Education Endowment Foundation

Thinking, Doing, Talking Science

Evaluation report and executive summary

December 2018

Independent evaluators:

Sami Kitmitto, Raquel González, John Mezzanote & Yongqui Chen

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education.

Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:
Danielle Mason
Head of Research

Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP
p: 020 7802 1679
e: danielle.mason@eefoundation.org.uk
w: **www.educationendowmentfoundation.org.uk**

About the evaluator

The project was independently evaluated by a team from the American Institute for Research:

Sami Kitmitto, Raquel González, John Mezzanote & Yongqui Chen.

The lead evaluator was Dr Sami Kitmitto.

Contact details:

Dr Sami Kitmitto
American Institute for Research
1000 Thomas Jefferson Street, NW
Washington, D.C. 20007

P: (202) 403-5000

E: skitmitto@air.org

Contents

Executive summary

The project

Thinking, Doing, Talking Science (TDTS) is a teacher training intervention aimed at improving Year 5 science outcomes by making science lessons more effective. Teachers are trained to develop and teach challenging, enquiry-based science lessons that incorporate more practical activities, deeper thinking and discussion, and sharply focused recording.

Following promising results from an efficacy trial of TDTS (Hanley et al., 2015), this effectiveness trial was commissioned to test its impact when implemented at scale. In this trial, science teachers received the training on four days spread across the school year, and the impact on the pupils in their science classes was measured at the end of the year.

The model of TDTS tested here differs from the previous model in three key ways:

- the number of teacher training days was reduced from five to four;
- funding for two additional days of preparation per teacher (in the form of cover costs) was cut; and
- the programme also used a ‘train the trainer’ model in which CPD was led by pairs of experts who were trained in the TDTS model by the programme authors, rather than the authors training teachers directly (trainers were delivering the training to schools for the first time during the trial).

The study used a two-arm cluster randomised controlled design; 205 schools were randomised to either receive TDTS or to be part of a ‘business as usual’ waitlist control group; 233 Year 5 teachers at 102 intervention schools received the training. The evaluation focused on the impact of TDTS on the educational attainment of pupils, as measured through a standardised science assessment.

The study also used a survey to measure pupil interest in, and self-efficacy toward, science. The process evaluation assessed whether the train-the-trainer model was successful and explored the perceived value of the programme for teachers.

The trial took place in schools between September 2016 and June 2017.

Key conclusions
1. There is no evidence that TDTS had an impact on pupils’ science knowledge attainment, on average. This result has a high security rating.
2. Among children receiving free school meals, those in TDTS schools made a small amount of additional progress compared to those in other schools. However, this finding is not statistically significant”. This means that the statistical evidence supporting the impact finding does not meet the threshold set by the evaluator to be convincing.
3. The programme led to small increases in pupil interest in science and self-efficacy for science, as measured by pupil surveys.
4. Teachers who received TDTS training reported confidence in their understanding of, and ability to apply, the strategies they had learned. They felt that those strategies required minimal extra time to implement.

EEF security rating

These findings have a high security rating. This trial was an effectiveness trial, which tested whether the intervention worked under everyday conditions in a large number of schools. The trial was a well-designed, two-armed randomised controlled trial and was well-powered.

Eleven percent of pupils who started the trial were missing the primary outcome variables or baseline KS1 test scores and so were not included in the final analysis. The pupils in TDTS schools were similar to those in the comparison schools in terms of prior attainment.

Additional findings

Although the evaluation found no evidence that TDTS had an impact on pupils' science knowledge on average, teachers in treatment schools were more likely to report that their pupils were engaged with, confident in, and made good progress in science.

Teachers in treatment schools were also 19 percentage points more likely to report confidence in adapting their teaching to engage pupils' interest, and 30 percentage points more likely to report confidence in providing challenging tasks to high-achieving pupils, than teachers in control schools. Conversely, teachers in control schools were more likely to report that they required pupils to write-up whole investigations.

The evaluation found that teachers in the treatment schools felt confident in their understanding of, and ability to apply, the TDTS strategies, and trainers also reported feeling that they were trained well by the intervention authors.

In comparing the results of this effectiveness trial with the original efficacy trial, some key considerations are that the programme was delivered differently (through a train-the-trainers model) and that this trial had one less day of CPD. In addition, teachers in the original trial were provided supply cover for two days of in-school time per teacher to plan together how to implement the TDTS strategies. All of these differences could lead to variations in the way the programme was delivered in this trial.

Cost

The average costs of the intervention for one teacher was around £1,854, or £29 per pupil per year when averaged over three years. Costs of implementation may vary, particularly if implemented at a smaller scale. The non-financial costs included four days of teacher supply time per teacher and laptops for trainers to use in the CPD sessions.

Table 1: Summary of impact on primary outcome

Outcome/ Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	*P*-value	EEF cost rating
Science	0.01 (-0.08; 0.1)	0	4 of 5 padlocks	7,806	0.791	£ (1 of 5)
Science (FSM)	0.05 (-0.07; 0.18)	1	N/A	2,000	0.386	£ (1 of 5)

Introduction

Background evidence

Thinking, Doing, Talking Science (TDTS) is an intervention aimed at providing teachers with skills and strategies to develop challenging enquiry-based science lessons that incorporate more practical activities, deeper thinking and discussion, and more focused recording. The hypothesis is that the TDTS professional development programme will support teachers in lesson planning so that they can better develop engaging lessons with a clear focus on encouraging higher order thinking among pupils. This would then improve teacher self-efficacy and teaching practices, which in turn would improve pupil engagement, content knowledge, and enquiry skills.

A 2013 Office for Standards in Education, Children's Services and Skills (Ofsted) report on science education found that the best science teaching in primary schools included scientific enquiry as a central focus of teaching and had sustaining pupil interest in science as a goal (Ofsted, 2013). The report's findings emphasised the importance of continuing professional development (CPD) for teachers of primary science, but found that the provision of such CPD was low.

The research literature recommends CPD for teachers to build skills in the teaching of science, particularly in teaching pupils how to conduct investigations (Murphy and Beggs, 2005). A synthesis of research on primary school science interventions (Slavin et al., 2012) concluded that interventions focused on supporting teacher instruction showed potential for improving science learning. Although enquiry-based interventions that involved a specific curriculum, sets of tools, or "kits" for instruction did not have significant effects, interventions that supported teachers through CPD focused on enquiry-based teaching did.[1]

One intervention featured in a widely used synthesis on science teaching was the direct precursor to TDTS (Mant et al., 2007). In this study, two teachers of Year 6 pupils in 16 schools (selected for treatment out of a matched pair with the non-selected school from the pair acting as a control school) were provided eight days and four twilight sessions of CPD dispersed throughout the school year. The CPD sessions were intended to help teachers design 'science lessons that had more practical work, more discussion, more thinking and less (but more focused) writing' (Mant et al., 2007). The national science test was used as the pre- and post-intervention outcome measure. The study found that the proportion of pupils scoring level 5, the highest level, in treatment schools increased 10% relative to the increase in control schools.

As a more formalised programme, the potential of TDTS as a promising CPD-based intervention was established in a peer-reviewed, 41-school efficacy trial funded by the EEF (Hanley et al., 2015). In that study, the programme was administered by programme developers to teachers of Year 5 pupils. Two teachers from each treatment school participated in the programme which was administered as five days of CPD spaced throughout the school year. The authors found promising positive impacts of the programme, including a positive impact on pupil science knowledge attainment (effect size +0.22) and on various measures of pupil attitudes towards science (no effect sizes reported). The Hanley et al. study had two limitations that the effectiveness trial reported here addressed: (a) the close reliance on developer involvement in teacher training and (b) the lack of statistical power to establish the statistical significance of important subgroup differences in impacts.

Intervention

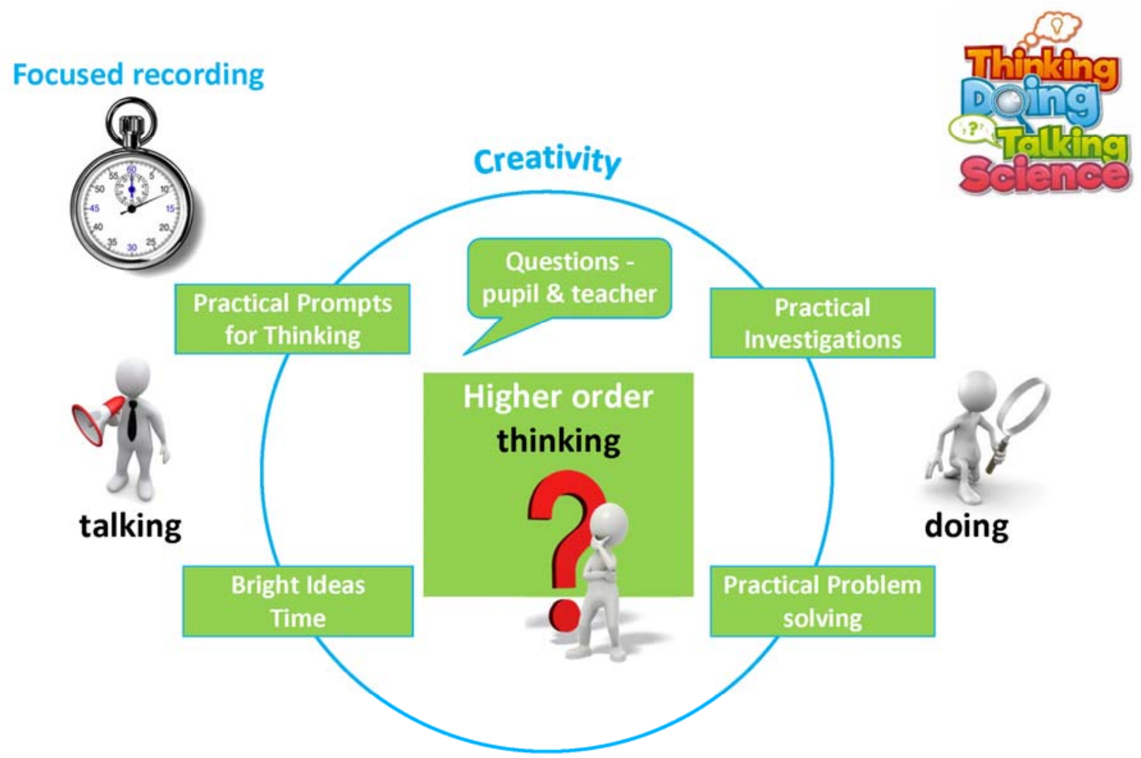
TDTS is designed to build the capacity of teachers to create and deliver creative, challenging, and engaging science lessons. The intervention trains teachers to incorporate strategies that are easy to

[1]Interventions that helped teachers incorporate technology also were found to have significant benefits.

plan and implement and are overtly focused on encouraging pupils' higher order thinking. The strategies are intended to help teachers spend more time in class encouraging pupils to talk about scientific concepts. The strategies seek to increase the time within their normal science lessons dedicated for discussion, problem solving, and investigation activities while streamlining and focusing pupils' time spent recording.

The central approach of TDTS is summarised in a conceptual graphic (Figure 1). This graphic acted as a point of reference when training trainers and teachers on the programme. It illustrates the 'talking' and 'doing' strategies that are intended to support higher order thinking. 'Practical Prompts for Thinking' are short, accessible teacher demonstrations that are designed to intrigue pupils and also act as discussion starters within science lessons. 'Bright Ideas Time' are slots of ten minutes or less dedicated to discussion of a prompt which is often a question with no correct answer that allows pupils to think deeply and creatively. An important aspect of TDTS is that when pupils conduct practical work (in the 'Practical Investigations' and 'Practical Problem Solving') they do not record everything about the activity but only what is necessary to demonstrate their progress towards the particular objective of the activity which might be 'to present the results in an appropriate manner', 'to plan a fair test', or 'to make a prediction with a scientific reason'.

Figure 1: Graphic of the TDTS approach



Source: TDTS authors.

In this trial, TDTS was delivered to 233 Year 5 teachers (see section on Participant Selection) through 4 CPD days (one fewer than the efficacy trial)[2] staggered throughout the school year and led by pairs of experts who were trained in the TDTS model by the programme developers. This train-the-trainer model for delivery was the key adaptation made in this effectiveness trial, allowing the intervention to be implemented in 102 schools (with an additional 98 schools used as a waitlist control group) compared with 21 schools (with an additional 20 school used as a waitlist control group) in the earlier efficacy trial (Hanley et al., 2015). The train-the-trainer approach was adopted as a model for how the programme might be implemented at scale in numerous schools.

For the delivery of TDTS in this evaluation, the programme developers (the 'Oxford team') recruited seven pairs of trainers (14 total) from local organisations in the following seven regions:

- Bath
- Dorset
- Hampshire
- Lancashire
- Lincolnshire
- London
- Teesside.

The Oxford team identified trainers for their expertise in primary science as well as their connections to schools in their local regions to facilitate recruitment for the intervention.

- All trainers had experience in either leading CPD sessions or coaching and mentoring teachers (12 of 14 had experience in both).
- Nine of the 14 trainers had a postgraduate certificate in education, a master's degree, or a doctorate.
- Nine of the 14 trainers had a BA or BS in science.
- Twelve of the 14 trainers had been teachers, and all 12 had experience teaching science in a primary classroom.

To prepare the trainers for the delivery of the TDTS programme, the Oxford team worked with the trainers during four two-day, face-to-face sessions (eight days total for each trainer). These sessions were staggered throughout the year and preceded each teacher CPD session. They began in April prior to the intervention year, and subsequent sessions were timed to occur just prior to the delivery of TDTS CPD to teachers. The first train-the-trainers session included a briefing on the evaluation and details on the process for the trainers to recruit schools from their home regions into the study.

In each train-the-trainer session, the Oxford team modelled the delivery of the upcoming CPD session, including engagement of the trainers in the science investigations and activities in the manner that they were to be used with teachers. Trainers were provided with, and trained on, materials for the upcoming CPD session: an agenda, presentation slides, trainer notes, and science investigation materials and instructions. The sessions were highly interactive, and the Oxford team leveraged the experience and knowledge of the trainers by soliciting feedback and applying that feedback in subsequent days of training. For example, the trainers encouraged the Oxford team to make explicit how each practical that was shared with the teachers linked to the key programme elements (illustrated in Figure 1). Discussions with the trainers also included thoughts on what to prioritise from the content of the latest national curriculum and how to infuse that content into the design of the CPD days. The intervention ethos and strategies remained the same, but the Oxford team updated the materials to be used with teachers based on feedback from the trainers.

[2] The programme was reduced by one day for two reasons. First, the programme authors felt that a five-day course presented a high burden for schools given that science often is not seen as having as high a priority as English and maths because it is not tested in the Year 6 SATs. Second, the fifth and final day of training in the efficacy trial was so late in the year that it would have little opportunity to impact teacher practice or pupil outcomes.

Face-to-face CPD was provided at locations central to the teachers in each study region. These locations included teacher training facilities in universities and hotel conference suites. The aim was to use high quality, professional facilities that could provide good catering, IT/AV equipment, and ample room for teachers to have discussions and practise investigations. Laboratory equipment was not needed, but the activities required, for example, room to drop an egg from a significant height and to launch a glider with some open space.

On average, each CPD session included approximately 30 teachers and were led by two trainers. Sessions were highly interactive and practical in nature, requiring equipment to be set out and cleared away throughout the day. Two trainers were used to maximise interactions with the teachers, to help enable all teachers to participate in the practical investigations, and to minimise time lost due to logistical transitions between sections of the session. In addition, the intervention team requested that at least two teachers from each school participate in the CPD sessions to facilitate ongoing collaboration and mutual support. The ethos the trainers sought to create with the teachers was that of a team working together to find the best possible practice. Because the sessions took the entire day, scheduled in term, and were co-ordinated across multiple schools, supply cover was necessary as opposed to using inset days or evening sessions.

In the CPD sessions, teachers were provided with a folder of reference materials to use when they returned to their classrooms and to assist with lesson planning. These materials included an agenda, the background to the strategies, exemplar teaching materials for each strategy, slides that could be used in the classroom, and instructions for all the investigations and problem-solving activities. All materials in the folder were available online through the programme’s secure web portal.

The CPD days were staggered throughout the school year and the training systematically introduced a range of strategies with the intention that teachers would then incorporate the gained knowledge and skills in the classroom. Teachers left each session with a gap task to try out and evaluate with their own classes prior to the next CPD session. The gap tasks were structured so that it would be straightforward for teachers to implement a TDTS strategy into their own teaching. In the following CPD session, teachers discussed how the gap task went with their pupils, reflected on their practice, and shared their ideas for implementing the strategies.

The first gap task was designed to be very straightforward and achievable to implement so that teachers could easily get started incorporating the strategies and quickly see the difference that higher-order thinking approaches make. After Day 1, the gap task was to use a Bright Ideas Time discussion activity called the Odd One Out (OOO) with their pupils. In OOO, the teacher presents three or four items and pupils are asked to choose one that they think is the odd one out and explain their reasoning. Teachers were asked to record the scenario that they presented to pupils, the pupil responses, and reflections on how it worked (for example, how it impacted learning, engagement, or pupil management). As the programme progressed, the gap tasks become more demanding in parallel with growing teacher confidence in the strategies. After Day 3, the gap task was to attempt both a Practical Prompt for Thinking and to provide feedback on a practical investigation with pupils.

The TDTS programme is designed to be easily incorporated into the lives of busy teachers. As such, trainers contacted teachers between sessions only to send a reminder about the gap tasks. There was no intention that trainers would act as coaches or mentors between sessions.

Evaluation objectives

The main research question for the evaluation was as follows:[3]

[3] The statistical analysis plan (SAP) and evaluation protocol can be found on the effectiveness trial website: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/thinking-doing-talking-science-effectiveness-trial/

When implemented at scale, what is the impact of the Thinking, Doing, Talking Science (TDTS) programme on the (a) science knowledge attainment and (b) attitudes toward science of participating pupils at the end of the school year?

This research question focused on two broad outcomes: pupil science knowledge attainment and attitudes toward science. TDTS is intended to help teachers create more engaging lessons that would result in increased pupil interest in, and self-efficacy toward, science. In addition, TDTS is focused on boosting science learning and enquiry skills. Hence, the ultimate outcome would be a meaningful increase in pupils' knowledge of science.

Although the evaluation's primary outcomes concerned the programme's impact on pupils, as an effectiveness trial the process evaluation examined:

How is Thinking, Doing, Talking Science implemented at scale?

Because the train-the-trainers model was a key part of implementing TDTS at scale, it was important to assess the success of that transfer of knowledge. This research question focused on trainer and teacher self-reported comprehension of, and confidence in, the TDTS model.

Additionally, TDTS works by changing the way in which teachers approach lesson planning and teaching. As such, the process evaluation also asked:

Were there any differences in teacher practice between those whose school was assigned to the control group and those whose school was assigned to the intervention group?

Ethics and trial registration

The American Institutes for Research (AIR) obtained approval for the study on behalf of the entire evaluation team from its Institutional Review Board (IRB), which is registered as a research institution (IORG0000260) with the U.S. Office of Human Research Protection at the U.S. Department of Health and Human Services. With approval of this IRB, AIR is authorised to conduct research under its own Federalwide Assurance (FWA00003952).

For this evaluation, AIR used opt-out consent, in which families received information about the study and instructions on how to opt out of the study. Sensitive personal data were not collected. Therefore, per interpretation of the Education Endowment Foundation (EEF), the opt-out procedures correspond with Section 2, Conditions 1 and 6, of the Data Protection Act.

