage of trial follow-up. (Please see Appendix A that summarises changes between the original and proposed rerun). Given the previous response rate and current school challenges, we propose to recruit 100 teachers to ensure that we achieve the full suite of test data (end-of-topic tests and PTS13) from an adequate number of randomised units to include them in the analysis (recognising that we may not get data from all 100 teachers).

Participants

Schools: All state-funded schools in England are eligible. There are no geographical constraints on schools or exclusions based on participation in other trials. Schools that provide pupil and teacher data at baseline will be put forward for random allocation.

Year 8 science teachers and their classes: Schools can nominate as many teachers as they wish but a minimum of two science teachers/classes per school is required. Participation will be limited to one class per teacher to reduce data collection burden for each participating teacher. As the teachers/classes are randomised, chosen trial topics will be taught in their entirety to the participating

class by only one teacher. Any student grouping approach can be used in the science classes (streaming, setting, mixed ability). NFER will not collect information about science grouping practices.

Topic eligibility: Prior to randomisation, each school will nominate two Year 8 science topics for the trial irrespective to the number of teachers. These topics will be selected such that the lesson starters can be used in at least six lessons. All participating classes within a school should take the same end of topic test². These are end of topic tests normally used by the school and administered by the teachers at the end of the topic and lesson starter delivery - once after Autumn 2 and the other after Spring 1. End-of-topic tests will have at least eight independent items/questions. If the test covers multiple science topics, it should have at least eight independent items/questions for the chosen trial topic.

Pupils: All Year 8 pupils taught only by the participating teachers will be eligible to take part except where pupils are taught the chosen trial topic by more than one teacher.

Sample size calculations

Unlike the original trial, NFER will not collect information about science grouping practices. Thus randomisation will not be stratified by ability groups.

The sample size calculations are run with assumptions of standard parallel-groups cluster design with the following parameters: pre-post correlation of 0.7³, expected effect size of 0.2 and class size of 30 students. If we achieved a class-level ICC of 0.2, we will need 90 classes in total; with a smaller class-level ICC of 0.17, we will need 80 classes in total and with a class-level ICC of 0.15, we will need 72 classes. Based on these calculations and the fact that the trial was not set-up to detect a specific effect size, it was decided that NFER would recruit 80 classes to take part in the trial. In order to achieve adequate number of classes with the follow-up data, we propose to recruit 100 teachers for this trial.

Table 3: Sample size calculations

		OVERALL
Minimum Detectable Effect Size (MDES)		0.20
Pre-test/ post-test correlations	level 1 (pupil)	0.70
Intraclass correlations (ICCs)	teacher	0.17
Alpha		0.05
Power		0.80
One-sided or two-sided?		Two-sided
Average cluster size		30
Number of teacher-class units		80
Number of pupils		2,374

Outcome measures

Baseline measures

² This criteria was added in the second run of the trial based on the learning from the previous round that schools may use different topic assessments for different classes depending on their perceived science ability.

³ https://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol_or_SAP/Pre-testing_paper.pdf

Originally, it was intended that the Key Stage 2 maths scaled score will be used as a baseline measure but the Key Stage 2 were cancelled in 2020 for the current participating students. Hence, it was decided that no baseline data will be collected for the trial cohort. Besides, the main aim of rerunning the trial is exploring the feasibility of running a cross-over analysis where a baseline measure is not essential and therefore no prior attainment or baseline attainment data will be collected.

Primary outcome

Schools will measure science attainment via schools' chosen end-of-topic science tests. These tests are likely to be purchased from the publishers or tests created by teachers or schools themselves. Participating teachers will administer a test at the end of each half-term and share pupil results with NFER. This way, each pupil will be tested twice on end-of-topic test during the trial- after they have been taught using each lesson starter. End-of-topic tests will be designed, delivered and marked by each teacher. They will share the total raw score for each end-of-topic test (or each topic) with NFER via a secure portal.

In addition to this, NFER will also administer the GL Assessment's PTS13 Spring/Summer version. PTS13 is linked to the science national curricula⁴ and it was trialled and standardised in 2015⁵. PTS13 will be administered with an aim to enable scaling or calibrating of all end-of-topic test results to allow comparison across different tests and topics. For further details on these, please see the analysis section.

