Thinking, Doing, Talking Science
(second re-grant – a two-armed, cluster randomised trial)

Evaluation report

June 2025

Lyn Robinson-Smith, Caroline Fairhurst, Louise Tracey, Imogen Fountain, Katie Whiteside, Tom Davill, Jess Hugill-Jones, Maria Turkenburg van-Diepen, Maja Brakovic-Thomas, Rachel Bottomley-Wise, Pam Hanley, Louise Elliott, and David Torgerson

The Education Endowment Foundation (EEF) is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, nurseries, and colleges to improve teaching and learning for 2–19-year-olds through better use of evidence.

We do this by:

- **Summarising evidence.** Reviewing the best available evidence on teaching and learning and presenting in an accessible way.
- **Finding new evidence.** Funding independent evaluations of programmes and approaches that aim to raise the attainment of children and young people from socio-economically disadvantaged backgrounds.
- **Putting evidence to use.** Supporting education practitioners, as well as policymakers and other organisations, to use evidence in ways that improve teaching and learning.

We were set-up in 2011 by the Sutton Trust partnership with Impetus with a founding £125m grant from the Department for Education. In 2022, we were re-endowed with an additional £137m, allowing us to continue our work until at least 2032.

For more information about the EEF or this report please contact:

Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team at the University of York that involved: Lyn Robinson-Smith, David Torgerson, Pam Hanley, Louise Elliott, Imogen Fountain, Caroline Fairhurst, Katie Whiteside, Rachel Carr, Jess Hugill-Jones, Louise Tracey, Maria Turkenburg van-Diepen, Maja Brakovic-Thomas and Rachel Bottomley-Wise. The co-principal investigators were Lyn Robinson-Smith and David Torgerson (from December 2022), Pam Hanley (until December 2022), and Louise Elliott (until August 2022).

Contact details:

Lyn Robinson-Smith

York Trials Unit

Department of Health Sciences

University of York

Email: lyn.robinson-smith@york.ac.uk

Acknowledgements

The evaluation team would like to thank all the schools, teachers, parents, carers, and pupils that were involved in this evaluation for giving their time and efforts in participating in the evaluation of Thinking, Doing, Talking Science (TDTScience).

We would also like to thank the delivery team Science Oxford for their cooperation in all aspects of this evaluation, and the Education Endowment Foundation and the Wellcome Trust for their funding and support.

This work was undertaken in the Office for National Statistics Secure Research Service using data from ONS and other owners and does not imply the endorsement of the ONS or other data owners.

Executive summary

The project

Thinking, Doing, Talking Science (TDTScience) is a continuing professional development (CPD) programme—and a pedagogical teaching approach—which aims to increase primary aged pupils' positive attitudes towards science and improve overall science attainment by providing more opportunities for the development of higher order thinking (HOT) skills through creative practical science and quality discussion. The CPD programme, created by Science Oxford and Oxford Brookes University, was delivered by Science Oxford through a network of accredited TDTScience trainers. It involved four and a half-days of face-to-face delivery to Year five primary teachers, delivered at five points across the year with gap tasks in between to encourage teachers to try out specific strategies and evaluate these in their own contexts. Teachers embedded the TDTScience approaches into their normal science lessons for the whole class and were provided with TDTScience course resources (physical and online) and inexpensive science equipment.

The effectiveness trial was a cluster randomised controlled trial, involving 180 primary schools from across England. Schools were randomised to receive either TDTScience or continue with usual teaching practice. The impact of the programme on science attainment was measured as a primary outcome, using a Year 5 Science Assessment. Science attitudes were assessed as a secondary outcome, using a Science Attitudes Questionnaire. Mixed methods research was used to explore the relationship between delivery and programme outcomes. This involved interviews, surveys with teachers and pupils, and observations of teacher training and lessons.

The evaluation took place between January 2022 and July 2024 and included a comprehensive train-the-trainer programme (which is evaluated in a separate report). The evaluation of the programme involved two cohorts of Year 5 pupils in subsequent years (Cohort 1 in 2022–2023 and Cohort 2 2023–2024). This report focuses on Cohort 1 delivery in 2022–2023 only. An addendum report will be issued in 2025, which includes the assessment of Cohort 2, exploring the impact of TDTScience once teachers have had time to embed TDTScience into their teaching practice for the full year. It will also include a longitudinal analysis following Year 5 pupils in Cohort 1 into Year 6, assessed using a Year 6 Science Assessment, and Key Stage 2 reading and maths scores. The evaluation was co-funded by the Education Endowment Foundation (EEF) and the Wellcome Trust, and undertaken by York Trials Unit, University of York.

Table 1: Key conclusions

Key conclusions
1. Pupils in TDTScience schools, made the equivalent of zero months' additional progress in science attainment, on average, compared to pupils in other schools. This result has a high security rating.
2. Pupils' attitudes to science in TDTScience schools were positively impacted, on average, compared to pupils in other schools. This positive change in attitudes was of a small size.
3. Pupils receiving free school meals (FSM) in TDTScience schools made the equivalent of zero months' additional progress in science attainment, on average, compared to pupils receiving FSM in other schools. These results may have lower security than the overall findings because of the smaller number of pupils.
4. Teachers changed their practice in accordance with TDTScience's theory of change. This was evidenced through lesson observation and noted by pupils in focus group discussions who indicated more time for discussion, practical work, and less writing. Teachers observed, and pupils reported, high levels of engagement and self-efficacy in science, which translated into improvements in pupils' attitudes to science, but was not reflected at the end of the academic year in science attainment.
5. TDTScience was implemented as intended, with high compliance and fidelity. Interviews and surveys reported that teachers felt TDTScience improved their lesson planning, and they were more confident teaching science, with clear strategies, techniques, and practicals embedded in the programme.

EEF security rating

These findings have a high security rating. This 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. Around 16% of pupils who started the trial were not included in the final analysis, primarily because pupils were absent during testing. There was a small difference between the intervention and control group at the start of the trial, but this was not considered a threat to validity and was controlled for in the impact analysis.

Additional findings

Pupils in TDTScience schools made, on average, zero months' additional progress compared to those in the control group equivalent. This is our best estimate of impact, which has a high security rating. As with any study, there is always some uncertainty around the result: the possible impact of this programme includes small negative effects of one month's less progress and positive effects of up to two months' additional progress. Pupils receiving FSM also made zero months' additional progress in schools receiving TDTScience compared to peers in control schools.

Pupils' attitudes to science in TDTScience schools were positively impacted, on average, compared to pupils in the control group equivalent. This change in attitudes was a small size. This suggests that TDTScience approaches can improve pupils' attitudes to science, without any overall negative impact on attainment. Teacher surveys indicate the time spent on science lessons, and practical science was similar across groups, which suggests that TDTScience approaches, and pedagogy were driving this change.

Teachers implemented TDTScience as intended, with high compliance and fidelity. Teachers reported that TDTScience improved their lesson planning and that they were more confident teaching science. Lesson observations also indicated that teachers changed their practice in accordance with the TDTScience ethos and this was supported by pupils in focus groups noting that science now had more time for discussions and practicals, and less writing.

Teachers reported perceived positive impacts on their Year 5 pupils and found TDTScience to be very inclusive to all groups of pupils, with teachers observing, and pupils reporting, high levels of engagement and self-efficacy in science, which aligns with the theory of change. Although, these factors did not translate into improvements in pupils' science attainment at the end of Year 5, they did result in improvements in attitudes to science. Overall, barriers to implementing TDTScience were minimal, with some teachers reporting the need to follow existing schemes of work, issues relating to aligning the programme with the content and volume of the national science curriculum, school culture, and/or curriculum time.

The findings in this evaluation relating to pupils' science attainment contrast with a previous small-scale efficacy trial involving Year 5 pupils in 41 schools (Hanley, Slavin, and Elliott, 2015), where pupils of teachers trained in TDTScience made three months' additional progress in science. The findings of this evaluation are consistent with those found in the first effectiveness trial of TDTScience in 205 schools (Kitmitto *et al.,* 2018), which showed no evidence of additional progress in science attainment for Year 5 for pupils of teachers trained in TDTScience. The main difference between this trial and the previous effectiveness trial (Kitmitto *et al.,* 2018) was the more comprehensive train-the-trainer model.

This evaluation will further explore the impact of TDTScience in an addendum report, providing the evidence of impact once teachers have had time to embed TDTScience into their teaching practices for the full year, to see whether there is a 'soak' effect on a second cohort of Year 5 pupils (Cohort 2). It will also include a longitudinal analysis following Year 5 pupils in Cohort 1 into Year 6, assessed using a Year 6 science assessment, and Key Stage 2 reading and maths scores.

Cost

The estimated cost of TDTScience is £9,210.90 per school over a three-year period, or £71.40 per child per year when averaged over three years. This figure is based on 43 pupils per school per year and includes costs for intervention delivery, training, staff travel, optional staff cover, and the purchase of optional additional materials and resources.

Impact

Table 2: Summary of impact on primary outcome

Outcome / group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils (intervention; control)	P-value	EEF cost rating
Cohort 1 Year 5 Science Assessment	0.02 (-0.07, 0.12)	0	🔒🔒🔒🔒	6,107 (3,175; 2,932)	0.66	**£** £ £ £ £
Cohort 1 Year 5 Science Assessment (FSM)	0.00 (-0.11, 0.10)	0	N/A	1,852 (916; 936)	0.96	N/A

Introduction

Study rationale and background

Government biennial sampling tests estimated that only 21.2% of pupils achieved the expected standard in science in 2018 (Standards and Testing Agency, 2019).

Markwick and Reiss (2023) surmised from studies conducted over the past two decades (prior to the COVID-19 pandemic) that primary school science in England had been in decline. In 2021, the Office for Standards in Education, Children's Services and Skills (Ofsted) also reported that there had been a decline in science teaching in primary schools due to the COVID-19 pandemic where many schools taught a reduced science curriculum. This was often so that maths and English teaching and achievement could be prioritised (Ofsted, 2021).

The primary science experience heavily influences subsequent subject attitudes, however, for many years primary teachers have expressed a lack confidence and feelings of being ill-equipped to teach science (e.g. see Markwick and Reiss, 2023; Ofsted, 2021; Harlen and Qualter, 2018; Slavin *et al.,* 2014). Jarvis and Pell (2004) reported that many primary school teachers lack the deep understanding of scientific concepts that are required to teach science well. The Parliamentary Office for Science and Technology (2003) identified that many primary school teachers tended to concentrate too much on teaching factual content in science and not enough on core scientific ideas. As such, Bianchi, Whittaker, and Poole (2021) reported that pupils regularly experienced 'fun [science] activities', however, they did not necessarily develop a deep understanding of the associated scientific concepts.

Thinking, Doing, Talking Science (TDTScience) is a continuing professional development (CPD) programme developed by Science Oxford and Oxford Brookes University that aims to enable the teachers to adapt their pedagogy to plan and confidently teach creative science lessons that overtly encourage their pupils' higher order thinking (HOT) within dedicated discussion times, challenging practical work and focused pupil recording. Together the aim is to increase pupils' engagement with science, their attitude towards it, and overall science attainment.

According to Bloom's revised taxonomy, the cognitive process is a continuum of thinking skills. At the start of that continuum are the lower order thinking skills of remembering and understanding, with applying, analysing, evaluating, and creating categorised as HOT skills (Aksela, 2005; FitzPatrick and Schulz, 2015). HOT skills essentially encourage pupils to engage more actively and understand more deeply, which in turn helps them to better remember the factual knowledge that they are being taught (Jensen *et al.,* 2014).

TDTScience encourages pupils to think and talk about scientific concepts in every science lesson through dedicated discussion slots linked to the topic being taught. For example, a teacher may show three pictures and ask, which one is the 'Odd One Out' inviting pupils to offer an answer along with their reasoning. Teachers are encouraged to allow pupils to contribute until all ideas have been saturated in a bid to promote thinking and creativity. Teachers facilitate their pupils' thinking through doing purposeful practical science, providing them with frequent opportunities for them to apply, analyse, and evaluate through creative investigations, problem-solving, and other types of enquiry activity. Pupils do not record everything they do in a science lesson, as the priority is 'focused recording' for learning directly relevant to the lesson's learning objectives, so that time for thinking, doing, and talking is prioritised. In a small-scale efficacy trial involving Year 5 pupils in 41 schools (Hanley, Slavin, and Elliott, 2015), pupils of teachers trained in TDTScience made three months' additional progress in science, with a particularly positive effect among girls and pupils with low prior attainment. There were indications that the approach might be especially beneficial for pupils eligible for free school meals (FSM), but this required further exploration. There was an apparent positive impact on attitudes towards science. However, a subsequent effectiveness trial in 205 schools (Kitmitto *et al.,* 2018) failed to show evidence of additional progress for most pupils, although pupils eligible for FSM made a small amount of additional progress and pupils' interest and self-efficacy in science showed a small improvement.

Other than scale, the main change between the two trials related to teacher training. In contrast to the efficacy trial, the previous effectiveness trial used a 'train-the-trainers' model, rather than the developers training the teachers directly, the CPD days were reduced from five days to four days, and the funding to cover two in-school preparation days/teacher was eliminated. There are several other instances of success in the Education Endowment Foundation (EEF) smaller trials not being replicated at scale. One commonly shared change on scale-up is the adoption of the 'train-the-trainer' model: practice shifts from training of the teachers being delivered directly by the developers in the first trial to delivery by trainers trained by the developers, in the second trial and often delivering the programme training for the first time.

Train-the-trainer is the theory that a group of individuals can be given training in a new concept and then go on to train a large group in this newly acquired skill (Ray *et al.,* 2012). The model is increasingly implemented in business, healthcare settings, and education (Gask, Coupe, and Green, 2019), as it can be used to reach a large audience in a relatively cost-effective way (Wedell, 2005). Although there is not a wealth of literature regarding this model, evidence suggests that this multi-level process can generate a number of problems. The main disadvantage is the dilution of the knowledge as it is passed down (Hayes, 2000). Reasons for this dilution include: knowledge transfer and the ability to train others (Turner, Brownhill, and Wilson, 2017); focus of knowledge at the uppermost levels and 'transmissive training' (Hayes, 2000); lack of social and cultural awareness (Bax, 2002); and lack of confidence of the trained to teach their new knowledge (Dichaba and Mokhele, 2012). Other problems include rate of staff turnover (Gask, Coupe, and Green, 2019) and the lack of 'proactive technical assistance' after the initial training of the trainer (Ray *et al.,* 2012).

After the first TDTScience effectiveness trial, the Science Oxford team proposed strengthening the train-the-trainers model by splitting training into three consecutive stages:

- trainers, acting as if they were teachers, receive training in TDTScience from the developers;
- trainers are trained by the developers to deliver TDTScience training to teachers; and
- trainers train pre-trial teachers in TDTScience.

Such recommendations were rolled-out in a one-year 'pre-trial', which ran from July 2021 to July 2022. The pre-trial offered the developers the opportunity to train-the-trainers prior to this main effectiveness trial, improve quality assurance, for instance by observing trainers delivering to the pre-trial schools, and to improve trainer resources. Standley *et al.* (2023) evaluated the pre-trial train-the-trainers model for efficacy and fidelity and reported that all three stages of the training stages were needed to produce confident and competent trainers for the main trial. Being trained in TDTScience as if they were teachers (Stage 1) meant trainers could familiarise themselves with the content and approach and make suggestions for possible improvements. The train-the-trainers element (Stage 2) was valued particularly for providing the opportunity to plan and practise with their delivery partners. The real-life delivery of TDTScience to teachers (Stage 3) acted as a rehearsal for the main trial. They could refine how to run the days, gain confidence, and identify where more practice and support was needed. Stage 3 also gave them an opportunity to explore the extent to which they could make the delivery their own (by inserting personal anecdotes, for instance) while maintaining fidelity to the TDTScience programme. Although the fundamentals of the programme were already firmly fixed, the developers were willing to consider trainers' suggested improvements at every stage of the pre-trial, and to adopt them if they were considered appropriate and easy to incorporate. During Stage 3, some trainers had ideas for amendments for delivery that did not affect the content of the training and could be adopted in their own practice in the main trial. There was little feeling among trainers that the training model could have been improved by eliminating or compressing any elements, even among those who had delivered TDTScience before, and the developers also expressed satisfaction with the process.

The main effectiveness trial, reported herein, ran from January 2022 to July 2023 and focused on the TDTScience training of the teachers and the experience, attitudes, and performance of their Year 5 pupils in terms of science attainment (Cohort 1). The primary outcome measure was a general science assessment developed by the Centre for Industry Education Collaboration (CIEC) and York Trials Unit, University of York that covered the Year 5 curriculum (Joshi *et al.,* 2022). The primary outcome measure used previously in the first TDTScience effectiveness trial was outdated; its creation (Abrahams *et al.,* 2014) preceded the new science curriculum (Department for Education, 2013) with its changed content and emphases (e.g. more focus on 'working scientifically'/science enquiry). Therefore, a new outcome measure that mapped to the current science curriculum was developed for use in this trial as not suitable, previously validated outcome was in existence. A secondary outcome of pupil science attitude questionnaire was also collected. This was a slightly modified version of the one used in the previous efficacy and effectiveness trials, which included two additional questions on practical work during science lessons. This trial aimed to assess the impact of TDTScience across two cohorts: at the end of the year teachers' received TDTScience training (Cohort 1 Year 5 pupils); and one year after receiving the training (Cohort 2 Year 5 pupils), to investigate whether the effect of TDTScience appears to be modified in any way after teachers have received the entire training package (which is delivered across the academic year) and had a greater opportunity to incorporate TDTScience in their science teaching.

The results of the primary and secondary outcome for Cohort 1 are reported here, along with the results of a thorough implementation and process evaluation (IPE), which also took place and included lesson observations, interviews, and surveys of trainers, teachers, and pupils.

In Autumn Term 2023, the next cohort of Year 5 pupils (Cohort 2) in participating main trial schools were recruited and in June 2024 to July 2024 these pupils were given the same science assessment and attitude questionnaire as Cohort

1. Participating schools were requested to complete an online survey as part of the follow-up IPE. Cohort 1 Year 5 pupils were also followed into Year 6 and requested to complete a science assessment at the end of Year 6 (June 2024 to July 2024) to assess any longitudinal impact of the programme. This assessment was developed by the York Trials Unit, University of York, and Manchester Metropolitan University and is designed to reflect the current curriculum, have a mix of question types, and have an emphasis on 'working scientifically'. Pupils were also followed up based on their attainment in Mathematics and Reading in the Year 6 Standard Assessment Tests (SATs), using data from the National Pupil Database (NPD) to identify any spill-over effects. The results of the longitudinal follow-up will be reported in an addendum report in 2025.

Intervention

TDTScience aims to develop teachers' delivery of science lessons so that they actively encourage their pupils' HOT. This is achieved through dedicated discussion slots, known as the 'Bright Ideas Time'. These are less than or equal to ten-minute slots dedicated to discussion of a prompt, which can be a question with multiple possible answers that allows pupils to think deeply and creatively. For example, the teacher may show pupils three or four photographs and ask: 'Which is the odd one out and why?' or another example is a 'Positive Minus Interesting (PMI)' scenario where pupils are asked to share their thoughts about the advantages and disadvantages of a situation. The discussion is linked to the topic being taught and through practical science, providing pupils with frequent opportunities for creative investigations and problem-solving.

In the first year of the trial, Year 5 teachers were invited to attend CPD sessions, four of which were spread throughout the first two terms, with a further half-day in the third (summer) term. Teachers were given 'gap' tasks/strategies to use with their classes between the sessions and encouraged to reflect on their implementation, discuss with their in-school colleagues, and then feedback at the next CPD session. Any Year 5 teachers who joined the school during the academic year (2022–2023) should have inherited the previous teacher's file and received input from the other participating teacher(s) in their school as well as attending any subsequent training sessions.

The second year of the trial (with a second cohort of Year 5 pupils plus following the first cohort of Year 5 pupils into Year 6) will examine the legacy of the TDTScience training and any effects of embedding of the TDTScience practices.

The Template for Intervention Description and Replication (TIDieR) in Table 3 below outlines the details for the procedure for the first year of the trial. Where relevant, differences in the second year are summarised in square brackets.

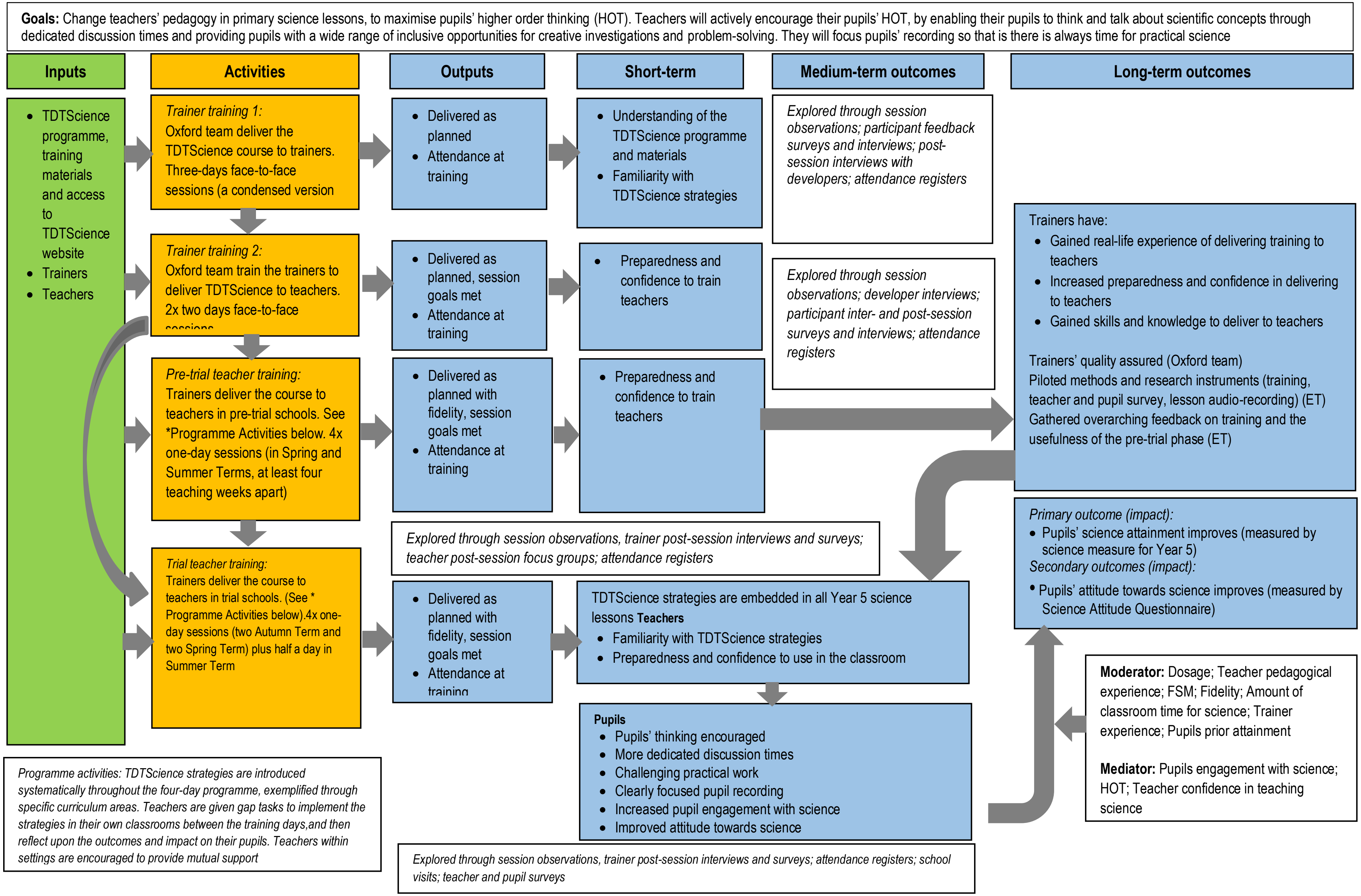
Further information about the TDTScience intervention is included in the logic model (**Error! Reference source not found.**), which predicts that, by encouraging HOT skills, pupils will engage more deeply and actively with science, develop an increased interest and self-efficacy, and ultimately increase their science attainment outcomes.