The study was registered, and the International Standard Randomised Controlled Trial Number is ISRCTN22499525.

Data protection

Pupil data from schools, from the science assessment, and from the NPD was collected and securely housed by NatCen. At recruitment, schools signed an MOU (Appendix C) by which they agreed to provide a list of their Year 4 pupils to NatCen and send a notice to the parents of all those pupils that included an opt-out form (Appendix D). All parents who opted-out by returning this form would not have their children's data included the study.

A study website was also set up and links included in correspondence with schools and parents. This provided information about how data would be used and processed in the study (Appendix E). The website had a link to NatCen's privacy notice (**http://natcen.ac.uk/help/privacy/**).

Because the AIR evaluation team was located outside of the U.K. and E.U., extra precautions were taken to keep data secure. NatCen anonymised pupil data prior to sharing with AIR by removing pupil and school identifiers. Additionally, NatCen aggregated sensitive data such as IDACI rank of schools and pupils to coarser levels than provided by the NPD. Prior to receiving pupil data from NatCen, AIR destroyed all data files with school names and identifiers (such as files used for randomization) and removed any identifiers from data collected directly by it, such as the trainer and teacher evaluation forms and teacher surveys. Original AIR data files were archived with NatCen in the event they needed to be referenced.

Project team

TDTS Developers Project Team (The Oxford Team)

The Oxford team included the experts who developed the TDTS programme, trained the trainers, worked with the trainers to make updates to the programme and materials, and managed the overall implementation of delivery from developers to trainers to teachers:

Bridget Holligan—Science Oxford: project leader;
Helen Wilson—School of Education, Oxford Brookes University: project leader; and
Catherine Aldridge—project manager.

TDTS Trainers

The trainers were trained on the TDTS programme by the Oxford team. They were responsible for the recruitment of, and liaison with, the schools in their region and delivered the TDTS CPD to teachers in participating schools.

Allie Beaumont and Sarah Earle—Bath Spa University (Bath region);
Caroline Galpin and Stuart Twiss—Early Years Science (Dorset region);
Julie Reynolds and Caroline Whittaker—Mathematics and Science Learning Centre, University of Southampton (Hampshire region);
Ruth Perkins and David Price—Science Made Simple (Lancashire region);
Joy Parvin and Jane Winter—Centre for Industry Education Collaboration, University of York (Lincolnshire region);
Esme Glauert and Jill Trevethan—UCL Institute of Education (London region); and
Jenny Harvey and Nicky Waller—Centre for Industry Education Collaboration, University of York (Teesside region).

Evaluation design and analysis partner: American Institutes for Research

AIR was the lead evaluator responsible for the overall trial design, instrument development, administering electronic data collection instruments (trainer evaluation forms, teacher evaluation forms, and the end-of-year survey to teachers in treatment and control schools), all data analysis, and authoring this evaluation report.
Sami Kitmitto, PhD—principal investigator;
Raquel González, PhD—project director and head of qualitative data;
John Mezzanote—programmer; and
Yongqui Chen—programmer.

Evaluation data collection partner: NatCen

NatCen was the evaluation partner responsible for collecting data and agreements from schools at the time of recruitment, administering the pupil assessment and survey (scheduling with schools, managing invigilators, scoring the pupil assessment, entering pupil responses into an electronic database), requesting data from the NPD, and anonymising all data they collected for sharing with AIR for analysis.
Rakhee Patel—project director;
Lesley Mullender—head of logistics;

Bryan Mason—jead of data; and
Michael Lumpkin—research administrator;
Claire Jones—field project manager;
Mags Anderson—senior researcher;
Tom Chadwick—researcher; and
Migle Aleksejunaite—data manager
Alessio Fiaccio—programmer.

Methods

Trial design

This study was a two-arm, blocked cluster randomised controlled trial, with treatment assignment at the school level within regions (blocks; see Table 2). Because the intervention was delivered to teachers, randomisation was conducted at the school level rather than the teacher or classroom level to prevent spillover (contamination) of the intervention from treatment to control teachers who might be in the same school and enable treatment group teachers to collaborate with their colleagues after receiving CPD. Schools were randomised within regions (blocks) to align with the train-the-trainers model, where trainers were responsible for recruiting schools and delivering TDTS to approximately half the schools in the first year (treatment schools) and the remaining schools in the second year (waitlist control schools).

Table 2: Summary of trial design

Trial type and number of arms		Blocked cluster randomised controlled trial with two arms.
Unit of assignment		Schools.
Variables used for treatment/control assignment		The percentage of pupils eligible for free school meals (FSM) in three categories: low (9% FSM or less), medium (between 9% and 24% FSM), and high (greater than 24% FSM). Number of Year 5 teachers in two categories: one or two and three or more.
Primary outcome	variable	Science assessment score.
	measure (instrument, scale)	Study-delivered science assessment with scores ranging from 0 to 41 points.
Secondary outcome(s)	variable(s)	Interest in science index; self-efficacy for science index.
	measure(s) (instrument, scale)	Study delivered pupil survey. Each index is a sum of responses: 13 items for the interest in science index and three items for the self-efficacy for science index.

Participant selection

The Oxford team began recruiting the TDTS trainers from different areas of the country in October 2015. As noted earlier, the trainers were primary science specialists who also were knowledgeable of the schools and systems in their respective areas.

With oversight and guidance from the Oxford team, the trainers were responsible for recruiting schools in their area. Recruitment took place from April to June 2016. Trainers were instructed to recruit a collection of schools that were broadly representative of their areas but with a special focus on schools whose proportion of pupils receiving FSM was higher than average. Trainers used their connections and previous knowledge of schools in their area to recruit into the study. Over 2000 schools were sent recruitment materials by mail. Additionally, using information about the schools, 380 were selected for personalised follow-up by the trainers to target schools with a high percentage of FSM pupils, low pupil attainment, and/or a low OfSTED rating. Along with the TDTS developers, the trainers held local recruitment conferences where teachers and senior leaders could find out about the project and ask questions

For both treatment and control schools, all Year 5 teachers were requested to participate in the study. In addition, the delivery of TDTS required that at least two teachers from each school participate. When a one form entry school was recruited into the study, then an additional teacher was identified to participate, including a science co-ordinator or the head of Key Stage. All pupils of the selected Year 5 teachers were eligible to participate in the evaluation.

To participate in both the programme and the evaluation, schools agreed to:

- be randomly assigned to experience the TDTS CPD programme in either the 2016/2017 school year or the 2017–18 school year;
- complete a brief school survey describing professional development opportunities available to Year 5 teachers before randomisation;
- allow all Year 5 teachers (a minimum of two teachers) to participate in the study, participate in the programme when it was offered, and use their newfound skills from the programme in their classrooms;
- provide contact information (teacher email addresses) for all participating Year 5 teachers to allow the evaluation team to send a survey link directly to the teachers;
- allow and encourage teachers in the study to participate in a brief online, end-of-year teacher survey about their practices;
- send an opt-out consent form to the parents of all eligible pupils in the study (that is, all Year 4 pupils at the time of randomisation, the year prior to implementation) and record and report to NatCen any opt-outs received so that the data for these pupils would not be passed to the evaluation team;
- provide pupil identification information (name, unique pupil number, date of birth) for all pupils in the study to the evaluation team; and
- allow the administration of a science assessment and pupil survey (jointly administered) by the evaluation team to all pupils in the study at the end of the 2016/2017 school year.

Outcome measures

The primary outcome measure—science knowledge attainment—was measured with a science assessment (Appendix F) administered as part of the evaluation by NatCen field staff trained in data collection. This study used the science assessment and scoring guide developed by a team from the Institute for Effective Education (IEE) at the University of York, which also conducted the earlier efficacy trial of TDTS. The items in the assessment were drawn from a larger bank of items developed in 2001 by Terry Russell and Linda McGuigan. The assessment included items that addressed the science curriculum content appropriate for the year group and represented a range of topics.[4] In analysis, we used raw scores that had a possible range from 0 to 41. The assessment was scored by independent contractors hired by NatCen, supervised under the direction of NatCen's Head of Data.

The secondary outcomes—interest in and self-efficacy toward science—were measured with a pupil survey administered at the same time as the science assessment (Appendix G). As with the science assessment, this study used the pupil survey developed for the efficacy trial of TDTS.[5] That survey was adapted from a questionnaire developed in 2007 by Kind, Jones, and Barmby. The items on the survey used a five-point Likert scale ranging from 1 ('agree a lot') to 5 ('disagree a lot').

For evaluation purposes, it was important to summarise the items into meaningful constructs. *A priori*, we interpreted these items to fall into the domains—interest in and self-efficacy toward science—and constructed indices of each to use as an outcome variable. Interest in science items included statements such as 'I look forward to my science lessons', 'I would like to do more science at school', and 'science is fun'. The interest in science variable included 13 items total. The self-efficacy index included three items: 'I find science difficult to understand', 'I am just not good at science', and 'I understand everything in my science lessons'. After reverse coding negative items and adding responses for each pupil across items, the indices were standardised across pupils to have a mean of

[4] AIR obtained permission to use this pupil science assessment from the developers and efficacy trial authors, IEE.
[5] AIR obtained permission to use this pupil survey from IEE.

0 and standard deviation of 1. Results from the exploratory and confirmatory factor analysis supporting the creation of these indices is provided in Appendix H.

Sample size

The choice of model selected for this study was a mixed multilevel model with three levels (pupil, school, and region) and treatment at the second level (school level). This choice was based on the following parameters for the study:

- schools would be recruited within regions;
- schools would be the unit of randomisation; and
- pupils would be the level of observation.

To establish the necessary sample size for the study, we conducted statistical power calculations using the PowerUp tool (Version: 22/01/2015). The following parameters were used in the power calculations:

- alpha level (α) = 0.05, two-tailed test, with power (1-β) = 0.80;
- interclass correlation (ICC) = 0.12;
- proportion of Level 2 units randomised to treatment = 0.50;
- number of school-level covariates = 2;
- number of pupils per school: 32 pupils, of which six would be FSM pupils;
- proportion of pupil variance explained by pupil covariates ($R^2{}_1$) = 0.30; and
- proportion of school variance explained by school covariates ($R^2{}_2$) = 0.40.

The alpha and power parameters are standard values used in designing randomised controlled trials. The interclass correlation, number of pupils per school, and number of FSM pupils per school were set based on findings from the prior efficacy trial of TDTS (Hanley et al., 2015). The proportion of pupil variance explained by pupil covariates were set conservatively. One benchmark used to set these values was the *Variance Almanac of Academic Achievement* (https://arcdata.uchicago.edu/), which reports that past studies found that using U.S. Grade 4 (age 9–10) data with a pre-test (one school-level and one pupil-level variable) 48% of pupil-level and 68% of school-level variation is explained.

Our goal was to choose a sample size to obtain a minimum detectable effect size (MDES) of 0.18 or less. We assumed that 30 schools would be recruited within each of six planned regions, and, on average, each school would have 32 Year 5 pupils of which six would be FSM eligible. Using these conservative parameters listed previously, six regions would generate an MDES of 0.176 for FSM pupils and 0.127 for all pupils.

In the recruitment stage, the sample sizes were larger than anticipated. First, the Oxford team was able to add an additional region and two trainers to the trial. Across the seven regions in the trial, 205 schools were recruited into the study. Second, the average number of pupils per school was larger than anticipated. On average, the schools had 44 Year 5 pupils, of which 11 were FSM eligible.[6]

At the time of analysis, five schools (2.4% of the schools; four treatment and one control) had dropped out of the study. However, these were small schools, and only 1.1% of the pupils were lost to follow-up. Final sample sizes and the resulting statistical power at each stage are discussed further below in the Impact Evaluation section.

Randomisation

Assignment to the treatment and control groups occurred at the school level and was conducted separately within each study region using statistical minimisation methods to improve the balance in

[6] This study used the NPD variable KS2_FSM6 to identify pupils who were ever FSM eligible.

background characteristics across the research groups. The minimisation included two school characteristics that were provided by the school at the time of recruitment:

- percentage of pupils eligible for FSM in three categories that were derived empirically from the data to create equally sized groups:
    - low (9% FSM or less);
    - medium (between 9% and 24% FSM); and
    - high (greater than 24% FSM); and
- number of Year 5 teachers in two categories:
    - one or two; or
    - three or more.

The MinimPY programme was used for school assignment.[7]

In total, 205 schools were recruited into the study, and in July 2016, 106 schools were assigned to the treatment group and 99 schools to the control group. The minimisation process often results in unequal numbers assigned to the treatment and control groups because the process probabilistically assigns each school one at a time.

The baseline measure of pupil achievement for this study was academic attainment at the end of Key Stage 1 for pupils in participating schools, as measured by the end of Key Stage 1 assessment. This measure was not available at the time of school assignment.

Statistical analysis

Intent-to-treat analysis

The intent-to-treat analysis of primary and secondary pupil outcome measures was conducted using a mixed multilevel regression model. The sample analysed included all pupils and teachers with non-missing data (see the Missing Data section) regardless of the level of participation of teachers (see the Implementation and Process Evaluation section for information on teacher attendance). To reflect the nested nature of the data of pupils located within schools, the intercept was specified to vary randomly across schools. The primary outcome model for science knowledge attainment included:

- an intervention indicator at the school level; and
- covariates for the reading/writing and mathematics KS1 score (dummy coded for the KS1 levels present in the data).

The secondary outcome models for interest in, and self-efficacy for, science included only the intervention indicator. This choice was made to keep the model parsimonious and eliminate covariates lacking a strong theoretical justification for their impact on the outcomes.

Analysis was conducted using Stata 15 and its MIXED command.

Missing data

Because more than 10% of the pupil observations analysed were missing the primary outcome variables or the baseline KS1 test scores, the pattern of missingness was investigated (Appendix I). To check the robustness of results to missing data, a fully conditional multiple imputation approach was used to complete cases where information was missing and primary models were estimated using imputed data. Only pupils in schools that did not drop out were included in the imputations and analysis. Of the pupils in schools that did not drop out, 9.6% were not given the assessment (see the Attrition

[7] In MinimPY, the 'biased coin' method was used with a base probability of 0.75.

section) and, hence, were missing the outcome measure. Also, 3.1% of the pupils were missing the pre-test KS1 measures. Only 0.07% were missing both post- and pre-tests.

Subgroup analysis

Subgroup analysis was conducted for the population of FSM pupils using the 'ever FSM eligible' variable (KS2_FSM6). For this analysis, the primary and secondary outcome analysis models were estimated using data limited to this sample. Furthermore, an interaction model was run on the entire data. This interaction model mirrored the primary analysis model but also included an indicator for 'ever FSM eligible' and the interaction between 'ever FSM eligible' and the treatment indicator as covariates.

Additional analyses

Additional analyses were conducted where the primary impact model was expanded to include additional covariates. First, the school-level variables used in treatment/control assignment as covariates: the school percentage FSM eligible category, the number of Year 5 teachers category, and school region indicators. Second, the model was further expanded to include pupil covariates: gender, age, ever FSM eligible, and IDACI category. Given potential imbalance of the pupil sample, these models were of interest as a gauge of the potential impact of any imbalance on results.

Effect sizes

Effect sizes were calculated using the unadjusted pooled variance of the outcome. The numerator was the coefficient on the school treatment indicator (a regression-adjusted impact estimate), and the denominator was calculated using the observations included in the regression according to the following formula (Education Endowment Foundation, 2018):

$$s^* = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Compliance analysis

A Complier Average Causal Effect (CACE) analysis of TDTS was conducted. For this evaluation, the Oxford team defined an appropriate minimum level of exposure to designate 'compliers' as schools that sent at least one teacher to at least three of the four sessions. The rationale for this exposure level was that if at least one teacher attended a session, he or she would be able to share knowledge from the session with colleagues who could then implement the programme.

The CACE estimates were obtained for the outcomes using an instrumental variables (IV) approach with treatment/control assignment predicting compliance in the first stage (Huang, 2018). The first stage of the two-stage least squares IV approach estimates the extent to which assignment to the treatment group predicts compliance. The second stage estimates the impact of compliance on the outcomes using the predicted values from the first stage as an instrument for compliance. This approach is recommended by EEF guidance (Education Endowment Foundation, 2018).

Implementation and process evaluation

To determine how well TDTS was implemented at scale and assess implementation fidelity to the programme model, we collected information from the Oxford team, trainers, teachers, and schools.

First, prior to treatment/control assignment and at the time of recruitment, schools were asked questions about their past offerings (2015/2016 school year) and future plans (2016/2017 school year) for CPD, and specifically CPD in science, for Year 5 teachers. This information was used to describe what 'business as usual' was in the study schools. The information was collected by NatCen as part of the initial data collection from schools at the time of recruitment.

Second, attendance records of teacher attendance at the CPD sessions were retained by the trainers and centralised by the Oxford team for the evaluation. These attendance records indicated the extent to which teachers participated in the TDTS programme and act as a measure of successful implementation. The attendance records were used to generate a 'compliance' measures which was used in the compliance analysis described in the Statistical Analysis section.

Third, given the train-the-trainers mode of delivery, it was important to gather information about the perceived effectiveness of the training at both levels: the training of trainers and the eventual CPD delivery by those trainers. Evaluation forms were developed for each type of session and delivered to trainers and teachers via an online survey platform. At the end of the first and last sessions, trainers were asked about the quality of the sessions and their confidence in their ability to deliver the TDTS programme. This information provided a measure of successful implementation of the train-the-trainers model. Trainers were asked to fill out the evaluation form by the Oxford team. The anonymous evaluation form included two clearly marked sections: one to be used by the evaluation team and one to provide direct feedback to the Oxford team to improve future sessions.

At the end of each CPD session, teachers were asked about the quality of their CPD sessions, the usefulness of the materials provided, and their confidence in their ability to implement new lesson plans in the classroom. This information is used in the evaluation to provide measures of the fidelity of implementation of the train-the-trainers model. Teachers were also asked about their perceptions of the impact of TDTS on their teaching, pupil engagement, and pupil learning. Reponses to these questions were used as measures of the perceived outcomes of the intervention. The trainers instructed the teachers on how to complete the online evaluation form.

Fourth, we gathered information on teacher attitudes and practice in the treatment and control schools through an end-of-year online survey. Although teachers may come out of CPD sessions feeling that they have learned useful skills, it was important to establish whether this translated into a shift in teaching practices and strategies. We used a teacher survey as the main data source for these teacher-level outcomes, using responses from teachers in the control schools to establish a counterfactual. The survey covered numerous questions about teacher attitudes towards, self-efficacy for, and practices used in teaching science. The teacher survey was administered via an online survey platform. At the time of administration of the pupil assessment and survey in schools, the teachers were provided a flyer with a link to the online survey by the visiting invigilator from NatCen. AIR followed up with reminder emails to the teachers with a link to the survey.

The teacher survey used in this evaluation was comprised mostly of items from the teacher survey used in the efficacy trial (Hanley et al., 2015) but additional content was borrowed (with permission from IEA) from the Trends in International Mathematics and Science Study (TIMSS) Fourth Grade Teacher Questionnaire (TIMSS 2015) and Science Teacher Questionnaire (TIMSS 1998).

For analysis, responses were transformed into binary indicators (for example, an indicator for those who responded 'agree slightly' or 'agree strongly') and the percentages among teachers in intervention and control schools were reported and compared using a chi-squared test. Prior to the delivery, the evaluation team asked the Oxford team to indicate, based on their theory of action for TDTS, where one would expect to see differences between teachers in treatment and control schools. As part of our analysis, we compared their expectations to our pattern of significant findings.