GL Assessment's PTS administration will be managed by NFER. This will take place after the second end-of-topic test, in late spring term 2022. We appreciate that the schools will be facing challenges during this time, and NFER will work closely with each school involved in the project to ensure that their participation in the assessment is as easy as possible, supporting them throughout to ensure the impact on pupils and teachers is minimal and that the experience is positive. We are proposing to use a flexible approach to the test administration which will allow schools to select whether they want the test administrator to come into the classroom or to simply liaise with the assigned school contact, to drop off and collect the scripts, ensuring all scripts are collected and that the attendance register is complete and matches the scripts. All test administrators will be provided with suitable PPE (if required at this point) and will receive specific health and safety training, in line with the latest government guidance on Covid-19. Once the testing is completed, the scripts will be returned securely to NFER, and then delivered back to GL Assessment who will mark the tests and send the item-level data back to NFER for analysis.

Compliance

Compliance in terms of 'how much' is defined for 'A Winning Start' trial as: teachers have used the allocated lesson starter approach for no more than ten minutes at the start of every lesson in their topic (unless unforeseen circumstances occur such as a fire alarm, or local lockdowns or teachers' self-isolating, in which case a N/A category might be required). This will be captured through a short teacher survey at the end of each half term.

The questions of 'how' and 'how well' will be explored as part of implementation fidelity and implementation support – as set out in the section on Implementation fidelity and discussed in the IPE section of this protocol.

⁴ <https://www.gl-assessment.co.uk/media/294658/pts13-links-to-national-curricula.pdf>

⁵ https://www.gl-assessment.co.uk/media/1872/pts-technical_information.pdf

Analysis

Participating schools will choose topics to take part in the trial. Due to this diversity of science topics, we will need to combine results across the topics. In discussion with EEF, we propose the following two analytical methods to create the primary outcome. Method 1 relies on analysing the total raw score for each end-of-topic test (or each topic) via a linear model for each school. Due to the variation in topics and tests, there will be at least two models per school- one for each end-of-topic test. We will combine results from each model using a meta-analysis approach as per Borenstein et al, 2009.

Method 2 relies on calibration of tests across topics within each school. As different topic tests are completed by the same students and all the students will take a common science assessment (PTS13), it will be possible to link their end-of-topic test scores via Item Response Theory (IRT). If IRT derives an ability scale for all pupils within a school, we will then use this scale as an outcome measure to analyse the cluster crossover design with a fixed school effect. There are numerous challenges in constructing an IRT model, the exact methods are specified in the TC1 Statistical Analysis Plan (Andrade et al, 2020). We will also update and publish the SAP for this trial. Table 4 presents advantages and disadvantages of the two proposed methods.

Table 4: Proposed analysis methods

	Advantage	Limitation
Method 1: Meta-analysis of individual test analyses	It will be possible to use all tests and combine across topics	Does not harness the power of cross-over design
Method 2: Calibration of tests across topics within each school; cluster cross-over analysis across whole sample	Allows power of cross-over design to be harnessed	Calibration of tests may not be possible in all schools

Presenting a comprehensive set of sample size parameters for future trials will be crucial. We will compile important parameters such as pre-post correlations and effect sizes that can be used in future trials.

Implementation and process evaluation

IPE research focus

The implementation and process evaluation (IPE) will focus on Research Questions 1 and 2 of the whole study. It will serve two purposes (adapted slightly in light of learning from the first 'A Winning Start' trial, and so as not to unnecessarily burden schools):

- **Feasibility of the RCT methodology:** To explore how existing (TC1) schools and new schools respond to recruitment and how well the teachers stay engaged in the trial. To understand how the cross-over trial methodology itself is followed by teachers, including adherence to randomisation conditions and allocated topics, and trial requirements such as testing and returning data; and views on the amount and nature of 'trial information' provided and whether the trial is manageable/feasible for teachers to take part in (i.e. the feasibility of this cross-over RCT as a methodology for Teacher Choices evaluations).
- **IPE of the choice/approach:** To understand how, and how well, the lesson starter approaches are implemented by teachers, how implementation varies, and views on the guidance and support provided.