Table 3: Description of the programme using the TIDieR checklist

Aspect of TIDieR	Exemplification relating to the evaluation
Brief name	Thinking, Doing, Talking Science (TDTScience).
Why: Rationale, theory and/or goal of essential elements of the intervention	TDTScience aims to improve Year 5 pupils' HOT skills and science outcomes by improving teachers' delivery of science lessons. Government biennial sampling tests estimate that only 21.2% of pupils achieved the expected standard in science in 2018. There has been previous efficacy (Hanley, Slavin, and Elliott, 2015) and effectiveness (Kitmitto *et al.,* 2018) trials of the intervention. This second effectiveness trial incorporated an evaluation of an amended train-the-trainer model to the one used in the first effectiveness trial (reported by Standley *et al.,* 2023) as well as evaluating the intervention itself.
Who: Recipients of the intervention	Teachers in all Year 5 classes will be invited to attend training. Where there is only one Year 5 class, another teacher, ideally the science subject lead, will also receive training. There will be no external delivery of the intervention to any new Year 5 teachers in the second year of the trial. They will be reliant on the teachers trained in the first year and accompanying physical/online materials for any learning about TDTScience. To caveat, the transfer of information from a trained Year 5 teacher to a new teacher can be assumed; this practice would likely vary considerably from school to school and be dependent on the teacher and their circumstances.
What: Physical or informational materials used in the intervention	Each teacher will receive hard copies of all TDTScience course resources in a ring binder and some low-value science equipment at the point of course delivery. They will also have ongoing access to online versions of TDTScience course resources via a dedicated website (https://TDTScience.org.uk/). [First year only, although online access will still be available in the second year to Year 5 teachers in the intervention arm.]

What: Procedures, activities, and/or processes used in the intervention	All Year 5 teachers in intervention schools will receive four one-day CPD sessions; these will be held towards the beginning and end of the first two terms of the academic year. There will be a further half-day during the third term to share good practice and provide advice on disseminating TDTScience within their schools. [First year only.] Between training sessions, teachers will be asked to try some strategies with their classes and then feedback and discuss at the next session. At least two teachers from each school will participate in the intervention and TDTScience teachers will be encouraged to provide informal peer support for each other within schools.
Who: Intervention providers/implementers	Qualified TDTScience trainers, certified to deliver the training course will deliver the TDTScience course to Year 5 classroom teachers. [First year only.]
How: Mode of delivery	Teachers attend group CPD sessions delivered face-to-face. Each session will be run once per region by a pair of trainers, with expected attendance of 20–40 teachers. [First year only.]
Where: Location of the intervention	CPD sessions will be run in each of the six regions. [First year only.] The regions will be spread across England as far as possible, but the final choice will be a pragmatic one based on the location and reach of the final team of trainers.
When and how much: Duration and dosage of the intervention	The CPD will consist of four one-day sessions spread over the first two terms of the academic year and a further half-day in the Summer Term. [First year only.]
Tailoring: Adaptation of the intervention	No adaptations anticipated.
How well (planned): Strategies to maximise effective implementation	Six of the 24 full-day CPD sessions will be observed by the evaluation team, one in each region, and teachers will be asked for feedback on the training in the teacher surveys and interviews as part of the process evaluation. The evaluation team will also use the teacher feedback designed by the delivery team (in consultation with the evaluators) and complete after each training session. Short interviews will be conducted with trainers after each observed training session, and they will also be asked to complete a brief survey to gather their feedback on the effectiveness of the session.

Figure 1: TDTScience logic model



Evaluation objectives

The latest version of the protocol (Hanley *et al*. 2023)and statistical analysis plan (SAP) (Fairhurst 2023) for the effectiveness trial can be found on the EEF project website.

Impact evaluation research questions

1. What is the impact of the TDTScience programme, in comparison to usual Year 5 provision, on the science attainment of Year 5 pupils? [primary outcome]

2. What is the impact of the TDTScience programme, in comparison to usual Year 5 provision, on pupils' attitudes towards science? [secondary outcome]

3. What is the impact of the TDTScience programme, in comparison to usual Year 5 provision, on the science attainment of Year 5 pupils who are eligible for FSM?

Additional research questions pertaining to Cohort 2 and the long-term impact of the TDTScience programme in Cohort 1 will be detailed in an addendum report to be published in 2025.

IPE research questions

1. To what extent was TDTScience implemented as planned?
    a. Training.
    b. Classroom practice.
2. What processes are involved for teachers and schools implementing TDTScience—what are the main facilitators and barriers?
3. What are the perceptions of teachers as regards TDTScience?
    a. What are their opinions about training and support, including cascading from colleagues where relevant?
    b. What are their views of TDTScience strategies and techniques?
    c. What impacts has TDTScience had on their classroom practice?
    d. How has it affected their engagement with and confidence in teaching science?
    e. How do they think it has impacted on pupils?
4. How do pupils respond to TDTScience?
    a. What is their experience of, and reaction to, the different TDTScience strategies?
    b. What is their experience of practical work in the science classroom?
    c. What is their engagement with science lessons?
5. How does TDTScience compare with practice in business-as-usual science lessons?
    a. What strategies and techniques are used in science lessons?
    b. How interested and engaged are teachers and pupils in science teaching and learning?
    c. What is the frequency and length of science lessons?
    d. What practical science takes place?
    e. How much training have Year 5 teachers received in science?

Ethics and trial registration

University of York Health Sciences Research Governance Committee (HSRGC) granted ethical approval for this study in May 2020. All outputs will be anonymised so that no setting or student will be identifiable in the report or dissemination of results. The statistical database will hold non-identifiable data. Confidentiality will be maintained and no-one outside of the evaluation team will have access to the database, which will be held securely on University of York servers.

All participating schools were asked to sign a Memorandum of Understanding (MoU) to cover the requirements of the project, along with a Data Sharing Agreement issued by University of York.

The trial's ISRCTN (International Standard Randomised Controlled Trial Number) Registration number is 52414724, available to view here: https://www.isrctn.com/ISRCTN52414724.

Data protection

All data collected for the trial was treated with the strictest confidence and processed and stored in compliance with the General Data Protection Regulation (GDPR) and Data Protection Act 2018. The University of York acted as data controller and data processor, as defined by the GDPR. The trial participants, including both pupils and teachers in participating schools and the TDTScience trainers, were the data subjects.

Personal data was processed under Article 6 (1) of the GDPR(Data Protection Act, 2018a) 'processing is necessary for the performance of a task carried out in the public interest' as the research was being conducted to support education provision in the UK and, if applicable, Special Category data under Article 9 (2) (j) (Data Protection Act, 2018b). A Data Protection Impact Assessment was conducted, and Data Sharing Agreements (see Technical Notes) were put in place with schools.

At the end of the study, data will be submitted to the Office for National Statistics (ONS) Secure Research Service (SRS) for archiving in the EEF data archive (managed by the EEF's archive manager) and will include data only individually identifiable to the Department for Education. For the purposes of research and archiving, the data will be linked with information about the pupils from the NPD and shared with the Department for Education, the EEF, the EEF's archive manager, the ONS, and potentially other research teams. Further matching to NPD and other administrative data may take place during subsequent research. The EEF will act as the data controller for the archive, which is managed on their behalf by the FFT and held in the ONS SRS.

Parents/carers were informed about the research though an information sheet (see Technical Notes) sent on behalf of the evaluation team by schools to parents/carers. Parents/carers were asked to return a signed 'withdrawal from research' form (see Technical Notes) if they did not wish to share their child's data with the evaluation team and/or they did not wish their child to take part in any assessments, surveys, or focus groups. This applied to both cohorts of pupils.

For the purposes of the research, details of participating pupils (i.e. name, date of birth, gender, and unique pupil number [UPN]) were collected from schools and further details from the NPD (FSM, early years foundation stage profile [EYFSP], and Key Stage 2 results). The details were fully specified in the Data Sharing Agreement, which was put in place with participating schools before any data transfer.

Data was transferred from schools to York Trials Unit on encrypted spreadsheets via the University of York's secure file transfer service (University of York DropOff). A unique trial identification (ID) number (Pupil ID) was generated for each participant when their details were entered into the trial database.

The trial database and all electronic data was held on secure servers with access limited to specified members of University of York staff. Paper documents and assessment papers were held securely in a controlled access area in locked cabinets. The dataset for statistical analysis contained pseudonymised data and no schools, teachers, or pupils will be identifiable in the report or dissemination of any results.

Individually identifiable data held by York Trials Unit will be disposed of five years after report publication (scheduled for 2025, giving a destruction date of 2030). Anonymised electronic data and paper documents may be kept indefinitely by York Trials Unit.

The University of York's data protection policy is publicly available at: https://www.york.ac.uk/records-management/dp/

Project team

Evaluation team

York Trials Unit, University of York

Dr Lyn Robinson-Smith (Co-Principal Investigator) is an assistant professor with extensive experience of leading and delivering large scale randomised controlled trials in education and health, many of which have been funded by the EEF. Lyn was co-principal investigator from December 2022 leading on the impact evaluation, with oversight of the entire trial, undertook some qualitative survey analysis and contributed towards writing the final report. Lyn was on maternity leave from February 2024.

Professor David Torgerson (Co-Principal Investigator) was director of the York Trials Unit. He provided consultancy on methodology and design and contributed towards writing the final report. He was co-principal investigator on the trial from August 2022.

Dr Pam Hanley (Co-Principal Investigator) has an extensive background in education research, including many randomised controlled trials at the University of York. Her experience with the EEF includes other science-related interventions in addition to the TDTScience efficacy trial. Pam was jointly responsible for the day-to-day management and coordination of the trial along with leading on the qualitative aspects of the project until December 2022.

Louise Elliott (Co-Principal Investigator) was jointly responsible for the day-to-day management and coordination of the trial and led on the impact evaluation until August 2022. She has been involved in a large number of trials, including several for the EEF as principal investigator. She has broad experience of education research and has worked on a wide range of trials covering science, including the efficacy trial of TDTScience, literacy, and mathematics.

Imogen Fountain is an experienced trial support officer and contributed to coordinating schools through recruitment, survey completion, and outcome testing.

Caroline Fairhurst is a senior statistician who has worked on many education trials for the EEF. She oversaw all statistical aspects of the trial, conducted the statistical analysis and cost evaluation and contributed to writing the report. Caroline was on maternity leave from April 2024.

Dr Katie Whiteside is an experienced trial coordinator and has worked on a number of randomised controlled trials evaluating education and health care interventions. Katie undertook data management and trial coordination oversight and duties for the trial from August 2022 and contributed towards report writing.

Tom Davill was a trial support officer for the evaluation from October 2022 and assisted with recruitment and delivering outcome testing.

Jess Hugill-Jones has worked on a number of randomised controlled trials evaluating education and health care interventions. Jess was a trial coordinator for the evaluation from March 2022 to January 2023. Jess contributed towards IPE data analysis and report writing.

Dr Rachel Carr has a background in health psychology and has experience in varied trials, including those involving children and parents, and health behaviours during the postpartum period. Rachel was a trial coordinator for the evaluation from August 2022 to May 2023.

School of Education, University of Leeds

Professor Louise Tracey is centre lead for Inclusion, Childhood and Youth at the School of Education, University of Leeds. She has extensive experience as a principal investigator and co-investigator on the EEF trials focusing on primary education including SPOKES (Supporting Parents on Kids Education in Schools), ReflectED, and Grammar for Writing. Louise led on the IPE case study visits and training observations/interviews during Cohort 1 including the analysis and write up.

Department of Education, University of York

Dr Maria Turkenburg van-Diepen is a research associate in the University of York Science Education Group in the Department of Education. She has experience of education research at primary, secondary, and tertiary level, including a Systematic Review of Primary Science, and a mixed method study of the impact of science CPD for primary school teachers. Maria assisted in IPE instrument design, IPE research assistant training, IPE quality assurance monitoring, fieldwork, IPE analysis, and write up.

Maya Brakovic-Thomas is a PhD student in the Department of Education at the University of York, investigating the development of critical environmental literacy among secondary school pupils. She also holds a Master of Arts (MA) and Master of Research (MRes) from Kings College London and is a qualified teacher with 12 years of experience working with children in various educational settings. Maya assisted in the IPE instrument design, IPE fieldwork, IPE analysis, and write up.

Rosie Lennon is a PhD Student in the Department of Education at the University of York, researching mental health and well-being for children. Rosie is an experienced educationalist, having worked in various teaching roles and as a headteacher. Rosie is also a qualified counsellor for both adults and children. Rosie conducted IPE fieldwork in case study schools and attended one of the training days.

Delivery team

The delivery team had responsibility for recruiting and training the trainers, coordinating the training of teachers, recruiting participants in cooperation with trainers/other local partners, and liaising with the evaluation team in order to ensure the smooth-running of the evaluation and associated data collection activities.

Bridget Holligan was the director of education and engagement for Science Oxford until November 2023 and has spent her career in the informal science learning sector, with a particular focus on working with primary teachers and pupils in science. She jointly developed and led the TDTScience projects (2013–2023) with Helen Wilson, funded by the EEF and others. She led the creation of the Science Oxford Centre for primary schools and families, which is founded on the TDTScience ethos, and which opened to the public in 2019.

Helen Wilson was an affiliate lecturer at Oxford Brookes University, having been a principal lecturer in Science Education there. She began her career as a secondary physics teacher and then moved into primary teaching. She then went into Initial Teacher Education, eventually leading the primary teacher training programmes at Oxford Brookes University. As a primary science consultant, she researched the links between creative, challenging primary science lessons and pupils' attitudes and attainment. She jointly developed and led the TDTScience projects (2013–2023), funded by the EEF.

Andrew Kensley was head of Education Outreach, Training, and Communities for Science Oxford until November 2024, having formerly been an engineer and project manager for National Grid (and science, technology, engineering, and mathematics [STEM] ambassador) and then a primary school teacher. He led on the development and delivery of Science Oxford's local CPD for teachers, including courses for STEM Learning and the Primary Science Quality Mark (PSQM) as well as TDTScience-based twilight sessions. He was a TDTScience-trained trainer and project manager (from 2021) for the TDTScience effectiveness trial 2020–2023.

This core team were joined by the following trainers:

Bryony Turford from Primary Science Geeks.

Wendy Precious from Precious Learning Ltd.

Rachael Webb from Lancashire County Council.

Sarah Earle from Bath Spa University.

Alison Trew from Primary Science Teaching Trust.

Allie Beaumont an independent consultant.

Mandy Hodgkinson from East Riding of Yorkshire Council.

Nicky Waller, Jane Winter, and **Joy Parvin** from the CIEC, University of York.

Mike Dennis an independent consultant.

Methods

Impact evaluation design

Table 4: Evaluation design

Trial design, including number of arms		Two-armed, cluster randomised controlled trial Cohort 1 Year 5 (main trial), 2022–2023 Year 6 (longitudinal follow-up), 2023–2024 (e.g. participating Year 5 pupils move to Year 6 and continue their participation in the trial) Cohort 2 (longitudinal follow-up) Year 5, 2023–2024 Longitudinal follow-up and Cohort 2 results collected in Summer Term 2024 and published in 2025 as part of addendum report
Unit of randomisation		School
Stratification variable(s) (if applicable)		Geographical region (six levels: Lancashire; Lincolnshire and East Midlands; North East; South West; Staffordshire and West Midlands; and Yorkshire) Percentage of pupils eligible for FSM in the school (taken at the time of recruitment from the latest census data, two levels: dichotomised at the median <24%; ≥24%)
Primary outcome	Variable	Science attainment at the end of Year 5 (Cohort 1 only)
	Measure (instrument, scale, source)	Year 5 Science Assessment, 15-item measure scored 0–45, CIEC and York Trials Unit, University of York
Secondary outcome(s)	Variables	1. Science Attitudes (Cohort 1) 2. Science Attitudes (Cohort 2, to be reported in longitudinal follow-up addendum report) 3. Science attainment at Year 6 (Cohort 1, to be reported in longitudinal follow-up addendum report) 4. Science attainment at Year 5 (Cohort 2, to be reported in longitudinal follow-up addendum report) 5. Key Stage 2 English Reading and Maths (Cohort 1, to be reported in longitudinal follow-up addendum report)
	Measures (instrument, scale, source)	1. Science Attitudes Questionnaire, 28-item measure, 5-point Likert scale, based on Kind, Jones, and Barmby (2007) standard score from total score 20–100 (Cohorts 1 and 2) 2. Year 6 Science Assessment, 12-item measure scored 0–43, York Trials Unit and Manchester Metropolitan University (Cohort 1) 3. Year 5 Science Assessment, 15-item measure scored 0–45, CIEC and York Trials Unit (Cohort 2) 4. Key Stage 2 SATS Reading (KS2_READSCORE, range 0–120), Maths (KS2_MATSCORE, range 0–120) (Cohort 1)
Baseline for primary outcome	Variable	EYFSP
	Measure (instrument, scale, source)	Average point score from 17 Early Learning Goals (ELGs) that make up the EYFSP, scored 1–3, NPD
Baseline for secondary outcome(s)	Variable	EYFSP
	Measure (instrument, scale, source)	Average point score from 17 ELGs that make up the EYFSP, scored 1–3, NPD

Table 4 provides an overview of the evaluation design, which includes the main trial (primary outcome) and the longitudinal follow-up. This was a two-armed, cluster randomised controlled, effectiveness trial with random allocation at the school level. Schools were randomly allocated 1:1 to one of two groups:

- intervention group: Schools allocated to receive the TDTScience programme; and

- control group: Schools allocated to business as usual.

For the first year (2022–2023), schools allocated to TDTScience were not offered a financial incentive because we anticipated that attrition would be low, as was the case in the previous effectiveness trial, which reported a 3.7% (4/106) attrition rate among the intervention group (Kitmitto *et al.,* 2018); however, they were given a resources grant, based on the number of teachers taking part, which could be used for equipment, etc., and some low-value science equipment to take away from the training days. Upon completion of the second year of the trial, due to receiving no further input from the TDTScience trainers, intervention schools were eligible for a financial incentive of £500 to help reduce attrition among this group. The control group schools were offered a total financial incentive of £1,500 for participating, payable in two amounts: £1,000 after completion of the requirements in the first year; and £500 after completing the requirements at the end of the second year.

Participant selection

Schools

The delivery team led the recruitment of schools to the trial, with support from the evaluation team. Recruitment began in January 2022 in six geographical areas across England (Lancashire, Lincolnshire and East Midlands, North East, South West, Staffordshire and West Midlands, and Yorkshire). The areas were chosen to provide a varied geographical spread across England and areas, which did not overlap with those used for the previous efficacy and effectiveness trials. As noted in Table 3, the final choice of region was a pragmatic one based on the location and reach of the final team of trainers. The 'lead trainer' in each of the six regions took the lead for recruitment in their region. A recruitment webinar for trainers was held on 31 July 2021. Recruitment methods included using existing contacts, conferences, publicity through third parties, and social media. Trainers used a range of methods to contact schools. These included:

- emailing schools (using contact lists available to trainers through their roles as primary science leads, school advisors, and training providers in the region);
- local network meetings, other CPD, and conferences;
- use of existing newsletters to teachers, science leads, or headteachers; and
- contacting individual schools by phone or email.

Recruitment materials produced and provided to trainers:

- draft letter/email text to send to schools;
- 'brochure' for schools;
- set of presentation slides;
- promotional video;
- school information sheet;
- draft MoU; and
- Expression of Interest form (some trainers opted to produce their own online version).

Schools were approached and asked to complete an Expression of Interest form, after which they were invited to attend an online recruitment conference. Fourteen online conferences were held across the six regions (range 1 to 3 per region). Interested schools were then asked to return a signed MoU to the trainer. At this point, the school was considered recruited, their details were shared with the evaluation team, and they were randomised. It was agreed between the EEF, evaluation team, and delivery team that flexibility was permitted on the number of schools recruited per region, as long as the total number of 180 schools was achieved. Once 180 signed MoUs were received, a small number of schools were put on a waiting list. Once schools were randomised, any schools on the waiting lists were advised that they would not be able to take part. The schools were recruited with an aim to being representative of their area while targeting those that are higher than average in the percentage of pupils receiving FSM.

School eligibility criteria:

- The school must be state-maintained.
- The school must have a minimum of one full class of Year 5 pupils (mixed year group classes were not eligible to take part).
- The school did not operate a two-year science curriculum that involves Year 5 pupils (i.e. either Year 4/Year 5 or Year 5/Year 6).
- The school would allow all Year 5 teachers to be available for the four and a half-days of training. If a school only had one Year 5 teacher, another teacher (ideally the science coordinator) would also need to attend the training.
- Schools within a Multi-Academy Trust (MAT) are eligible to participate on the understanding that schools within the same MAT must agree that they either do not usually, or will not during the period of the trial, collaborate on science teaching. The MAT must accept that their schools will be randomised individually and so may be allocated to different groups. Alternatively, a MAT can nominate just one school to take part.
- The school or individuals involved were not involved in the previous trials of TDTScience, or trained in TDTScience, or taken part in the pre-trial. If the school was part of a MAT then none of the schools within the MAT would have taken part in the pre-trial.
- The school was not involved in the EEF Stop and Think trial as this is also a science CPD for Key Stage 2.
- The school was not involved in the EEF Focus for Teacher Assessment in Primary Science (Focus4TAPS) as this is also a science CPD for Key Stage 2.
- The school agreed to all requirements outlined in the Information for Schools and MoU documents (including commitment to keep same Year 5 teachers across the two years wherever possible).

Pupils

The TDTScience programme was delivered at a whole-class level, meaning all the Year 5 pupils within the school were able to participate in the trial. At the beginning of the academic year 2022–2023, parents/carers were informed about the research through an information sheet (see Technical Notes) sent by schools on behalf of the evaluation team. Parents/carers were then asked to return a ‘withdrawal from research’ form if they were unwilling to share their child’s data with the evaluation team and/or they did not wish their child to take part in any assessments, surveys, or focus groups. Schools were then asked to securely share the details of participating pupils with the evaluation team.

Outcome measures

Baseline measures

The baseline measure was the average point score from the 17 ELGs that make up the EYFSP; these data were accessed via the NPD in early 2024. This baseline measure was chosen as an alternative to the Key Stage 1 English Reading and Maths scores used in the previous effectiveness trial, as Key Stage 1 results are not available for this cohort of pupils in this trial, who would have been in Year 2 during the academic year 2019–2020 when national Key Stage 1 assessments were cancelled due to the COVID-19 pandemic. Key Stage 1 results would have been the preferred choice for the baseline measure as this would have allowed a direct comparison of results with the previous effectiveness trial. Additionally, it is likely that the correlation between Key Stage 1 results and the outcomes in the trial would have been higher than with EYFSP results as these were assessed long ago, which may have resulted in the trial having a lower minimum detectable effect size (MDES) for the same power and sample size.

Within the EYFSP (for the academic year 2017–2018 when the pupils in this trial would have been in Reception), for each ELG, the pupil’s learning and development was rated as:

- best described by the level of development expected at the end of the early years foundation stage (EYFS) (expected);

- not yet at the level of development expected at the end of the EYFS (emerging); and
- beyond the level of development expected at the end of the EYFS (exceeding).

These are scored from 1 to 3 (1 = emerging, 2 = expected, 3 = exceeding), and all 17 scores were summed and averaged (to produce a total score ranging from 1 to 3).