Costs

Cost information was collected through documentation by the Oxford team and conversation with the AIR evaluation team to determine how to categorise costs (that is, distinguish between start-up and steady-state running costs).

The TDTS programme as delivered in this evaluation followed a train-the-trainers model. This involved two levels of start-up costs: one for the trainers and one for the teachers. To properly distribute the fixed costs of an intervention across time, we assumed that a trained teacher would stay at a school for three years. In this train-the-trainers model, we extended this assumption to the fixed cost of developing trainers by assuming that once they were trained on the TDTS programme, those fixed costs could be spread across three annual cohorts of teachers.

Timeline

Table 3 summarises the key dates related to the delivery of the TDTS programme in this evaluation, including the training of trainers and the provision of CPD, as well as the evaluation, including recruitment dates, when treatment was assigned, and when the final pupil assessment and survey were administered. The study protocol anticipated that the pupil assessment and teacher survey would be administered in April, however this was revised to June to align with regular field administration of end-of-year assessments by NatCen. This additionally allowed time for the final TDTS CPD session to be scheduled and for teachers to incorporate the CPD into their teaching.

Table 3: Timeline

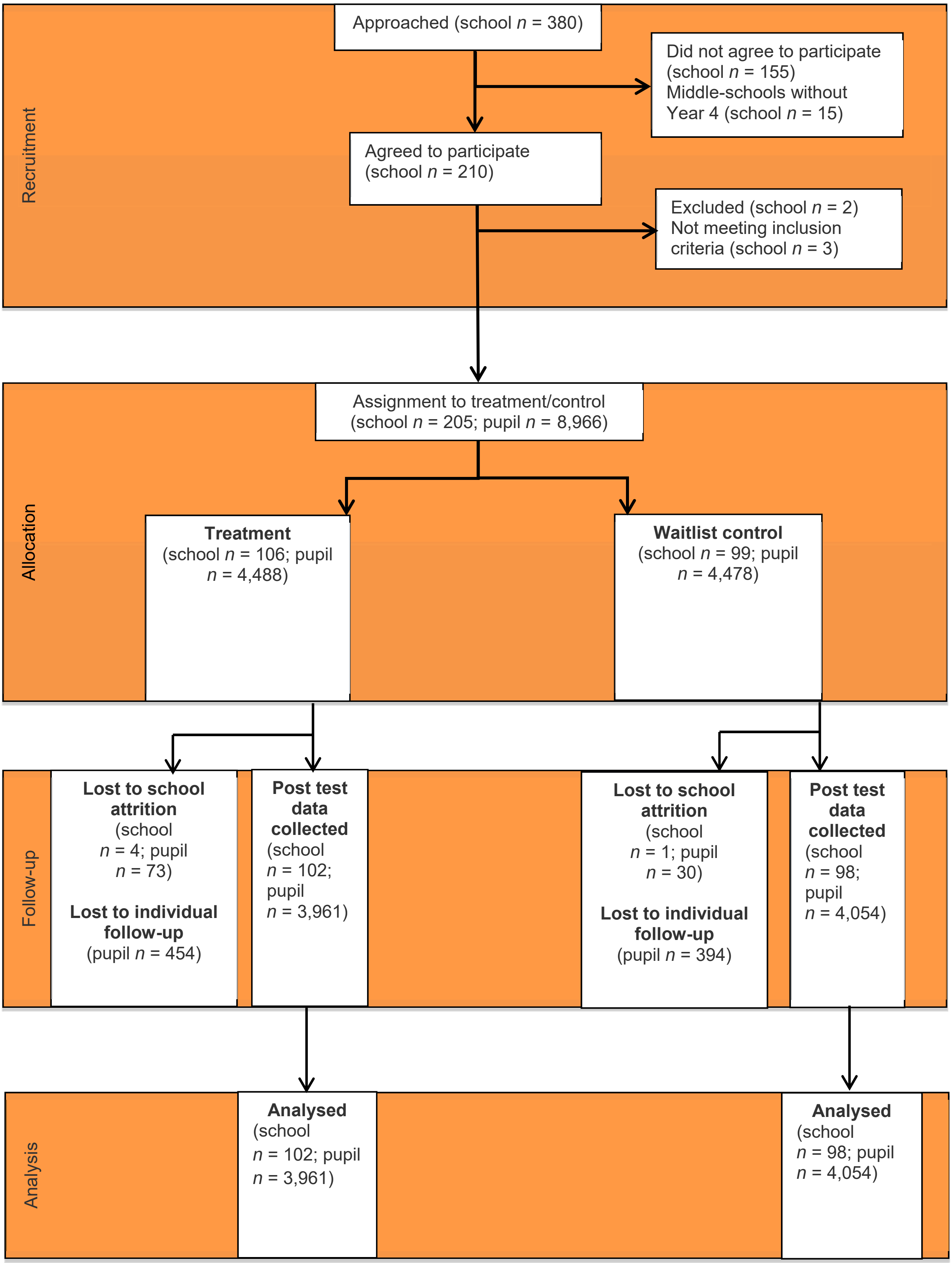
Date	Activity
Winter 2016	**Start of recruitment of trainers by Oxford team:** the number of regions/pairs of trainers was expanded from six to seven.
April 2016	**First train-the-trainers session:** trainers were introduced to the evaluation and provided instructions to start school recruitment.
July 2016	**School recruitment closed:** schools signed a memorandum of understanding, provided lists of Year 4 pupils, and sent parents of the pupils the opt-out form. **Assignment to treatment or waitlist control groups:** using minimisation, schools were assigned to treatment and control in three batches to allow schools the maximum amount of time allowable to meet the administrative requirements of study participation.
September/ October 2016	First CPD session.
November 2016	Second CPD session.
February 2017	Third CPD session.
April/May 2017	Fourth CPD session.
June 2017	Pupil science assessment and survey administered in schools by NatCen; teachers directed to online, end-of-year survey administered by AIR.

Impact evaluation

Participant flow including losses and exclusions

Figure 2 shows the flow of schools and pupils into and out of the study. At the time of assignment to the treatment or waitlist control groups, 205 schools and 8,996 pupils were included in the study. Four treatment schools and one control school were lost to follow-up. Of the four lost treatment schools, two dropped out before the TDTS CPD sessions were started. One treatment school dropped out because it was closed after assignment due to low enrolment. Two treatment schools dropped out due to staff changes and one due to a decision by the head that science was not a priority. The control school that dropped out did so due to budget concerns and staff changes. Although the evaluation lost 2.4% of schools (five of 205) after assignment, the lost schools represented only 1.1% of the pupils because they were small schools compared with the schools remaining in the sample. Of those pupils not lost to school attrition, 454 of the 4,415 pupils in treatment schools and 394 of the 4,448 pupils in control schools were not included in any of our primary analysis due to missing data. See the Attrition section for additional discussion.

Figure 2: Participant flow diagram



The statistical power at each stage of the study is given in Table 4. Power calculations in the original trial protocol and at the time of school assignment to treatment/control reflect a model that was later altered, as described in the statistical analysis plan (SAP), to remove pupil and school covariates from the models estimating impact. The power calculations in the analysis columns reflect the final model used for analysing the primary outcome, pupil scores on the science assessment, the final sample size for that model, and the proportion of variance explained by the prior KS1 test scores.

Table 4: Minimum detectable effect size at different stages for the primary science assessment outcome

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.127	0.176	0.115	0.140	0.161	0.183
Proportion of level 1 (pupil) variance explained by level 1 covariates (R^2)		0.30	0.30	0.30	0.30	0.40	0.40
Proportion of level 2 (school) variance explained by level 2 covariates (R^2)		0.40	0.40	0.40	0.40	N/A	N/A
Intracluster correlations (ICCs)	level 2 (class)	0.12	0.12	0.12	0.12	0.15	0.16
Alpha		0.05	0.05	0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two	Two	Two
Average cluster size		32	6	43.7	11.4	44.3	11.6
Number of schools	intervention	90	90	106	106	102	102
	control	90	90	99	99	98	98
	total	180	180	205	205	200	200
Number of pupils	intervention	2,880	540	4,488	1,228	3,843	1,033
	control	2,880	540	4,478	1,108	3,963	967
	total	5,760	1,080	8,966	2,336	7,806	2,000
	Number of blocks	6	6	7	7	7	7

Note. N/A = Not applicable. After randomisation, the decision was made, as documented in the statistical analysis plan, to remove level 2 explanatory variables from the primary model.

Attrition

Overall, approximately 89% of the pupils included at the stage of randomisation were administered the pupil science assessment and survey. Some pupils (1.1%) were lost because a school dropped out of the study. These pupils were not included in our analysis sample. The remaining 9.4% of pupils were not administered the assessment and survey for a variety of reasons, including disability, long-term illness, moving to a different school, and other reasons. These pupils were included in our analysis sample. Details on missing outcome data are provided in Table 5.

Table 5: Missing Outcome Data

	Treatment schools		Control schools		All schools	
	N	Pct.	*N*	Pct.	*N*	Pct.
Administered pupil science assessment and survey						
Pupil attended session	3,961	88.3%	4,055	90.6%	8,016	89.4%
Not administered pupil science assessment and survey						
Long-term illness or long-term absence	22	0.5%	4	0.1%	26	0.3%
Special educational need or disability	42	0.9%	35	0.8%	77	0.9%
Moved schools before September 2016	83	1.8%	76	1.7%	159	1.8%
Moved schools after September 2016	188	4.2%	147	3.3%	335	3.7%
Other	119	2.7%	131	2.9%	250	2.8%
School dropped out						
School dropped out of study	73	1.6%	30	0.7%	103	1.1%
Total	4,488	100.0%	4,478	100.0%	8,966	100.0%

Pupil and school characteristics

Schools were assigned to treatment or control status within regions using the school-reported number of Year 5 teachers (in two categories) and the percentage of FSM-eligible pupils (in three categories) as minimisation variables. Additional variables were obtained from the NPD and EduBase. For each measure, treatment schools were not significantly different from control schools (Table 6).[8]

Table 6: Baseline comparison of schools in the analysis sample

School level (categorical)	Intervention group		Control group	
	n/*N* (missing)	Percentage	*n*/*N* (missing)	Percentage
Region				
A	12/102 (0)	11.8%	12/98 (0)	12.2%
B	17/102 (0)	16.7%	17/98 (0)	17.3%
C	19/102 (0)	18.6%	17/98 (0)	17.3%
D	17/102 (0)	16.7%	16/98 (0)	16.3%
E	13/102 (0)	12.7%	14/98 (0)	14.3%
F	14/102 (0)	13.7%	13/98 (0)	13.3%
G	10/102 (0)	9.8%	9/98 (0)	9.2%
Number of Year 5 teachers (at time of randomisation)				
1 to 2	82/102 (0)	80.4%	80/98 (0)	81.6%
3 or more	20/102 (0)	19.6%	18/98 (0)	18.4%

[8] Significance was tested using a chi-squared test (Stata TABULATE with the CHI2 option). Ordered categorical variables also were tested using the two-sample rank sum test (Stata RANKSUM). Results did not differ qualitatively between the two.

School level (categorical)	Intervention group		Control group	
	n/N (missing)	Percentage	*n/N* (missing)	Percentage
School percentage FSM (at time of randomisation)				
Low (less 9%)	34/102 (0)	33.3%	34/98 (0)	34.7%
Medium (9% to 24%)	34/102 (0)	33.3%	33/98 (0)	33.7%
High (greater than 24%)	34/102 (0)	33.3%	31/98 (0)	31.6%
IDACI rank				
Lowest tercile	35/102 (0)	34.3%	41/98 (0)	41.8%
Middle tercile	36/102 (0)	35.3%	21/98 (0)	21.4%
Highest tercile	31/102 (0)	30.4%	36/98 (0)	36.7%
Percentage of pupils FSM eligible				
Low (less than 33%)	94/102 (0)	92.2%	90/98 (0)	91.8%
Medium (33% to 66%)	8/102 (0)	7.8%	8/98 (0)	8.2%
High (greater than 66%)	0/102 (0)	0.0%	0/98 (0)	0.0%
Percentage of pupils White British ethnic origin				
Low (less than 33%)	10/102 (0)	9.8%	11/98 (0)	11.2%
Medium (33% to 66%)	14/102 (0)	13.7%	8/98 (0)	8.2%
High (greater than 66%)	78/102 (0)	76.5%	79/98 (0)	80.6%
Percentage of pupils whose first language is other than English				
Low (less than 33%)	83/102 (0)	81.4%	84/98 (0)	85.7%
Medium (33% to 66%)	11/102 (0)	10.8%	9/98 (0)	9.2%
High (greater than 66%)	8/102 (0)	7.8%	5/98 (0)	5.1%
Ofsted Rating				
Requires improvement	7/87 (15)	8.0%	12/84 (14)	14.3%
Good	67/87 (15)	77.0%	55/84 (14)	65.5%
Outstanding	13/87 (15)	14.9%	17/84 (14)	20.2%
Establishment type				
Academies	30/102 (0)	29.4%	30/98 (0)	30.6%
Local authority maintained schools	72/102 (0)	70.6%	68/98 (0)	69.4%
Special schools	1/102 (0)	1.0%	0/98 (0)	0.0%
Urban/Rural School Type				
Urban	80/102 (0)	78.4%	77/98 (0)	78.6%
Rural	22/102 (0)	21.6%	21/98 (0)	21.4%
School level (continuous)	***n* (missing)**	**Mean (*SD*)**	***n* (missing)**	**Mean (*SD*)**
Number of Year 4 pupils at recruitment	102 (0)	42.46 (24.4)	98 (0)	45.51 (26.27)

At the pupil level, several statistically significant differences occurred between the treatment and control groups (Table 7). First, pupils in the treatment schools were more likely to be ever FSM eligible (27.4%)

than pupils in the control schools (24.7%).[9] Second, pupils in the treatment schools had slightly lower KS1 test scores.[10] For example, the percent of pupils in treatment schools scoring 2A and above was 52.2% in mathematics and 55.2% in reading and writing compared with 54.6% in mathematics and 57.8% in reading and writing among pupils in control schools. The effect sizes of the differences at this cut-point (scoring 2A and above)—, 2.4 percentage points in mathematics and 2.5 in reading and writing—, are both approximately 0.06 using Cox's Index (U.S. Department of Education, 2014). Finally, pupil IDACI also was lower in the treatment schools (46.9% in the low category) than in the control schools (40.3% in the low category).[11] The additional analyses of outcome measures included these variables as covariates in the model.

Table 7: Baseline comparison of pupils in the analysis sample

Pupil level (categorical)	Intervention group		Control group	
	n/N (missing)	Percentage	*n/N* (missing)	Percentage
Ever eligible for FSM	1,210/4,415 (0)	27.4%	1,100/4,448 (0)	24.7%
Female	2,044/4,272 (143)	47.8%	2,129/4,326 (122)	49.2%
Age				
9 years old	1,003/4,415 (0)	22.7%	994/4,448 (0)	22.3%
10 years old	3,402/4,415 (0)	77.1%	3,444/4,448 (0)	77.4%
11 years old or older	10/4,415 (0)	0.2%	10/4,448 (0)	0.2%
KS1 maths				
W	27/4,263 (152)	0.6%	27/4,323 (125)	0.6%
1	313/4,263 (152)	7.3%	246/4,323 (125)	5.7%
2C	521/4,263 (152)	12.2%	506/4,323 (125)	11.7%
2B	1,175/4,263 (152)	27.6%	1,180/4,323 (125)	27.3%
2A	1,225/4,263 (152)	28.7%	1,312/4,323 (125)	30.3%
3	1,001/4,263 (152)	23.5%	1,051/4,323 (125)	24.3%
4	1/4,263 (152)	0.0%	1/4,323 (125)	0.0%
KS1 reading and writing				
W	50/4,263 (152)	1.2%	41/4,323 (125)	0.9%
1	528/4,263 (152)	12.4%	439/4,323 (125)	10.2%
2C	303/4,263 (152)	7.1%	287/4,323 (125)	6.6%
2B	1,028/4,263 (152)	24.1%	1,057/4,323 (125)	24.5%
2A	1,058/4,263 (152)	24.8%	1,161/4,323 (125)	26.9%
3	1,295/4,263 (152)	30.4%	1,337/4,323 (125)	30.9%
4	1/4,263 (152)	0.0%	1/4,323 (125)	0.0%
IDACI rank				
Low (less than 33%)	1,985/4,228 (187)	46.9%	1,734/4,301 (147)	40.3%
Medium (33% to 66%)	1,222/4,228 (187)	28.9%	1,384/4,301 (147)	32.2%
High (greater than 66%)	1,021/4,228 (187)	24.1%	1,183/4,301 (147)	27.5%

Outcomes and analysis

Primary intent-to-treat analysis

Using our primary model specifications, we did not find large (in terms of effect size) or statistically significant effect size impacts of TDTS on the primary outcome (pupil science knowledge attainment)

[9] Chi-squared *p*-value of 0.004.
[10] Chi-squared *p*-values of 0.013 in mathematics and 0.013 in reading and writing. Rank sum *p*-values of 0.061 in mathematics and 0.021 in reading and writing.
[11] Chi-squared *p*-value < 0.000; rank sum *p*-value < 0.000.

but did find statistically significant effects with modest effect sizes on pupil attitudes toward science. These later effects included both interest in, and self-efficacy toward, science.

The mean science assessment score among pupils in the treatment schools was 19.3, and 19.5 among pupils in the control schools. As noted, pupils in the treatment schools had somewhat lower attainment to start. Accounting for differences in KS1 mathematics and reading/writing attainment, the estimated impact of TDTS on the science assessment was an effect size of +0.01.

More promising results were found when looking at pupil attitudes toward science. The mean of the index of pupil interest in science in the treatment schools was 0.06, whereas among pupils in the control schools it was -0.06—a difference of 0.12. The primary impact model did not include any covariates, and because the outcome measure was standardised, an impact estimate of 0.12 translated into an effect size that also was 0.12. This estimate was statistically significant (p = 0.013).

Similarly, the mean of the index of pupil self-efficacy for science in the treatment schools was 0.04, whereas among pupils in the control schools it was -0.04. The effect size was 0.09 (rounded to the second digit), which also was statistically significant (p = 0.004). Table 8 presents the full results of the primary analysis and details of the effect size calculations are in Table 9. (Regression results are provided in Appendix J and sample analysis code is provided in Appendix K.)

Table 8: Primary analysis results

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; Control)	Hedges g (95% CI)	*p*-value
Science assessment (controlling for KS1 scores)	3,961 (454)	19.48 (19.29; 19.68)	4,055 (393)	19.62 (19.43; 19.82)	7,806 (3,843; 3,963)	0.01 (-0.08; 0.1)	0.791
Interest in science	3,850 (565)	0.06 (0.03; 0.09)	3,927 (521)	-0.06 (-0.09; -0.03)	7,777 (3,850; 3,927)	0.12 (0.03; 0.22)	0.013
Self-efficacy for science	3,933 (482)	0.04 (0.01; 0.07)	4,021 (427)	-0.04 (-0.07; -0.01)	7,954 (3,933; 4,021)	0.09 (0.03; 0.16)	0.004

Table 9: Primary analysis effect size estimation

			Intervention group		Control group		
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Science assessment (controlling for KS1 scores)	-0.14	0.08	3,843 (572)	39.02	3,963 (485)	39.55	39.29
Interest in science	0.12	0.12	3,850 (565)	0.92	3,927 (521)	1.07	1.00
Self-efficacy for science	0.08	0.09	3,933 (482)	0.98	4,021 (427)	1.02	1.00

Due to the extent of missing data (see Appendix I for an analysis of missing data) primary analysis models were also estimated on data completed using multiple imputation. Results from the estimation of models using imputed data are presented in Appendix L.