Research methods

The IPE will be informed by the EEF's latest IPE guidance, with a focus on compliance (although collected differently to that in the previous trial), fidelity and usual practice data collection. We will check that the logic model stands, and adapt in collaboration with EEF if needed. The IPE will also be

informed by learning from the first trial of 'A Winning Start', to help both streamline and focus the IPE for the re-run. This includes:

- formative findings on the guidance and support – which suggests we should continue to collect initial data from EEF's support emails and ad hoc calls (if required by the school), in order to check if schools have any initial preparation queries, but no need to collect consultation feedback on the guidance document itself
- observed variations in starter practice – which suggests value in gathering views in addition to the surveys. We suggest using telephone interviews here, rather than case studies, in order to avoid burden on schools and any issues with external visitors during the academic year 2021/22
- challenges for teachers in completing the lesson logs – suggesting a revised approach for collecting compliance data
- the burden on schools of providing starter materials, when we need this trial to focus on collecting test data for the cross-over design – we suggest collecting starter materials from telephone interview schools only, where possible.

As before, the IPE will also be guided by the other dimensions in EEF's IPE guidance (Humphrey et al., 2017), focusing on: responsiveness, implementation adaptation and on implementation support factors. These are the dimensions and factors that will help us explore the extent to which teachers are able to implement the choice, whether they adapt it to their context, what their usual practice is, and the extent to which this approach and design are feasible in the classroom.

The IPE for the second run of 'A Winning Start' will involve:

1. Two short online surveys to gather teachers' reports of compliance, and their views on starter implementation, and guidance and support. The first survey will be completed at the end of Autumn term 2 (and will take about five minutes to complete), and the second at the end of Spring term 1 (this will also include a bank of questions on the manageability of taking part in the trial, and will take about ten minutes to complete). As in TC1, the survey will include mainly rating scale questions and a small number of open responses).
2. Telephone interviews towards the end of Spring 1, in four schools (two that are new to the trial, and two that took part previously), including interviews with the head of department and up to two teachers in each school. Instead of collecting starter materials from all teachers (which was burdensome for schools in TC1), we will collect starter materials for quizzes and discussions and schemes of work from these telephone interview schools only.
Schools will be asked in survey 1 to indicate their willingness to take part in telephone interviews; the four schools will be selected randomly from those that are willing, to cover two that are new and two that took part previously.

In addition, the EEF will provide NFER with a record of their support emails and ad hoc calls offered to each school, and any issues raised in emails (or calls, if any) with teachers about implementing the lesson starters or about the trial. NFER and EEF will collaboratively develop a recording sheet for this purpose.

Table 5 below provides an updated IPE methods/research questions overview, in line with the latest EEF IPE guidance.

Table 5: IPE methods/RQs overview

Research methods	Data collection methods	Participants/ data sources (type, number)	Data analysis methods	Research questions addressed	Implementation dimensions/logic model relevance
IDEA-light discussion	Revise TIDIER framework, logic model and Teacher guidance document	EEF team, NFER team	N/A	N/A	N/A
Short feedback survey 1	Survey for teachers (during Autumn 2)	Send to 100 teachers	Basic frequencies; thematic analysis; compliance report	RQ1; RQ2	Compliance; fidelity; responsiveness; adaptation; usual practice
Short feedback survey 2	Survey for teachers (during Spring 1)	Send to 100 teachers	Basic frequencies; thematic analysis; compliance report; formative views	RQ1; RQ2	Compliance; fidelity; responsiveness; adaptation; usual practice; perspectives on trial methodology
Telephone interviews	Semi-structured interviews	4 schools (2 that are new to the trial, 2 that took part previously); interview the head of dept. and up to 2 teachers per school	Within-site/within-approach analysis; triangulation	RQ1; RQ2	Fidelity Implementation and support quality Usual practice Responsiveness Adaptation Cost/resource information
Samples of quiz and discussion starters and schemes of work	Document analysis	Collect from 4 telephone interview schools	Content analysis; categorisation of quiz and discussion starters matching to	RQ1; RQ2	Fidelity, adaptation, quality