Primary outcome

Science attainment

A new measure, the Year 5 Science Assessment, was used as the primary outcome measure. This was recently developed by the CIEC and York Trials Unit, University of York (Joshi *et al.,* 2022) and designed to be suitable to be administered to Year 5 pupils as a meaningful outcome measure in future evaluations. It was originally developed for use in two of the EEF-funded randomised controlled trials in 2020. However, both these trials were delayed because of school closures due to the COVID-19 pandemic. This new measure has been developed to better reflect the current curriculum, have a mix of question types, and have greater emphasis on 'working scientifically' than the alternatives. Details of the development and validation of this measure are published in Joshi *et al.* (2022). The measure previously used for both the efficacy (Hanley, Slavin, and Elliott, 2015) and effectiveness (Kitmitto *et al.,* 2018) trials of TDTScience was no longer fit for purpose. Its creation preceded the new science curriculum (Department for Education, 2013) with its changed content and emphases (e.g. more focus on 'working scientifically'/science enquiry). The main alternative (GL Progress Test in Science) was not considered to be a varied enough test (e.g. it is predominantly multiple choice) to be an adequate replacement. While the development of this new test was necessary, we are yet to know if the test results are predictive of later science attainment, as per other standardised Key Stage 2 assessments.

The Year 5 Science Assessment is a 15-section measure, each section is comprised of one to five questions (i.e. a, b, c, d, or e) and is worth between 1 and 5 marks (three sections are worth 1 mark, one section is worth 2 marks, seven sections are worth 3 marks, one section is worth 4 marks, and three sections are worth 5 marks), and incomplete sections are given a score of 0. Section scores are summed to produce a total score from 0 to 45, where a higher score indicates a better outcome.

The assessments were administered by independent, blinded research invigilators who were recruited and trained by the evaluation team. All research invigilators had an enhanced Disclosure and Barring Service check and underwent relevant safeguarding and data protection training. The evaluation team advised schools that a teacher or teaching assistant should be present during the assessment to assist pupils who may require pastoral care but requested that teaching staff did not help pupils to complete the assessment.

Teachers were advised that pupils who they considered would not be able to engage with the assessment in its current format (a 45-minute continuous, whole-class written assessment) should not complete the assessment. Teachers were informed that pupils can complete the assessment with the aid of support staff if support staff usually scribe or read assessment questions to them, but no help should be given by the support staff to the pupil when they are answering the actual questions.

Schools were advised that the assessment session for each class would last approximately 75-minutes, which includes time to: settle the pupils; distribute assessment papers; complete the assessment (45-minutes); and collect completed papers. To avoid disruption to the normal school day, the evaluation team advised schools that participating pupils completed the science assessment in their usual classroom; no special seating arrangements were required.

Assessments were marked according to a detailed mark scheme by a team recruited and trained by the evaluation team. At least 5% of the assessments for each school were second-marked to ensure consistency. Marks for each individual question (within each section), total marks for each section, and the total mark for the assessment were each manually entered and, to confirm input accuracy, were cross-checked against total marks calculated using software.

Both invigilators and markers were blind to group allocation as the evaluation team did not provide them with access to group allocation for each school. If a research invigilator became unblinded (e.g. the school mentioned that they had been following the TDTScience programme or that they had not received the training), they were asked to inform the evaluation team so this could be recorded; there were 18 known instances of this (6 instances in control schools and 12 instances in intervention schools).

Secondary outcomes

The science attitudes instrument used in both the efficacy trial (Hanley, Slavin, and Elliott, 2015) and the previous effectiveness trial (Kitmitto *et al.,* 2018) contained 23 items asking about interest, self-efficacy, and activity in science lessons. In the TDTScience pre-trial four new items were added to the instrument to strengthen the self-efficacy scale (Standley *et al.,* 2023). This 28-item, self-reported Science Attitudes Questionnaire was administered to 103 pupils from two schools in the pre-trial. Each item is scored from 5 (agree a lot) to 1 (disagree a lot), with negatively worded items reverse scored. Factor analysis on data from the pre-trial indicated that 20 of these items can be incorporated into a scale that measures 'interest and self-efficacy' (Standley *et al.,* 2023). The 28-item instrument was completed in class, supervised by class teachers, at the end of Year 5 for Cohort 1 in the TDTScience main trial. Responses to the 20 items identified by the factor analysis were summed to generate a total score from 20 to 100, where a higher score indicates a greater interest in science. The score was standardised to a mean of 0 and a standard deviation (SD) of 1 by subtracting the sample mean from each pupil's score and dividing it by the sample SD. The remaining items, not used in this scale are summarised separately. Pragmatics dictated the attitudes survey be teacher-administered, rather than being completed with the trained invigilators during visits to complete the primary outcome, because otherwise the session would have been too long for pupils of this age (45+ minutes for the science assessment plus this survey). Teachers were given instructions about how to administer the Science Attitudes Questionnaire and facilitated a session where the students completed the survey. Teachers were advised that the survey should be administered prior to the science assessment, where possible, and that the completion of the survey should not immediately follow assessment completion. The survey was being administered by teachers, as opposed to trained invigilators, and was tried and tested during the previous two TDTScience trials.

The following secondary outcomes will be reported in the addendum report:

- **Cohort 2 Year 5 Science Assessment.** The same 15-item science attainment assessment used for the primary outcome will be administered to Cohort 2 at the end of Year 5.
- **Cohort 2 Year 5 Science Attitudes**. The same 28-item science attitudes instrument used for the secondary outcome will be administered to Cohort 2 at the end of Year 5.
- **Cohort 1 Year 6 Science Assessment.** At the end of Year 6 for Cohort 1, we intend to collect the secondary outcome of science attainment, assessed via a new measure, which is aligned to the curriculum and age appropriate, the Year 6 Science Assessment, developed by York Trials Unit and Manchester Metropolitan University. This new measure is comprised of 12 items and is scored from 0 to 43. The measure reflects the current curriculum, has a mix of question types, and an emphasis on 'working scientifically'.
- **Cohort 1 Year 6 English Reading and Maths**. We will assess for any impact on Maths and English on Cohort 1 at the end of Year 6 by considering attainment based on pupils' Key Stage 2 results (English Reading and Maths), which will be obtained from the NPD in Autumn Term 2024. These will be measured via scaled assessment scores, using the variables KS2_READSCORE and KS2_MATSCORE, both scored on a scale from 0 to 120.

Sample size

From protocol

The following assumptions were made in the protocol and are outlined in Table 5:

Based on the previous TDTScience trials we assumed an intracluster correlation coefficient (ICC) of 0.15, and an average year group (cluster) size of 45 at randomisation. In the efficacy TDTScience trial (Hanley, Slavin, and Elliott, 2015), the observed correlation between the pre-test (Science Knowledge Questionnaire administered in Year 4) and outcome (Science Knowledge Questionnaire administered at the end of Year 5) was 0.51. In the first effectiveness trial (Kitmitto *et al.,* 2018), the analysis model for the outcome (Science Knowledge Questionnaire) included achievement at Key Stage 1 in reading/writing and maths as a covariate (as a measure of prior attainment). The proportion of variance explained by level 1 covariates (R2) was 0.4, suggesting a pre- and post-test correlation of around 0.6. In this trial, the average score from the 17 ELGs of the EYFSP (obtained via the NPD) was the measure of prior attainment. This was similarly used in the EEF Stop and Think trial (Roy *et al.,* 2019), for which the post-test was GL Assessment's Progress Test in Science ten measured at the end of Year 5, and the correlation between pre- and post-test was 0.53. Based on these estimates but acknowledging the differences in outcome measures used as pre- and post-tests, we conservatively assumed a pre- and post-test correlation of 0.5 for this calculation. Hence, to detect an effect size of 0.15 with 80% power and two-sided alpha of 0.05, assuming pupil-level attrition of 15%, we calculated that a total of 180 schools would be required (8,100 pupils per year group).

In January 2020, 17.3% of pupils were eligible for FSM. We therefore, assumed to recruit 180 schools and 8,100 pupils each year of the trial, that there would be approximately 1,400 pupils eligible for FSM each year (approximately eight per school). Under the same assumptions as above, an MDES of 0.19 would be detectable.

Table 5: Sample size calculation at protocol stage

		Protocol	
		Overall	FSM
MDES		0.15[a]	0.19[a]
Pre-test/post-test correlations	Level 1 (pupil)	0.5	0.5
	Level 2 (class)	–	–
	Level 3 (school)	–	–
ICCs	Level 2 (class)	0.15	0.15
	Level 3 (school)	–	–
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two	Two
Average cluster size		45	~8
Number of schools	Intervention	90	90
	Control	90	90
	Total	180	180
Number of pupils	Intervention	4,050	700
	Control	4,050	700
	Total	8,100	1,400

[a] *Accounting for 15% attrition.*

At randomisation

At randomisation, there were 180 settings. For logistical reasons, schools had to be randomised and informed of their allocation at the end of the academic year 2021–2022 so intervention schools could begin to make arrangements to attend the training. The evaluation team could only collect pupil details from participating schools at the start of the academic year 2022–2023 when schools knew, which pupils would be in the class, this was post-randomisation. Twelve schools withdrew during this process, and so the number of schools from which we received participating pupil details was 168 (7,261 pupils; intervention, n=3,790; control, n=3,471). This is an average of 43.2 pupils per school. Assuming 80% power, a pre- and post-test correlation of 0.5, an ICC of 0.15, and 15% pupil-level attrition, the MDES with this sample size would be approximately 0.16 (see Table 8).

A total of 2,168 of the randomised pupils were eligible for EVER6FSM (average of 13 per school). With this sample size, under the same assumptions, the MDES would be approximately 0.17 (see Table 8).

Randomisation

A statistician at York Trials Unit randomised schools 1:1 to either the intervention arm (offering the TDTScience CPD programme) or the control arm (continuing with usual provision for the duration of the evaluation).

A dedicated computer program, MinimPy (Saghaei and Saghaei, 2011), was used for randomisation via minimisation using the following factors:

- School region—six levels: Lancashire, Lincolnshire and East Midlands, North East, South West, Staffordshire and West Midlands, and Yorkshire for logistical reasons, to ensure a balanced spread of intervention and control schools in each area.
- School deprivation level (i.e. the percentage of pupils eligible for FSM in the school based on latest available data)—two levels: dichotomised at the median for the 171 schools that were randomised in the first batch (see below <24%, ≥24%) to ensure balance between the randomised groups, since this school characteristic and individual pupil deprivation may moderate outcomes.

Randomisation was carried out in batches (groups of schools that were ready to be randomised at that time) to avoid delays in programme induction and to maximise programme delivery for as many schools as possible. Deterministic (i.e. without the use of a random element) minimisation was used (Altman and Bland, 2005). This was deemed to be sufficient as the allocations were conducted in batches, rather than one-by-one prospectively, meaning predictability was not a concern and hence a random element was not required.

Statistical analysis

Analysis followed the EEF's (2022) most recent guidance,[1] and is detailed in full in the published SAP.[2] Analyses were conducted in STATA version 17 (StataCorp LLC, College Station, Texas, USA). The trial statistician was not blind to group allocation. All analyses were conducted on an intention-to-treat (ITT) basis, where data were available, including all schools and pupils in the group to which they were randomised irrespective of whether or not they actually received the intervention, using two-sided tests at the 5% significance level. A CONSORT (Consolidated Standards of Reporting Trials) diagram shows the flow of schools and pupils through the trial.

School and pupil characteristics and outcome measures assessed at baseline are summarised descriptively by randomised group both as randomised and as analysed in the primary analysis. No formal comparison of the baseline data were undertaken, except for a comparison of the difference in prior attainment (average EYFSP score) between the groups, reported as a Hedges' g effect size with a 95% confidence interval (CI).

Outcome data are summarised descriptively by trial arm. The correlation of outcome measures and average EYFSP score are presented with a 95% CI. Effect sizes based on the difference between the groups at the outcome assessment point are presented as adjusted mean differences for continuous outcomes, and odds ratios (ORs) and difference in proportions for dichotomous outcomes, with their associated 95% CI and p-value. Treatment effects are also presented as (estimated) Hedges' g effect sizes.

Primary analysis

The primary analysis investigated any difference in Year 5 Science Assessment score between the two arms. A linear mixed effects regression model at the pupil level was used to estimate the adjusted mean difference in scores. Group allocation, average EYFSP score, and the minimisation factors (region, FSM) were included as fixed effects, and school as a random effect. FSM was dichotomised at the school level for use as a minimisation factor in the randomisation (<24% / ≥24%, see 'Randomisation' section above), but was entered into the analysis model as a dichotomous variable at the *pupil* level (EVERFSM_6_P from the NPD) as this provides more granular information. Robust standard errors were specified to account for any potential heteroscedasticity.

---

[1]https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design
[2]https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/thinking-doing-talking-science-effectiveness-trial-2

Model assumptions were checked as follows: the normality of the standardised residuals was checked using a qq plot. If the model assumptions were in doubt, sensitivity analyses were conducted in which transformations of the outcome and/or covariate data were tried to improve the model fit.

Subgroup analyses

A subgroup analysis was conducted for the primary outcome of Year 5 Science Assessment considering FSM status (EVERFSM_6_P), first by retaining the whole analytic sample and including an interaction between FSM and group allocation in the primary analysis model, and second by repeating the primary analysis only within the restricted FSM subgroup (following the Statistical analysis guidance for EEF evaluations; EEF, 2022).

Analysis in the presence of non-compliance

Compliance was measured as a binary outcome as a class level, as opposed to at the school level. The delivery team kept registers of attendance at each training session, which they provided to the evaluation team. When schools sent their participating pupils' details, they indicated what class the pupil was in and who their teacher is. Teacher changes throughout the year were recorded by the delivery team and passed to the evaluation team.

For Cohort 1, compliance is defined as the following: The class was taught by a teacher who attended at least three out of the four full days of training This included a class that (because of long-term sick leave, resignations, etc.) has been taught by two teachers who together attended three additional training days. For example, a class was considered compliant if the Year 5 teacher attends two days training in the Autumn Term then leaves the school; then the new teacher attends at least one further full-day TDTScience training session.

A CACE (complier average causal effect) analysis (Dunn, Maracy, and Tomenson, 2005) was conducted for the Year 5 Science Assessment outcome using the dichotomous compliance measures described above. This analysis used a Two Stage Least Square (2SLS) approach with group allocation as the instrumental variable for the compliance indicator, with cluster standard errors to account for clustering at the school level. Results for the first stage (of the 2SLS process) are reported alongside: i) the correlation between the instrument and the endogenous variable (presented as the partial R^2 statistic from the first stage estimation); and ii) a F-test (F-statistic and p-value). The F-statistic should exceed ten for inference based on the 2SLS estimator to be reliable when there is one endogenous regressor, as in this case (Bound, Jaeger, and Baker, 1995; Stock, Wright, and Yogo, 2002).

Missing data analysis

The amount of missing baseline and outcome data is summarised, and reasons for missing data explored and provided in the report where available. Since the percentage of missing cases exceeded 5%, a multi-level logistic regression model was used to model presence or absence of the primary outcome including all available pupil- and school-level baseline data as fixed effects, and school as a random effect. Significant predictors are discussed.

The impact of missing data on the primary analysis was additionally assessed using multiple imputation (MI) including all available pupil-level baseline variables (i.e. gender, EVER6FSM status, and EYFSP average point score) and school-level baseline variables (i.e. allocation, location, whether rural or urban, type of school, and latest Ofsted rating).

A 'burn-in' of ten was used, which means that the first ten iterations of the imputation were not used to allow the iterations to converge to a stationary distribution, and 30 imputed datasets were created. The primary analysis was then rerun within the imputed datasets and Rubin's rules (Rubin, 1987) used to combine the multiply imputed estimates. The resulting intervention effect size was compared for magnitude and statistical significance to that from the primary analysis model to assess whether missing data had biased the results.

Secondary analysis

Standardised scores from the 20-item 'interest and self-efficacy' scale of the Science Attitudes Questionnaire (at the end of Year 5) were compared between the two trial arms. As for the primary analysis, a linear mixed effects regression model was used to estimate the adjusted mean difference in scores. School was included as a random effect and group allocation, average EYFSP score, and minimisation factors (as in the primary analysis) were included as fixed effects.

The remaining items, not used in the scale, are summarised separately.

Estimation of effect sizes

Hedges' g effect sizes were calculated by dividing the adjusted mean difference between the intervention and control group (accounting for baseline measures and the minimisation factors) by the pooled unconditional SD obtained from the model run without these covariates. A 95% CI for the effect size were calculated by dividing the 95% confidence limits for the adjusted mean difference by this same denominator. All parameters used in these calculations are provided in this report.

$$\mathrm{ES} = \frac{(\overline{\mathrm{Y}}_{\mathrm{T}} - \overline{\mathrm{Y}}_{\mathrm{C}})_{\mathrm{adjusted}}}{sd_{pooled}}$$

where, $(\overline{\mathrm{Y}}_{\mathrm{T}} - \overline{\mathrm{Y}}_{\mathrm{C}})_{\mathrm{adjusted}}$ denotes the difference in means between trial groups adjusting for pre-test score and the minimisation factors, from the multi-level analysis model; and sd_{pooled} denotes the pooled, unconditional SD of the two groups (square root of the sum of the within- and between-cluster variances) (Xiao *et al.,* 2016).

Estimation of ICCs

The ICC associated with school for the outcomes (both pre- and post-test where available) are presented alongside a 95% CI. The ICC at post-test was computed for the analysis model, and also for an empty model (i.e. one without covariates). The ICC at pre-test was calculated for a linear model with pre-test as the outcome and school as a random effect.

IPE design

The IPE aimed to explore the relationship between delivery and programme outcomes, in particular to provide greater context and understanding of the results of the impact evaluation. In order to do so the IPE focused on compliance (as defined in the impact evaluation), implementation fidelity, perceived impact of the programme, and comparison with the control conditions/business as usual.

Research methods and analysis

The IPE adopted a mixed methods approach including teacher surveys, pupil questionnaires, case study school visits, observations of training, surveys, and interviews with trainers, and interviews with the delivery team. The instruments for each method of data collection were developed and refined during the pre-trial phase. Further details of each are provided below and in Table 6.

Teacher surveys

All teachers involved in the evaluation were asked to complete an online teacher survey between October 2022 and December 2022 to establish a baseline of school and teacher contextual factors, current science provision (both the amount of science teaching and the strategies used), and teacher attitudes towards and confidence in teaching science. Follow-up surveys were administered to all Year 5 teachers in June 2023 to July 2023 at the end of intervention delivery, with additional questions for teachers in intervention schools to explore feedback about training sessions, use of the TDTScience approach in the classroom, and the effect on their confidence and practice of teaching science. Teachers were also asked about the perceived effects on pupils, including engagement and confidence in their understanding of science.

Pupil questionnaires

At the end of the first year, Year 5 pupils were asked to complete a science attitude measure (similar to the instrument used in the previous TDTScience trials) along with a questionnaire about their science lessons (to compare TDTScience with business as usual and triangulate against teacher feedback and observations).

Qualitative school sub-study

Two intervention schools in each geographical region were selected to be visited by trained researchers. School selection considered Year 5 cohort size and number of pupils in receipt of FSM. One school in each region was intended to be visited twice during the academic year 2022–2023, in Autumn and Summer Terms, to allow direct comparison of the experience and perception of the TDTScience programme towards the beginning and end of the intervention period.

The other was to be visited once, in the Spring Term, to pick up on any aspects of the intervention that might be particular to the midway point. The alternative of visiting each of six schools three times was considered but it was deemed to be too burdensome on individual schools and would restrict the evaluation to a more limited spread of settings. In the event, due to staffing issues the initial school visits were delayed until the beginning of the Spring Term but then proceeded as planned across the Spring and Summer Terms of 2023. In total, 17 visits occurred with one school not receiving a return visit after two attempts were cancelled due to unforeseen circumstances outside of the school and researcher control. Two school visits were conducted by two researchers together for quality assurance purposes.

Lesson observations, teacher interviews, and pupil focus groups occurred at each visit to assess implementation fidelity and the attitudes/engagement of teachers and pupils. Observations were designed to allow the researcher to understand implementation of learning from the programme in a Year 5 science lesson with a focus on TDTScience strategies and techniques as well as pupil engagement in the lessons. Interviews with Year 5 teachers were designed to ask about TDTScience training, implementation of learning from the programme, and perceived impact of the programme on pupils. Where possible, more than one TDTScience teacher was interviewed within each school, and in most cases, interviews were conducted individually. Where a teacher was observed, the interview tended to take place after the observation in order to provide the teacher with the opportunity to reflect on using TDTScience in practice. In total, 24 teacher interviews were conducted with 22 teachers in total (with some teachers interviewed twice in separate visits) across the 12 schools. Twelve of these teachers also had their lesson observed, the remaining interviews were with another Year 5 teacher in the same school (n=9) and one interview with the school science lead (who taught another year group). All teachers interviewed had attended the TDTScience training. Eighteen lesson observations took place over 17 school visits with two lessons by different teachers being observed in one school. It had been anticipated that if recordings of lessons were successful in the pre-trial phase of the study, a sample of lessons in another six schools would also be recorded to capture shifts across time and minimise ‘hothouse’ effects. However, following trialling of this method it was not deemed to be viable and so did not occur during the main trial.

Focus groups consisted of around six pupils,[3] selected by the teacher(s), with parental consent to participate. Teachers were encouraged to select pupils to participate in the focus groups from a range of socio-demographic backgrounds and prior attainment levels. Focus groups tended to take place after the observation and participating pupils were generally all from the observed class. However, occasionally the focus group also contained pupils from the other Year 5 class(es) and when the sequence (of observation followed by focus group) was unable to take place due to other school requirements (e.g. the school timetable) the researchers complied with whatever arrangement worked best for the school. The focus groups were designed to encourage pupils to discuss their experiences of and attitudes towards learning science in school and, in particular, the TDTScience strategies implemented by their Year 5 teacher. Pupils were informed by their teachers and the researchers that they were being asked about their current science lessons although many knew that their teacher was attending science-related training and were aware that they were taught using the TDTScience programme. In total, 17 focus groups were conducted across the 12 schools selected for the in-depth qualitative sub-study. Where two focus groups were held within a school at different timepoints (i.e. Spring and Summer Terms 2023), it varied between schools as to whether or not the same pupils participated.

Photographs were also taken by the researchers of a sample of pupils’ work to assess the move to more focused recording of investigations. The selection of this sub-sample was pragmatic and either chosen by the researcher from the selection of science books within the observed class, pre-selected by the teacher, or were taken from the books of pupils who participated in the focus group. Approximately, four to five pages from five to six pupils across all 12 schools and 17 visits were recorded. Additionally, participating Year 5 teachers from three control schools provided samples of pupils’ work, consisting of between one and five pages per pupil.

Copies of the case study visit research instruments are provided in the Technical Notes.

Training observations

One training session was observed per region, meaning that each pair of trainers was observed at least once. The observations were spread across the first four days of the four and a half-days programme. Each training day was visited once with training days Day 3 and Day 4 visited twice (once in two different regions). The observation focused on implementation fidelity and teacher engagement. One training day was attended by two researchers for quality

[3] We originally anticipated four to five pupils per focus groups but teachers’ and pupils were keen for us to speak to pupils’ about TDTScience so most focus groups had five to six pupils.

assurance purposes. An additional training day (Day 1) was attended by a third researcher with the principal investigator to understand more about the training prior to the observation visits occurring.

Trainer surveys and interviews

After each training day observed, both trainers who delivered the session were interviewed (together, due to time constraints) in order to get their feedback on how they felt the training day went. The evaluation team recognise that the data trainers provided may have been different had they been interviewed separately. In total, six interviews were conducted, one with the training team for each geographical area in the trial. A short online survey was administered to all the remaining trainers after each day to obtain feedback. Copies of the training day observation schedule and trainer interview schedule can be found in the Technical Notes.

Developer focus group

Towards the end of the first year, two members of the delivery team were interviewed together to obtain their views of how the intervention was implemented. The interview lasted approximately one hour and was conducted over video conferencing.