Subgroup analysis of ever FSM pupils

When limiting the sample to only those pupils who were ever FSM eligible, the magnitude of the estimated impact on pupil science attainment was larger, an estimated effect size of +0.05 compared with the 0.01 estimate for all pupils, but still statistically insignificant (*p* = 0.386). The interaction model found that the impact estimate for FSM pupils was not statistically significant from the estimate for non-FSM pupils (*p*=0.330).

Looking at pupil attitudes, both impact estimates for FSM pupils were similar in magnitude to the estimates for all pupils. For interest in science index, the effect size was +0.15 (*p* = 0.018) for FSM pupils compared with +0.12 for all pupils. For the self-efficacy for science index, the effect size was +0.06 (but not statistically significant, *p* = 0.266) for FSM pupils compared with +0.09 for all pupils. For both indices, the difference between the impact estimates for FSM and non-FSM pupils was not statistically significant (*p* = 0.810 for interest index; *p* = 0.594 for self-efficacy index). Tables 10 and 11 present the results for the subgroup analysis. Interaction results with the test of differential impact for ever FSM pupils are in Appendix J.

Table 10: Subgroup analysis of ever FSM pupils

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model (intervention; Control)	Hedges g (95% CI)	*p*-value
Science assessment (controlling for KS1 scores)	1,033 (177)	16.99 (16.64; 17.35)	968 (132)	17.06 (16.7; 17.43)	2,000 (1033; 967)	0.05 (-0.07; 0.18)	0.386
Interest in science	989 (221)	0.09 (0.03; 0.15)	918 (182)	-0.03 (-0.1; 0.04)	1,907 (989; 918)	0.15 (0.03; 0.28)	0.018
Self-efficacy for science	1,021 (189)	-0.01 (-0.08; 0.05)	959 (141)	-0.07 (-0.14; -0.01)	1,980 (1021; 959)	0.06 (-0.04; 0.16)	0.266

Table 11: Subgroup analysis of ever FSM pupils effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Science assessment (controlling for KS1 scores)	-0.07	0.31	1,033 (177)	33.82	967 (133)	33.9	33.86
Interest in science	0.11	0.16	989 (221)	0.94	918 (182)	1.16	1.05

Self-efficacy for science	0.06	0.06	1,021 (189)	0.98	959 (141)	1.00	0.99

Additional analyses

Additional analyses were conducted that added covariates to the models. Given the imbalance at baseline, these analyses give some indication as to whether that imbalance might impact the results. Table 12 reports impact estimates for the primary outcome (pupil science knowledge attainment) as additional covariates were added to the model. Although the point estimates are higher, they are still very close to zero. Adding these covariates did not substantially change the estimated effect size. Details of the effect size calculations are reported in Table 13.

Table 12: Impact on science assessment when including additional covariates

		Effect size		
Outcome	**Model covariates**	**n in model (intervention; Control)**	**Hedges g (95% CI)**	***p*-value**
Science assessment	KS1 scores, school FSM, school N Year 5 teachers, region	7,806 (3843; 3963)	0.02 (-0.06; 0.1)	0.668
Science assessment	KS1 scores, school FSM, school N Year 5 teachers, region, pupil age, pupil gender, pupil FSM, pupil IDACI	7,763 (3818; 3945)	0.02 (-0.06; 0.1)	0.635

Table 13: Impact on science assessment when including additional covariates effect size estimation

			Intervention group		Control group		
Outcome	**Model covariates**	**Adjusted differences in means**	**n (missing)**	**Variance of outcome**	**n (missing)**	**Variance of outcome**	**Pooled variance**
Science assessment	KS1 scores, school FSM, school N Year 5 teachers, region	0.11	3,843 (572)	39.02	3,963 (485)	39.55	39.29
Science assessment	KS1 scores, school FSM, school N Year 5 teachers, region, pupil age, pupil gender, pupil FSM, pupil IDACI	0.12	3,818 (597)	38.96	3,945 (503)	39.46	39.21

Additional analysis results for the secondary outcomes—the indices of pupil attitudes toward science—are reported in Table 14 for the interest in science index (details of the effect size calculations in Tables 15) and Table 16 for the self-efficacy for science index (details of the effect size calculations in Tables 17). The primary model for these outcomes did not include prior KS1 test scores. Neither adding these prior scores as covariates nor adding other pupil and school covariates changed the effect size estimates or their statistical significance.

Table 14: Impact on pupil interest in science when including additional covariates

		Effect size		
Outcome	Model covariates	n in model (intervention; Control)	Hedges g (95% CI)	*p*-value
Interest in science	KS1 scores	7,572 (3,734; 3,838)	0.12 (0.02; 0.22)	0.015
Interest in science	KS1 scores, school FSM, school N Year 5 teachers, region	7,572 (3,734; 3,838)	0.12 (0.03; 0.21)	0.010
Interest in science	KS1 scores, school FSM, school N Year 5 teachers, region, pupil age, pupil gender, pupil FSM, pupil IDACI	7,531 (3,710; 3,821)	0.12 (0.03; 0.21)	0.010

Table 15: Impact on pupil interest in science when including additional covariates effect size estimation

			Intervention group		Control group		
Outcome	Model covariates	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Interest in science	KS1 scores	0.12	3,734 (681)	0.93	3,838 (610)	1.08	1.01
Interest in science	KS1 scores, school FSM, school N Year 5 teachers, region	0.12	3,734 (681)	0.93	3,838 (610)	1.08	1.01
Interest in science	KS1 scores, school FSM, school N Year 5 teachers, region, pupil age, pupil gender, pupil FSM, pupil IDACI	0.12	3,710 (705)	0.93	3,821 (627)	1.08	1.01

Table 16: Impact on pupil self-efficacy for science when including additional covariates

		Effect size		
Outcome	Model covariates	n in model (intervention; Control)	Hedges g (95% CI)	*p*-value
Self-efficacy for science	KS1 scores	7,748 (3,816; 3,932)	0.09 (0.03; 0.15)	0.006
Self-efficacy for science	KS1 scores, school FSM, school N Year 5 teachers, region	7,748 (3,816; 39,32)	0.09 (0.03; 0.15)	0.003
Self-efficacy for science	KS1 scores, school FSM, school N Year 5 teachers, region, pupil age, pupil gender, pupil FSM, pupil IDACI	7,705 (3,791; 3,914)	0.09 (0.03; 0.15)	0.003

Table 17: Impact on pupil self-efficacy for science when including additional covariates effect size estimation

			Intervention group		Control group		
Outcome	Model covariates	Adjusted differences in means	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Self-efficacy for science	KS1 scores	0.09	3,816 (599)	0.98	3,932 (516)	1.01	1.00
Self-efficacy for science	KS1 scores, school FSM, school N Year 5 teachers, region	0.09	3,816 (599)	0.98	3,932 (516)	1.01	1.00
Self-efficacy for science	KS1 scores, school FSM, school N Year 5 teachers, region, pupil age, pupil gender, pupil FSM, pupil IDACI	0.09	3,791 (624)	0.98	3,914 (534)	1.01	1.00

Compliance analysis

Using the definition of compliance provided by the programme developers, the Oxford team, 98 of the 102 treatment schools (96.1%) containing 97.3% of pupils in treatment schools were found to have complied with the programme. Table 18 provides results from the CACE analysis using instrumental variables (details of effect size estimation are in table 19). With rounding, the estimates do not differ from the primary analysis estimates reported in Table 8. (Full regression results are presented in Appendix M.)

Table 18: Compliance primary analysis

	Effect size		
Outcome	n in model (intervention; Control)	Hedges g (95% CI)	*p*-value
Science assessment (controlling for KS1 scores)	7,806 (3843; 3963)	0.01 (-0.08; 0.11)	0.791
Interest in science	7,777 (3850; 3927)	0.13 (0.03; 0.22)	0.013
Self-efficacy for science	7,954 (3933; 4021)	0.09 (0.03; 0.16)	0.004

Table 19: Compliance primary analysis effect size estimation

Outcome	Unadjusted differences in means	Adjusted differences in means	Intervention group		Control group		Pooled variance
			n (missing)	Variance of outcome	n (missing)	Variance of outcome	
Science assessment (controlling for KS1 scores)	-0.14	0.08	3,843 (572)	39.020	3,963 (485)	39.551	39.290
Interest in science	0.12	0.13	3,850 (565)	0.922	3,927 (521)	1.070	0.997
Self-efficacy for science	0.08	0.09	3,933 (482)	0.980	4,021 (427)	1.017	0.998

Cost

Using the assumptions listed earlier (the trainers train three cohorts of teachers; each teacher uses the training with three years of pupils), the financial costs of the intervention were £1,671 per teacher or £29 per pupil (Table 20).

Table 20: Financial costs to deliver TDTS to one cohort of teachers

Item	Type of cost	Cost per teacher	Total cost per teacher across 3 years	Total cost per pupil per year across 3 years
Teacher training: four CPD sessions in 1 year	Start-up cost for each teacher	£730	£730	
Annual costs for each teacher	Running cost per teacher	£0	£0	
One-off training of trainers	Start-up cost for each trainer per teacher (spread across three cohorts of teachers)	£170	£170	
Annual TDTS programme costs	Running costs for the programme, incurred once per teacher cohort	£771	£771	
Total			£1,671	(£1,671/3 years/19 pupils per year) = £29

The non-financial costs included four days of teacher supply time and laptops for trainers to use in the CPD sessions. Because sessions are a full day in term, and co-ordinated across schools that may have different schedules, supply cover was necessary for teacher attendance as opposed to sessions outside teaching hours or during school inset days. The intervention is intended to change how teachers plan and deliver lessons. This could potentially just be a shift in how teachers are using their time (that is, not add additional time to planning and teaching) or it could require additional time. A high percentage of teachers reported that the intervention required ‘a little’, ‘very little’, or ‘no’ additional time to incorporate the strategies into their lesson planning (87%) and teaching (86%; see Implementation and Process Evaluation section).

In this evaluation, 233 teachers and 4,448 pupils were recruited into the study—about 19.26 pupils per teacher. These numbers were used to calculate training costs per teacher and final costs per pupil.

The training of each cohort of teachers on the TDTS programme was assumed to be implemented in a one-year time frame via four CPD sessions that included the following costs:[12]

- trainer preparation and delivery of the CPD sessions;
- venue and catering for each CPD session;
- trainer travel;
- materials (printing and investigation consumables);
- marketing and school recruitment; and
- quality assurance visits by the programme developers

Once trained, teachers incurred no costs for using the TDTS programme with pupils in subsequent years. It was assumed that, once trained, each teacher would use the TDTS programme for three years—this follows the standard approach for calculating cost in EEF trials and is not an assumption specific to this programme.

The cost of developing the TDTS trainers was assumed to be spread out across three cohorts of teachers, all assumed to be from different schools. These costs included the following:

- trainer recruitment;
- programme developer preparation and delivery of the train-the-trainer sessions;
- venue and catering for each training session;
- programme developer and trainer travel;
- materials (printing and investigation consumables); and
- equipment for each trainer to use in the CPD sessions

Annual costs of maintaining the TDTS programme included general logistical and financial management, including website hosting.

One-time costs associated with programme as delivered in this evaluation that were not included in the previous estimates included the following:

- development of the train-the-trainer sessions;
- development of the tutor folders for trainers;
- revisions to the teacher folders; and
- programme website and branding setup.

Costs specifically associated with the evaluation were not included in these estimates. These costs included the following:

- equipment gifts/grants to schools for participating in the evaluation;
- programme developer time devoted to the evaluation; and
- trainer time devoted to learning and assisting with the evaluation.

Finally, it should be noted that in future delivery of TDTS, the costs will likely vary from those reported here. Reasons for variation in costs include the following:

- travel and venue costs will vary depending on location;
- costs will be higher if implementation is at a smaller scale when—
    - training new trainers,
    - printing materials, and
    - purchasing equipment; and
- recruitment of schools and retention of teachers will vary depending on location.

[12] The evaluation covered only two cohorts—the treatment cohort and the waitlist control cohort. The assumption of three cohorts per trainer was only for calculating costs per teacher and per pupil.

Implementation and process evaluation

The implementation and process evaluation examined how the TDTS programme was implemented at scale through the 'train-the-trainers' model of delivery to teachers. First, trainers reported that they had been trained well by the Oxford team and were confident in their ability to deliver TDTS to teachers. Second, teachers in treatment schools reported confidence in their understanding of, and ability to apply, the strategies that they learned in the TDTS CPD sessions. These are indications of a successful transfer of TDTS strategies from developer to trainer to teacher. Additionally, teachers felt that those strategies required little to no extra time to implement in their teaching.

Teachers in treatment schools reported differences in pupil outcomes and teaching practice that were hypothesised to be impacted by the programme; they often reported that their pupils were engaged with, confident in, and made good progress in science. Teachers in treatment schools reported using more practical work and more small-group work while teachers in control schools reported more often that they required pupils to write-up whole investigations. This is consistent with the intent of TDTS, which encourages teachers to use more focused recordings rather than writing about whole investigations which may include recording that is less relevant to learning. Teachers, however, did not report differences in practices of assessing pupil work as had been hypothesised.

Implementation

The first step in implementing TDTS at scale was to train the trainers who would deliver the programme to teachers. At the end of the last train-the-trainers session, the trainers indicated that they had a firm understanding of the programme and were confident in their ability to deliver the programme. All trainers 'agreed' or 'strongly agreed' that they had a good understanding of TDTS, felt confident teaching/training others on the programme, and successfully did so. All trainers felt that the Oxford team did a 'good' or 'very good' job of preparing them (Table 21).

Table 21: Trainer perceptions of the train-the-trainer sessions

'Please rate the level to which you agree with the following statements about ALL the trainings received from the Oxford team.'	Strongly agree	Agree	Somewhat agree	Disagree	Completely disagree	Total
I feel confident teaching/ training others in the use of the TDTS programme.	12	2	0	0	0	14
I have a good understanding of the TDTS programme.	13	1	0	0	0	14
I was able to successfully train teachers on TDTS.	11	3	0	0	0	14
'Please rate the following aspects of ALL the trainings received from the Oxford team.'	**Very good**	**Good**	**Average**	**Poor**	**Very poor**	**Total**
The overall training.	13	1	0	0	0	14

One measure of successful implementation was teacher attendance at the CPD sessions. While just over half (58.8%) of the schools sent at least two teachers to every session, almost all schools (87.3%) had at least one teacher at every session (Table 22). In addition, seven schools sent extra teachers, with one school sending an extra teacher to all four CPD sessions. The Oxford team made attempts to contact schools and teachers who did not attend a session and recorded reasons for absences and all appear to be unavoidable. In some instances, an individual teacher did not attend due to being ill, being on leave (medical or pregnancy), or had left the school. In four instances, all teachers in a school were absent from a CPD day due to Ofsted inspections for which there is very little advance notice.

Table 22: School-level participation in the TDTS programme

School attendance category	*N*	Percentage
All teachers from school attended all CPD sessions	48	47.1%
Attendance from school was two or more teachers at all four CPD sessions	12	11.8%
Attendance from school was one or more teacher(s) at all four CPD sessions	29	28.4%
Attendance from school was one or more teacher(s) at three of four CPD sessions	10	9.8%
Other	3	2.9%

In response to an open-ended question asking how TDTS compared to other CPD models, 6 of the 14 trainers specifically praised the delivery approach of spacing the four CPD sessions with gap tasks in between.[13] This feedback indicates that this approach may be an important condition of success to maintain when delivering TDTS in the future.

Fidelity

Trainers all reported ('agree' or 'strongly agree') that they followed the TDTS tutor files, covered all the material, and were successful in effectively training teachers on the key elements of TDTS (Table 23).

[13] It may be that other trainers also found this important but did not mention it. Of the eight who did not mention the importance of staggered CPD with gap task, four responded to the question with very brief *and* very positive response.

Table 23: Trainer perceptions of their delivery of TDTS to teachers

'Please rate the level to which you agree with the following statements about ALL the trainings received from the Oxford team.'	Strongly agree	Agree	Somewhat agree	Disagree	Completely disagree	*N*
I followed the tutor file very closely during the CPD trainings.	11	3	0	0	0	14
During the teacher training sessions, I was able to cover the agenda provided by the Oxford team.	8	6	0	0	0	14
I was able to effectively train the teachers on how to use focused recording during science lessons.	6	3	5	0	0	14
I was able to effectively train the teachers to use the Bright Ideas Time.	12	2	0	0	0	14
I was able to effectively train the teachers on Practical Prompts for Thinking.	11	3	0	0	0	14
I was able to effectively train the teachers on practical investigations.	12	2	0	0	0	14
I was able to effectively train the teachers on practical problem solving.	12	2	0	0	0	14

These sentiments were shared by teachers. More than 90% of the teacher respondents after each CPD session reported being well prepared to use TDTS strategies in their classrooms and understanding how to implement key elements of the TDTS programme (Table 24). The one exception was teachers' understanding of how to implement more focused recordings in the classroom, although this item still had more than 75% of the teachers agreeing that they knew how to implement it. Unlike the other items, the percentage who agreed that they knew how to use more focused recording increased over time.

Table 24: Teacher perceptions of the effectiveness of CPD sessions in teaching them how to implement TDTS

'Rate the level to which you agree with the following statements about today's TDTS training.'	Percentage of teachers who agree or strongly agree			
	CPD Session 1	CPD Session 2	CPD Session 3	CPD Session 4
The training received has prepared me well for using the TDTS strategies in my classroom.	95%	96%	93%	96%
The materials provided increased my understanding of the TDTS strategies.	94%	96%	92%	95%
I understand how to implement Bright Ideas Time in my classroom.	94%	94%	96%	97%
I understand how to implement Practical Prompts for Thinking in my classroom.	95%	94%	93%	97%
I understand how to implement practical investigations in my classroom.	98%	98%	97%	99%
I understand how to implement practical problem solving in my classroom.	92%	96%	95%	97%
I understand how I would use focused recordings in my classroom during my science lessons.	75%	86%	87%	89%

Outcomes

Teachers overwhelmingly 'agreed' or 'strongly agreed' (more than 90%) that the TDTS strategies that they learned in each CPD session were useful to them, and that they were excited to use the strategies in their classrooms (Table 25). Teachers also were asked to reflect on their implementation of the training. More than 75% of the teachers reported ('agreed' or 'strongly agreed') that they had used the strategies regularly in their classes, and more than 89% reported that they were able to implement the strategies successfully. Teachers were asked about the perceived impact on pupils (Table 25). Almost all (more than 96%) felt that pupils enjoyed the TDTS activities, and more than 93% felt that the TDTS strategies increased pupil engagement. Many teachers also felt that TDTS helped pupils understand science content. This percentage increased from 69% when asked after CPD Session 2 to 85% after CPD Session 4.