Research methods	Data collection methods	Participants/ data sources (type, number)	Data analysis methods	Research questions addressed	Implementation dimensions/logic model relevance
			schemes of work		
Support emails/ ad hoc calls data (collated by EEF)	Excel worksheet	Record of emails/ ad hoc calls with schools and teachers	Summary of themes/issues raised	RQ1; RQ2	Compliance; fidelity; implementation' support

Process analysis

NFER will analyse their own recruitment, trial retention data and survey response rate data to understand how well schools and teachers responded to recruitment and whether they were engaged in the trial. NFER will analyse the teacher survey, compliance questions, and interviews (and the EEF telephone support data summary) to inform both teachers' ability to follow the trial design, and to implement the lesson starter approaches – according to the compliance, implementation fidelity, and support descriptors set out in the Implementation fidelity section above.

We will explore the nature of fidelity and any adaptations through interviews, surveys, content analysis of materials, and EEF email/ad hoc call support data (if required by the school). Views/responses from the process data will be analysed thematically – against relevant research questions (bearing IPE dimension and supporting factors in mind as appropriate to each RQ).

Cost information

During the telephone interviews with heads of department and teachers, we will ask them about the time involved in preparing their lesson starters and any other additional support they had to put in place to implement their choice. Should there be any minimal monetary costs involved, we will ask teachers about these in interviews.

Our cost evaluation will be exploratory and qualitative. As each choice is implemented for only a small number of weeks, we do not envisage calculating a cost per pupil per year over three years of implementation of such as choice.

Ethics and registration

The trial will be designed, conducted and reported to CONSORT standards (<http://www.consort-statement.org/>). This trial is set up within a feasibility study to understand and explore the best designs and methods to evaluate teacher classroom practices, therefore it will not be registered. The evaluation will be conducted in accordance with NFER's Code of Practice, available at:

http://nfer.net/media/3029/code_of_practice_final_march_2019.pdf.

Ethical agreement for participation in the trial will be provided by the school headteacher via signing a Memorandum of Understanding. Head of Science in the school will act as a key contact person for the trial. Year 8 science teachers and parents of Year 8 pupils will be provided with full details about the evaluation and will be given an opportunity to withdraw their child from data processing if they have objections to this.

Data protection

All data gathered during the trial will be held in accordance with the data protection framework created by the Data Protection Act 2018 and the General Data Protection Regulation 2016/679. It will be treated in the strictest confidence by the NFER and EEF. No individual, teacher or school will be identified in any report. NFER will share names and contact details of all school staff involved in the trial with EEF. EEF will use this data to contact the schools and support them in delivering the interventions. Once the GL Assessment's PTS13 is administered, NFER will send the pupil tests to GL Assessment for marking, who will share the pupil results with NFER. NFER will not share any pupil data with EEF but schools will be able to see their pupils' PTS13 scores on the portal set up by GL Assessment.

NFER is the data controller for this evaluation, with the EEF as the data processor for school staff contact details and GL Assessment as data processor for GL PTS13 data

The legal basis for processing personal data is covered by:

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

We have carried out a legitimate interest assessment, which demonstrates that the evaluation fulfils one of NFER's core business purposes (undertaking research, evaluation and information activities) and it has broader societal benefits. It is therefore in our legitimate interest to process and analyse personal data for the administration of the RCT. We have considered and balanced any potential impact on the data subjects' rights and find that our activities will not do the data subject any unwarranted harm. NFER and EEF will sign a Data Sharing Agreement to set out the roles and responsibilities for this trial. This will include a description of the nature of the data being collected and how it will be shared, stored, protected and reported by each party. NFER will also publish a privacy notice intended for research participants (school staff and parents).

The headteacher of participating schools will sign a Memorandum of Understanding, which collects the details of the Head of Science who becomes the key contact for the trial. During the trial, schools will also provide details of Year 8 science teachers nominated and agreed to take part in the study, and details of their Year 8 pupils (where parents have not objected to their child's data being processed).

For heads of science and participating science teachers the personal data collected is names, contact details, job role and their science lessons resources along with their attitudes and views about science lessons. For each Year 8 student, the following personal data will be collected: names, date of birth, FSM eligibility, end-of-topic test results and GL Assessment's Progress Test in Science score.