Analysis

Table 6: IPE methods overview

Research methods	Data collection methods	Participants / data sources	Data analysis methods	IPE research questions addressed	Implementation / logic model relevance
Trainer feedback (after observed sessions)	Paired semi-structured interview	Six interviews (12 trainers)	Combination of inductive and deductive analysis	1a	Implementation activity (four-day professional development programme)
Trainer feedback (after each delivery phase)	Survey	Five timepoints (12 trainers per occasion)	Descriptive analysis	1a	Implementation activity
Teacher feedback on training (Collected by developers)	Survey	All teachers (n=180) attending TDTScience training (collected after each training event)	Frequency counts; descriptive/thematic analysis	1a 3a 3c	Implementation activity
Baseline	Survey	All participating teachers (c360) in intervention and control schools		3c 5a-d	Usual practice
Follow-up at the end of the first	Survey	All participating teachers (c360) in intervention and control schools		1b 2 3 a-e 5 a-d	Comparison of pre- and post-intervention; TDTScience and control
Sub-study school lessons	Lesson observations	Sub-study schools / teachers (n=18)	Descriptive analysis (of schedule and fieldnotes)	1b 2 4a-c	Whether teachers are implementing strategies in lessons; confidence; pupil engagement
Sub-study school teachers	Semi-structured interviews	Teachers (n=36)	Combination of inductive and deductive analysis	2 3 a-e 5 a-d	Whether teachers are implementing strategies in lessons; confidence
Sub-study pupil feedback	Focus groups	Pupils in sub-study schools (18 groups of five to six)		1b 4a-c	Pupil response to TDTScience strategies; triangulation of practices, etc.
Sub-study pupil written recording	Examination of samples of pupils' work (five pupils x five pages/visit)	Pupils in sub-study schools (c100)	Descriptive analysis (proforma and fieldnotes)	1b 4c 5b	Evidence of TDTScience affecting written work (e.g. focused recording)

Interview and focus group data were audio-recorded and transcribed with the exception of one interview where the teacher requested not to be recorded. In this instance, extensive notes were taken at the time to ensure this was an accurate record and these notes were then coded in the same way as the interview transcripts. The data was analysed thematically using NVivo software using a mixture of deductive and inductive coding. This was to ensure that the data was used to address the specific IPE research questions, as well as to seek out any additional information contained in the data not originally anticipated by the evaluation team or contained in the logic model. Observation data (from sub-study school visits and attendance at training days) were analysed descriptively, with the emphasis on implementation fidelity, and engagement (teacher engagement on training days, pupil engagement with TDTScience lessons). Excerpts from Year 5 pupil science books were analysed descriptively to: i) understand more about how focused reporting worked in practice; and ii) to compare with those samples of student work provided by the control school. We recognise that this analysis is limited due to the small number of samples from pupils in the control condition and that they were all obtained from the same school.

Interview, focus group, and survey data were triangulated to enable a more thorough understanding of the different perspectives of trainers, teachers, and pupils. Findings from the observations were also fed back into this triangulation to provide a more holistic understanding of each data source. While it was anticipated that by visiting sub-study schools twice over the implementation year we would be able to encapsulate some indication of 'change' given that the original visits were delayed (see 'Qualitative school sub-study' section above) it became apparent that this was not as feasible as originally intended. This was compounded by the fact that given that the training days were dispersed throughout the academic year, school visits tended to observe a lesson based on a recent training day and therefore, was often closely aligned with the content from that training day rather than from the programme as a whole. However, where a school was visited twice, we have indicated whether or not the interview data or observation was from a first or second visit (by labelling the excerpts from the data V1 or V2; Visit 1 of Visit 2). Similarly, the observed teacher is noted as T1 in each sub-study, with any additional interviewees within a school being labelled as T2.[4]

In contrast to the in-depth qualitative sub-study schools, all data relating to the trainers (i.e. trainer interviews and training day observations) have been completely anonymised given the small sample size. Therefore, no reference is made to individual trainers or the region they are responsible for. However, the views expressed, and the excerpts used, from the interviews cross the whole cohort of trainers to ensure that there is balance in the reporting.

Costs evaluation design

The cost analyses followed the 'ingredients method' (Levin *et al.,* 2017) to account for the costs of implementation at key stages.

Sources of data

Surveys were administered to all Year 5 teachers in June 2023 to July 2023 at the end of intervention delivery, with additional questions for teachers in intervention schools to capture data on intervention implementation and costs associated with delivering the TDTScience intervention. The survey captured whether, and for how many hours, the school had to arrange for paid cover for the teacher to: i) attend the TDTScience in-person training; ii) receive other support from the TDTScience team; iii) share TDTScience training with other members of staff within their school who missed the in-person training sessions; iv) receive the training from another staff member within their school; and v) complete other activities related to the TDTScience programme. The cost of travel to the venue(s) to attend in-person TDTScience training was also collected. If teachers indicated they had spent time completing TDTScience training and/or other-related activities outside of normal working hours/in their own time, the survey asked for a breakdown of time spent on each activity. The survey also requested information on what existing resources, if any, the teacher had used as a result of implementing TDTScience programme (e.g. computer, internet connection, learning resources) and any additional resources, with the quantity and cost, they had bought.

Attendance at the in-person training was captured for each teacher in registers collected by the Science Oxford team and provided to the evaluation team.

[4] Please note a maximum of two teachers was interviewed per school.

Cost data was also collected directly from the delivery team in order to estimate the cost of delivering the intervention. Two sets of costs were provided: i) the actual cost of programme set-up, intervention development, recruitment, training, and intervention delivery incurred in the trial; and ii) the 'real-world' costs of delivering the intervention outside the trial.

Analysis

The total time devoted by school personnel for training and intervention delivery/engagement are summarised descriptively. The average cost to the school of taking part in the intervention was calculated by summing the constituent parts (cost of teacher cover, travel to training venues, and purchase of additional resources) and dividing by the number of schools who responded to the surveys.

Per pupil costs were determined by summing the total costs per school and dividing by the average number of pupils per school eligible for inclusion in this evaluation (n=43).

The cost to the Science Oxford team of the development, set-up, and delivery of the intervention are summarised and presented as an average cost per school and an average cost per pupil per school year across three years.

Assumptions

For the costs directly incurred by schools, the following assumptions were made:

- Data for the costs are based on responses to the surveys. Not all schools provided data and it is likely that not all teachers within the schools that were represented may have provided data. However, for the purposes of these analyses, we made the assumption that the teachers and schools responding to the surveys represent the whole sample of intervention schools and so the costs and time commitments reported by them are generalisable to all randomised schools.
- For the purpose of calculating the costs over a three-year period, we assumed that the costs incurred by schools of staff cover and travel to attend in-person training would be a one-off but the purchase of additional resources would be recurring based on the assumption that some materials may need to be replaced annually or additional, similarly-priced materials may need to be purchased. These costs are considered as optional extras given the inconsistent uptake (i.e. purchasing, if at all, will be different between schools) by schools, except for travel costs as it is assumed that at least one teacher per school would have to travel to in-person training (either taking other teachers as passengers and/or to subsequently be able to cascade the training to other teachers).
- For the purposes of calculating the average cost of paid cover for teachers to complete training and other TDTScience-related activities, we assumed that the cost of staff cover is £18.21 per hour, which is the average hourly rate of a primary school teacher in the UK.

For the costs provided by the delivery team of delivering the intervention in this trial, the following assumptions/considerations were made:

- Costs were taken from Science Oxford's final spend, only removing the following:
    - costs relating to online course development started due to impact of COVID-19 but in the end not required;
    - 50% of the costs relating to recruitment of randomised controlled trial schools (but NOT recruitment costs relating to pre-trial schools, which have been included); and
    - payments to control schools.
- Arguably some of the project management and admin costs will also relate to control schools, but these are difficult to disentangle so have not been adjusted to provide a more conservative estimate.
- Additionally, costs include delivery of pre-trial courses as, again, these are difficult to disentangle.

For the 'real-world' costs of delivering the intervention outside of the trial provided by the delivery team, the following assumptions were made:

- Two new trainers, both completing the full 'train-the-trainer' journey (three phases, as below), ending with the delivery of a TDTScience course that they deliver together for 30 teachers:

- Phase 1: New trainers (x2) attend a TDTScience course as a participant (an existing course, with course delivery costs covered by participant fees).
- Phase 2: New trainers attend a train-the-trainer course (two days with one already accredited trainer).
- Phase 3: New trainers each deliver TDTScience course with an already accredited trainer (these would both be existing courses, with course delivery costs covered by participants).

We estimated the average cost per pupil per year for schools receiving the TDTScience programme following the EEF costing guidance issued in 2023 (Education Endowment Foundation 2023). A year is defined as a year of implementation. This may not align with the calendar or academic year. This costing model estimated costs based on the mean number of eligible pupils per school randomised into the evaluation (n=43). Given that this is a staff CPD programme, all pupils being taught by the trained teacher are likely to be impacted, thus, the per pupil costs would be dependent on the actual size of the class and could be reduced if they are spread across a greater number of pupils. Additionally, teachers who have attended a TDTScience course may impact more than one class, as the TDTScience pedagogy that the course enables them to implement would continue with new classes. If this assumption (to some extent) was also factored in, then this would bring the per pupil cost down further still.

Timeline

Table 7 reports the timeline for data collection, analysis, and reporting of the main trial (academic year 2022–2023 only). The second year of the TDTScience trial runs 2023–2024, the results of which are to be reported in 2025; the full timeline for the second year of the TDTScience trial can be found within the TDTScience protocol (Fairhurst 2023).

Table 7: TDTScience evaluation timeline (Cohort 1, academic year 2022–2023)

Dates	Activity	Staff responsible / leading
May 2020	Ethical approval granted	Evaluation team
Early-mid 2020	Recruitment of trainers	Developer
April 2021	ISRCTN registration	Evaluation team
January 2022–June 2022	Recruitment of trial schools	Developer
June 2022	Randomisation of trial schools	Evaluation team
September 2022–December 2022	Collect pupil details	Evaluation team
September 2022–June 2023	Intervention period (delivery of TDTScience to teachers in trial schools)	Trainers
September 2022–April 2023	Evaluation of trainers delivery to trial schools	Evaluation team
October 2022–December 2022	Teacher baseline survey data collection	Evaluation team
January 2023–May 2023	Case study visits to schools	Evaluation team
June 2023–July 2023	Pupil outcome assessments	Evaluation team
June 2023–July 2023	Pupil survey and attitude to science questionnaire collection	Evaluation team
June 2023–July 2023	Teacher survey follow-up data collection	Evaluation team
January 2023–September 2023	Code/analyse IPE data	Evaluation team
July 2023–March 2024	Assessment marking, data entry, and analysis	Evaluation team
March 2024	Draft report submitted to the EEF (reporting primary outcome and IPE data collected within 2022–2023)	Evaluation team

Dates	Activity	Staff responsible / leading
July 2024	Submission to the EEF of final report (reporting primary outcome and IPE collected within 2022–2023)	Evaluation team

Impact evaluation results

Participant flow including losses and exclusions

School recruitment and attrition

Between January 2022 and June 2022, 4,402 schools were approached to take part in the trial across the six geographical regions, of which 343 (7.8%) returned an Expression of Interest form (Figure 2). Signed MoUs were returned by 180 schools, which were then considered recruited and were randomised (intervention, n=90; control, n=90). It was necessary to randomise schools before the summer holidays so that communication to intervention schools could occur before the start of the 2022–2023 academic year. In September 2022, randomised schools were asked to provide pupil details for their eligible pupils. Pupil details were returned from 168 schools (intervention, n=89; control, n=79) between September 2022 and December 2022; the remaining 12 schools fully withdrew from the study before providing these details. All bar one of the 168 participating schools provided post-test Year 5 Science Assessment data; one school in the intervention group was not asked to complete post-tests as the evaluation team became aware that the school was running a rolling curriculum and Year 5 pupils had completed Year 6 topics, which made them ineligble for the trial.

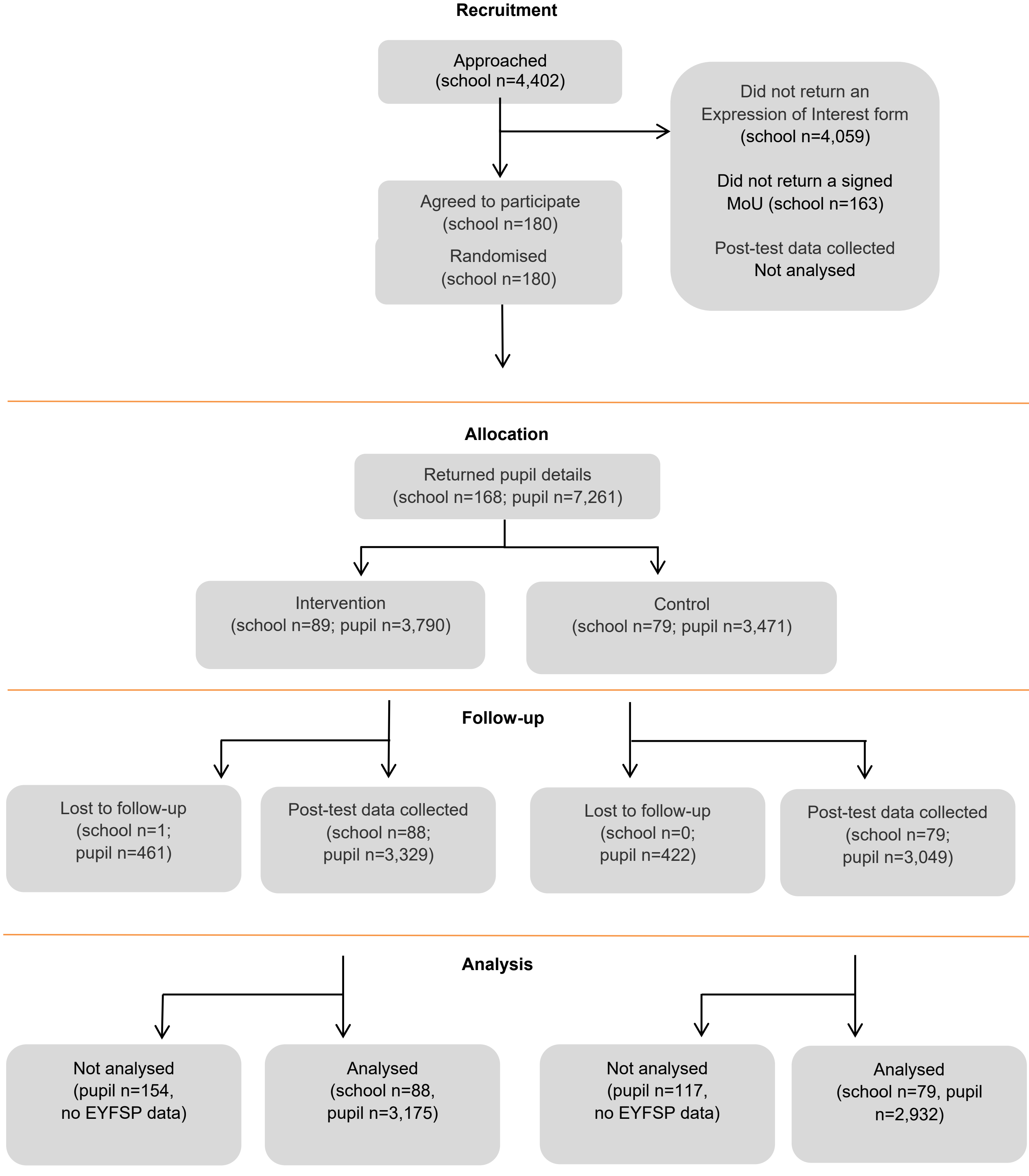
Cohort 1 pupil recruitment and attrition

For Cohort 1, at the start of the academic year 2022–2023, randomised schools were asked to provide details of eligible pupils in Year 5 whose parents had not withdrawn them from the evaluation. Pupil details were returned from 168 schools (n=7,261 pupils, intervention, n=3,790; control, n=3,471; mean per school 43.2, SD 20.7, median 33, range 15 to 121). These pupils form our randomised sample.

Post-testing with the Year 5 Science Assessment was due to take place for 7,227 pupils across 167 schools (intervention, n=3,756/3,790, 99.1%; control, n=3,471/3,471, 100%). Post-testing was not due to take place with 34 pupils, all in the intervention group, as the evaluation team had been informed that they had left the school (n=3) or their school had been withdrawn from the trial (n=31 pupils) before this was due.

A post-test science assessment score was avalable for 6,378 pupils from 167 schools (intervention, n=3,329/3,790, 87.8%; control n=3,049/3,471, 87.8%; mean per school 38.2, SD 18.5, median 32, range 14 to 110). Reasons for data being unavailable for the remaining 849 include: pupil was absent on day of testing (n=505); pupil had left school (n=138); pupil had withdrawn (n=127); and other/unknown reason (n=79).

Figure 2: Participant flow diagram (two arms)



At randomisation, there were 180 settings. For logistical reasons, schools had to be randomised and informed of their allocation at the end of the academic year 2021–2022 so intervention schools could begin to make arrangements to attend the training. The evaluation team could only collect pupil details from participating schools at the start of the academic year 2022–2023 when schools knew, which pupils would be in the class. Twelve schools withdrew during this process, and so the number of schools from which we received participating pupil details was 168 (7,261 pupils; intervention, n=3,790; control, n=3,471). This is an average of 43.2 pupils per school. Assuming 80% power, a pre- and post-test correlation of 0.5, an ICC of 0.15, and 15% pupil-level attrition, the MDES with this sample size would be approximately 0.16 (Table 8).

The actual observed ICC at the school level obtained from the primary analysis model was 0.10 (95% CI: 0.08 to 0.13). The correlation between the pre- and post-test scores of the participants included in the primary analysis was 0.50. Based on the number of pupils included in the primary analysis model (n=6,107), and the observed ICC and baseline/post-test correlation, the estimated MDES at analysis for the primary outcome was 0.13.

A total of 1,852 of the randomised pupils were eligible for EVER6FSM and could be included in analyses (average of 11 per school). The correlation between baseline and post-test scores within the FSM sample was 0.48 and the ICC associated with school was 0.05. With this sample size, the MDES was therefore, approximately 0.14.

Table 8: MDES at different stages

		Protocol		Randomisation		Analysis	
		Overall	FSM	Overall	FSM	Overall	FSM
MDES		0.15[1]	0.19[1]	0.16[a]	0.17[a]	0.13	0.14
Pre-/post-test correlations	Level 1 (pupil)	0.50	0.50	0.50	0.50	0.50	0.48
	Level 2 (class)	–	–	–	–	–	–
	Level 3 (school)	–	–	–	–	–	–
ICCs	Level 2 (class)	0.15	0.15	0.15	0.15	0.10	0.05
	Level 3 (school)	–	–	–	–	–	–
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		45	8	43	13	36	11
Number of schools	Intervention	90	90	89	89	88	88
	Control	90	90	79	79	79	79
	Total	180	180	168	168	167	167
Number of pupils	Intervention	4,050	700	3,790	1,086	3,175	916
	Control	4,050	700	3,471	1,100	2.932	936
	Total	8,100	1,400	7,261	2,168	6,107	1,852

[a] *Accounting for 15% attrition.*

Attrition

A total of 6,107 of the 7,261 randomised pupils (intervention, n=3,175/3,790, 83.8%; control n=2,932/3,471, 84.5%) from 167 schools had valid baseline EYFSP average point score and post-test Year 5 Science Assessment data and were included in the primary analysis. This equates to an overall attrition rate of 15.9% (Table 9).

Table 9: Pupil-level attrition from the trial (primary outcome)

		Intervention	Control	Total
Number of pupils	Randomised	3,790	3,471	7,261
	Analysed	3,175	2,932	6,107
Pupil attrition (from randomisation to analysis)	Number	615	539	1,154
	Percentage	16.2	15.5	15.9

Pupil and school characteristics

Characteristics for the 180 randomised schools and 7,261 participating pupils are presented in Table 10 and are similar between the two groups. No formal hypothesis testing was performed on baseline data (Senn, 1994), so comparisons are made visually only.

Half of the participating pupils were male (intervention, n=1,873/3,790, 49.4%; control, n=1,813/3,471, 52.2%) and, overall, 30.1% were eligible for FSM, though there was some missing data in this variable (6.8%) so the proportion eligible for FSM out of those with valid data was 32.3%.

One source we could identify for providing some national-level data to assess the representativeness of our recruited sample of schools and pupils is a government report published annually on schools, pupils, and their characteristics, the latest publication being June 2023.[5] Data indicate that the percentage of pupils in state-funded primary schools ever eligible for FSM in the last six years in 2022–2023 was 24.0%, which is lower than the average percentage of ever FSM pupils in our recruited schools at the time of enrolment into the trial (approximately a third).

Another source of national data is the 'Get Information about Schools (GIAS)' government service. According to data accessed in October 2022, restricting to primary schools in England, just under half are community schools (47.0%), 16.9% are academy converter, 15.2% are voluntary aided schools, 10.2% are voluntary controlled schools, 6.1% are academy sponsor led, 3.7% are foundation schools, and 0.9% are free schools. Three-quarters are in urban locales (73.1%), 25.3% are rural, and 1.6% are missing these data. An Ofsted rating was unavailable for 38.3% of schools (N.B. common reasons why a school does not have an Ofsted rating include the school having recently converted to an academy or it has been newly established within the past three years), but for those that did have these data, 15.6% were rated 'Outstanding' in their latest inspection, 75.4% were rated 'Good', and 9.0% were rated 'Requires improvement'.

In our randomised sample, relative to the national picture, we have a lower proportion of community schools (28.3%) and a higher proportion of academy converter schools (36.7%); we have a higher proportion of schools in urban areas and a lower proportion in rural areas (82.8% and 17.2%, respectively); and, of the schools for which the latest Ofsted inspection rating was available, we have a lower proportion of schools rated 'Outstanding' (9.5%) and a higher proportion of schools rated 'Good' (81.0%).

EYFSP data, to be used as a measure of prior attainment, was obtained from the NPD. NPD data were requested for 7,120 randomised pupils (98.1%; intervention, n=3,716/3,790, 98.1%; control, n=3,404/3,471, 98.1%) since 141 pupils withdrew from the evaluation throughout the course of the trial, which invalidated their consent for the evaluation team to access their NPD data. NPD data could be matched for 6,792 pupils of which 6,785 had valid EYFSP data. This equates to 93.4% of the randomised population (intervention, n=3,527/3,790, 93.1%; control, n=3,258/3,471, 93.9%).

[5] https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics

The mean EYFSP average point score was slightly higher in the intervention group than the control group (2.07, SD 0.45, compared with 2.03, SD 0.44; Hedges' g effect size between the groups 0.09; 95% CI: 0.04 to 0.13).

Table 10: Baseline characteristics of randomised schools and pupils, by group, as randomised

School level (categorical)	Intervention group (n=90)	Control group (n=90)
	Count (%)	Count (%)
Location		
North East	22 (24.4)	22 (24.4)
Lancashire	18 (20.0)	17 (18.9)
Yorkshire	16 (17.8)	15 (16.7)
Staffordshire and West Midlands	12 (13.3)	13 (14.4)
South West	12 (13.3)	12 (13.3)
Lincolnshire and East Midlands	10 (11.1)	11 (12.2)
%FSM[a]		
≥24%	44 (48.9)	44 (48.9)
<24%	46 (51.1)	46 (51.1)
Rural urban classification		
Urban	72 (80.0)	77 (85.6)
Rural	18 (20.0)	13 (14.4)
Type of school		
Academy converter	33 (36.7)	33 (36.7)
Community school	25 (27.8)	26 (28.9)
Voluntary aided school	12 (13.3)	13 (14.4)
Academy sponsor led	8 (8.9)	11 (12.2)
Voluntary controlled school, or Foundation school or Free school[b]	12 (13.3)	7 (7.8)
Ofsted rating		
Outstanding	8 (8.9)	7 (7.8)
Good	60 (66.7)	68 (75.6)
Requires improvement	8 (8.9)	7 (7.8)
Missing	14 (15.6)	8 (8.9)
Cohort 1	N=3,790	N=3,471
Pupil level (categorical)	Count (%)	Count (%)
Gender		

Male	1,873 (49.4)	1,813 (52.2)	
Female	1,917 (50.6)	1,658 (47.8)	
Eligible for FSM			
Yes	1,086 (28.7)	1,100 (31.7)	
No	2,432 (64.2)	2,151 (62.0)	
Missing	272 (7.2)	220 (6.3)	
Pupil level (continuous)	Mean (SD) n=3,527	Mean (SD) n=3,258	Effect size
EYFSP average point score	2.07 (0.45)	2.03 (0.44)	0.09 (0.04, 0.13)

[a] Based on data provided by schools at the start of the evaluation, dichotomised for the randomisation at the median level for the recruited schools. FSM based on % of pupils eligible for FSM at any time during the past six years.
[b] Rows combined to prevent statistical disclosure relating to cell counts for schools less than three.