Table 25: Teacher perceptions of the impact of TDTS on their teaching and on pupil engagement and learning

	Percentage who agree or strongly agree			
'Rate the level to which you agree with the following statements about today's TDTS training.'	CPD Session 1	CPD Session 2	CPD Session 3	CPD Session 4
The TDTS strategies have been useful to me for teaching science.	93%	91%	93%	93%
I am excited to use TDTS strategies in my teaching.	94%	96%	93%	92%
'Rate the level to which you agree with the following statements about implementing the TDTS strategies in your teaching after the last training session you attended.'	CPD Session 1	CPD Session 2	CPD Session 3	CPD Session 4
I used the TDTS strategies regularly in my teaching.	N/A	75%	77%	81%
I was able to implement the TDTS strategies effectively in my teaching.	N/A	89%	89%	96%
Using the TDTS strategies increased pupils' engagement.	N/A	93%	93%	94%
My pupils seemed to understand the science content better after I used the TDTS strategies in science lesson(s).	N/A	69%	76%	85%
My pupils enjoyed the TDTS activities I used in my classroom.	N/A	96%	96%	97%

It is important to note that some of the measures point to an increased understanding of the programme over time. The percentage who agreed with the statement that they 'used TDTS strategies regularly' increased from 75% at the second CPD session to 81% at the fourth CPD session. The percentage who agreed that they were 'able to implement TDTS strategies effectively' increased from 89% at the second CPD session to 95% at the fourth CPD session.

Teachers reported that implementing TDTS took little to no additional time in lesson planning or teaching (Table 26).

Table 26: Teacher perceptions of how much additional time was required to use TDTS

	Percentage who responded a little, very little, or no additional time			
'How much additional time did using the TDTS strategies add to the following activities?'	CPD Session 1	CPD Session 2	CPD Session 3	CPD Session 4
Lesson planning	N/A	93%	89%	87%
Science teaching	N/A	88%	89%	86%

After the final CPD session, almost all teachers reported that they would continue to use the TDTS strategies in the following school year (98% 'agreed' or 'strongly agreed') and would recommend the CPD training and TDTS strategies to colleagues (93% 'agreed' or 'strongly agreed').

Results from the end-of-year teachers survey administered to teachers in both treatment and control schools provided additional information from the perspective of teachers about the potential impact of TDTS on themselves and their pupils. The response rates to the teacher survey were 50.6% in treatment schools, 59.9% in control schools, and 55.2% overall. As hypothesised by the Oxford Team, teachers in treatment schools were more confident in using science practicals, adapting teaching to pupils' interests, and helping pupils discuss science ideas (Table 27).

Table 27: Teacher confidence in teaching activities in treatment and control schools

'In teaching science to your class, how would you characterise your confidence in doing the following?'	Percentage who responded 'High' or 'Very high'					
	Treatment teachers	**N Treatment**	**Control teachers**	**N Control**	**Difference**	***p-value***
Inspiring pupils to learn science [T]	86%	118	76%	133	10%	0.054
Explaining science concepts or principles by doing science practicals [T]	80%	118	65%	133	14%	0.012
Providing challenging tasks for the highest achieving pupils	65%	118	35%	133	30%	0.000
Adapting my teaching to engage pupils' interest [T]	92%	118	73%	133	19%	0.000
Assessing pupil learning in science	49%	118	39%	133	10%	0.109
Improving the understanding of struggling pupils	55%	118	54%	133	1%	0.880
Helping pupils discuss scientific ideas [T]	88%	118	71%	133	17%	0.001

Note. Group hypothesised by the Oxford team to have the higher rate indicated in square brackets:

T = treatment teachers; C = control teachers; blank = no hypothesised difference.

Interestingly, though there was no difference hypothesised by the Oxford team for this trial, treatment teachers reported being more confident in providing challenging tasks to high-achieving pupils by a large margin (+30 percentage points). Considering this finding, it should be noted that TDTS is delivered as a whole-classroom approach to teaching and there is no evidence that it is more effective with high achieving pupils. In the TDTS efficacy trial, Hanley et al. (2015) estimated impacts for pupils above and below the median pre-test and found smaller effect sizes for those above the median (+0.22) than for those below the median (+0.30), though the difference was not statistically significant.

Teachers in treatment schools also more often reported that their pupils were engaged with, confident in, and made good progress in science, in line with hypotheses (Table 28). Treatment teachers also more often reported that their pupils could work independently and come up with their own scientific ideas. Also, in line with expectations, teachers in control schools more often felt that their pupils 'do a lot of writing' in science.

Table 28: Teacher reflections on teaching science in the past year in treatment and control schools

'Reflecting on your experience of teaching science this year, how much would you agree or disagree with the following statements?'	Percentage who responded 'Agree slightly' or 'Agree strongly'					
	Treatment Teachers	N Treatment	Control Teachers	N Control	Difference	*p*-value
My pupils have enjoyed their science lessons [T]	99%	118	97%	133	2%	0.222
My pupils have made good progress in science [T]	99%	118	93%	133	6%	0.017
My pupils are confident in science [T]	95%	118	87%	133	8%	0.035
My pupils can work independently in science [T]	95%	118	74%	133	20%	0.000
My pupils come up with their own scientific ideas [T]	91%	118	74%	133	17%	0.001
My pupils do a lot of writing in science [C]	27%	118	47%	133	-20%	0.001
I have changed the way I teach science [T]	92%	118	52%	133	40%	0.000
I enjoy teaching science [T]	98%	118	91%	133	7%	0.012
I have a good knowledge of science	97%	118	92%	133	5%	0.076
My pupils have been engaged with science [T]	99%	118	93%	133	6%	0.017

Note. Group hypothesised by the Oxford team to have the higher rate indicated in square brackets: T = treatment teachers; C = control teachers; blank = no hypothesised difference.

Finally, teachers in treatment schools also reported that their practices were different than those in control schools in ways that were hypothesised (Table 29). Teachers in treatment schools reported using more practical work and more small-group work. Teachers in control schools reported more often that they required pupils to write-up whole investigations. One area where differences in practice were hypothesised but not observed in the teacher survey was how teachers assess pupils (Table 30). The expectation was that teachers would more often use pupil work on practicals, observation of pupils, and pupil responses in class than in control schools but no significant differences were found. Teachers in treatment schools did, however, report less use of tests and worksheets produced outside the school as hypothesised.

Table 29: Teacher practice in teaching science in treatment and control schools

'How often do you do the following in science lessons?'	Percentage who responded 'Quite often' or 'Very often'					
	Treatment Teachers	**N Treatment**	**Control Teachers**	**N Control**	**Difference**	***p*-value**
Pupil practical work [T]	96%	118	83%	133	12%	0.002
Teacher demonstration	56%	118	61%	133	-5%	0.425
Pupil discussion in whole class [T]	99%	118	93%	133	6%	0.017
Pair or small group practical work [T]	95%	118	83%	133	12%	0.003
Pair or small group discussion [T]	99%	118	89%	133	10%	0.001
Give pupils time to think [T]	95%	118	89%	133	6%	0.077
Teach scientific facts	80%	118	86%	133	-6%	0.204
Ask pupils to solve scientific problems [T]	85%	118	60%	133	25%	0.000
Design or plan practical investigations [T]	90%	118	73%	133	17%	0.001
Carry out practical investigations [T]	94%	118	82%	133	12%	0.004
Present results from practical	70%	118	63%	133	7%	0.229
Interpret results from practical investigations [T]	79%	118	67%	133	12%	0.035
Use results from practical investigations to support conclusions [T]	79%	118	70%	133	9%	0.109
Require the pupils to write up the whole investigation [C]	18%	118	41%	133	-24%	0.000
Require the pupils to memorise facts and principles	20%	118	29%	133	-8%	0.131

Note. Group hypothesised by the Oxford team to have the higher rate indicated in square brackets: T = treatment teachers; C = control teachers; blank = no hypothesised difference.

Table 30: Teacher practice in assessing pupil work in treatment and control schools

'In assessing the work of the pupils in your science class, how much weight do you give each of the following types of assessment?'	Percentage who responded 'Quite a lot' or 'A great deal'					
	Treatment Teachers	N Treatment	Control Teachers	N Control	Difference	*p*-value
Tests and worksheets produced outside the school [C]	14%	118	23%	133	-10%	0.048
Teacher-made short answer tests that require pupils to describe or explain their reasoning	38%	118	46%	133	-8%	0.216
Teacher-made multiple choice, true-false and matching tests [C]	16%	118	22%	133	-6%	0.252
Assess how well pupils undertake practical scientific inquiry [T]	84%	118	82%	133	2%	0.683
Observations of pupils [T]	97%	118	95%	133	2%	0.469
Responses of pupils in class [T]	98%	118	94%	133	4%	0.081
Assessment through pupils' recording in their books	87%	118	86%	133	2%	0.716

Note. Group hypothesised by the Oxford team to have the higher rate indicated in square brackets: T = treatment teachers; C = control teachers; blank = no hypothesised difference.

Control group activity

At the time of recruitment (end of the 2015/2016 school year), schools were asked about their past and future plans for CPD. Responses from treatment and control schools were compared using t-tests. Schools responses of 'refused' and 'don't know' were coded as missing. There were no significant differences between treatment and control schools in the average number of science CPD days, nor in the percentage planning any science CPD days in the coming year (Table 31).

Table 31: School plans for Science CPD

	Treatment		Control			
	N (missing)	Average	N (missing)	Average	difference	*p*-value
The number of days of CPD Year 5 teachers…						
Received in science in 2015-2016	90 (12)	1.1	88 (10)	1.0	0.12	0.553
Will receive in science in 2016-2017	73 (29)	1.3	74 (24)	1.1	0.29	0.276
	N (missing)	**Pct.**	**N (missing)**	**Pct.**	**difference**	***p*-value**
Science CPD days planned for 2016/2017 greater than zero	73 (29)	53%	74 (24)	45%	9%	0.287

Over two thirds of schools had participated in at least one specific science intervention in 2015/2016 or 2016/2017 (Table 32). However, there were no significant differences in participation between treatment and control schools.

Table 32: School participation in specific science interventions

	Treatment		Control			
	N	Pct.	N	Pct.	difference	*p*-value
Has your school been involved in any specific science intervention this academic year or next?						
Primary Science Quality Mark	102	16%	98	13%	2%	0.629
Primary Science Teaching Trust	102	5%	98	5%	0%	0.949
STEM Learning/Science Learning Centre courses	102	26%	98	21%	5%	0.406
Association of Science Education (ASE) courses	102	5%	98	7%	-2%	0.507
Other	102	36%	98	45%	-9%	0.216
Don't know	102	28%	98	22%	6%	0.334
Prefer not to answer	102	5%	98	2%	3%	0.273
'Yes' to at least one of the options	102	67%	98	76%	-9%	0.170

Conclusion

Key conclusions
1. There is no evidence that TDTS had an impact on pupils' science knowledge attainment, on average. This result has a high security rating.
2. Among children receiving free school meals, those in TDTS schools made a small amount of additional progress compared to those in other schools. However, this finding is not statistically significant". This means that the statistical evidence supporting the impact finding does not meet the threshold set by the evaluator to be convincing.
3. The programme led to small increases in pupil interest in science and self-efficacy for science, as measured by pupil surveys.
4. Teachers who received TDTS training reported confidence in their understanding of, and ability to apply, the strategies they had learned. They felt that those strategies required minimal extra time to implement.

Interpretation

TDTS was not found, in this evaluation, to have a strong impact on pupil science knowledge attainment. This differs from results from the efficacy trial. Consistent with the conclusions from the efficacy trial, the programme was found to have an impact, though modest, on pupil interest in science and pupil self-efficacy for science.

In comparing the results in this trial with those from the efficacy trial, three differences need to be considered that could explain the difference in findings.

First, the programme was delivered differently: in the efficacy trial teachers were trained directly by the programme developers while in this evaluation a train-the-trainers model was used.

Second, in the efficacy trial teachers were provided supply cover for two days of in-school time per teacher to plan together how they would implement the strategies. Both of these differences could lead to lower fidelity of implementation at the training and classroom levels. In contrast, the feedback from trainers and teachers was that the programme was delivered effectively. This feedback is, however, subjective. More objective measures of implementation, such as systematic observation of teacher classrooms, were not included in the study.

Third, the programme was delivered in this study over four CPD sessions as opposed to five used in the efficacy trial. It is unclear, however, if this change would have an impact on results. In the efficacy trial, the fifth and final day of training occurred late in the school year and, hence, it was felt that it may have had little impact on measured outcomes. If true, the difference in the number of CPD sessions would not explain the difference in findings.[14] Alternatively, one fewer day may be important. Though the intent was to cover the same amount of material, it may be that either less material was covered with teachers or material was covered in less depth. Either could lead to a dilution of the programme's impact.

Another factor that may cause dilution of the programme's impact is less than full teacher attendance. This might not explain differences in results as there did not appear to be a substantial difference in teacher attendance between the efficacy trial and this study: in the efficacy trial, 81% of the schools had one or more teachers attend each of the five sessions while in this study, 87% of the schools had one or more teachers at each of the four sessions. Nevertheless, in this study the ideal situation of having at least two teachers from the school at each session occurred in only 59% of the schools.

[14] In this trial, the fourth and final CPD session was delivered at different times in the different regions from 4 April to 23 May. Testing was conducted from 5–23 June. It is possible that in some schools in some regions the date of the final CPD session was as little as two weeks from testing.

Results from the implementation and process evaluation paint a very positive picture of the programme. Feedback from teachers about the programme was positive with over 80% of teachers in the programme reporting using the strategies regularly by the time of the final CPD session. Some of the process questions point to an increasing understanding and use of TDTS strategies over time. It is possible that the programme is more effective once teachers have been exposed to the full content of the programme. However, in this evaluation the final CPD session does not occur until soon before outcomes are measured. Still, it is unclear how a longer period of observation would allow for different results. On one hand, a longer period of evaluation may allow for the programme to have a stronger impact on pupil attainment. On the other hand, if the TDTS programme is not reinforced over time, the impact on teacher practice may fade over time and, hence, one might not observer a greater impact on pupil attainment. Nevertheless, it is possible that in this evaluation we attempted to measure impacts too soon after completion of the CPD sessions and before teachers had fully internalised and operationalised their delivery of the TDTS approaches to pupils.

In summary, this evaluation found that TDTS is a low-cost intervention that does not have an immediate impact on pupil attainment but does have a small impact on pupil interest in, and self-efficacy towards, science. Available evidence indicates that the programme can be implemented at scale through a train-the-trainers model, that it is valued by teachers exposed to the programme, and that it changes their teaching practices in a manner consistent with hypotheses.

Limitations

There are number of limitations to this study that should be considered when interpreting results.

First, as noted previously, the process evaluation did not include objective measures of implementation such as systematic observation of teacher classrooms. Though feedback from trainers and teachers about their respective training on the TDTS programme is valuable, it is not as reliable a measure of successful implementation as direct observation of teachers’ use of TDTS in the classroom.

Second, the primary outcome measure of science attainment is derived from a set of assessments that are standardised and meant to measure general science attainment. The TDTS programme is meant to impact science attainment in general, but there is also a focus on enquiry skills. It is possible that the TDTS programme could have a significant impact on enquiry skills that are not illuminated by a general science assessment.

Finally, caution should be exercised in generalising the results of this study. This trial was conducted with one pupil age group, Year 5 pupils and their teachers,[15] and in seven regions—though schools were not selected to be representative of those regions. Additionally, given that some pupils were not administered the study’s science assessment and questionnaire due to special educational needs or disability, results are not readily generalizable to that group.

Future research and publications

This evaluation suggests a few specific research questions that could be addressed by future research:

1. Using measures that are not self-reported, for example, classroom observations, do teachers exposed to the TDTS programme use its strategies in the classroom? Do they use them with fidelity?
2. How does teachers’ use of TDTS strategies change over the course of the year TDTS is offered to them? Does their use diminish once the programme is completed?
3. Does the programme have a focused impact on pupil enquiry skills?

[15] The precursor to the TDTS programme was studied with Year 6 pupils and teachers (Mant et al., 2007).

References

Education Endowment Foundation (2018) 'Statistical analysis guidance for EEF evaluations', London: EEF.

Hanley, P., Slavin, R. and Elliott, L. (2015) 'Thinking, doing, talking science: evaluation report and executive summary', London: EEF.

Huang, F. L. (2018) 'Using Instrumental Variable Estimation to Evaluate Randomized Experiments with Imperfect Compliance', *Practical Assessment, Research and Evaluation*, 23 (2). Available online: http://pareonline.net/getvn.asp?v=23&n=2

Kind, P., Jones, K. and Barmby, P. (2007) 'Developing attitudes towards science measures', *International Journal of Science Education,* 29 (7), pp. 871–893.

Mant, J., Wilson, H. and Coates, D. (2007) 'The effect of increasing conceptual challenge in primary science lessons on pupils' achievement and engagement', *International Journal of Science Education,* 29 (14), pp. 1707–1719.

Murphy, C. and Beggs, J. (2005) *Primary science in the UK: a scoping study,* London: Wellcome Trust.

Ofsted (2013) 'Maintaining curiosity: A survey into science education in schools', Manchester: Ofsted: <https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/379164/Maintaining_20curiosity_20a_20survey_20into_20science_20education_20in_20schools.pdf> [accessed 30 May 2018].

Russell, T. and McGuigan, L. (2001) *Science assessment series 1: teachers' guides and assessment units KS1 and KS2*, Berkshire: NFER-Nelson.

Slavin, R. E., Lake, C., Hanley, P. and Thurston, A. (2012) 'Effective programs for elementary science: a best-evidence synthesis', Baltimore: Johns Hopkins University, Center for Research and Reform in Education.

StataCorp. (2017) *Stata statistical software: release 15*, College Station, TX: StataCorp LLC.

TIMSS (1998) *Science Teacher Questionnaire Main Survey*, International Association for the Evaluation of Educational Achievement (IEA), Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

TIMSS (2015) *TIMSS 2015 User Guide for the International Database*, International Association for the Evaluation of Educational Achievement (IEA), Boston: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

U.S. Department of Education (2014) *What Works Clearinghouse: Procedures and standards handbook (version 3.0)*, Washington, DC: Institute of Education Sciences.

Appendix A: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention in 3 years. More information about the EEF's approach to cost evaluation can be found **here**. Cost ratings are awarded as follows:

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

1. Criteria for rating: in each column highlight the relevant cell in green.
2. Initial score: write how many padlocks the trial has received based on the first 3 columns ("x🔒") and highlight in green (initial score is the lowest rating from the first three columns – see guidance on security classification for more detail).
3. Adjust: record adjustment for balance and threats for validity in the adjust column
4. Final score: write the number of padlocks ("x🔒") in the relevant cell and highlight in green
5. Provide a brief summary of your classification, following the bullet point prompts below

Rating	Criteria for rating			Initial score	Adjust	Final score
	Design	Power	Attrition[16]			
5 🔒	Well conducted experimental design with appropriate analysis	MDES < 0.2	0-10%			
4 🔒	Fair and clear quasi-experimental design for comparison (e.g. RDD) with appropriate analysis, or experimental design with minor concerns about validity	MDES < 0.3	11-20%	4	Adjustment for Balance [0]	4
3 🔒	Well-matched comparison (using propensity score matching, or similar) or experimental design with moderate concerns about validity	MDES < 0.4	21-30%			
2 🔒	Weakly matched comparison or experimental design with major flaws	MDES < 0.5	31-40%		Adjustment for threats to internal validity [0]	
1 🔒	Comparison group with poor or no matching (E.g. volunteer versus others)	MDES < 0.6	41-50%			
0 🔒	No comparator	MDES > 0.6	over 50%			

- **Initial padlock score:** lowest of the three ratings for design, power and attrition = The design is a randomised controlled trial which has been powered to 0.176 for FSM pupils and 0.127 for all pupils. The pupil attrition was 11%. However, the multiple imputation analyses were very similar. Therefore, this trial should have 4 padlocks.
- **Reason for adjustment for balance** (if made): The authors state there are differences at baseline with pupils in the treatment schools more likely to be ever FSM eligible (27.4%) than

[16] Attrition should be measured at the pupil level (even for clustered trials) and from the point of randomisation to the point of analysis.

pupils in the control schools (24.7%). Second, pupils in the treatment schools had slightly lower KS1 test scores (effect size 0.06). However, as the majority of the missing data appears to be by chance there is no need to drop a padlock for balance.