All the personal data and teachers' resources (such as quiz questions, discussion materials and schemes of work) will be transferred via NFER secure data portal. Teacher surveys will be administered via Questback. Pupil PTS13 assessments (completed on paper) will be returned to NFER using a secure courier.

All personal data will be deleted within one year of publication of the final report currently expected to be published in December 2022.

Personnel

EEF will be responsible for developing the teacher guidance and will act as a delivery provider. NFER will be the independent evaluator for the Teach Choices trials.

Table 6: Personnel

Name	Institution	Role
Robbie Coleman	EEF	Intervention design and producing guidance for teachers
Emily Yeomans, Joe Collin (in his previous role) and Igraine Rhodes	EEF	Leading on learning for EEF and programmes team
Rob Coe	EEF	Advising on evaluation design, statistical analysis and testing
Ben Weidmann	EEF	Advising on evaluation design and statistical analysis
Florentina Taylor and Arnaud Vaganay (in their previous roles) and Faizaan Sami	EEF	Leading on EEF evaluation side
Ben Styles	NFER	Trial director
Palak Roy	NFER	Trial manager
David Sims	NFER	Process evaluation director
Jennie Harland	NFER	Process evaluation researcher
Maria Galvis	NFER	Process evaluation researcher
Joana Andrade	NFER	Trial statistician
Kathryn Hurd and Kathryn Gallop	NFER	Operations lead
Priscilla Antwi	NFER	Operations researcher

Risks

Table 7: Risk assessment

Risk	Likelihood	Impact	Mitigation actions
Insufficient schools or teachers recruited to the study due to local Covid-19 lockdowns/school closures	MAY HAPPEN	SIGNIFICANT	Discuss the possibility of drawing a top-up sample; communicate with early EOI schools and speak to EEF and NFER's Communications team to tweet about the trial. Also, discuss the possibility of shifting the timetable for the whole trial in case sufficient schools are not recruited.
Local Covid-19 lockdowns/school closures	MAY HAPPEN	SIGNIFICANT	A local lockdown will have different effects depending on when it occurs. Recruitment and data collection: Schools may be unable to take part in the trial and supply the data. Approaching schools who were engaged with the first trial will ensure we have some existing schools willing to take part and aware of the trial requirements. Submission of pupil and teacher data will be a pre-requisite of randomisation. To mitigate the risks of low response rate, we have increased the sample size from 80 to 100 teachers between first and second run. Choice implementation: Severe disruption in a subset of schools. NFER's ongoing communication with schools will ensure if lesson starters are disrupted due to Covid-19, compliance records reflect this. Dropout from or insufficient IPE data: The data collection activities are adjusted from TC1 so as not having to visit participating schools and to reduce burden on teachers. If insufficient schools want to take part in telephone interviews, expand the sample of schools. Follow-up measurements: Severe disruption in a subset of schools. We would want to delay measurement until they re-open if possible. Covid-19 disruption unlikely to introduce bias.
Tight timetable for recruitment, baseline and randomisation	LIKELY	SIGNIFICANT	Discuss strategies to increase the rate at which each activity takes place including more frequent internal catch-ups and communication with EEF.
Tight analysis and reporting timetable	MAY HAPPEN	SIGNIFICANT	Discuss strategies to increase the rate at which each activity takes place including more frequent internal catch-ups and communicate to EEF. If follow-up data collection is disrupted due to Covid-19 local lockdowns/school closures, adjust analysis and reporting timetable accordingly to collect retrospective data.
School/teacher drop out from the intervention. That is, schools do not implement the assigned Teacher Choice	MAY HAPPEN	MODERATE	NFER to contact schools to inquire school status and gain feedback.
Key contact attrition (e.g. key contact becomes unresponsive or leaves the school)	UNLIKELY	VERY LOW	Establish contact with the headteacher who signed the MOU