Table 11: Baseline characteristics of randomised schools and pupils, by group, as included in the primary analysis

School level (categorical)	Intervention group (n=88)	Control group (n=79)	
	Count (%)	Count (%)	
Location			
North East	21 (23.9)	20 (25.3)	
Lancashire	18 (20.5)	16 (20.3)	
Yorkshire	16 (18.2)	14 (17.7)	
Staffordshire and West Midlands	11 (12.5)	11 (13.9)	
South West	12 (13.6)	9 (11.4)	
Lincolnshire and East Midlands	10 (11.4)	9 (11.4)	
%FSM[a]			
≥24%	43 (48.9)	41 (51.9)	
<24%	45 (51.1)	38 (48.1)	
Rural urban classification			
Urban	72 (81.8)	69 (87.3)	
Rural	16 (18.2)	10 (12.7)	
Type of school			
Academy converter	32 (36.4)	29 (36.7)	
Community school	24 (27.3)	22 (27.8)	
Voluntary aided school	12 (13.6)	13 (16.5)	
Academy sponsor led	8 (9.1)	10 (12.7)	
Voluntary controlled school, or Foundation school, or Free school[b]	12 (13.6)	5 (6.3)	

Ofsted rating			
Outstanding	8 (9.1)	7 (8.9)	
Good	59 (67.0)	58 (73.4)	
Requires improvement	7 (8.0)	6 (7.6)	
Missing	14 (15.9)	8 (10.1)	
Cohort 1	N=3,175	N=2,932	
Pupil level (categorical)	Count (%)	Count (%)	
Gender			
Male	1,553 (48.9)	1,512 (51.6)	
Female	1,622 (51.1)	1,420 (48.4)	
Eligible for FSM			
Yes	916 (28.9)	936 (31.9)	
No	2,259 (71.1)	1,996 (68.1)	
Missing	0 (0.0)	0 (0.0)	
Pupil level (continuous)	Mean (SD)	Mean (SD)	Effect size
EYFSP average point score	2.08 (0.45)	2.05 (0.43)	0.08 (0.03, 0.13)

[a] *Based on data provided by schools at the start of the evaluation, dichotomised for the randomisation at the median level for the recruited schools. FMS based on % of pupils eligible for FSM at any time during the past six years.*
[b] *Rows combined to prevent statistical disclosure relating to cell counts for schools less than three.*

Characteristics for the 167 schools and 6,107 pupils included in the primary analysis are presented in Table 11, and are very similar to the randomised population.

Outcomes and analysis

Primary analysis

A valid baseline average EYFSP point score was available for 6,785 randomised pupils (intervention, n=3,527; control, n=3,258), from 168 randomised schools. The ICC associated with school for the baseline score is 0.10 (95% CI: 0.08 to 0.13). In total, a valid post-test science assessment score was obtained for 6,378 randomised pupils (87.8%; intervention, n=3,329/3,790, 87.8%; control, n=3,049/3,471, 87.8%), from 167 schools (intervention, n=88; control, n=79). A mean of 23.1 (95% CI: 22.8 to 23.4) was observed in the intervention arm and 22.7 (95% CI: 22.4 to 22.9) in the control arm. The unadjusted mean difference is 0.45 (95% CI: 0.06 to 0.84). Histograms of the baseline and post-test scores show the science assessment score to be reasonably normally distributed, while the EYFSP average point score has a symmetric distribution but with a clear mode of 2 (Figure 3). Baseline and post-test scores were available for 6,107 pupils (84.1%; intervention, n=3,175/3,790, 83.8%; control, n=2,932/3,471, 84.5%). The correlation between the baseline and post-test scores was 0.50 (95% CI: 0.48 to 0.52). As a check of the analysis model assumptions, the distribution of the standardised residuals was checked using a QQ plot and was shown to be normal (

Figure 4).

The adjusted mean difference in post-test score between the intervention and control groups was 0.17 (95% CI: -0.58 to 0.92, p=0.66, Appendix Table 1). The estimated Hedges' g effect size was 0.02 (95% CI: -0.07 to 0.12; Table 12), which does not relate to any months' additional progress in the intervention group. The total variance used to calculate the effect size was 63.27; the sum of 54.94 (random variation between pupils, within-cluster variance) and 8.33 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.10 (95% CI: 0.08 to 0.13). The ICC for the empty model (i.e. without covariates) was 0.13 (95% CI: 0.10 to 0.17).

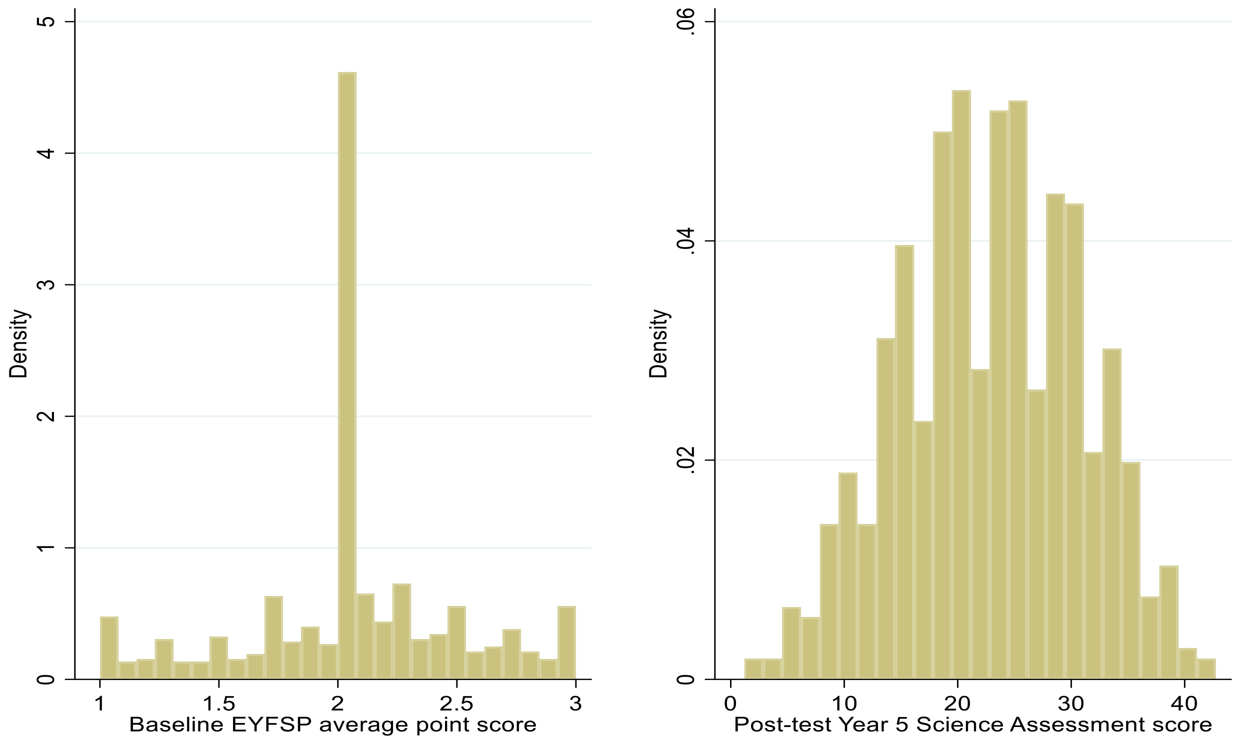


Figure 3: Histograms of baseline EYFSP average point and post-test Year 5 Science Assessment scores (scores have been aggregated to present the mean score for every ten pupils to prevent statistical disclosure)

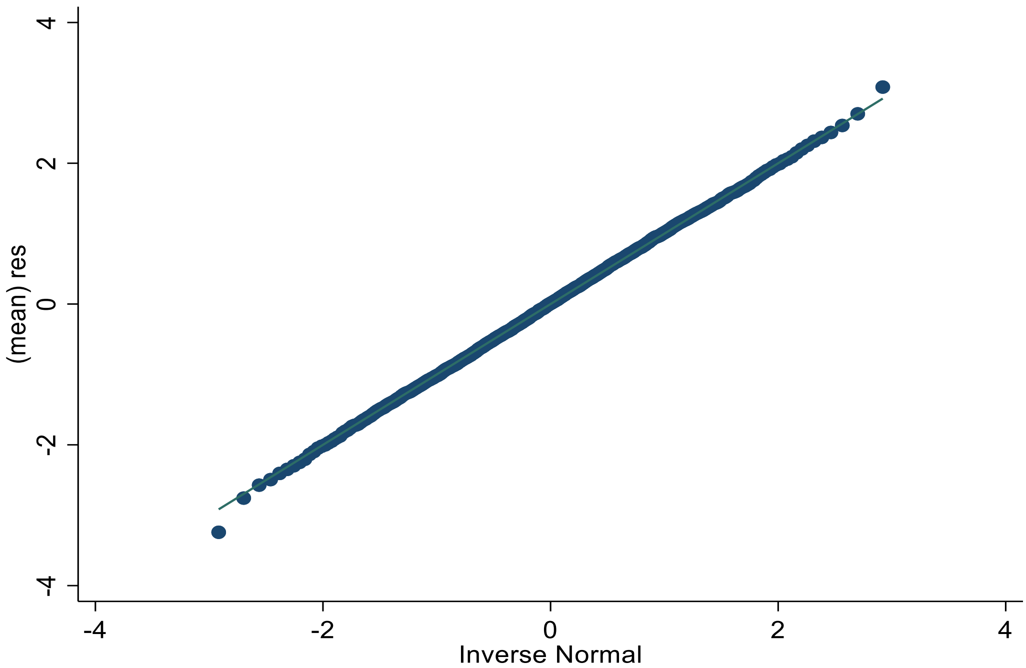


Figure 4: QQ plot of the standardised residuals from the primary analysis model to assess the normality assumption (residuals have been aggregated to present the mean residual for every ten pupils to prevent statistical disclosure)[6]

Table 12: Primary analysis (primary outcome)

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
Cohort 1 Year 5 Science Assessment (primary analysis)	3,329 (461)	23.1 (22.8, 23.4)	3,049 (422)	22.7 (22.4, 22.9)	6,107 (3,175; 2,932)	0.02 (-0.07, 0.12)	0.66
Cohort 1 Year 5 Science Assessment (eligible for FSM)	916 (170)	19.8 (19.4, 20.3)	939 (161)	19.7 (19.2, 20.2)	1,852 (916; 936)	0.00 (-0.11, 0.10)	0.96

Subgroup analyses

Summary statistics for the post-test Year 5 Science Assessment score are presented by FSM eligibility (Table 13). These summaries indicate that, in general, pupils eligible for FSM performed slightly worse on the post-test than those not eligible for FSM. In an adjusted regression analysis that an included interaction effect, there was evidence of an interaction between trial allocation and FSM eligibility (-1.79, 95% CI: -2.28 to -1.30, p<0.001).

Within the randomised sample, there was 2,186 pupils from 168 schools who were eligible for FSM (median per school 11, range 0 to 41 in both groups). Within the restricted sample of pupils eligible for FSM, the correlation between the baseline and post-test scores was 0.48 (95% CI: 0.44 to 0.51). The adjusted mean difference in post-test score between the intervention and control groups within this sample was -0.02 (95% CI: -0.80 to 0.76, p=0.96) and the estimated Hedges' g effect size was 0.00 (95% CI: -0.11 to 0.10), which does not relate to any additional months' progress in the

---

[6] QQ plot demonstrates how the distribution quantiles of our model residuals (y axis) line up with the normal distribution (x axis)—the dots should not deviate too far from the reference line.

intervention group. The total variance used to calculate the effect size (from a model without covariates) was 53.65—the sum of 50.43 (random variation between pupils, within-cluster variance) and 3.22 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.05 (95% CI: 0.03 to 0.09). The ICC for the empty model (i.e. without covariates) was 0.06 (95% CI: 0.03 to 0.11).

As a sensitivity check, the effect size for the FSM-eligible subgroup from the interaction effects model was extracted and compared with the effect size derived from the restricted sample. The adjusted mean difference in post-test score between the intervention and control groups within this sample was -0.06 (95% CI: -0.88 to 0.75, p=0.88) and the estimated Hedges' g effect size was -0.01 (95% CI: -0.11 to 0.09), which does not relate to any additional months' progress in the intervention group.

Table 13: Subgroup summary scores for the post-test Year 5 Science Assessment score

	Raw means			
	Intervention		Control	
FSM eligibility	n	Mean (SD)	N	Mean (SD)
Yes	916	19.8 (7.5)	939	19.7 (7.2)
No	2,259	24.5 (7.8)	1,996	24.1 (7.6)

Analysis in the presence of non-compliance

Ninety schools were randomised to the intervention group. One school did not provide pupil details and so was withdrawn from the trial and a further two schools formally withdrew from the TDTScience programme and did not send any teachers to the training sessions. Training attendance data for the four training sessions was provided by the delivery team for 199 teachers from 87 intervention schools (median 2, range 1 to 4, per school). Two (1.0%) of the teachers did not attend any sessions, 15 (7.5%) of the teachers attended one session, 23 (11.6%) of the teachers attended two sessions, 36 (18.1%) of the teachers attended three sessions, and 123 (61.8%) of the teachers attended all four sessions.

Schools were asked to inform the evaluation team of the teachers that would be teaching Year 5 science and the class they would be teaching in the academic year 2022–2023. Seven schools (7.7% of the randomised sample) had participating classes for whom a teacher was not provided, so these could not be matched to any teacher attendance data for the training sessions. Pupils in these classes were deemed to be non-compliant for the purposes of the CACE analysis. Of the randomised intervention pupils with Year 5 assessment data, 81.0% were deemed compliant.

The CACE estimate of the treatment effect on the pupils' attainment was a predicted increase of 0.21 points in Year 5 Science Assessment score (95% CI: -0.72 to 1.15, p=0.65; effect size 0.03, 95% CI: -0.09 to 0.14). The partial R^2 from the first stage of the CACE analysis was 0.68 and the F-statistic was F(1, 166 = 645.9, p<0.001); these indicate a high correlation between the instrumental variable (random allocation) and the endogenous variable, and that the inference of the CACE estimate is reliable.

Missing data analysis

A mixed effect logistic regression model suggested that a 'missing at random' (MAR) assumption was reasonable, meaning that MI by chained equations was an appropriate approach. Among the randomised population, pupils eligible for FSM were almost twice as likely to be missing from the primary analysis model than non-FSM pupils (adjusted OR 1.92, 95% CI: 1.60 to 2.30, p<0.001). Additionally, pupils with higher EYFSP average point scores were less likely to be missing (OR 0.54, 95% CI: 0.45 to 0.66, p<0.001). Type of school, overall, was also associated with missingness.

To investigate the impact of missing data, the primary analysis was repeated using MI by chained equations. The adjusted mean difference in Year 5 Science Assessment score following MI was 0.15 (95% CI: -0.54 to 0.83, p=0.68) and the Hedges' g effect size was 0.02 (95% CI: -0.07 to 0.10). This is a slightly smaller effect than the primary analysis with a slightly wider CI.

Secondary analysis

Science attitudes

In total, a valid post-test Year 5 Science Attitudes score was obtained for 6,372 randomised pupils (87.8%; intervention, n=3,320/3,790, 87.6%, control, n=3,052/3,471, 87.9%), from 167 schools (intervention, n=88; control, n=79). A mean of 0.07 (95% CI: 0.04 to 0.11) for the standardised score was observed in the intervention group and -0.08 (95% CI: -0.12 to -0.04) in the control group. The unadjusted mean difference is 0.15 (95% CI: 0.10 to 0.20). Histograms of the raw and standardised science attitudes score show them to have a slightly left-skewed distribution (

Figure 5). Baseline EYFSP average point and post-test science attitudes scores were available for 6,042 pupils (83.2%; intervention, n=3,127/3,790, 82.5%; control, n=2,915/3,471, 84.0%). The correlation between the baseline and post-test scores was 0.06 (95% CI: 0.03 to 0.08). One pupil did not have valid FSM data and so the analysis model included 6,041 pupils (83.2%; intervention, n=3,127/3,790, 82.5%; control, n=2,914/3,471, 84.0%). As a check of the analysis model assumptions, the distribution of the standardised residuals was checked using a QQ plot and was shown to be normal (Figure 6).

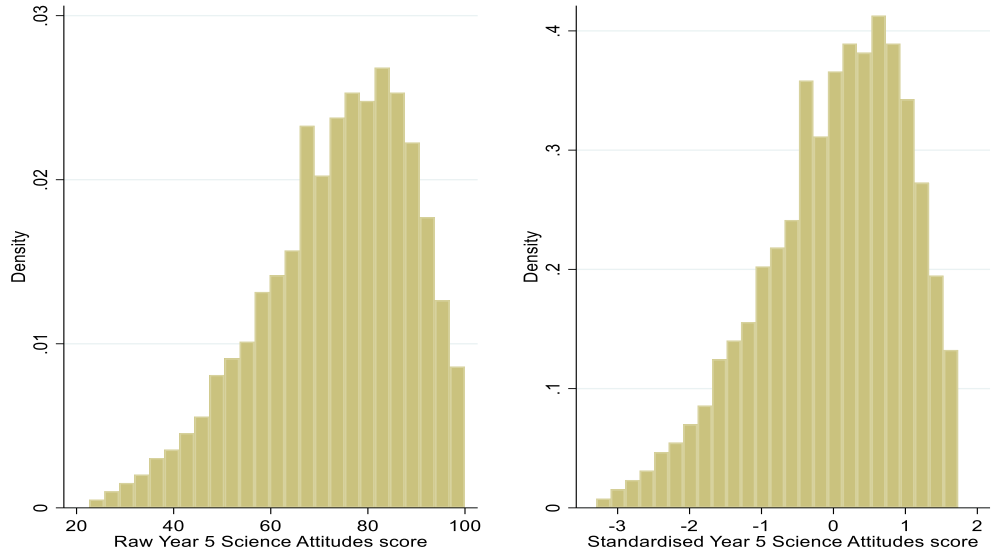


Figure 5: Histograms of raw and standardised post-test Year 5 Science Attitudes scores (scores have been aggregated to present the mean score for every ten pupils to prevent statistical disclosure)

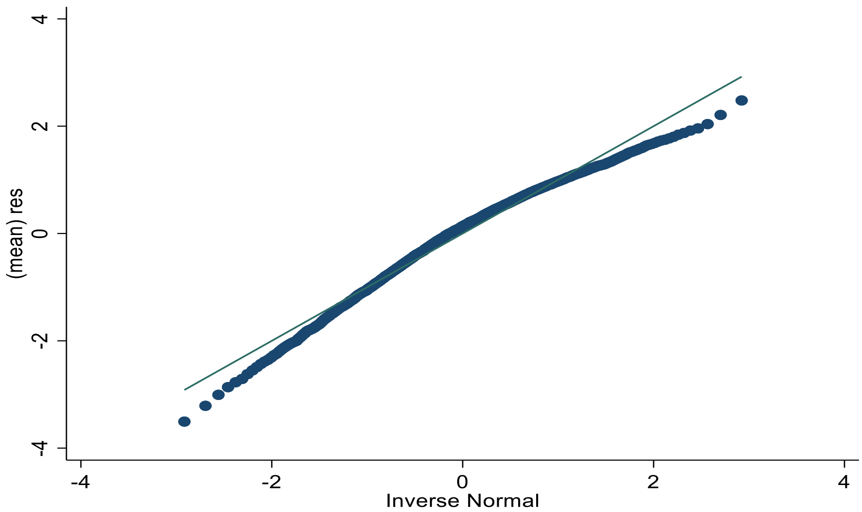


Figure 6: QQ plot of the standardised residuals from the secondary analysis model to assess the normality assumption (residuals have been aggregated to present the mean residual for every ten pupils to prevent statistical disclosure)

Table 14: Secondary outcome analysis

	Unadjusted means				Effect size		
	Intervention group		Control group				
Outcome	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
Cohort 1: Year 5 Science Attitudes Questionnaire	3,320 (470)	0.07 (0.04, 0.11)	3,052 (419)	-0.08 (-0.12, 0.04)	6,041 (3,127: 2,914)	0.12 (0.01, 0.24)	0.04

Table 15: Summary of remaining science attitudes items not used in scale

Science Attitudes Questionnaire, n (%)	Intervention (n=3,320) n (%)	Control (n=3,052) n (%)	Overall (n=6,372) n (%)
We spend a lot of time in science copying from the board			
Agree a lot	347 (10.5)	422 (13.8)	769 (12.1)
Agree a bit	616 (18.6)	632 (20.7)	1,248 (19.6)
Not sure	778 (23.4)	741 (24.3)	1,519 (23.8)

Disagree a bit	847 (25.5)	736 (24.1)	1,583 (24.8)
Disagree a lot	716 (21.6)	499 (16.3)	1,215 (19.1)
Missing	16 (0.5)	22 (0.7)	38 (0.6)
I think science is more for boys			
Agree a lot	81 (2.4)	84 (2.8)	165 (2.6)
Agree a bit	79 (2.4)	79 (2.6)	158 (2.5)
Not sure	364 (11.0)	355 (11.6)	719 (11.3)
Disagree a bit	217 (6.5)	211 (6.9)	428 (6.7)
Disagree a lot	2565 (77.3)	2312 (75.8)	4877 (76.5)
Missing	14 (0.4)	11 (0.4)	25 (0.4)
We often have discussions in science lessons			
Agree a lot	1978 (59.6)	1669 (54.7)	3647 (57.2)
Agree a bit	899 (27.1)	918 (30.1)	1817 (28.5)
Not sure	295 (8.9)	311 (10.2)	606 (9.5)
Disagree a bit	86 (2.6)	85 (2.8)	171 (2.7)
Disagree a lot	34 (1.0)	46 (1.5)	80 (1.3)
Missing	28 (0.8)	23 (0.8)	51 (0.8)
We do a lot of writing in science lessons			
Agree a lot	683 (20.6)	971 (31.8)	1654 (26.0)
Agree a bit	1133 (34.1)	1096 (35.9)	2229 (35.0)
Not sure	664 (20.0)	537 (17.6)	1201 (18.8)
Disagree a bit	628 (18.9)	358 (11.7)	986 (15.5)
Disagree a lot	200 (6.0)	79 (2.6)	279 (4.4)
Missing	12 (0.4)	11 (0.4)	23 (0.4)
I have made good progress in science this year			
Agree a lot	1157 (34.8)	1013 (33.2)	2170 (34.1)
Agree a bit	959 (28.9)	834 (27.3)	1793 (28.1)
Not sure	853 (25.7)	821 (26.9)	1674 (26.3)
Disagree a bit	169 (5.1)	195 (6.4)	364 (5.7)
Disagree a lot	177 (5.3)	184 (6.0)	361 (5.7)
Missing	5 (0.2)	5 (0.2)	10 (0.2)
We already know what will happen when we do science practical work			
Agree a lot	266 (8.0)	243 (8.0)	509 (8.0)
Agree a bit	554 (16.7)	539 (17.7)	1093 (17.2)
Not sure	1057 (31.8)	963 (31.6)	2020 (31.7)
Disagree a bit	714 (21.5)	662 (21.7)	1376 (21.6)
Disagree a lot	722 (21.7)	637 (20.9)	1359 (21.3)
Missing	7 (0.2)	8 (0.3)	15 (0.2)
We can decide for ourselves how to do science practical work			
Agree a lot	496 (14.9)	347 (11.4)	843 (13.2)
Agree a bit	873 (26.3)	711 (23.3)	1584 (24.9)
Not sure	1124 (33.9)	1081 (35.4)	2205 (34.6)
Disagree a bit	487 (14.7)	427 (14.0)	914 (14.3)
Disagree a lot	302 (9.1)	453 (14.8)	755 (11.8)
Missing	38 (1.1)	33 (1.1)	71 (1.1)
We do practical work in most science lessons			
Agree a lot	741 (22.3)	560 (18.3)	1301 (20.4)
Agree a bit	1113 (33.5)	917 (30.0)	2030 (31.9)

Not sure	791 (23.8)	776 (25.4)	1567 (24.6)
Disagree a bit	468 (14.1)	544 (17.8)	1012 (15.9)
Disagree a lot	173 (5.2)	236 (7.7)	409 (6.4)
Missing	34 (1.0)	19 (0.6)	53 (0.8)

The adjusted mean difference in post-test score between the intervention and control groups was 0.12 (95% CI: 0.01 to 0.24, p=0.04, Appendix Table 1). The estimated Hedges' g effect size was 0.12 (95% CI: 0.01 to 0.24; Table 14), corresponding to a small effect. The total variance used to calculate the effect size was 1; the sum of 0.88 (random variation between pupils, within-cluster variance) and 0.13 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.12 (95% CI: 0.10 to 0.15). The ICC for the empty model (i.e. without covariates) was 0.13 (95% CI: 0.10 to 0.16).