- **Reason for adjustment for threats to validity** (if made): No threats to validity are present so no adjustment is needed.
- **Final padlock score:** initial score adjusted for balance and internal validity = The padlock rating would indicate a 4 padlock trial due to attrition.

Appendix C: Memorandum of understanding with schools

The Thinking, Doing, Talking Science Project
Funded by the Education Endowment Foundation

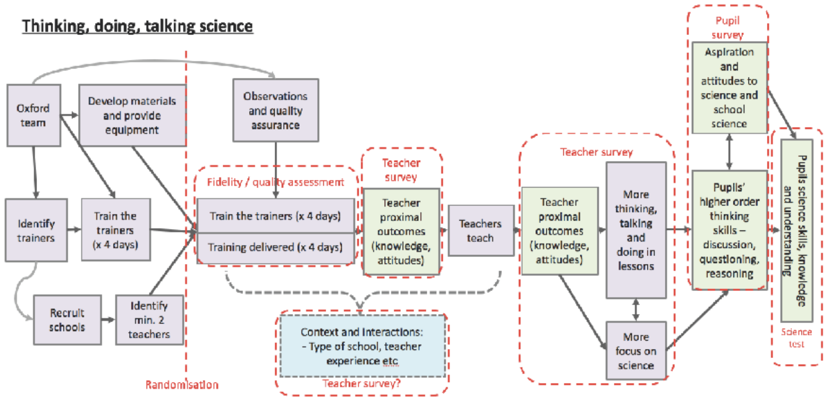
Background
The first phase of this project (the efficacy trial) was delivered by The Oxford Trust, whose public brand is Science Oxford, and Oxford Brookes University and evaluated by the Institute for Effective Education at the University of York. It resulted in a statistically significant positive impact on pupils' attainment, as well as on their attitudes to science.
https://educationendowmentfoundation.org.uk/evaluation/projects/thinking-doing-talking-science/

The project has been extended to an 'effectiveness trial' to see if the approach still has an impact with a wider model of delivery (and reduced financial support to schools).

The research design is a randomised controlled trial and it will involve 180-210 primary schools through 7 partners in different parts of England, each of whom aims to recruit 30 schools for a four-day training programme, with the four days spread across a school year. It will then be a randomised controlled trial for each region – where half the schools make up the intervention group and have their training in 2016-17 and the other half forms the control group. All schools (control and intervention) are required to administer a science assessment and attitude questionnaire at the end of the intervention year in July 2017. The control schools then have exactly the same training in 2017-18.

The project is focused on Year 5 pupils, so it will be important that all Year 5 teachers in the intervention schools attend the four-day training course. The project design is such that two teachers from each school are required to attend because of the beneficial mutual support that this allows. If there are two Year 5 classes in a school then both those teachers attend, but if there is only one Year 5 class the school can send any other teacher of their choice. If large schools are included and there are more than two Year 5 teachers, it is important that they all come to the training.

Schools will need to agree to a Memorandum of Understanding and provide some pupil data up front, including the names, gender, dates of birth and unique pupil numbers (UPNs) of all their Year 5 pupils and will need to complete a short proforma. The EEF has appointed an external evaluator and this is the American Institutes for Research (AIR). They will be responsible for the selection of the assessment and development of the attitude questionnaires. They will be working with NatCen Social Research, who will be administering the assessment in schools. AIR will also complete the statistical work and analysis and will write the final EEF research report. An overview of the project's logic model is shown below:



1

School Partnership Agreement: For all schools

Please read & sign this agreement & return to <training partner> with the required data at the address below by 20 June 2016.

The project aims to work between September 2016 and September 2018 to conduct a trial of the impact of a programme that aims to help primary schools to improve attainment in science by focussing on creative and challenging science lessons that encourage pupils to use higher order thinking skills.

Schools joining the project by signing this MOU will be informed in mid-July as to whether they have been randomly selected to be part of the intervention group (receiving training in 2016/17) or the control group (receiving training 2017/18).

This agreement is between <training partner>, AIR, NatCen Social Research and <school name>.

Under this partnership agreement <training partner> agrees to:

- Inform schools whether they have been randomly selected to be part of the intervention group (receiving training in 2016/17) or the control group (receiving training in July 2017 and 2017/18) by 13 July 2016.
- Deliver 4 training days
    - in September/early October, November, February and April/May of the 2016/17 school year to the intervention group
    - in July 2017, September/early October 2017, November 2017, February 2018 to the control group
- Provide equipment (worth £200), that complements the training, to the school in the year that the training is delivered.
- Provide a grant to the school on completion of the four training days (£400 for two teachers, £700 for three teachers and £1000 for four teachers).
- Provide the school with a report of the project findings. N.B. These will be overall project findings and schools will not receive individual assessment results.

Under this partnership AIR and NatCen Social Research agree to:

- Keep all information collected confidential. Our study has been approved by AIR's Institutional Review Board (IRB), which is responsible for ensuring the protection of study participants including plans to keep data secure. Reports of results will never include the name of any administrator, teacher, or pupil in the study.
- Do their best to accommodate your school's needs in scheduling data collection activities. We will schedule the pupil assessments at times that are as convenient as possible for school staff.
- Communicate clearly with you regarding study plans. We will let you know immediately if there are any changes to the plan and answer your questions promptly.
- Reduce burden on school staff as much as possible. We have limited pupil assessment and survey questions to information deemed essential to gather to answer the critical research questions this study is addressing.

Under this partnership agreement <school name, Address 1, Address2, Address 3, Town, Post code> agrees to:

- Provide a named contact and their contact details.
- Send an opt-out consent form to the parents of all Year 5 pupils (currently in Year 4) and keep a record of the number of pupils who are opted out by their parents.
- Complete a survey provided by NatCen Social Research. The survey will collect the UPN, first name, surname, and date of birth for all current Year 4 pupils (whose parents have not opted out). It also will ask about professional development opportunities available to Year 5 teachers and the number of and structure of Year 5

2

classes (i.e. number of Year 5 classes, if any Year 5 classes are multi-year). The survey must be submitted at the latest by Friday 1 July 2016.

- Send all Year 5 teachers to each of the 4 training days arranged in 2016-17 (intervention group) or in July 2017 and 2017-18 (control group). If there is only one Year 5 teacher, another teacher must also attend.
- Use the approaches developed during the training days with their Year 5 pupils and provide evaluation feedback through a brief teacher survey about their practice.
- Provide access to the researchers from NatCen Social Research to administer a science assessment and attitude questionnaire to all 2016-17 Year 5 pupils in summer term 2017. Before the assessment takes place, the School will take part in a telephone call with NatCen staff to schedule and plan the assessment visit. The school will ensure that a member of staff is present on the day of the assessment to support the pupils and manage their behaviour. NatCen will distribute and collect assessments and surveys.

Signature: Training partner name: Training partner organisation: Date:	Signature: Print Name: School name: Name of main contact: Telephone number: E-mail: Date:
Signature: Sami Kitmitto, Ph.D. Principal Investigator American Institutes for Research Date: 11 May 2016	Signature: Rakhee Patel NatCen Social Research Date: 11 May 2016

Please return one signed copy to: <Trainer Name, Email or by post to <Trainer Name, Organisation, Address, Post code>

3

Appendix D: Parent notification letter and pupil opt-out form

Dear Parent

Re: Thinking, Doing, Talking Science project for current Year 4 pupils

I am writing to let you know about an exciting project in which your child's school is participating next year. Thinking, Doing, Talking Science is a professional development programme for teachers with a focus on creative, practical and challenging science lessons. It was developed by Science Oxford and Oxford Brookes University. Through a trial in Oxfordshire, funded by the Education Endowment Foundation (EEF), this programme has been shown to significantly improve pupils' attainment in and attitudes towards science. The current project will evaluate whether the programme continues to be effective when it is delivered by other primary science experts across England. The project will be independently evaluated by American Institutes for Research (AIR).

AIR will be randomly assigning participating schools to either an 'intervention' or a 'control' group. Teachers in the 'intervention group' will receive training in 2016/17 while teachers in the 'control group' will receive training in 2017/18.

Towards the end of their time in Year 5, pupils in all participating schools will complete a short science assessment and a questionnaire about their views of science that will be used to evaluate the programme. The test and questionnaire will be part of normal classroom activities and take approximately 50 minutes in total to complete. The results from the assessment will only be used for the research and will not alter the teacher's assessment of your child's progress. The assessments will be invigilated, collected and processed by NatCen Social Research. The results will only be accessible to the researchers in the NatCen and AIR teams working on the project. The data will be kept confidential, in accordance with the Data Protection Act. Only group results of the programme evaluation will be published. We will not use your child's name or the name of the school in any report arising from the research.

We are asking for your permission to obtain information about your child from the school (including name, date of birth and Unique Pupil Number (UPN); please, see attached form). Information provided will be shared with a small group of researchers at NatCen Social Research and AIR. We will use the information provided to link with the National Pupil Database (which is held by the Department for Education and shared with the Department for Education, EEF, EEF's data contractor FFT Education and in an anonymised form to the UK Data Archive). Once your child's information is included in the data set, the data will be anonymised and no one will be able to identify individual pupils.

Complete the slip overleaf only if you do not want your child's school to share your child's information with the evaluators and wish your child to opt-out of the project evaluation.

If you would like more information about this project, you can contact me, the principal researcher Sami Kitmitto from AIR at tdts@air.org or Rakhee Patel from NatCen at tdts@natcen.ac.uk or 0207 549 7002. The project has been approved by AIR's Institutional Review Board, the ethics board for AIR. If you have concerns or questions about the ethics of this project, please contact the Chair of AIR's Institutional Review Board at IRBChair@air.org.

Kind regards,

Sami Kitmitto
Principal researcher, American Institutes of Research

The Thinking, Doing and Talking Science Project
Opt out form

Please discuss this project with your child.

If you are both happy for your child to participate in the evaluation and for information about your child to be used in this project, you do not need to do anything.

If you would rather your child's school did not share information about your child for use in this evaluation and would rather your child did not participate in the evaluation, please tick the box below and return it to your child's teacher.

☐ I DO NOT consent for my child to participate in the project evaluation nor for his/her information to be shared.

Parent name (BLOCK CAPITALS) ……………………………………………………

Parent signature: ………………………………………………………………………

Date ………………………………………………

(Please return the completed form to your child's class teacher)

Appendix E: Information on data linkage provided via a study website

More information about linking data

What is linking data?

Data linking is when information about one person which is kept in different places is brought together. Data linking helps to better understand if taking part in programmes helps people.

Why is linking data important for this study?

Linking your child's data will help us, the evaluation team made up of AIR and NatCen, understand if the programme has any benefits. It will also let us compare different groups of pupils and look at whether children experience benefits later on in their life as a result of the programme.

What data will be linked?

We want to link school data with data in the National Pupil Database.

1. **School Data:** Your child's school holds data, such as date of birth and test results, about your child.

2. **The National Pupil Database (NPD)**: the Department of Education manages this database on all pupils in England. We will request these data:
    - test and exam results
    - age
    - gender
    - free school meal entitlement
    - whether a child is looked after/ in care

3. **Unique Pupil Number (UPN):** Every child who is in school in England has a Unique Pupil Number. Schools use this number to collect and record information about pupils. This number stays with a child throughout their education, even if they change schools.

We will use your child's UPN to link his/her school data to data held on the NPD. How we will do this is explained below.

How does the data linkage happen?

This is how the process works for the Thinking, Doing and Talking Science Project Evaluation.

1. If you agree, your child's school will send us your child's Unique Pupil Number and other information such as test results.

2. Your child's Unique Pupil Number will be sent to the Department of Education, who manage the National Pupil Database.

3. They'll use this to find your child's records.

4. Your child's school records will then be sent to us. When we receive them, we will link them to your child's test results to create a new set of data to be used by researchers. This will include the information from your child's records listed above. It will **not** include your child's name, or your postal address. No children will be identified in our report.

5. At the end of the project, the data will be submitted to two archives in an anonymised format. These are:

- The Education Endowment Foundation's data archive managed by FFT Education
- The National Data Archive

How is data transferred and stored?

Schools will send us your child's data securely, either in a secure email or by saving the data to a secure website.

We will save all data relating to your child on a secure system. Only researchers working on the project will be able to access it.

At the end of the project, we will securely delete all the data we hold on your child.

What do I have to do?

If you agree to let us link your child's data to our study, you do not have to do anything. If you do not want your child's data to be linked to our study, please complete the Thinking, Doing and Talking Science Project Opt-Out Form and return it to your child's teacher. If you complete the opt-out form we will not link your child's data to this study.

What if I change my mind?

If you change your mind, you can agree to data linkage and you can withdraw your permission to data linkage at any time. Please just inform your child's teacher.

Appendix F: Pupil science assessment

Science Knowledge Questionnaire

YEAR 5

First name ______________________________

Last name ______________________________

School______________________________

- The test is 40 minutes long
- You will need: pen, pencil, rubber, ruler, protractor and calculator
- The test starts with easier questions
- Try to answer all of the questions
- Write all your answers on the test paper—do not use any rough paper
- Check your work carefully
- Ask your teacher if you are not sure what to do

For marker's use only	Total Marks __________

01

What three things do ALL animals do?
Tick THREE boxes.

Move	☐	Swim	☐
Reproduce	☐	Run	☐
Smile	☐	Grow	☐
Read	☐	Cry	☐

Total out of 3 ______

SCIENCE KNOWLEDGE QUESTIONNAIRE | Year 5

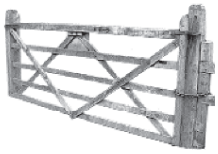
1

02

Peter sorted these objects into sets according to the material from which they are made. He called this set 'wood'.

table

gate



chair

wood

a. Write the name of the material for this set.

envelope

tissue

comic

1 mark ______

b. Write the name of the material for this set.

key

ring

coin

1 mark ______

03

Dylan and Gemma want to know if the root always grows downwards from a seed.

They place some sunflower seeds in different positions on a paper towel. They water them.

pointing upwards **pointing downwards** **pointing sideways**

This is how the seeds grew.

a. Complete this table to show their results.

starting positions of seeds	pointing upwards	pointing ______________	pointing ______________
direction roots grew	down	______________	______________

2 marks ______

b. Dylan said they could use just one seed pointing each way.

Gemma said it was better to use three seeds in each position.

Why is Gemma's suggestion better?

____

1 mark _______

c. What did they find out about the starting position of the seeds and the direction in which the roots grew?

____

1 mark _______

Total out of 4 _______

04

Olivia wanted to find a good material for blocking out sounds.

She held tubes over her ears and filled them with different materials.

tissue cotton wool newspaper straw

a. Write down the question that she is investigating.

____

____

1 mark ______

b. Michael made the sounds to see whether Olivia could heat them.
Why did he tell Olivia to close her eyes as he made the sounds?

____

____

1 mark ______

05

Colin pushed his car twice. It went much further the second time.

Explain why the car went further on the second push.

Tick ONE box.

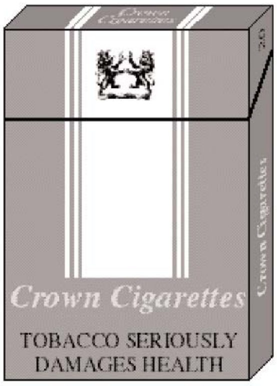
Colin pushed harder the second time.	☐
The car was heavier the first time.	☐
Cars sometimes do that.	☐
The wheels on the car are not straight.	☐

Total out of 1 ________

06

Dan saw this warning printed on a cigarette packet.

TOBACCO SERIOUSLY DAMAGES HEALTH



Which body organ is most likely to be damaged by contact with inhaled tobacco smoke?

____

Total out of 1 _______

SCIENCE KNOWLEDGE QUESTIONNAIRE | Year 5

7

07

One year a farmer killed all the foxes on his land because they were killing his chickens. Later in the year he found that his lettuces were being eaten by rabbits.

Why were there so many more rabbits on his land later in the year?

____

____

____

Total out of 1 ______

08

Sam and Hannah put 150g of each crushed rock into a separate jar. They covered each material with water.

Next they remove the materials from the water and measure the mass of each again.

a. What property of the materials were they investigating?
Tick ONE box.

absorbency	☐	hardness	☐
solubility	☐	transparency	☐

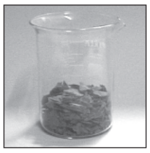
1 mark ______

b. What would they need to keep the same to make sure their test was fair?
Tick ONE box.

Time of day the investigation is done	☐
The shape of the containers	☐
How long the rocks are kept in the water	☐
The colour of the containers	☐

1 mark ______

Results

material		mass when dry (grams)	mass when wet (grams)
	**brick**	150	180
	**sandstone**	150	165
	**slate**	150	155
	**marble**	150	160

c. What was the mass of the brick when wet?

____ grams

1 mark ______

d. Which building material would you predict is least permeable to water?
Tick ONE box.

brick ☐ slate ☐

sandstone ☐ marble ☐

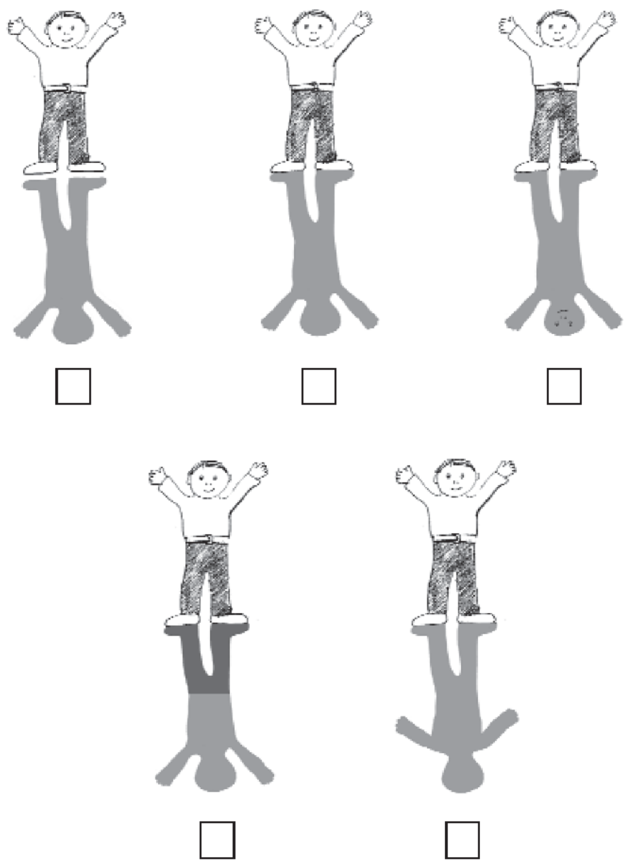
1 mark ______

Total out of 4 ______

09

Which of these drawings correctly shows a boy and the shadow he makes standing in the playground on a sunny day?

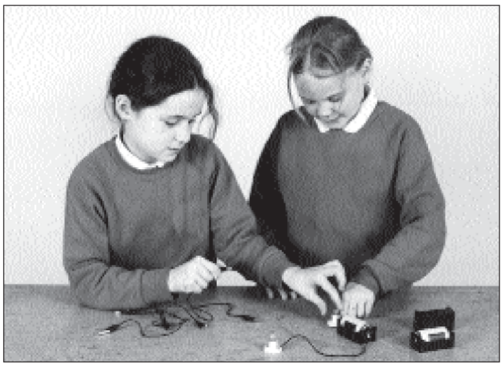
Tick ONE box.