Risk	Likelihood	Impact	Mitigation actions
during the course of the trial)			
Intervention (Teacher Choice) is not well implemented	MAY HAPPEN	MODERATE	EEF and NFER put the processes in place in terms of teacher guidance and attend to email queries received
Active control group adopts aspects of intervention / contamination	MAY HAPPEN	MODERATE	EEF and NFER put the processes in place in terms of teacher guidance and attend to email queries received
Schools do not complete the data on Implementation and fidelity; or provide responses influenced by the evaluation, the choice or by social responsibility bias	MAY HAPPEN	MODERATE	NFER to communicate to schools the importance of such data, and the importance of honesty and openness in responses (however much or little they used the approach, their experience and reasons are important).
Drop out from/ insufficient IPE data	MAY HAPPEN	MODERATE	NFER to communicate to schools the importance of such data
Loss of taped interview recordings.	UNLIKELY	VERY SEVERE	Only NFER equipment is used. Recordings deleted once notes are written up.
School's key contact details not shared securely between NFER and EEF	UNLIKELY	SIGNIFICANT	Only use EEF or NFER's secure data portal to exchange personal data

Timeline

Table 8: Timeline for 'A Winning Start' trial

Dates	Activity
May-June 2021	- Set up the trial and prepare document revisions - Draft and publish protocol
June-September 2021	- School recruitment and data collection - MoU signed - Schools upload teacher and pupil data
October 2021	- Randomisation of lesson starter approach - NFER to communicate randomisation results; send Teacher Guidance and data collection templates to participating teachers - Teachers to prepare lesson plans for Spring 1 and Spring 2 terms using the guidance document - EEF Support emails/ad hoc calls
November – December 2021 (Autumn 2 term)	- Teachers implement the choices by delivering the allocated lesson starter approach - NFER receives the data templates from schools with end-of-topic test results for each pupil (for the first topic) - Teachers complete first mini-survey, including self-report of compliance

Dates	Activity
January – mid-February 2022 (Spring1 term)	- Schools swap the lesson starter approach and deliver the choices in the second topic - NFER receives the data templates from schools with end-of-topic test results for each pupil (for the second topic) - Teachers complete second mini-survey, including self-report of compliance and overall views on trial manageability
Mid-February – March 2022	- NFER conducts telephone interviews and collects materials from four schools - NFER administers GL Assessment's PTS to Year 8 classes participating in the trial
April 2022	- Incentive payment to schools - GL Assessment's PTS marking
May-June 2022	Trial analysis
July-Aug 2022	- Learning and discussion workshop- NFER and EEF - NFER sends the first draft report to EEF
September-December 2022	Report revisions and final report publication

References

- Agarwal, P. (2019). 'Retrieval practice and bloom's taxonomy: Do students need fact knowledge before higher order learning?' *Journal of Educational Psychology*, **111**, 2, 189–209 [online]. DOI 10.1037/edu0000282.
- Andrade, J., Styles, B. and Roy, P. (2020). *A Winning Start Statistics Analysis Plan* [online]. Available:
https://educationendowmentfoundation.org.uk/public/files/Projects/A_winning_start_SAP.pdf [16 August, 2021].
- Borenstein, M., Hedges, L. V., Higgins, J. P. and Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester: John Wiley and Sons Ltd.
- Demack, S. (2019). *Does the Classroom Level Matter in the Design of Educational Trials? A Theoretical and Empirical Review* [online]. Available:
https://educationendowmentfounddation.org.uk/public/files/Publications/Does_the_classroom_level_matter.pdf [20 August, 2020].
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J. and Willingham, D. T. (2013). 'Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology', *Psychological Science in the Public Interest*, **14**, 1, 4–58 [online]. DOI 10.1177/1529100612453266.
- Education Endowment Foundation (2020a). *Teacher Choices trial: A Winning Start Evaluation Protocol* [online]. Available:
https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Teacher_Choices_Pilot_1_Trial_Protocol.pdf [16 August, 2021].
- Education Endowment Foundation (2020b). *Teacher Choices: The Story Time Trial Evaluation Protocol* [online]. Available:
https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Protocols/Trial_2_Evaluation_Protocol.pdf [16 August, 2021].
- Roediger, H. L., and Butler, A. C. (2011). 'The critical role of retrieval practice in long-term retention'. *Trends in Cognitive Sciences*, **15**, 1, 20–27 [online]. DOI 10.1016/j.tics.2010.09.003
- Rohrer, D., Dedrick, R. F., Hartwig, M. K. and Cheung, C. N. (2019). 'A randomized controlled trial of interleaved mathematics practice', *Journal of Educational Psychology*, **112**, 1, 40–52 [online]. DOI 10.1037/edu0000367.
- Rowland, C. A. (2014). 'The effect of testing versus restudy on retention: a meta-analytic review of the testing effect', *Psychological Bulletin*, **140**, 6, 1432–1463 [online]. DOI 10.1037/a0037559.
- Van Gog, T. and Sweller, J. (2015). 'Not new, but nearly forgotten: the testing effect decreases or even disappears as the complexity of learning materials increases', *Educational Psychology Review*, **27**, 247–264 [online]. DOI 10.1007/s10648-015-9310.