The remaining items, not used in the scale, are summarised separately (Table 15). Here, pupils' responses in both groups were broadly similar for the majority of science attitudes items with the exception of 'we do a lot of writing in science lessons'. In response to this statement, a greater proportion of pupils in the control group selected that they 'agree a lot' with the statement in comparison to the intervention group (control, 31.8%; intervention 20.6%). Similarly, a greater proportion of pupils in the intervention group selected they 'disagree a bit' with the statement in comparison to the control group (control, 11.7%; intervention 18.9%).

Implementation and Process Evaluation

Summary of IPE data collected by method

Teacher surveys

Table 16 summarises the IPE related surveys sent to participating schools over the course of the trial, including the number of valid responses received for each by allocation. Reminder emails were sent to non-responders to encourage completion at each timepoint. Survey responses were cleaned at each timepoint to remove invalid responses. Responses from unintended responders (e.g. a staff member who was not a Year 5 teacher) were not included within the analysis. Where duplicate responses were received, the first complete response from the correct intended respondent was used for analysis. Only valid responses were analysed.

Table 16: Summary of valid responses to IPE surveys at each timepoint by allocation

Survey type / intended respondent	Intervention	Control	Total
Pre-randomisation baseline survey / Year 5 teacher (October 2022–December 2022)	115[a]	83[b]	198
Follow-up survey / Year 5 teacher (June 2023–July 2023)	76[c]	75[d]	151

[a]115 valid responses were received from 80 intervention schools.
[b] 83 valid responses were received from 62 control schools.
[c] 76 valid responses were received from 61 intervention schools.
[d]75 valid responses were received from 57 control schools.

Trainer surveys

A short, online feedback survey was sent to the 12 relevant trainers following each of the five training days. All 12 trainers completed all five surveys at the relevant timepoint.

Case studies

Table 17: Comparison of characteristics between randomised intervention schools and IPE schools

School level (categorical)	Intervention (randomised, n=90)	Intervention (IPE sub-study sample, n=12)
	Count (%)	Count (%)
%FSM[a]		
≥24%	44 (48.9)	5 (41.7)
<24%	46 (51.1)	7 (58.3)
Rural urban classification		
Urban	72 (80.0)	≥10 (≥75%)
Rural	18 (20.0)	<3 (<25%)
Type of school		
Academy converter	33 (36.7)	5 (41.7)
Community school, or Foundation school, or Free school[a]	31 (34.4)	4 (33.3)
Voluntary aided school, or Academy sponsor led, or Voluntary controlled school[a]	26 (28.9)	5 (41.7)
Ofsted rating		

Outstanding or Good[a]	68 (75.6)	9 (75.0)
Requires improvement or Missing[a]	22 (24.4)	3 (25.0)

[a] Rows combined to prevent statistical disclosure relating to cell counts for schools less than three.

The total number of schools randomised to the intervention was 90, so the 12 schools (two from each of the six recruiting regions) that participated in case studies represents 13%. Table 17 above, provides a comparison of school-level characteristics between all schools randomised to the intervention group and the group of schools that participated in the IPE. Case study schools were broadly representative of all randomised intervention schools.

Fidelity

This section will address the IPE research questions relating to fidelity, both in terms of training and in terms of classroom practice. In doing so it addresses the following research questions:

1. To what extent was TDTScience implemented as planned?

    a. Training.
    b. Classroom practice.

2. What processes are involved for teachers and schools implementing TDTScience—what are the main facilitators and barriers?

3. What are the perceptions of teachers as regards TDTScience?

    a. What are their opinions about training and support, including cascading from colleagues where relevant?
    b. What are their views of TDTScience strategies and techniques?
    c. What impacts has TDTScience had on their classroom practice?

In order to address these research questions this section draws on data collected from the sources:

- training day observations;
- trainer interviews;
- trainer feedback surveys;
- case study school visits, primarily the classroom observations and interviews with teachers;
- teacher survey feedback on training (collected by the developers); and
- developer interview.

Training

There were four and a half-days training in total, distributed over the academic year to enable teachers to embed their practice. The first four days each exemplified the TDTScience approach using a topic from the Year 5 science curriculum.[7] The remaining half-day was designed to discuss further embedding of the programme, facilitate teachers developing their own resources, and discuss next steps. The overview of the training days and when they took place during the school year is provided in Table 18. Each training day was delivered once in each area with the exception of the North East where it was delivered twice to two cohorts of teachers (due to the large number of recruited schools in the North East).

Table 18: Training day schedule

Training day[a]	Focus	Month of delivery
Day 1	Materials	September 2022–October 2022
Day 2	Forces	November 2022–December 2022

[7] See: https://www.gov.uk/government/publications/national-curriculum-in-england-science-programmes-of-study/national-curriculum-in-england-science-programmes-of-study#year-5-programme-of-study

Day 3	Earth and Space	January 2023–February 2023
Day 4	Living Things	March 2023
Day 5	Review and Embedding, Ready for Next Steps	May 2023

[a] Each training session was a scheduled full-day training (six hours + 30 minutes for registration) with the exception of Day 5, which was a half-day (approximately three and a half-days, four hours).

Researchers visited one of each training day and one in each region, meaning that two days were visited twice in two different areas (Day 3 and Day 4). Therefore, a total of six visits were made to training sessions, for the purposes of observation and to interview the trainers delivering that day's session.

The training all followed the timetable and slides provided in the TDTScience trainer folder. It was designed to be co-delivered by two trainers. As the researcher notes for one training day recorded:

> *Trainers working fluently together, particularly within changeovers: Trainers contribute to each other's sections well and it doesn't feel disjointed when this happens, they work fluently together*. (Training day observation notes)

This well reflects the training days attended by the researchers and the trainers in their interviews discussed how they divided up who would present, which sections based on their individual strengths and, which aspects of the training they felt most comfortable with.

> *The previous year we had done sort of alternating and then thought actually, no, that's more my strength, that's more your strength but I think we're comfortable with what we're doing now.* (Trainer interview)

During the developer interview, it was acknowledged that there was variability between the training pairs in the programme and noted that some pairs needed some extra support. Nonetheless, the developers felt that the pairs worked well together in the training and would experiment with how best to share the delivery and improve their performance as trainers. Within the feedback survey administered following each training day, trainers were asked how confident they felt in training teachers in TDTScience. All 12 (100%) trainers strongly agreed that they felt confident on Days 1, 2, 4, and 5. On Day 3, 11 out of 12 (91.7%) trainers strongly agreed, and one trainer indicated they 'agreed slightly' with the statement.

Following delivery of each training day, trainers were also asked 'how well prepared were you to deliver the training'. The responses were very positive, and all trainers felt prepared, as summarised in Table 19. The trainers mostly felt 'very well prepared' to deliver the training to the teachers on each day, while some trainers felt 'quite well prepared'.

Table 19: Trainers feedback survey responses relating to preparation post-delivery (n=12 trainers)

Training day	Very well prepared *n* (%)	Quite well prepared *n* (%)	Not very well prepared *n* (%)	Not at all prepared *n* (%)
Day 1	9 (75.0)	3 (25.0)	0 (0.0)	0 (0.0)
Day 2	10 (83.3)	2 (16.7)	0 (0.0)	0 (0.0)
Day 3	11 (91.7)	1 (8.3)	0 (0.0)	0 (0.0)
Day 4	12 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Day 5	11 (91.6)	1 (8.3)	0 (0.0)	0 (0.0)

The number of slides per training day was extensive (approximately 100 slides per full training day) and this meant that the delivery tended to be '*pacey*' (Teacher interview, S6–T1) with some slides given more time than others. This appeared to be roughly in line with the guides provided with the training slides. As one trainer explained:

> *I mean it's so tight, ridiculously tight and, you know, if you were running your own course you might think, I'll do a bit more of that this afternoon and I'll pull back on that, but you can't. They've got the ring binder. So you can't even think, oh I'll miss a couple of those slides out. Every single bit of it has got to be included.* (Trainer interview)

However, as another trainer explained, the pressure was on the trainers to maintain the correct pace, rather than making the teachers feel rushed:

> *So we're not rushing them and we are helping them feel like they've had enough time to consider things, even though we're like paddling ferociously under the water (laughs), like nice and calm from the top.* (Trainer interview)

Within the feedback survey, trainers were asked how much they agreed or disagreed with the statement 'we were able to cover the whole agenda'. All 12 (100%) trainers strongly agreed that they were able to cover the whole agenda on training Days 1, 2, 3, and 5. On Day 4, 11 out of 12 (91.7%) of trainers strongly agreed, and one trainer indicated they 'disagreed slightly' with the statement.

The trainers, in their interviews, all commented that they were very aware of their responsibility to replicate the developers' training as faithfully as possible, particularly as the delivery was part of a randomised controlled trial:

> *It is part of a trial that a lot of money has gone into, we're very aware of that and that our colleagues are in different geographical regions and we're supposed to be doing the same thing so we we'd better not tweak too much*. (Trainer interview)

Trainers generally referred to minor 'tweaking' of the training when asked about any changes made to delivery. Adding aide memoires to training day slides and adapting anecdotes and examples was seen as acceptable, but changing emphasis or adding something new, was not. Any such 'tweaks' were often as a result of experiences of previous training delivery or drawing on their own experiences:

> *We added a couple of slides to clarify…the curvature of the earth…we found* [what] *we thought was a picture that showed the top of the sail more clearly than the one that they showed where they were all different, but we kept their slide in, it was just simply to make it a little bit clearer we felt.* (Trainer interview)

> *We kept to the script but we…yeah, you embellish it with your own examples every now and again*. (Trainer interview)

It should also be noted that any (minor) additions were not perceived to have changed the overall pace of the slide transitions, rather to allow trainers to clarify their explanations of difficult concepts more clearly based on their own experience.

There was also a need to tailor the training to the particular group of teachers at the training day, for example, in order to address any questions they had:

> *I remember a time when I'd, '[Name], right, I'll whizz through these next slides'. I thought these are quite straightforward ones, we'll gain a bit of time. And then I hadn't anticipated that one of the participants then asked a question, which turned into quite a lengthy question…and then you're like, 'Oh God', that's more important, isn't it, to talk through some issues that they've raised.* (Trainer interview)

Within the trainer feedback surveys, respondents were asked to indicate how closely 'they followed the tutor file'. All 12 (100%) trainers agreed strongly that they had followed the tutor file on all five days that they delivered the training. Trainers were also asked how much they agreed or disagreed that they 'were able to effectively/train teachers on key TDTScience components' (Table 20 above). Across all training days, all trainers strongly considered they could effectively teach the underlying principles and ethos of the TDTScience programme, except on Day 3, where one trainer suggested they only agreed slightly with this statement. Similarly, across all training days, all trainers either strongly agreed or slightly agreed that they could effectively teach Bright Ideas Time. 'Practical prompts for thinking' had slightly more mixed opinions, with trainers feeling more strongly that they could teach this on Day 1 and Day 2, the agreement reducing for Day 3 and Day 4. A similar pattern was reported for 'teaching practical investigations'. For 'focused recording' the responses ranged between agree strongly and agree slightly. For 'practical problem-solving' 75% of trainers agreed 'strongly' on Day 1 and Day 3 that they could effectively deliver it, but by Day 4 this reduced to 50% of trainers.

The dynamic of the group of teachers attending training was also felt to impact on delivery:

> *The group dynamics definitely, I think the conversations in the groups were possibly better this time than we've had in the past.* (Trainer interview)

When timing did slip a little it was generally made up in other parts of the day. In one observation where slides took less time than scheduled at the start of the day the trainers allowed for more time in the practical. In another example, extra time was spent focusing on HOT. This form of balancing was not felt by the researchers to be out of keeping with the programme, rather it was responsive to teacher needs. For example, in one training session teachers argued that they did not have time for asking pupils what they would like to know and the trainer suggested an alternative, an '*ask it basket'* (Training day observation notes).

They agreed that the train-the-trainer model had worked well and having already pre-delivered the training provided them with confidence:

> *Of course, the whole fact that we did the pre-trial running it before I think is really beneficial because once you do it the second time round you kind of feel a lot more confident.* (Trainer interview)

Table 20: Trainers perceptions on how effectively they delivered TDTScience

	Agree strongly *n* (%)	Agree slightly *n* (%)	Disagree slightly *n* (%)	Disagree strongly *n* (%)
The underlying principles and ethos of the TDTScience programme				
Day 1	12 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Day 2	12 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Day 3	11 (91.7)	1 (8.3)	0 (0.0)	0 (0.0)
Day 4	12 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Day 5	12 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Bright Ideas Time				
Day 1	11 (91.7)	1 (8.3)	0 (0.0)	0 (0.0)
Day 2	11 (91.7)	1 (8.3)	0 (0.0)	0 (0.0)
Day 3	12 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Day 4	11 (91.7)	1 (8.3)	0 (0.0)	0 (0.0)
Day 5	n/a	n/a	n/a	n/a
Practical prompts for thinking				
Day 1	11 (91.7)	1 (8.3)	0 (0.0)	0 (0.0)
Day 2	12 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)
Day 3	8 (72.7)	3 (27.3)	0 (0.0)	0 (0.0)
Day 4	6 (50.0)	6 (50.0)	0 (0.0)	0 (0.0)
Day 5				
Practical investigations				
Day 1	11 (91.7)	1 (8.3)	0 (0.0)	0 (0.0)
Day 2	11 (91.7)	1 (8.3)	0 (0.0)	0 (0.0)
Day 3	8 (66.7)	4 (33.3)	0 (0.0)	0 (0.0)
Day 4	6 (54.5)	5 (45.4)	0 (0.0)	0 (0.0)
Day 5	n/a	n/a	n/a	n/a
Focused recording				
Day 1	5 (41.7)	7 (58.3)	0 (0.0)	0 (0.0)
Day 2	10 (83.3)	2 (16.7)	0 (0.0)	0 (0.0)
Day 3	8 (66.7)	4 (33.3)	0 (0.0)	0 (0.0)
Day 4	11 (91.7)	1 (8.3)	0 (0.0)	0 (0.0)
Day 5	n/a	n/a	n/a	n/a
Practical problem-solving				
Day 1	9 (75.0)	3 (25.0)	0 (0.0)	0 (0.0)
Day 2	10 (83.3)	2 (16.7)	0 (0.0)	0 (0.0)
Day 3	9 (75.0)	3 (25.0)	0 (0.0)	0 (0.0)
Day 4	5 (50.0)	4 (40.0)	1 (10.0)	0 (0.0)
Day 5	n/a	n/a	n/a	n/a

n/a, not applicable.

This phase had also allowed them to refine their delivery, for example, in terms of when and how often they would switch roles across the training day. Trainers also discussed learning from each other:

> *[Trainer from another region] outlined how she ran the gestation period cards sort which I wasn't happy with last time and I used the model of how [the trainer] approached that. And I must admit, it did seem to work better, just the first part of it.* (Trainer interview)

This also explains some of the 'tweaks' discussed above. For example:

> *We had a bit of feedback from other trainers didn't we that was shared in the course of the trial about the big question for example and we all realise that that's one that teachers find harder, as I think we all do.* (Trainer interview)

Trainers also felt that the delivery team had been responsive to their feedback during this year and made some changes to the training accordingly.

> *We'd given them lots of our feedback already and we gave them feedback after the first year and there's a real sense of they absolutely take onboard what you say.* (Trainer interview)

In the researcher notes and in the trainer interviews, high levels of teacher engagement with the programme were noted:

> *Teachers clearly enjoyed the training…[the trainers] were friendly and approachable, and there was a sense that the teachers were at ease with them. Most were eager to engage and share their thoughts or answer questions*. (Training Day observation notes)

> *And they certainly engaged in the things that we'd given them to do from before. They get it, they get what we're trying to do, and they absorb themselves in the activities in a way that makes you feel like they're going to do them in school.* (Trainer interview)

Teachers' perceptions of training and support

In the sub-study school visits, teachers were also asked about the training and support they received through the programme. Teachers' response to the training was overwhelmingly positive during interviews, and this response was also observed in teacher feedback surveys issued at the end of every training day by the developers. Teachers described the training days as both '*exciting*' and '*enjoyable*'. Overall, teachers agreed that the structure, length, flow, and pace of the training sessions were good; one teacher described the training as '*fast paced but not too fast'* (Teacher feedback survey respondent, Day 2). The majority of interviewees felt that the training was particularly successful in providing great new ideas for teaching, it offered perspectives and approaches they never thought about before, refreshed their teaching, and compared favourably with other training experiences.

> *So quite a lot of it was quite new to me and a lot of the practical activities were very, very new to me and I think that's probably why they were quite exciting as well...It's just really enjoyable to go and do as well and you know everything they're doing's transferable to doing in school as well.* (S3–T1V1).

> *Most of the time when you get to go on a course, it's just stuff that you kind of already know, especially if you're quite an experienced teacher, it's just reinventing the wheel, but this is new stuff that I've never really thought of before and then straight away we're, like, 'Can we do it in our school?'…I'm more than happy, it's over my expectations*. (S1–T1V1)

Teachers stated that they were also looking forward to attending further training days:

> *I actually enjoyed going to them. I look forward to going to them.* (S7–T2V1)

Teachers appreciated the opportunities to share knowledge and ideas (learn from other group's successes and mistakes and share ideas of what worked and didn't with others during the training):

> *What's useful is the feedback at the end of each experiment, where each table can feed back, and again, you're learning as an adult what works and what doesn't work. So we do that as a class now, where we didn't do that before*. (S10–T2V1)

However, some teachers felt that their own lack of experience sometimes prevented them from participating in those discussions as much as they would have liked.

All of the teachers interviewed also felt further supported by the resources provided via the training, both physical (for experiments/often referred to as ‘goodies’ or ‘freebies’) and the paper-based and online resources (TDTScience folders and links to relevant websites).

> *No, I think it was just right, because this is the most I’ve ever got from a course, if that makes sense. It all being in one place and being able to refer back to the resources, it was just the right amount.* (S2–T1)

Teachers indicated that the training had a good balance between theory and practice, with some explicitly stating that the TDTScience training made its ethos and core ideas clear, making the theory behind it accessible and relevant:

> *I think that’s one of the biggest things, sometimes you go to training sessions and you see they go, ‘You’re going to do this; this is the best way forward’. But you don’t necessarily get their reasoning why is that beneficial. Whereas we have the why is this beneficial through it, like the lines of questioning; the practicals, the big question at the start of lessons that you can use. They’ve all got validity behind them due to the reasons behind them.* (S5–T1)

Some teachers spoke about training being useful in reflecting on their own practice and helping them to gain a better understanding of the aims of the curriculum, what pupils really need to know, what learning attention is, and what knowledge is being gained through an activity. Furthermore, many felt that training was useful in giving them engagement tools, ‘little tips and tricks’ for having more discussions in the classrooms, improving questioning techniques including HOT, ‘chunking the lessons down’ and giving pupils more independence in the lessons.

When asked about, which aspects of the training they enjoyed the most, many talked about practicals and experiments, not only because they were engaging and interesting, but because they were useful in helping teachers think more carefully about the activities, how they link to the topics, and providing a valuable experience of the steps involved, knowing what can go wrong, and how it can be made right. A common comment was that one of the most interesting and useful aspects of doing the practicals was being put in the position of a pupil, getting their perspective, realising the processes that pupils are going through, and remembering how pupils learn and feel in the classroom. The delivery team felt that having the teachers be able to experience the practical science activities during the training, as pupils will do in the lessons, was beneficial for deepening the teachers’ understanding of what the pupils may experience. Generally, teachers felt that the activities were easily translated into the classroom, but they did not all feel the same about the extent to which these activities need to be adapted and changed when delivered to the pupils:

> *It’s really fun, really engaging. But it took a room full of adults and science teachers the best part of 45 minutes. And I think one group managed to get theirs to fly. Whereas if you’re doing that in a classroom, you double that time, and then the majority of them could be unsuccessful. So I think some of the practicals are maybe not a 100%. The idea behind them is wonderful, but I think some of them are not maybe best suited for this age of children.* (S10–T2V1)

> *Yes, I think there were elements where I used the same activities that we did. For example, we did the egg dropping activity...I remember thinking about the needs of my class. I know that I’ve got some with SEN [special educational needs], and thinking how would they…?...I tried to find the balance of letting them explore it, but at the same time trying to guide them as well really. I did adapt it slightly, but I tried to keep the lesson as it was*. (S9–T2)

Overall, teachers agreed that the trainers did a very good job; they were ‘knowledgeable’, ‘supportive’, and modelled well how to do thinking, talking, and how to teach it to pupils. Furthermore, many felt that the general atmosphere created by trainers was welcoming and conducive to learning and sharing, helping build confidence.

> *Once we got going in a session, lots of people start talking …they ask questions, and because [the trainers] had the approach of when they’re teaching children, there’s no really wrong or stupid answer, you’re allowed to answer whatever, like, we asked questions and sometimes someone might ask a question and I’ll be, like, ‘Oh yeah, no, I’ve never thought of that’, and I think it’s quite good in terms of just having that discussion, and they’re always open to, ‘If you need anything, if you need any support, just email us…’*. (S7–T2V1)

As indicated above, some teachers did not feel all of the activities could be transferred directly into the classroom. However, this was not necessarily a criticism of the programme as a whole. Likewise, there were few negative comments relating to the training and these were mostly framed as recommendations for improvement or linked to teachers' own situations—namely, lack of confidence in certain areas, which made some elements more difficult for them. It is worth noting, however, that there were comments about the Space unit (Day 3) where a few teachers felt that it was not as successful or exciting as the other units, but those teachers pointed out that they understand that Space is a more difficult unit when it comes to practical activities.

Specific recommendations for improving the training and support included understanding more fully what pupils' misconceptions may be and more examples to enable tailoring to a particular class:

> *Apart from just showing us, now you can use this experiment to teach the concept, teaching us where is the part that we might need to explain more or we support a little bit more, where the misconception might be that maybe we need to invest more time on it*. (S12–T2V1)

> *I'd like even more ideas because there's some that you can see and you're, like, for us in this area, it's not a feasible idea to do in a classroom, or, like, it could be anything, previous attainment, it could be the background of the children, it could be funding, it's, like, we can't do that type of lesson, so maybe a few more examples of how to teach a certain one.* (S1–T1V1)

A couple of teachers also suggested that it would be beneficial if the programme could be extended—either by doing it across different year groups, by doing presentations at the school (by trainers/developers), or by making it possible/easier for more people to attend the training.