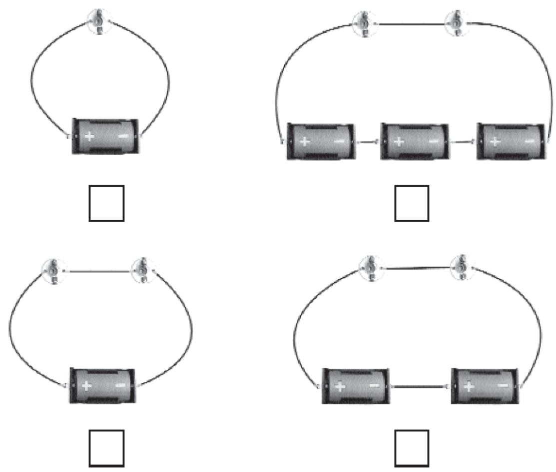
Total out of 1 ______

10

Claire and Jane test the brightness of the light with one bulb and then two bulbs in their circuits.



a. Which TWO circuits should they use to make their test fair? Tick TWO boxes.



1 mark ______

b. What is the ONE and ONLY thing they should change as they carry out their test?

____

1 mark ______

11

Yeast is a micro-organism that is useful to humans.

Tick ONE correct way in which it is used.

- Spread on cuts ☐
- Use in baking bread ☐
- Use for sunburn ☐
- Makes hair grow ☐
- Use for plant fertiliser ☐

Total out of 1 ______

12

Look at the photograph of the flower below.

a. Label the part of the flower at A.

1 mark ______

b. Seeds need water to germinate. Yet in Britain many seeds produced by wild plants in the autumn do not germinate until the following spring. Why do seeds germinate and grow better in spring?

____

1 mark ______

c. What must happen to the pollen in order for a flower to be pollinated?

____

1 mark ______

Total out of 3 ______

13

Look at the table below. Some of these changes are reversible and others are not.

Complete the table to show the reversible changes and the irreversible changes.

The first one has been done for you.

	Reversible Change	or	Irreversible Change
Burning paper	☐		☑
Melting an ice cube	☐		☐
Boiling an egg	☐		☐
Stretching elastic	☐		☐

Total out of 1 ______

14

Megan is plucking a string of her guitar to make a sound.

a. How does plucking the string cause it to make a sound?

____

____

1 mark ______

b. Tick TWO boxes to show how she could make a sound with higher pitch.

She could pluck the same string harder. ☐

She could pluck the same string more softly. ☐

She could tighten the same string before she plucks it. ☐

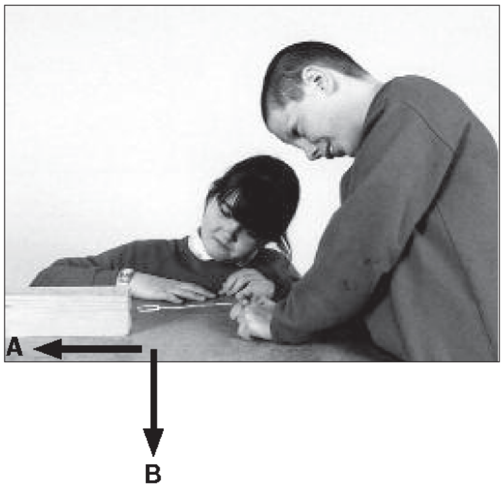
She could pluck one of the thicker strings. ☐

She could pluck one of the thinner strings. ☐

1 mark ______

15

The arrows on the photograph show two of the forces acting on the block when it is being pulled across the table.



Complete the following sentences to name the forces A and B.

a. Force A is

____

1 mark ______

b. Force B is

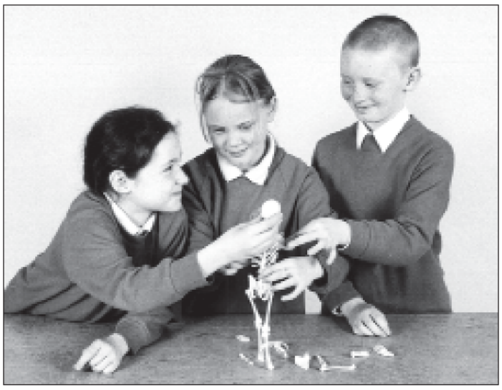
____

1 mark ______

Total out of 2 ______

16

One of the functions of your skeleton is to protect some of your body's organs.



Complete these sentences.

a . Your skull protects your ____

b . Your ____________________________ protect your heart and your lungs.

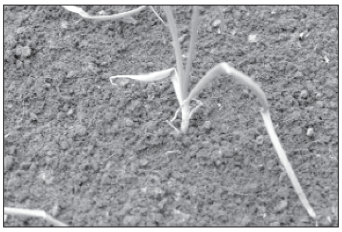
Total out of 1 _______

17

Look at the two photographs of the same part of a garden where an onion had been planted.

All the weeds were cleared.

first photo



second photo

The second photograph was taken after one week.

Many small plants have grown, although none of them were planted.

Where did all the small plants come from?

____

____

1 mark _______

18

The children observe water droplets forming inside the window.

a. What is the name of the process by which droplets of water form inside the window?

____

1 mark _______

b. Where do the droplets of water come from?

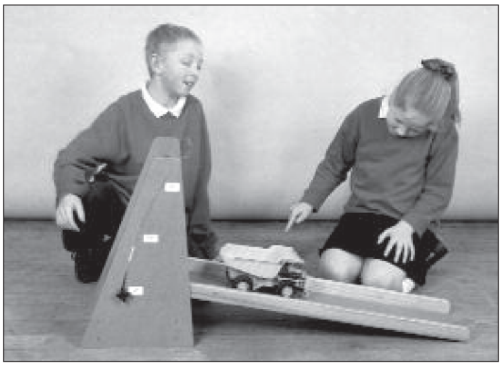
____

1 mark _______

Total out of 2 _______

19

The children investigate the effect of changing the angle of the slope on the distance the car travels.



a. They want to be sure that their test is fair.

What is the ONE and ONLY thing they should change as they do their investigation?

____

1 mark ______

b. Write TWO things they should keep the same.

i ____

1 mark ______

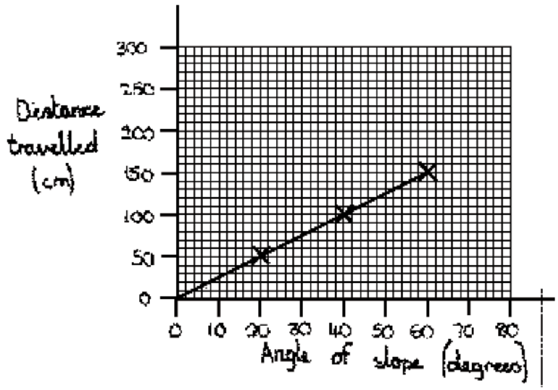
ii ____

1 mark ______

c. **Results**

angle of slope (degrees)	distance travelled (cm)
20	50
40	100
60	150

They made a line graph of the results.



Use the graph to predict the angle of the slope which would make the car travel 200cm.

__________________________ degrees

1 mark ______

d. Mary suggests that they should measure the distance travelled three times for each angle.

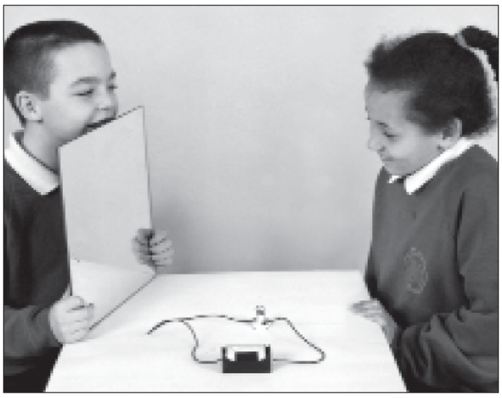
Why is it a good idea to measure the distance three times?

____

1 mark ______

20

Yasmin can see the bulb reflected in the mirror.



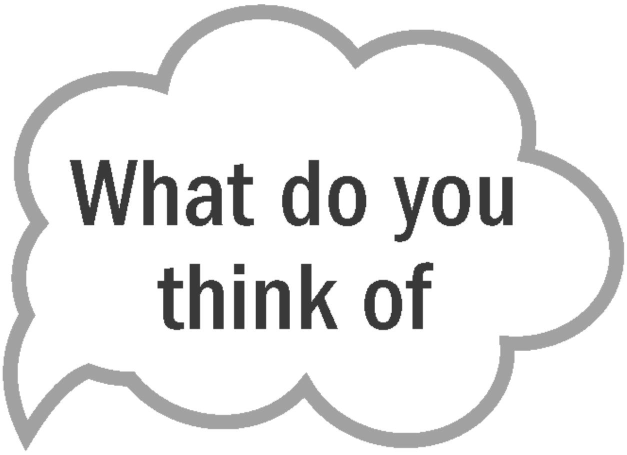
Draw arrows on the picture to show how light travels to allow her to see the reflection of the bulb.

Total out of 2 ______

0950a_04/17

END OF TEST

Appendix G: Pupil survey



This booklet asks questions about you and your interest in science.

There are no right or wrong answers. We want to know what you think.

My name is ________________________________

My teacher is ________________________________

My school is ________________________________

01

Learning science at school

Do you agree with these views?
(Please tick only one box in each row.)

	Agree a lot	Agree a bit	Not sure	Disagree a bit	Disagree a lot
a. Science lessons make me think	☐	☐	☐	☐	☐
b. I look forward to my science lessons	☐	☐	☐	☐	☐
c. Science lessons are interesting	☐	☐	☐	☐	☐
d. I would like to do more science at school	☐	☐	☐	☐	☐
e. Science is fun	☐	☐	☐	☐	☐
f. We spend a lot of time in science lessons copying from the board	☐	☐	☐	☐	☐
g. I enjoy discussions in science lessons	☐	☐	☐	☐	☐
h. Science lessons are boring	☐	☐	☐	☐	☐

SCIENCE KNOWLEDGE SURVEY 1

02

Learning science at school

Do you agree with the following statements?
(Please tick only one box in each row.)

	Agree a lot	Agree a bit	Not sure	Disagree a bit	Disagree a lot
a. I find science difficult to understand	☐	☐	☐	☐	☐
b. I am just not good at science	☐	☐	☐	☐	☐
c. I think science is more for boys	☐	☐	☐	☐	☐
d. I understand everything in my science lessons	☐	☐	☐	☐	☐
e. We often have discussions in science lessons	☐	☐	☐	☐	☐
f. We do a lot of writing in science lessons	☐	☐	☐	☐	☐
g. It is important that we learn science	☐	☐	☐	☐	☐
h. I like thinking about scientific ideas	☐	☐	☐	☐	☐

03

About practical work in school science

Do you agree with these views?
(Please tick only one box in each row.)

	Agree a lot	Agree a bit	Not sure	Disagree a bit	Disagree a lot
a. Doing practical work in science lessons is fun	☐	☐	☐	☐	☐
b. We already know what will happen when we do science practical work	☐	☐	☐	☐	☐
c. I can decide what to do for myself in science practical work	☐	☐	☐	☐	☐
d. We do practical work in most science lessons	☐	☐	☐	☐	☐
e. I look forward to doing science practicals	☐	☐	☐	☐	☐
f. Practical work in science is boring	☐	☐	☐	☐	☐
g. Solving science problems is enjoyable	☐	☐	☐	☐	☐

04

Finally, is there anything you would like to say about your science lessons this year?

Did you answer every one? You did? Then, you have finished!

0950b_04/17

Thank you very much

SCIENCE KNOWLEDGE SURVEY 3

Appendix H: Supporting tables for indices on pupil attitudes toward science

Exploratory and confirmatory factor analysis was conducted to support the creation of indices derived from the pupil survey responses. A priori, we hypothesised that there were three domains represented in the questions: pupil interest in science, pupil self-efficacy for science, and reports of science activities. In the statistical analysis plan we proposed analysing the first two as they are aligned with the theory of action for TDTS.

In exploratory factor analysis (EFA), we found that items loaded as hypothesised with very little cross loading (Table H1).

Table H1: Exploratory factor analysis Factor loadings for a three-factor model

TDTS pupil survey	Factor 1	Factor 2	Factor 3
Factor interpretation	Interest	Self-efficacy	Activity
Do you agree with these views?			
01a Science lessons make me think.	0.5401		
01b I look forward to my science lessons.	0.7811		
01c Science lessons are interesting.	0.7871		
01d I would like to do more science at school.	0.7280		
01e Science is fun.	0.8206		
01f We spend a lot of time in science lessons copying from the board.			0.3508
01g I enjoy discussions in science lessons.	0.6042		
01h Science lessons are boring.	0.6499		
Do you agree with the following statements?			
02a I find science difficult to understand.		0.6772	
02b I am just not good at science.		0.6068	
02c I think science is more for boys.			
02d I understand everything in my science lessons.		0.5025	
02e We often have discussions in science lessons.	0.3007		
02f We do a lot of writing in science lessons.			
02g It is important that we learn science.	0.6259		
02h I like thinking about scientific ideas.	0.6719		
Do you agree with these views?			
03a Doing practical work in science lessons is fun.	0.5878		
03b We already know what will happen when we do science practical work.			0.4269
03c I can decide what to do for myself in science practical work.			
03d We do practical work in most science lessons.			
03e I look forward to doing science practicals.	0.6451		
03f Practical work in science is boring.	0.4493		0.3510
03g Solving science problems is enjoyable.	0.6473		

Note. Factor loadings less than 0.3 were suppressed.

Based on the EFA results, a final model was chosen selecting items that had loadings greater than 0.3. For “interest in science” 13 items were selected and for “self-efficacy for science” 3 items were selected. Confirmatory factor analysis was conducted using this model and statistics revealed an acceptable fit (Table H2). For RMSEA and SRMR, lower values indicate better fit and generally 0.08 is a cut-off for acceptable fit. For CFI and TLI, higher values indicate better fit and generally 0.90 is a cut-off for acceptable fit.

Table H2: Confirmatory factor analysis model fit for two-factor models: Interest and self-efficacy

Interest and self-efficacy	Fit statistic
Root Mean Square Error of Approximation (RMSEA)	0.082
Standardized Root Mean Square Residual (SRMR)	0.054
Comparative Fit Index (CFI)	0.903
Tucker Lewis Index (TLI)	0.887

Indices were constructed by summing responses across items and standardised to mean zero and standard deviation of one. Reliability was estimated using Cronbach’s alpha (Table H3). The interest in science index had very high reliability while the self-efficacy index had moderate reliability which is likely due to the small number of items included in the index.

Table H3: Reliability coefficients for interest and self-efficacy indices

	Cronbach’s alpha
Interest	0.9141
Self-efficacy	0.6788

Appendix I: Analysis of missing data

To analyse the patterns of missing data, we focused on our primary outcome, pupil science assessment score, and investigated whether a missing outcome was related to our pre-test measures, KS1 scores in mathematics and reading/writing or pupil FSM status. In addition, we investigated whether the relationships of these measures with a missing outcome differ by treatment and control schools. To conduct this investigation, we estimated a series of logit regression models and used the likelihood ratio test (LRT) chi-squared to examine our hypotheses.

Table I1 reports the *p*-values of the LRTs for the KS1 test scores and for pupil FSM status. Tests including covariates are also reported. For the KS1 test score model, the covariates included: pupil FSM, region indicators, and the two variables used for minimisation, school FSM category and an indicator for whether the school had 3 or more year 5 teachers. For the pupil FSM model, the covariates included: the KS1 test scores, region indicators, school FSM category, and an indicator for whether the school had 3 or more year 5 teachers. The sign on the coefficients and LRT results indicate significant relationships: missingness in the outcome variable is negatively related to KS1 test scores and is positively related to pupil FSM status. In other words, low achieving and FSM pupils are more likely to be missing outcome data.

Table I1: LRT of models predicting missing pupil science attainment, KS1 test scores and pupil FSM status

Test Conducted	Likelihood Ratio Test *p*-value
KS1 test scores vs. Null	0.000
KS1 test scores with covariates vs. only covariates	0.000
Pupil FSM vs. Null	0.000
Pupil FSM with covariates vs. only covariates	0.000

Table I2 reports *p*-values of the LRTs examining whether the relationship of missing science assessment with KS1 test scores and pupil FSM status, respectively, differ by treatment and control schools. In this analysis, all models additionally include the treatment indicator as a covariate. The contrast is between a model including an interaction between the treatment indicator and the variable of interest compared to the model without the treatment interaction. This is a test of whether that interaction is statistically significant which would indicate a different relationship in treatment and control schools. The LRT results do not find a significant difference when the interaction terms is included. This indicates that missingness in the outcome variable is not differentially related to KS1 test scores and pupil FSM status across treatment and control schools.

Table I2: LRT of models predicting missing pupil science attainment, KS1 test scores and pupil FSM status interacted with Treatment

Test Conducted	Likelihood Ratio Test *p*-value
KS1 test scores X Treatment vs. KS1 test scores	0.920
KS1 test scores X Treatment with covariates vs. KS1 test scores with covariates	0.908
Pupil FSM X Treatment vs. Pupil FSM	0.524
Pupil FSM X Treatment with covariates vs. Pupil FSM with covariates	0.545

In summary, we found that missing the science assessment outcome was related to KS1 test scores and pupil FSM status, but not differentially in treatment and control schools. Given the level of missing data in the outcome variable and the relationship with key covariates we employed multiple imputation to fill in missing values and estimated our primary analysis models on these data to check the robustness of results.

Appendix J: Supporting tables for outcome analyses

Tables J1 through J6 provide detailed regression results from the primary and additional analysis models of pupil science attainment, interest in science, and self-efficacy for science. The tables report estimated coefficients on covariates which generally exhibited patterns expected. For example, higher KS1 test scores (reference category is ”W”) were positively related to attainment on the study science assessment as was pupil IDACI (reference category is ”Low”) while pupil FSM status was negatively related (reference category is ”not ever FSM”). Higher school FSM levels were negatively associated with pupil scores on the study science assessment (reference category is “low”). The coefficient reported for the “intervention” variable is the regression adjusted impact estimate and one can see how this changes (or not) as covariates are added to the model.

The tables also report the variance estimated at the pupil (level 1) and school (level 2) levels, the total variance, and the ICC.