Appendix A: Summary changes between the original and second run of 'A Winning Start' trial

Trial activity	Original trial (TC1)	TC1 second run
Recruitment target	80 teachers	100 teachers- due to various challenges posed and uncertainty around local coronavirus outbreaks and government's response to this, with a larger sample we can be more confident about collecting sufficient follow-up data in case of school withdrawals.
Unit of randomisation	Teacher-class units (as a teacher could teach more than one participating class)	Teacher or class (as only one class per participating teacher will take part in the trial) Randomisation will not be stratified by class-level grouping practices as we will not collect this data.
Selection of classes and topics	School nominated at least two teacher-class units to the trial where some teachers taught more than one class	• Secondary state-funded schools can nominate as many teachers as they wish but at least two teachers/classes. • Only one class being taught by each participating teacher. • Chosen topics to be taught in its entirety to participating class by only one teacher (no joint-teaching). • Chosen topics must have at least six lessons to deliver the starter. • End-of-topic tests to have at least eight independent items/questions. If the test covers multiple science topics, it should have at least eight independent items/questions for the chosen trial topic.
Baseline measure	Year 7 science assessment	No baseline assessment data will be collected and hence no pre-test covariate in analyses models.
Teacher Guidance document	We consulted with non-trial teachers on EEF's teacher guidance document	Since the guidance document has been used in TC1, we will not run pre-trial consultation with teachers. We will however, update the guidance document with minor adjustments based on TC1 and on the needs of this trial, in consultation with EEF.
Trial activity logs (compliance)	Teachers were asked to fill this log to indicate the usage and length of allocated lesson starters for each science lesson	To reduce daily/weekly burden on teachers to complete activity logs, a mini survey at the end of each half-term will be used for teachers to indicate compliance to delivering allocated lesson starter.
Online teacher survey (compliance questions)	Survey also gathered teachers' self-reports on usage / compliance	
Online teacher survey (fidelity, support and trial manageability)	The survey gathered teachers' views on delivering each starter, on support, and on taking part in the trial	Administer as two mini surveys, at the end of each half-term. The first on compliance/fidelity/support with delivering starter 1; the second survey on compliance/fidelity/support with delivering starter 2 and on taking part in the trial.
End-of-topic tests	We collected item-level data to analyse psychometric properties of these tests	As this is analysed and reported in TC1, we will not burden teachers by asking them to submit item-level data. We will only collect total scores for each participating student.

Trial activity	Original trial (TC1)	TC1 second run
Other data collection	We collected test papers to explore the content validity of the tests. We collected lesson starter materials and schemes of work.	Since these were collected and reported in TC1, we will not collect end-of-topic tests. We will only collect lesson starter materials and schemes of work in telephone interview schools.
Administration of the standardised test	Although this did not take place due to nationwide school closures in March 2020, the plan was to send NFER's experienced test administrators to administer these tests. They would have visited the classrooms to administer this test on a pre-agreed date.	We are proposing to use a flexible approach to the test administration, which will allow schools to select whether they are happy for the test administrator to come into the classroom or to simply liaise with the assigned school contact, to drop off and collect the scripts.
Interview/observation data collection with schools	Case studies involving observations of lesson starters in practice, interviews with teachers and heads of department, and discussion groups with students.	TC1 has rich data from four case study schools including the student perspective. To reduce visitor burden on schools, we will instead conduct telephone interviews with staff in four schools – with a revised focus, two schools that have previously taken part in TC1 and two schools that have not.