Teachers' perceptions of cascading

Interviewees saw a number of opportunities for improving and refining practice within their school through cascading their learning from the programme. They discussed sharing with their colleagues including cascading to other year groups and subjects (cross-curricular cascading), primarily through applying TDTScience strategies and techniques in those different contexts—'*the pedagogy rather than the actual matter* [concentrating on] *the talking and the discussion, key ideas and key concepts and get the children doing things or even a little like starter activities*' (S7–T1V1):

> *Like I say, I think it's had a really good impact on us here, I'm really excited to kind of roll it out whole school and I think the idea behind it is really good and like I've mentioned, it's definitely brought back a love of science here, yeah, really good*. (S9–T1)

They saw leadership support, the ability to go to training with others, and the whole-school approach as facilitators for successful cascading, while time, subject knowledge, curriculum restrictions, and teacher 'authority' were seen as potential barriers.

> *Because the thing is I think sometimes you need…obviously we are teachers and we are important, but I think sometimes you need someone a bit higher up [i.e. in the senior leadership team] to see it, to implement it.* (S7–T2V1)

During the interview, the delivery team acknowledged that some schools and/or Trusts stipulate science schemes of work that need to be taught. The delivery team felt that it was an additional challenge for teachers to navigate these pressures along with implementing the principles of TDTScience, although this shouldn't necessarily be a barrier to implementation given that TDTScience can be applied to different schemes of work.

Classroom practice

This section focuses on fidelity to the programme as identified through classroom observations and teacher observations. We recognise the limitations of classroom observations, given that they were one-off occurrences (although some schools were visited twice) and there may have been observer effects. The findings relating to the observations are therefore, supplemented by information elicited via the teacher interviews relating to changes in practice as a result of the TDTScience programme. Taken together they focus on answering research question 3c 'What impacts has TDTScience had on their classroom practice?' Additional findings from the interviews are also presented regarding research question 2 'What processes are involved for teachers and schools implementing TDTScience—what are the main facilitators and barriers?'. However, first we present findings relating to research question 3b ''What are their views of TDTScience strategies and techniques?' in order to provide a broader context to implementation fidelity.

Teachers' views of the TDTScience techniques and strategies

In the interviews, teachers were overwhelmingly positive about the overall TDTScience approach to teaching science. They were predominantly positive as well about the strategies and techniques embedded in the programme. Where concerns were expressed, they tended to be in the minority, but we include them where appropriate below for the sake of transparency.

Strategies designed to get pupils thinking during Bright Ideas Time, such as Odd One Out, the Big Question, and PMI were all singled out for praise, not only for the promotion of HOT but also because they encouraged discussion of science:

> *And I love the Odd One Out one, because I love that everyone wants to be involved because everyone wants to know what the odd one out is. So, when they can, so even you're maybe not the most confident scientist they will have a go and you get the rudimentary responses first and then that's when the higher-level thinking gets exposed. And I found that really successful*. (S5–T1)

However, some teachers sensed that pupils prefer 'the right answer', which meant that TDTScience techniques needed to be chosen carefully and strategically at every stage of a topic. Teachers also felt that over time the technique improved with pupils' practice. In addition, a small number of interviewees pointed out that they did not fully understand how to use the Big Question. There was also an appetite from some teachers for more thinking techniques, beyond the ones currently available, and also for TDTScience for other year groups.

Similarly, while interviewees saw the value of the open nature of some of the experiments demonstrated during training some felt that pupils demanded more guidance: '*Is she not going to tell us what we have to do?*' (S11–T1). The training seemed to be trying to build teachers' confidence in letting them '*take the reins off*' (S6–T2) in larger discussions/debates and investigations, and did not suit all teachers equally well.

Teachers would have also liked a wider bank of practicals to draw from although the final half-day training (which took place after the interviews were completed) was designed to enable teachers to move forward and plan their own lessons based on their learning from the previous training and experience of implementing in the classroom.

Perhaps the issue that teachers most struggled with was focused recording. Schools, with their accountability, tend to encourage extensive recording of learning, which has a bearing on how much physical evidence teachers and pupils produce (not least of which for Ofsted purposes). However, allowing the pupils a certain amount of freedom in how and how much to record was seen as something '*the children buy into*' (S10–T2V1). For many interviewees this was a new approach and one, which could challenge their existing thinking:

> *I've not even considered before that we didn't have to have all this evidence in books, so when it was [Trainers Name] talking and [they were] saying that you don't need to, you know, it doesn't have to be page after page of pieces of work, I was thinking, 'Well how could we do it?*' (S10–T2V1)
>
> *It's shown me that you can write the key bit that links to what you want them to show. That's something that we're developing, but I think that's something that we're on a journey with to make sure that we're capturing the main bit' (S2–T1)*

However, for some teachers the emphasis on focused recording confirmed their own, already existing belief that schools are over-reliant on extensive recording in pupils' exercise books. In such cases TDTScience was perceived to allow teachers to provide their senior management with further justification to reduce physical evidence in pupil books by focused recording of appropriate evidence, for example:

> *I was always in the frame of mind that I didn't like that we had to do all the recording, and it's just, we moved away from that as a school a couple of years ago. This validated, actually, we don't need to do any recording. In our head we've got a really specific learning objective. By the end of the lesson we can actually assess whether the children have got that.* (S3–T2V2)

The delivery team felt that there were two levels to judge implementation through classroom practice. First, how much the teachers implemented the TDTScience strategies (e.g. gap tasks) in the classroom, but also asking teachers about their confidence and understanding in the TDTScience approach. For both of these areas, the delivery team judged that the teachers were overall implementing TDTScience as planned:

*Yes, they were certainly doing what was required in the course and getting positive feedback, but they also indicated that they had gained confidence in the understanding of the approach. (*Delivery team, 2023)

Nonetheless, the delivery team acknowledged that it was hard to accurately judge the classroom implementation. The training session discussions proved to be a useful way to evaluate the teacher's evidence of understanding, particularly how much teachers contributed to these conversations and understood how the different elements of TDTScience fit together.

Implementation in the classroom

In total, 17 lesson observations took place, across 12 schools. The focus of the observed lessons is provided in Table 21. The majority of lessons took place in the afternoon and were approximately 45 minutes to one hour in length.

In summarising the observations, it should be noted that these were single snapshots of single lessons at one point in time.

Table 21: Lesson observations

Lesson focus	Sub-study school visit	Visit month	Related training day
Conductors and insulators	S4–V1	March 2023	Materials, Day 1 September 2022 and October 2022
Hardness of materials	S7–V1	January 2023	
Gravity	S10–V1	February 2023	Forces, Day 2 November 2022 and December 2022
How can I understand the effect of gravity and friction on a moving object	S6	March 2023	
Investigating forces (specific focus not made explicit)[a]	S12–V1	January 2023	
Investigating levers	S1	January 2023	
Investigating pulleys	S5	March 2023	
Levers	S11	March 2023	
To investigate how water resistance can affect objects moving through water	S3–V1	January 2023	
Explain and model the movement of the Earth and moon	S2	March 2023	Earth and Space, Day 3 January 2023 and February 2023
Qualities of the planets in the solar system	S9	March 2023	
To compare and contrast planets in our solar system	S10–V2	April 2023	
Asexual reproduction—how to clone a potato	S4–V2	June 2023	Living Things, Day 4 March 2023
Describe and compare living things[a]	S8	March 2023[b]	
Life cycles—gestation	S12–V2	April 2023	
To describe adaptations for habitats	S3–V2	May 2023	
To research naturalists and behaviourists, and what they do	S7–V2	May 2023	

[a] *Observed by two researchers.*
[b] *Observed prior to the associated training day.*

Overall, the lessons implemented learning from the TDTScience programme well. The teachers were well prepared, as would be expected, especially when expecting an external visitor. Most of the lesson focuses were related to training days already attended by teachers, although the observations did highlight that teachers' did not always teach topics in the order in which they were delivered in training (see 'Barriers to implementation' section for further details) and one lesson focused on a topic prior to the related training day. This was something the trainers were also aware of:

> *Some schools had signed up this time last year, all of the schools had signed up by July, before July, and every single letter I send them I tell them what topics we'll be doing when and some of them go, 'Oh, yeah, well we did that last term', and you're just like… 'Why? Why?' It would have made your life so much easier.* (Trainer interview)

It should, however, also be noted that after training day of Day 1, teachers did have the TDTScience teacher folder with resources for all training days to draw from.

The focus of each lesson tended to be a practical (or, in the example of gestation, a card sorting activity) closely modelled on the activities in the training (only in one observation was there no practical or similar activity observed). In addition, in the main there was little deviation from the examples provided in the training and this can be seen to be a result of teachers aligning their content with the training while embedding the programme principles. However, where teachers were over-reliant on the training for practical work the implementation was not necessarily always felt by the researcher to be as successful as for those teachers who used the programme ideas and adapted or embedded them in their own practice more thoroughly. This was particularly apparent where experiments were quite open with pupils being provided with materials and left to determine their own approach to a topic. For example, in some cases pupils were observed to spend a lot of time sorting through materials or figuring out how to start the task. This concern with the 'openness' of some of the practicals and the potential difficulty of applying such approaches in class was discussed in the interviews with teachers (see above). However, in general, the pupils tended to work in groups, were well engaged in the lessons, and appeared to be familiar with the style and content (see 'Pupils engagement' section below).

In line with the more open nature of some of the practicals and the TDTScience training some teachers also left their learning objectives more open, for example, 'investigating levers', or did not specify them at the start of the lesson, although they became clear to the pupils as the lesson progressed (e.g. S12–V1). A large number of questions from teachers were noted within the observations. Although there is no specific requirement for a Big Question, Odd One Out, or PMI in each lesson (and these were not always observed) teachers asked a variety of questions within the lessons. Where these were most effective was when they appeared to be carefully thought through and planned to be inserted at appropriate points in the lesson rather than just as devices to get pupils talking. Overall, lessons rated highly tended to have the following characteristics, as noted by the researchers:

> *[Teacher allows] children to think and talk, before, during and after the practical activity.* (Researcher notes, S1–V1)

> *Lots of discussion, questions, thoughts, ideas.* (Researcher notes, S4–V1).

The lessons were generally well paced but with quite large amounts of content (with at least one observed lesson having to be carried over to the next day). While this did keep the pupils engaged this also could impact on teachers' ability to implement the strategies effectively:

> *The pace of the lesson made it hard to implement the strategies... [teacher management of the lesson meant that] some activities towards the end aimed at thinking and discussion but did not give enough opportunity [for the pupils] to do that as it was all very brief.* (Researcher notes, S6)

Finally, the researcher observed different strategies for focused recording within the observed classes, including mind maps, iPads, graphs, and flip charts, as well as focused written work in exercise books. However, in some classes, the amount of writing was more extensive, sometimes driven by the pupils themselves.

Changes in practice as a result of TDTScience training

In the interviews, the teachers identified a number of ways in which they felt the programme had changed their practice. These related to improving their planning, increasing the amount of time spent in the classroom 'thinking, talking, and doing' science, a change in the way they recorded science lessons and changes in their teaching style, each of which is discussed in turn below.

<u>Improved planning</u>

Teachers reported that TDTScience made planning quicker, easier, and more enjoyable. Rather than looking through different resources and websites such as Twinkle, they felt that the training had provided them with a bank of ideas backed up by good demonstrations and visualisation provided during the training days. As a result, this streamlined their lesson preparations and saved them time.

> *So, I mean if you look at our planning now, TDTS slide type things are used in every single lesson in so far as we tend to use the activate starter type things such as either an Odd One Out, a Big Question, a PMI. They are fully integrated within every science lesson we do. It's almost that forms the basis and the structure.* (S6–T2)

<u>More 'thinking, doing, talking science'</u>

Teachers talked about giving pupils more time to think during science lessons and spending more time to question and discuss their answers.

> *[It used to be] 'We've got an hour to get this done', and now it's, like, 'Let's try not to give it a time limit, let's try to give the children a chance to think about things and explore'.* (S1–T2V1)

The majority of teachers interviewed felt that their lessons were more practical as a result of TDTScience, both because they feel more confident about them and see pupils' enjoyment (see section 'Impacts of TDTScience' below). Rather than having displays such as PowerPoints or demonstrating 'to' pupils, they saw pupils as being part of the process and having more autonomy within science.

> *So every lesson is a practical lesson now whereas it's not just me doing a practical part it's the children are all involved in the practical part and they understand what it is that they're doing, they can talk about different variables like accuracy, using the vocabulary so it's increased my confidence overall in how I can teach things and how I can demonstrate them, but also it's helped the children be able to physically see how to do something and understand a process further along.* (S3–T1V1)

Teachers commented that there was much more discussion than prior to the programme and that TDTScience had given them tools to facilitate discussion. This was particularly apparent in questioning, both in terms of quantity and quality. Some teachers talked about actually planning questions more carefully——how to use them for differentiation or developing HOT skills, instead of following a schedule and preset order.

> *I don't want to say dismiss, but we didn't have a conversation about what the children were saying, but now we spend time and say, 'Oh, that's interesting that you've said that, why did you…' and trying to get it out of them why they think that, we're definitely doing a lot more of that now.* (S11–T1)

<u>Focused recording</u>

Another key change in practice frequently mentioned in the interviews was focused recording. This is a key message of the TDTScience training. Many teachers talked about realising/accepting that not everything done in lessons has to be in a pupil's exercise book, and they felt that TDTScience took pressure off to fill the books, record for the sake of recording, and overly focus on 'evidence'. Consequently, this resulted in more time in class for investigation, exploration, and discussion. Some talked about using more mind maps, writing just a few words, or even using sticky notes.

> *You want evidence in books, blah, blah, blah, but this course isn't about that, it's about just recording what you need to record* (S1–T1V1)

> *More focused recording rather than writing up an experiment for writing up an experiment's sake, which was like, the norm, as it* were. (S7–T1V2)

There were many examples of focused recording in the samples of pupils' work in the intervention group, supporting teachers' discussions of implementing more focused recording. One school (S12) in particular, evidenced a variety of ways for students to record their work. These, and examples from other settings, are illustrated below (Images 1 to 10).

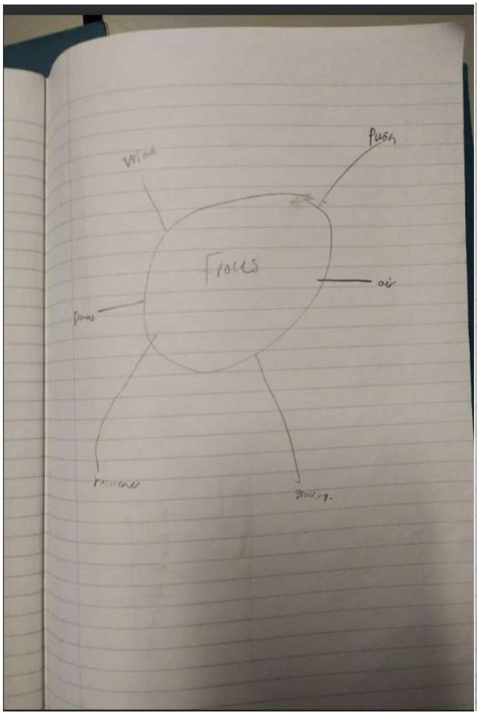


Image 1: Focused recording through a mind map about Forces (S12–I–B–V1.a)

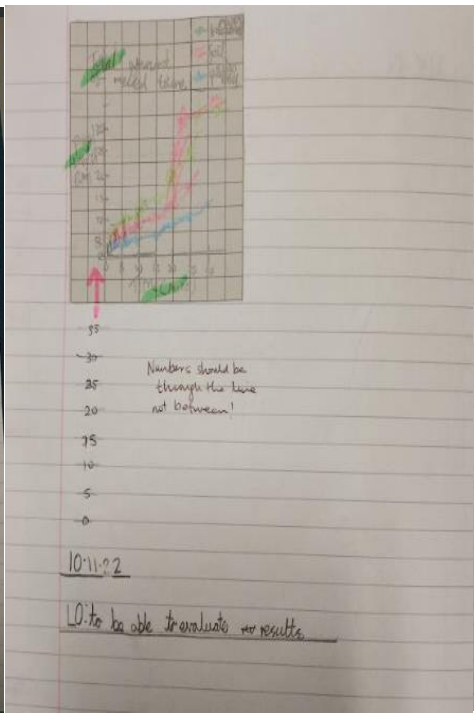


Image 2: Focused recording through graph about Insulation (melt puddle size vs time) (S12–I–B–V1.b)

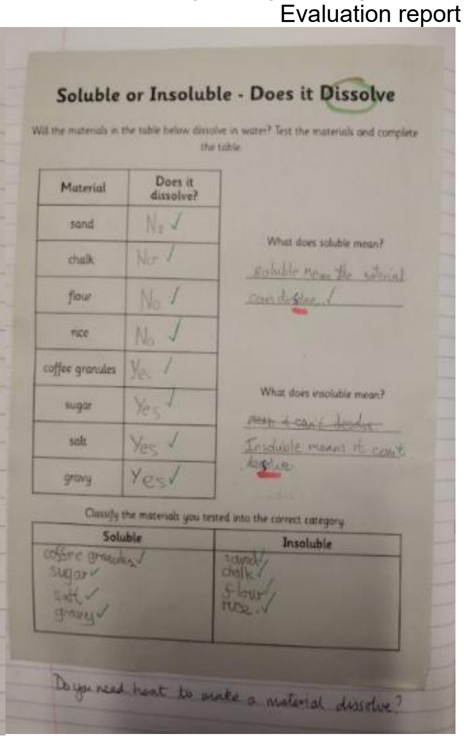


Image 3: Focused recording through table about Soluble / Insoluble (S12–I–B–V1.c)

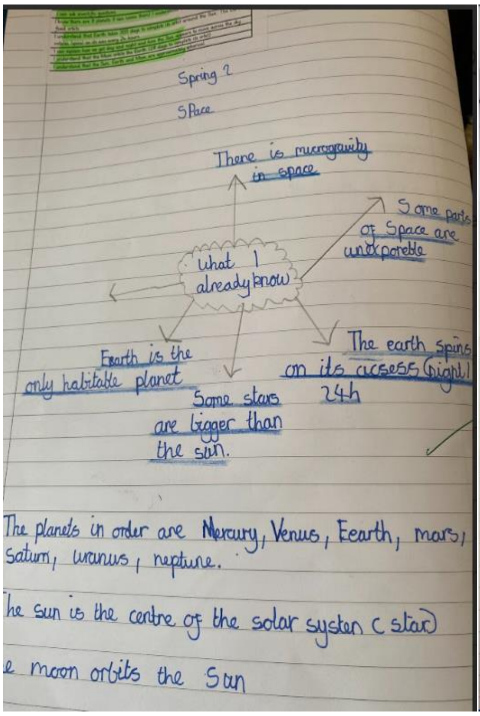


Image 4: Focused recording through mind map about prior knowledge of Gravity and Space (S10–I–B–V2)

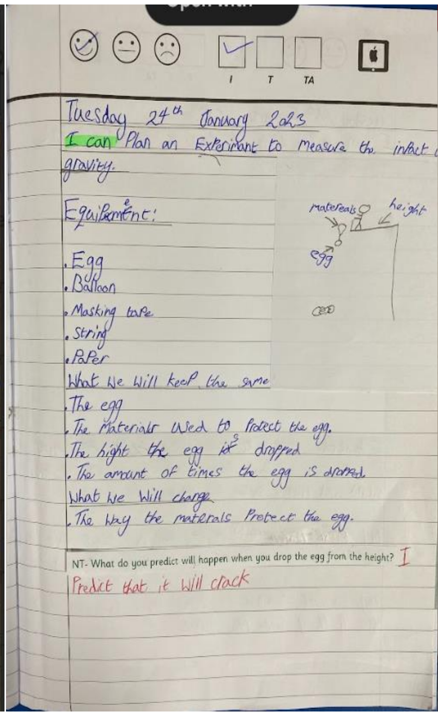


Image 5: Focused recording of equipment list, and prediction, during experiment planning (S9–I–B)

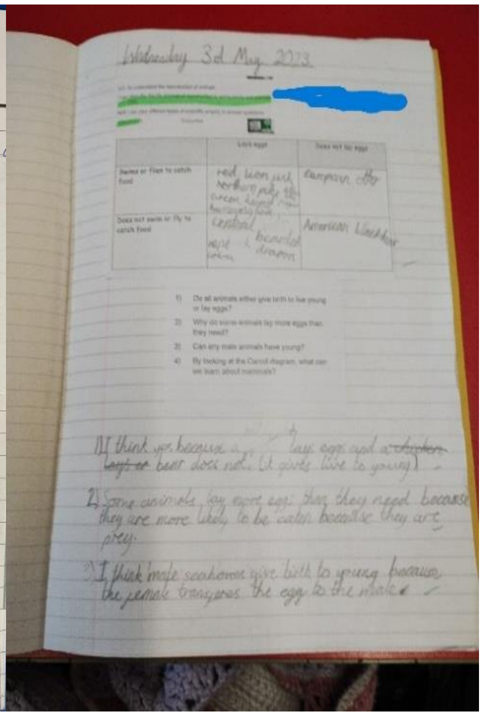
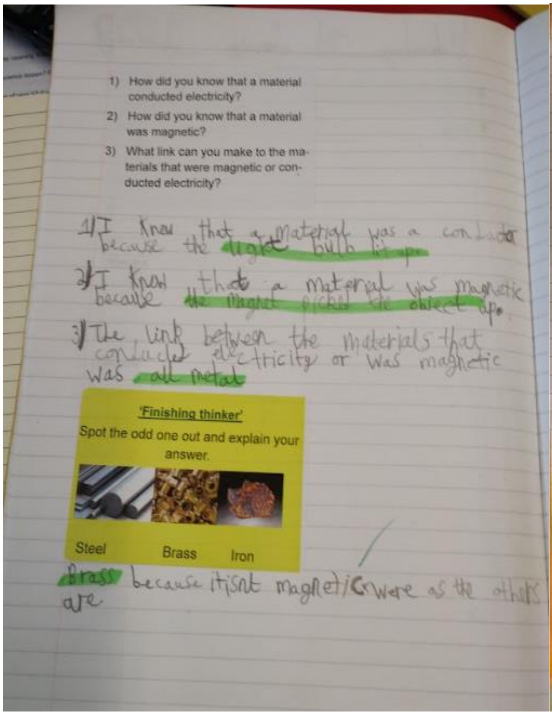


Image 6: Focused recording during cover lesson; table about Reproduction; planning through written questions (S7–I–B–V2)

In addition, there was also extensive evidence of TDTScience strategies, such as Odd One Out, being used in the Year 5 pupils' science books.