Table J1: Impact of TDTS on pupil science attainment

	Primary model								
	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value
Intervention	0.08	0.295	0.791	0.11	0.256	0.668	0.12	0.248	0.635
KS1 ReadWriting_1	-0.21	0.843	0.807	-0.31	0.843	0.715	0.53	0.882	0.545
KS1 ReadWriting_2C	0.37	0.875	0.676	0.25	0.875	0.771	1.04	0.912	0.253
KS1 ReadWriting_2B	1.55	0.868	0.074	1.41	0.868	0.104	2.15	0.906	0.018
KS1 ReadWriting_2A	2.99	0.874	0.001	2.85	0.874	0.001	3.54	0.913	0.000
KS1 ReadWriting_3	5.64	0.881	0.000	5.47	0.881	0.000	6.11	0.921	0.000
KS1 ReadWriting_4	9.70	3.340	0.004	9.55	3.338	0.004	10.05	3.339	0.003
KS1 Maths_1	1.41	1.154	0.223	1.44	1.153	0.212	0.86	1.181	0.465
KS1 Maths_2C	3.35	1.174	0.004	3.38	1.173	0.004	2.83	1.199	0.018
KS1 Maths_2B	4.99	1.178	0.000	5.02	1.177	0.000	4.47	1.203	0.000
KS1 Maths_2A	6.57	1.183	0.000	6.59	1.182	0.000	6.06	1.208	0.000
KS1 Maths_3	8.70	1.189	0.000	8.72	1.188	0.000	8.17	1.216	0.000
KS1 Maths_4	10.89	3.424	0.001	10.83	3.423	0.002	10.50	3.423	0.002
Sch_FSM_Med				-1.29	0.319	0.000	-1.14	0.310	0.000
Sch_FSM_High				-2.36	0.332	0.000	-1.93	0.331	0.000
Sch_Yr5 Teachers_3More				-0.38	0.338	0.263	-0.35	0.326	0.278
Sch_Region_B				-0.16	0.512	0.760	-0.19	0.495	0.700
Sch_Region_C				0.00	0.484	0.992	0.02	0.468	0.958
Sch_Region_D				0.59	0.503	0.239	0.53	0.487	0.280
Sch_Region_E				-0.43	0.511	0.403	-0.31	0.495	0.537
Sch_Region_F				-0.51	0.530	0.333	-0.54	0.513	0.292
Sch_Region_G				-0.22	0.567	0.702	-0.33	0.548	0.548
Pup_Female							0.16	0.107	0.134
Pup_Age_10							-0.15	0.124	0.233

	Primary model								
	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value
Pup_Age_11							-0.56	1.180	0.633
Pup_FSM							-0.57	0.132	0.000
Pup_IDACI_Med							0.45	0.145	0.002
Pup_IDACI_High							0.55	0.169	0.001
Constant	10.80	0.990	0.000	12.27	1.064	0.000	11.81	1.085	0.000
Model Statistics	**Estimate**	**Std. error**		**Estimate**	**Std. error**		**Estimate**	**Std. error**	
L2 var.estimate	3.64	0.438		2.59	0.329		2.38	0.311	
L1 var.estimate	20.28	0.329		20.28	0.329		20.13	0.327	
Total variance	23.92			22.87			22.51		
ICC	0.15			0.11			0.11		
Number of observations	7,806			7,806			7,763		

Table J2: Impact of TDTS on pupil science attainment for FSM pupils

	Primary model			Interaction Model		
	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value
Intervention	0.31	0.360	0.386	0.02	0.293	0.944
KS1 ReadWriting_1	1.13	1.457	0.437	-0.06	0.843	0.943
KS1 ReadWriting_2C	1.86	1.511	0.218	0.47	0.874	0.591
KS1 ReadWriting_2B	3.04	1.501	0.043	1.65	0.867	0.057
KS1 ReadWriting_2A	4.81	1.515	0.001	3.06	0.873	0.000
KS1 ReadWriting_3	7.00	1.539	0.000	5.68	0.880	0.000
KS1 ReadWriting_4				9.55	3.335	0.004
KS1 Maths_1	0.66	1.782	0.710	1.29	1.152	0.262
KS1 Maths_2C	2.16	1.816	0.234	3.24	1.172	0.006
KS1 Maths_2B	3.64	1.826	0.046	4.84	1.176	0.000
KS1 Maths_2A	5.12	1.839	0.005	6.41	1.181	0.000
KS1 Maths_3	7.06	1.866	0.000	8.50	1.188	0.000
KS1 Maths_4				10.68	3.420	0.002
Pupil FSM				-0.87	0.181	0.000
Pupil FSM X Treatment				0.25	0.254	0.330
Constant	9.70	1.382	0.000	11.11	0.988	0.000
Model Statistics	**Estimate**	**Std. error**		**Estimate**	**Std. error**	
L2 var.estimate	3.47	0.620		3.37	0.412	
L1 var.estimate	18.63	0.618		20.22	0.328	
Total variance	22.10			23.59		
ICC	0.16			0.14		
Number of observations	2,000			7,806		

Table J3: Impact of TDTS on pupil interest in science

	Primary model											
	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value
Intervention	0.12	0.049	0.013	0.12	0.050	0.015	0.12	0.047	0.010	0.12	0.046	0.010
KS1 ReadWriting_1				0.32	0.187	0.090	0.33	0.186	0.078	0.47	0.197	0.018
KS1 ReadWriting_2C				0.30	0.193	0.122	0.31	0.193	0.108	0.46	0.203	0.024
KS1 ReadWriting_2B				0.32	0.192	0.098	0.33	0.192	0.084	0.48	0.202	0.017
KS1 ReadWriting_2A				0.30	0.193	0.118	0.32	0.193	0.101	0.48	0.203	0.019
KS1 ReadWriting_3				0.42	0.194	0.029	0.44	0.194	0.023	0.61	0.205	0.003
KS1 ReadWriting_4				1.06	0.707	0.134	1.11	0.707	0.117	1.29	0.711	0.069
KS1 Maths_1				-0.75	0.258	0.004	-0.76	0.258	0.003	-0.82	0.266	0.002
KS1 Maths_2C				-0.73	0.262	0.005	-0.74	0.262	0.005	-0.80	0.270	0.003
KS1 Maths_2B				-0.69	0.263	0.008	-0.70	0.263	0.007	-0.77	0.271	0.005
KS1 Maths_2A				-0.71	0.264	0.007	-0.72	0.264	0.006	-0.79	0.272	0.004
KS1 Maths_3				-0.65	0.265	0.014	-0.66	0.265	0.012	-0.74	0.273	0.007
KS1 Maths_4				-0.60	0.728	0.410	-0.60	0.728	0.410	-0.71	0.732	0.328
Sch_FSM_Med							0.09	0.058	0.126	0.07	0.058	0.224
Sch_FSM_High							0.09	0.061	0.145	0.06	0.063	0.355
Sch_Yr5 Teachers_3More							-0.12	0.061	0.047	-0.12	0.060	0.050
Sch_Region_B							-0.22	0.094	0.017	-0.23	0.093	0.014
Sch_Region_C							-0.14	0.089	0.111	-0.15	0.088	0.092
Sch_Region_D							-0.22	0.092	0.016	-0.22	0.092	0.015
Sch_Region_E							-0.26	0.093	0.006	-0.28	0.093	0.003
Sch_Region_F							-0.30	0.097	0.002	-0.31	0.097	0.001
Sch_Region_G							-0.36	0.103	0.001	-0.36	0.103	0.000

	Primary model											
	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value
Pup_Female										-0.04	0.023	0.116
Pup_Age_10										0.02	0.027	0.441
Pup_Age_11										0.07	0.250	0.765
Pupil FSM										0.02	0.029	0.581
Pup_IDACI_Med										-0.01	0.031	0.733
Pup_IDACI_High										-0.07	0.036	0.052
Constant	-0.04	0.035	0.235	0.31	0.216	0.153	0.48	0.228	0.034	0.43	0.235	0.064
Model Statistics	**Estimate**	**Std. error**		**Estimate**	**Std. error**		**Estimate**	**Std. error**		**Estimate**	**Std. error**	
L2 var.estimate	0.09	0.012		0.09	0.012		0.08	0.011		0.08	0.011	
L1 var.estimate	0.91	0.015		0.91	0.015		0.91	0.015		0.91	0.015	
Total variance	1.00			1.00			0.98			0.98		
ICC	0.09			0.09			0.08			0.08		
Number of observations	7,777			7,572			7,572			7,531		

Table J4: Impact of TDTS on pupil interest in science for FSM pupils

	Primary model			Interaction Model		
	Estimate	**Std. error**	***p*-value**	**Estimate**	**Std. error**	***p*-value**
Intervention	0.16	0.066	0.018	0.13	0.051	0.014
Pupil FSM				0.01	0.038	0.864
Pupil FSM X Treatment				-0.01	0.054	0.810
Constant	-0.04	0.047	0.380	-0.04	0.036	0.234
Model Statistics	**Estimate**	**Std. error**		**Estimate**	**Std. error**	
L2 var.estimate	0.08	0.019		0.09	0.012	
L1 var.estimate	0.97	0.033		0.91	0.015	
Total variance	1.05			1.00		
ICC	0.08			0.09		
Number of observations	1,907			7,777		

Table J5: Impact of TDTS on pupil self-efficacy for science

	Primary model											
	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value
Intervention	0.09	0.032	0.004	0.09	0.033	0.006	0.09	0.031	0.003	0.09	0.031	0.003
KS1 ReadWriting_1				-0.07	0.186	0.723	-0.05	0.186	0.769	0.07	0.196	0.721
KS1 ReadWriting_2C				-0.10	0.193	0.611	-0.08	0.193	0.661	0.05	0.202	0.787
KS1 ReadWriting_2B				-0.10	0.191	0.596	-0.09	0.191	0.647	0.07	0.201	0.734
KS1 ReadWriting_2A				-0.09	0.192	0.628	-0.08	0.192	0.684	0.09	0.202	0.646
KS1 ReadWriting_3				0.06	0.194	0.741	0.08	0.194	0.679	0.27	0.204	0.179
KS1 ReadWriting_4				0.03	0.718	0.968	0.07	0.718	0.925	0.30	0.720	0.675
KS1 Maths_1				-0.02	0.250	0.923	-0.04	0.250	0.859	-0.04	0.257	0.869
KS1 Maths_2C				0.00	0.254	0.986	-0.02	0.254	0.952	-0.02	0.261	0.930
KS1 Maths_2B				0.14	0.255	0.579	0.12	0.255	0.630	0.10	0.261	0.708
KS1 Maths_2A				0.20	0.256	0.441	0.18	0.256	0.480	0.14	0.262	0.596
KS1 Maths_3				0.37	0.257	0.150	0.35	0.257	0.167	0.29	0.264	0.280
KS1 Maths_4				0.06	0.737	0.935	0.03	0.737	0.966	-0.09	0.738	0.899
Sch_FSM_Med							0.03	0.038	0.468	0.02	0.039	0.520
Sch_FSM_High							0.03	0.040	0.466	0.04	0.044	0.399
Sch_Yr5 Teachers_3More							-0.05	0.038	0.158	-0.06	0.038	0.144
Sch_Region_B							-0.17	0.063	0.008	-0.17	0.063	0.007
Sch_Region_C							-0.16	0.060	0.007	-0.17	0.060	0.006
Sch_Region_D							-0.19	0.063	0.003	-0.19	0.063	0.002
Sch_Region_E							-0.12	0.062	0.051	-0.13	0.062	0.037
Sch_Region_F							-0.14	0.066	0.028	-0.15	0.066	0.020
Sch_Region_G							-0.32	0.068	0.000	-0.33	0.068	0.000
Pup_Female										-0.11	0.023	0.000

	Primary model											
	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value
Pup_Age_10										0.04	0.027	0.127
Pup_Age_11										-0.15	0.253	0.561
Pupil FSM										-0.01	0.028	0.636
Pup_IDACI_Med										0.02	0.030	0.460
Pup_IDACI_High										-0.02	0.034	0.503
Constant	-0.04	0.023	0.080	-0.18	0.209	0.396	-0.03	0.215	0.906	-0.13	0.222	0.573
Model Statistics	**Estimate**	**Std. error**		**Estimate**	**Std. error**		**Estimate**	**Std. error**		**Estimate**	**Std. error**	
L2 var.estimate	0.02	0.005		0.02	0.005		0.02	0.004		0.02	0.004	
L1 var.estimate	0.97	0.016		0.94	0.015		0.94	0.015		0.94	0.015	
Total variance	1.00			0.97			0.96			0.96		
ICC	0.02			0.03			0.02			0.02		
Number of observations	7,954			7,748			7,748			7,705		

Table J6: Impact of TDTS on pupil self-efficacy for science for FSM pupils

	Primary model			Interaction Model		
	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value
Intervention	0.06	0.050	0.266	0.10	0.035	0.004
Pupil FSM				-0.06	0.038	0.124
Pupil FSM X Treatment				-0.03	0.054	0.594
Constant	-0.07	0.036	0.048	-0.03	0.025	0.308
Model Statistics	**Estimate**	**Std. error**		**Estimate**	**Std. error**	
L2 var.estimate	0.02	0.009		0.02	0.005	
L1 var.estimate	0.97	0.032		0.97	0.016	
Total variance	0.99			1.00		
ICC	0.02			0.02		
Number of observations	1,980			7,954		

Appendix K: Examples of analysis code

All analysis was conducted using STATA 15 (StataCorp., 2017). The primary models for science attainment (pup_SA_total), interest in science (pup_interest), and self-efficacy (pup_selfeff) for science were estimated using the following STATA commands:

```
mixed pup_SA_total sch_treatment i.pup_readwrit i.pup_maths || xSchool_ID:, mle variance
mixed pup_interest sch_treatment || xSchool_ID : , mle variance
mixed pup_selfeff sch_treatment || xSchool_ID : , mle variance
```

The science attainment model includes the categorical KS1 test scores in reading and writing (pup_readwrit) and mathematics (pup_maths) as covariates. All models include school random-effects (xSchool_ID is the school identifier in the data).

Appendix L: Multiple imputation analysis

Because more than 10% of the pupil observations analysed were missing the primary outcome variables or the baseline KS1 test scores, a fully conditional multiple imputation approach was used to complete cases where information was missing to examine the robustness of results to missing data. Only pupils in schools that did not drop out were included in the imputations and analysis.

Stata's MI IMPUTE CHAINED command was used to generate five imputed values for missing observations in any variable used in the analysis. The CHAINED imputation procedure allows for appropriate imputation models to be used based on the type of variable: the science assessment score was imputed using a truncated regression model; the pupil survey responses, the KS1 score categories, and the pupil IDACI level were imputed using an ordered logit model; and gender was imputed using a logit model. Variables with complete information that were used in the imputation included pupil age, pupil FSM, school percentage FSM, school IDACI level, school percentage White British level, and school percentage English as an additional language level. Impact models were estimated using imputed data but complete case results were also generated for comparison.

Results for primary models for are presented in tables L1 and L2 for all pupils and tables L3 and L4 for FSM pupils. Results are very similar though slightly attenuated compared to those reported using complete cases.

Table L1: Primary analysis using multiple imputation

	Raw means					Effect size	
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model	Hedges g (95% CI)	*p*-value
Science assessment (controlling for KS1 scores)	4,415 (0)	19.3 (19.11; 19.49)	4,448 (0)	19.52 (19.32; 19.73)	8,863	0.00 (-0.08; 0.09)	0.917
Interest in science	4,415 (0)	0.06 (0.03; 0.08)	4,448 (0)	-0.05 (-0.08; -0.01)	8,863	0.10 (0.01; 0.19)	0.030
Self-efficacy for science	4,415 (0)	0.04 (0.01; 0.07)	4,448 (0)	-0.03 (-0.06; 0)	8,863	0.08 (0.02; 0.14)	0.011

Table L2: Primary analysis using multiple imputation effect size calculations

			Intervention	Control	
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	n (missing)	Pooled variance
Science assessment (controlling for KS1 scores)	-0.23	0.03	4,415 (0)	4,448 (0)	39.805
Interest in science	0.10	0.10	4,415 (0)	4,448 (0)	1.013
Self-efficacy for science	0.07	0.08	4,415 (0)	4,448 (0)	1.002

Table L3: Primary analysis using multiple imputation for FSM pupils

	Raw means					Effect size	
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	n in model	Hedges g (95% CI)	*p*-value
Science assessment (controlling for KS1 scores)	1,210 (0)	16.85 (16.5; 17.19)	1,100 (0)	16.93 (16.57; 17.3)	2,310	0.05 (-0.07; 0.16)	0.412
Interest in science	1,210 (0)	0.06 (0; 0.13)	1,100 (0)	-0.03 (-0.11; 0.04)	2,310	0.12 (-0.01; 0.25)	0.069
Self-efficacy for science	1,210 (0)	-0.02 (-0.08; 0.04)	1,100 (0)	-0.07 (-0.14; -0.01)	2,310	0.05 (-0.05; 0.15)	0.287

Table L4: Primary analysis using multiple imputation for FSM pupils effect size calculations

			Intervention	Control	
Outcome	Unadjusted differences in means	Adjusted differences in means	n (missing)	n (missing)	Pooled variance
Science assessment (controlling for KS1 scores)	-0.09	0.28	1,210 (0)	1,100 (0)	34.028
Interest in science	0.10	0.12	1,210 (0)	1,100 (0)	1.076
Self-efficacy for science	0.05	0.05	1,210 (0)	1,100 (0)	0.989

Appendix M: Supporting tables for complier analyses

Tables M1 and M2 present results for the two stages of the instrumental variable complier analysis. The first stage estimates the extent to which assignment to the treatment group predicts compliance using the same covariates used in the second stage. The second stage estimates the impact of compliance on the outcome using the predicted values from the first stage as an instrument for compliance. The coefficients on the variable "Predicted Compliance" in table M2 provide these estimates.

Table M1: First stage regression results for CACE analysis

	For Science Assessment Model			For Attitude Index Models		
	Estimate	Std. error	*p*-value	Estimate	Std. error	*p*-value
Intervention	0.97	0.002	0.000	0.97	0.002	0.000
KS1 ReadWriting_1	-0.01	0.017	0.766			
KS1 ReadWriting_2C	0.00	0.018	0.797			
KS1 ReadWriting_2B	-0.01	0.018	0.690			
KS1 ReadWriting_2A	0.00	0.018	0.962			
KS1 ReadWriting_3	-0.01	0.018	0.505			
KS1 ReadWriting_4	0.00	0.083	0.971			
KS1 Maths_1	0.03	0.022	0.115			
KS1 Maths_2C	0.03	0.022	0.173			
KS1 Maths_2B	0.03	0.022	0.202			
KS1 Maths_2A	0.03	0.023	0.252			
KS1 Maths_3	0.03	0.023	0.142			
KS1 Maths_4	0.05	0.084	0.565			
Constant	-0.02	0.016	0.151	0.00	0.002	1.000
Number of observations	8,586			8,863		

Table M1: Second stage regression results for CACE analysis

	Science Assessment			Interest in science			Self-efficacy for science		
	Est.	Std. error	*p*-value	Est.	Std. error	*p*-value	Est.	Std. error	*p*-value
Predicted Compliance	0.08	0.303	0.791	0.13	0.050	0.013	0.09	0.033	0.004
KS1 ReadWriting_1	-0.21	0.843	0.807						
KS1 ReadWriting_2C	0.37	0.875	0.676						
KS1 ReadWriting_2B	1.55	0.868	0.074						
KS1 ReadWriting_2A	2.99	0.874	0.001						
KS1 ReadWriting_3	5.64	0.881	0.000						
KS1 ReadWriting_4	9.70	3.340	0.004						
KS1 Maths_1	1.40	1.154	0.224						
KS1 Maths_2C	3.35	1.174	0.004						
KS1 Maths_2B	4.99	1.178	0.000						
KS1 Maths_2A	6.57	1.183	0.000						
KS1 Maths_3	8.69	1.189	0.000						
KS1 Maths_4	10.89	3.424	0.001						
Constant	10.80	0.989	0.000	-0.04	0.035	0.235	-0.04	0.023	0.080
Model Statistics	Est.	Std. error		Est.	Std. error		Est.	Std. error	
L2 var.estimate	3.64	0.438		0.09	0.012		0.02	0.005	
L1 var.estimate	20.28	0.329		0.91	0.015		0.97	0.016	
Total variance	23.92			1.00			1.00		
ICC	0.15			0.09			0.02		
Number of observations	7,806			7,777			7,954		

Education Endowment Foundation

The Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP
www.educationendowmentfoundation.org.uk