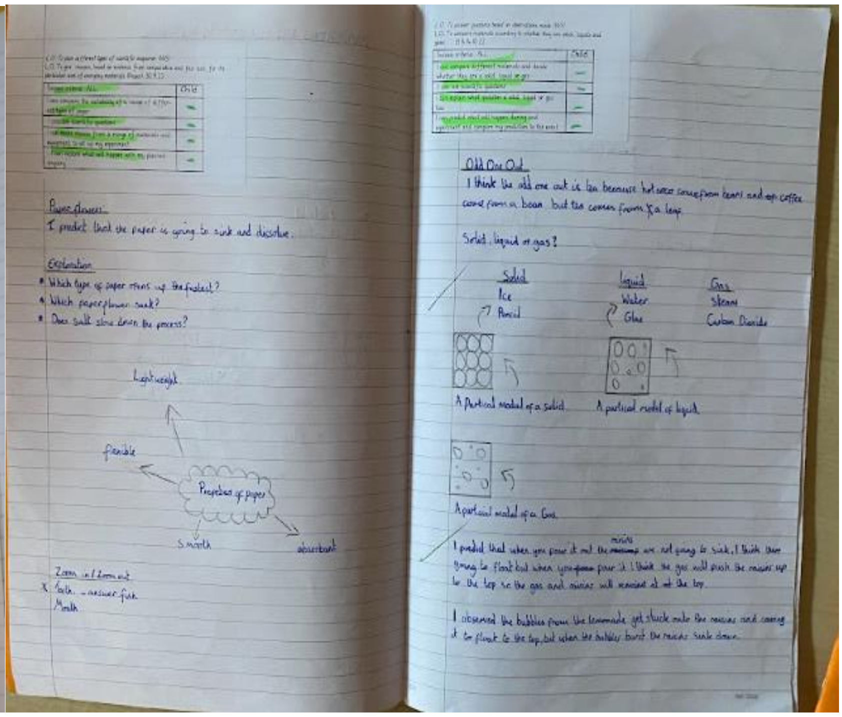
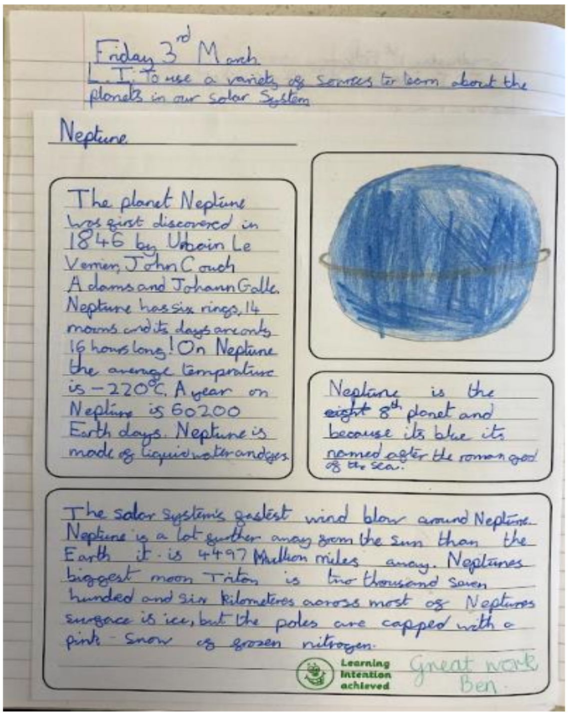


Image 7: Example of use of Odd One Out at the end of a lesson on Properties of materials (S7–I–B–V1)

Image 8: Example of use of Odd One Out at the start of a lesson on observations of solids, liquids, and gases (S10–I–B–V1)

However, in some exercise books there were also examples of extensive writing, as illustrated below (Images 9 and 10), although the format suggests that, in some cases at least, this was the pupils' preference and other examples of more focused recording were found within the same group of pupils (S2 below).



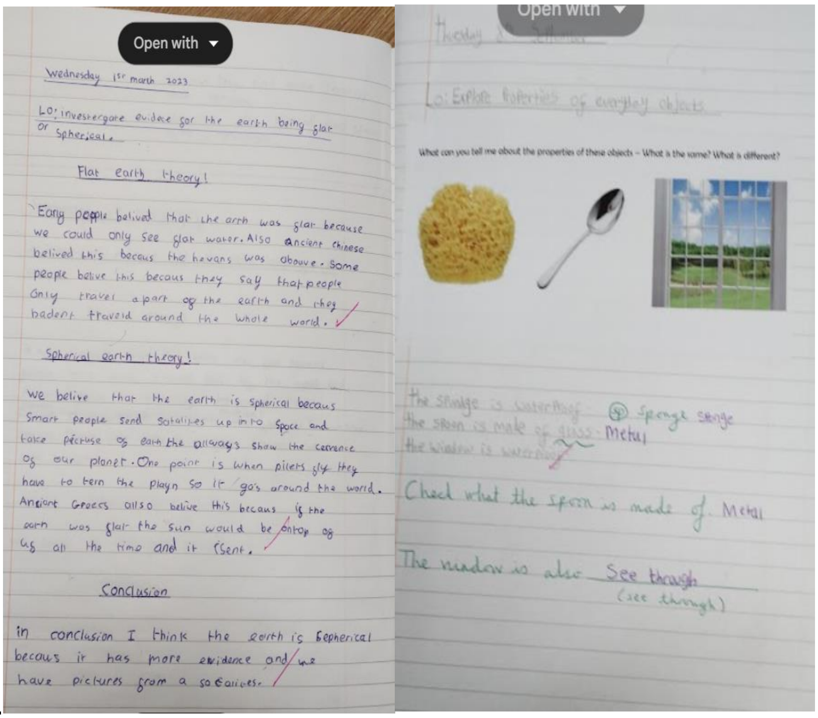


Image 9: Example of extensive writing in a lesson about Earth and Space (S8–I–B)

Image 10: Example of extensive writing in a lesson about Earth and Space (S2–I–B.a)

Teaching style

Many interviewees felt that their teaching style and classroom practice changed for the better as a result of their involvement with TDTScience. Some talked about being better at demonstrating things to pupils; rather than using visuals (PowerPoints, cartoons), they felt more able to use more 'real-life' examples and were better at moving away from the abstract to being able to explain things in a more accessible way. Teachers also mentioned feeling that they were moving away from the rigid/formulaic way of teaching science, going through steps 'for the sake of doing it'.

> '*So what do you think we might be doing today?' So rather than, 'No, wait a minute, I'll tell…' it's, like, 'Build on it, let's get that thought process going, let's build on the curiosity', whereas before it would have been, like, 'Stop, I want you to do that now, don't you worry about that, we'll come to that after'.* (S1–T2V1)

They talked about considering quality, not just quantity, having more focus on process and not just the end result, as well as having a more systematic approach, linking between different units and building on existing knowledge:

> *I try and link them into either previous learning or sometimes even just what we're going to be learning the next time, because then it's like a bit of a baseline of, like, 'Right, they misunderstood that, next week I can do a misconception and then do a core teaching*'. (S1–T1V1)

Finally, some interviewees stated that TDTScience had changed their practice in other subjects outside of their science teaching as well:

> *So for me I think it's made me a better teacher overall because I've also done that in maths and in DT [Design and Technology and things like that, been much more practical in that way, so not just within science.* (CS3–T1V1)

Facilitators to implementing TDTScience

In the interviews teachers discussed a number of factors, which enabled them to implement the programme. These operated at the school, class, and the teacher level. They included already being aligned to a similar philosophy to that adopted by the TDTScience approach, an enabling school culture, the school leadership, and the resources provided by the programme.

Alignment with TDTScience philosophy and practices

Some interviewees expressed the view that having experience with similar approaches to that of TDTScience made implementation of TDTScience easier. In particular they identified Philosophy for Children (P4C, https://www.sapere.org.uk/), use of mind maps, Explorify, and PSQM (https://www.herts.ac.uk/for-business/skills/psqm) as being similarly supportive in developing pupil's thinking skills and questioning—some teachers referred to it as 'finetuning' or 'refining of what they are already doing'. For some, they felt that moving to TDTScience was not a major shift but rather a subtle difference, which enabled them to build better on existing practice: 'bumping it up', 'edit what I have and make it better' or 'streamlining it'.

> *They take turns in listening to one another, and I think that sort of approach in P4C works really well with the science, so when they're having their discussions, they can take it turns, and you know, it might be, 'Well, I want to build on your idea', or, 'I disagree with that approach so I'm going to try this way'. So yeah, I think it works really well.* (S10–TV1)

> *It's kind of just a step on from what we were already doing by using Explorify and the ReachOut CPD, and the ASE [Association of Science Education] plans. Those were our main sort of focused resources and documents as in current best practice, and so it's just given us extra strategies and extra resources on the same kind of path we were on. It's just kind of bumped it up hopefully to the next level.* (S2–T2)

School culture

Many teachers felt that their school culture was also an important facilitator for implementation. For some, this meant working in a school willing to embrace new ideas:

> *I think in our school, we're really good at letting people trial things, so we haven't got a staff that are very rigid and won't try things. And our head, you know…they're really good at embracing new ideas*

> *and new ways and trialling things. And what we do, we have, in our staff meeting, we have a 'what went well' section at the beginning where you can share what's gone well.* (S10–T2V1)

Some teachers also discussed the importance of collaborative approach practices, such as attending training with colleagues and discussing and sharing what they learned:

> *When you're at the course, you can discuss it. It's been valuable to go with the science lead, because we can talk about it before I go and try it out. It's a very collaborative process. You don't feel on your own with it.* (S2–T1)

The support from other teachers across the school was also seen as an essential aspect of school culture that made the implementation of TDTScience easier.

> *We have a very supportive staff here, yeah, definitely have a very supportive staff here, and they're willing to give things a go, and they know when I'm running around like a headless chicken, but yeah*. (S10–T2V1)

Finally, some described a supportive **school leadership** team, particularly the science lead, as a crucial element of school culture in enabling the successful implementation of TDTScience. This included leadership being open-minded and flexible, trusting and confiding in teachers, and having good lines of communication.

> *I feel like we're very lucky here…the school has been very open to letting us do it the TDTScience way. I feel like our head [and] lead science were very open to us, 'Okay, well, that's what you're doing in TDTScience, that's really good, this is what you're doing now'.* (S10–T2V2)

Resourcing

The free science resources given out during the training were also mentioned by the majority of interviewees as important facilitators in implementation and was something teachers were overwhelmingly positive about. Furthermore, a few teachers mentioned that TDTScience experiments can often be done with *'what they can find around the school'* (S6–T1) and therefore, didn't require lots of additional spending.

Other

Finally, other facilitators, albeit mentioned less frequently, included that TDTScience allowed for flexibility with assessment, was easy to use for the purposes of differentiation, helped with planning and time management (due to the clear structure of a TDTScience lesson), knowing that the programme is backed by science and research, not having to worry about Ofsted (if they just had an inspection) and having pupils in school that were more academically able and skilled in discussions.

Barriers to implementing TDTScience

The barriers to implementation were generally the converse of the facilitators and primarily related to conflicting pressures acting on schools, which teachers felt were often out of their control.

Resources

The most mentioned barrier was the lack of resources and limited budgets, including having to share resources with other year groups or teachers' spending their own funds on buying supplies. However, this appeared to be a more general view in terms of barriers to teaching science in principle—most teachers added that the free resources provided through the TDTScience training were a considerable help in addressing and easing this concern:

> *For us in this school, for me the biggest problem would be resources as in good high quality resources and for everybody luckily today it was quite easy to get hold of, you know, however there's times when it's not and it's not very easy to be getting bits and bobs sorted and it's relied on either you to probably go out and get it or like share between groups of five or six type of thing. So I'd say for our school it's probably that, getting the resources for it but luckily [TDTScience] supplied us which is nice.* (S7–T1V1)

Existing schemes of work

The delivery team raised the potential issue that not all teachers taught science topics the same; this may have contributed to more variation in the depth of teachers' understanding of TDTScience. However, the delivery team did mention that they gave teachers some prior warning of the topics that would be covered in the TDTScience training, so that it could be possible for teachers to build it into their planning/teaching. Despite this, teachers reported that a key barrier was the issue of existing planning, which was described as being sometimes very rigid and leaving little 'wriggle room' to fully implement TDTScience. Some specifically referred to limitations set by their MAT:

> *...a bit of a dichotomy with what the [Trust] are providing us with our schemes of work and short-term plans and what that TDTScience is. ...and that's a decision that's way above my head as a class teacher,...I'm just a little cog in the wheel if you pardon the pun*. (S12–T1V1)

This barrier was felt to be particularly restrictive where existing medium- and long-term planning was also in place, which did not take TDTScience into account and in some cases meant that topics had already been delivered or partially delivered in class prior to the topic being addressed in TDTScience training (although it should be noted that teachers were informed on the topic/training day schedule prior to the first training day occurring).

<u>The national curriculum</u>

Alongside this concern with existing planning, others felt that there were some constraints to implementing TDTScience due to the requirements of the national curriculum, in particular how TDTScience fit with national curriculum objectives, even though the programme was designed to align with the national science curriculum. Teachers talked about the demands of the national curriculum in relation to both the content and the volume of the current requirements:

> *The curriculum at the moment is highly constraining, I feel. We've had this conversation. We cannot find a way through it, because we've got the national curriculum objectives that we have to cover, but I don't know when and where.* (S3–T2V2)

> *And obviously, then we've got the rest of the curriculum to squeeze in around. So sometimes that, and especially because we have to squeeze that around, sometimes I would really like to spend a full afternoon on a science lesson so we can have that full length of discussion. And I think sometimes we're cutting those sessions short*. (S5–T1)

<u>School culture</u>

While school culture could be a facilitator to implementation, for some teachers, it was perceived to be a barrier; they expressed concern that they were doing something different from others in their school or that the expectations that everything has to be written down would make TDTScience less acceptable to the rest of the school. One teacher in particular felt that the lack of unity in approach across the school was making implementation harder.

> *I will say it's often difficult because it feels like it clashes a little bit with the school's curriculum…sometimes it feels like myself and [the other Year 5 teacher] are doing one thing and the school is doing a slightly different thing, which is very similar to TDTScience, but it is a little bit different. So, I feel like often if the school implemented it fully, potentially we'd get more out of it.* (S10–T1V1)

<u>Curriculum time</u>

One further barrier mentioned by some teachers was an issue with scheduling and timing; they felt that they did not have enough time to implement the TDTScience fully as they were often 'losing' science time to extracurricular activities, as it was, unlike core subjects, scheduled for afternoons where these activities would take place.

Summary of fidelity key findings

A summary of the key 'fidelity' findings are presented below linked to the relevant research questions.

1. To what extent was TDTScience implemented as planned?

a. **Training**
   The train-the-trainer model worked well, with trainers reporting high levels of fidelity. They also expressed confidence in delivering the training, which was in line with the logic model. Teachers were highly engaged at the training days.

b. Classroom practice
Implementation was high in the classroom.

2. What processes are involved for teachers and schools implementing TDTScience—what are the main facilitators and barriers?

The main facilitators to implementation included the alignment of TDTScience with approaches that were already being implemented, school culture being in a supportive and collaborative school that is open to embracing new ideas, and the free resources provided at training.

Where there were barriers to implementation, these were identified as school culture, a lack of science resources, the need to follow existing schemes of work, the needs of the national curriculum in terms of both content[8] and volume, school culture, and curriculum time.

School culture and resources were identified as been both a facilitator and barrier to implementation and can be modifiable factors. As such, they have been added as moderators to a revised logic model, see Appendix E.

3. What are the perceptions of teachers as regards TDTScience?

a. What are their opinions about training and support, including cascading from colleagues where relevant?
Teachers liked the format of the training days, which provided accessible new ideas and approaches for teaching science and the opportunity to share knowledge. Teachers were overwhelmingly positive about the training and support they received and particularly valued the practicals and experiments embedded in the training. Specific recommendations for improving the training and support included understanding more fully what pupils' misconceptions may be and more examples to enable tailoring to a particular class.

Teachers considered that science teaching in their school could be improved and refined by cascading their learning from TDTScience.

b. What are their views of TDTScience strategies and techniques?
Teachers were positive about the TDTScience approach to science teaching and the strategies and techniques embedded in the programme. Strategies designed to get pupils thinking during Bright Ideas Time, such as Odd One Out, the Big Question, and PMI were all singled out for praise, not only for the promotion of HOT but also because they encouraged discussion of science. However, some teachers sensed that pupils prefer 'the right answer', which meant that TDTScience techniques needed to be chosen carefully and strategically at every stage of a topic. Teachers had mixed opinions as to how easily activities could be easily translated into the classroom without adaptation.

c. What impacts has TDTScience had on their classroom practice?
Teachers reported a change in their classroom practice. These changes included improvements to planning, the inclusion of focused recording, and changes to their teaching style and delivery, for example, giving the pupils more time to think and discuss during science lesson.

Impacts of TDTScience

The interviews with teachers in the qualitative sub-study schools explored not only changes in practice but also perceived impacts on teachers and pupils. This section focuses first on answering:

3. What are the perceptions of teachers as regards TDTScience?

d. How has it affected their engagement with and confidence in teaching science?

e. How do they think it has impacted on pupils?

4. How do pupils respond to TDTScience?

[8] As previously noted, however, the programme was developed to align with the content of the national science curriculum.

a. What is their experience of, and reaction to, the different TDTScience strategies?

b. What is their experience of practical work in the science classroom?

c. What is their engagement with science lessons?

In order to address these IPE research questions this section draws on data collected from the sources:

- qualitative sub-study school visits, primarily interviews with teachers, and pupil focus groups; and
- teacher surveys.

Teacher engagement and confidence in teaching science

One of the major changes teachers reported as a result of TDTScience was increased confidence in teaching science. During interviews, some said that they felt more confident with doing practical activities because they did them in training, they understood how they work (or don't) and felt they were better equipped to deal with possible misconceptions or what might not go to plan. The baseline and follow-up survey asked teachers 'How confident do you feel doing the following (specific statements in Figure 7) when teaching science to your Year 5 class?' Teachers in the intervention group tended to report being more confident in teaching science than those in the control group at follow-up, see Figure 7.

Figure 7: Teachers confidence in teaching science at baseline and follow-up, by allocated group

Some talked about confidence in terms of feeling that it is ok not to do things they were worried about before, such as 'not knowing everything' or having 'silent moments in the classroom'.

> *I think before I would have panicked if a child said, 'Well what about this?' whereas now I'd say, 'Well, shall we find out together? What way could we find that information together?'* (S10–T1V1)

The concept of validation was also discussed, in that teachers now felt that it was ok to do or not to do certain things:

> *I think as well going on things like that does validate what you're doing because you think, oh, actually they're doing it as well. We're doing the same, we're doing it right, and a lot of it was validation.* (S2–T2)

> *Yeah, I just think it's being brave enough to give yourself that much time and think, 'Right', even, like, this week, where you say, 'I think the next topic should be two weeks, I'm not starting it until tomorrow', now I'm confident enough to do that thinking, 'No, I'm going to finish, I'm going to give more time and we're going to start it tomorrow'.* (S1–T3)

Some teachers also reported being more engaged, they felt more creative as teachers, enjoyed teaching science more and had a reinvigorated and refreshed approach to science:

> *Hmm, I guess it's given…for somebody who's been teaching 32 years and been a science specialist all that time, it's actually reinvigorated the way I think about science teaching*. (S12–T1V1)

> *I think this approach has revitalised [my teaching] and it's made me want to teach it [i.e. science] and where the children are excited.* (S10–T1V2)

Teachers' perceptions of impact on students

In the interviews, teachers were asked about any observed impacts of the programme on the Year 5 pupils in their classes. The predominant theme was that TDTScience had improved student engagement with science and science lessons. This was mainly considered to be due to the 'hands on' and discussion-based aspects of the programme as well as focused recording:

> *They're now participating in discussions, they're keener*. (S3–T2V2)

> *I think within science they do seem to have more confidence because they're working in that group and they're doing a practical.* (S3–T1V1)

> *It can turn a lot of children off if you say, 'Now I'm expecting you to write all this, I want you to set it out in this table', whereas if they can choose, they can even do it verbally or they could present it as a poster, they can choose that way of sharing their knowledge and understanding, so I think the children buy into that*. (S10–T1V1)

In turn, these aspects of the programme were also perceived to have improved pupil's confidence and knowledge:

> *They're not watching a video, they're doing it themselves. And it's that sort of sticky knowledge, where they know that themselves, it sticks and they remember it.* (S10–T2V1)

> *Now, because we can really concentrate that focus with a well-resourced and practical lesson, I feel like it sits in their brain for a longer time.* (S1–T1V1)

One teacher also reported that the discussions engendered by the programme had helped the students uncover existing knowledge in the context of their science lessons:

> *It's not something I think we were aware of how good their subject knowledge was but because of all these discussions that we have we're able to see that now.* (S3–T1V1)

Many teachers discussed the way in which the programme had improved pupil's thinking skills:

> *They're also relating their thinking back to prior knowledge as well…they're thinking back and making links.* (S9–T2)

> *They are getting better at thinking first before they jump in…and the other thing they have developed is not always having to know the right answer and to try things out*. (S12–T1V1)

Finally, the programme was perceived to encourage a more exploratory and 'scientific' approach in a range of aspects including 'thinking scientifically' and using 'scientific vocabulary':

> *We're doing it, and then we're discussing it, do you know what I mean? We're being scientists, we're not sat writing about science*. (S1–T2)

Teachers were additionally asked if they felt the programme had any particular impact on subgroups of pupils. The main theme emerging from the analysis was that of the inclusivity of the programme to all groups of pupils due to the emphasis on practical work and discussion as well as focused recording. Teachers mentioned this in relation to pupils with English as an Additional Language (EAL), those with lower prior attainment, those from disadvantaged backgrounds, and those with SEN:

> *I think the focus on vocabulary and talk is really good for our catchment with regards to EAL…all of the practical, visual resources and things are really good to support them*. (S2–T2)

> *Because it's based on a practical activity, they want to get involved, they don't have to have any knowledge from home*. (S1–T1)

> *It doesn't exclude a group of children*. (S10–T2V1)

This was also linked to the higher overall levels of engagement and confidence in science (see above). One teacher specifically mentioned that the emphasis on practical work engaged the boys in their class who were reluctant writers.

TDTScience was also seen to have positive impacts on those pupils with high prior attainment, primarily because the programme allowed for flexibility and enabled them to engage in more HOT:

> *It's helped the greater-depth children, the ones who are working at a high level, push their ceiling higher so they can attain more*. (S1–T1)

In contrast to the above, only one teacher specifically suggested that SEN children were not always catered for within the programme:

> *Sometimes there are some SEN where the programme, it's not that it doesn't work for them, it's just that they can't access it as well as they could if there was something else in place for them*. (S2–T1)

Pupil responses to TDTScience

The pupils focus groups were designed to answer research question 4. 'How do pupils respond to TDTScience?' with particular reference to TDTScience strategies, practicals, and engagement with science lessons. Each of these is dealt with in more detail below.

<u>TDTScience strategies</u>

Generally, pupils felt that science teaching and learning in Year 5 were different to the previous year. In some instances, pupils were able to describe how the teacher's practice had changed and how that made learning more accessible, facilitating a better understanding of what was being taught and fostering better independence.

> *I think Miss [teacher] this year breaks it down a bit more into smaller pieces so we can do it one at a time and shows us a bit more videos so we can understand better.* (S9–FG)

> *I think it is better, because they are making you be more independent, which is good for you. …If you are independent, then you have more chance. You are by yourself, and you are going to get more chance of finding out something new instead of just someone else doing all the work for you.* (S11–FG)

In one school, pupils were able to explicitly link the change in teaching style to their teacher's involvement with TDTScience.

> *So I think her course* [is] *called like Talking, Thinking, Doing, Science. And I think the courses she's been doing is affecting how she's teaching 'cause we're thinking more, we're doing more and we're, like, talking more to each other about these lessons and she's making us think and talk more and do actions. And I think it's because she's been doing these courses that we are doing all these great things about talking and doing about these lessons so we can explore, like, what our thoughts are'.* (S12–FGV2)

Pupils also talked about noticing that there is more talking and discussing in science compared to previous years.

> *We definitely discuss a lot more than in Year 4. We do more group activities definitely even in other lessons, we discuss more. So, we get to talk to each other about our theories and opinions more, even in other lessons and science.* (S12–FGV1)

Some pupils made connections between the increase in discussion and practical work and changes in focused recording. In most cases, pupils felt the new approach resulted in less writing.

> *But in Year five, the teachers have said we should…instead of writing, it doesn't help very much, you're just writing a paragraph. Instead, we should do more talking and experiments.* (S8–FG)

> *I feel like it's just like writing, just with your words and instead of writing down everything that you've done in the lesson. So you would tell the teacher what you've done in the lesson, then she'll, like, put it down on the board in a way that, like, everyone has had their little bit in it. And then we would write down that sentence.* (CS–9–FG)

> *I was shocked when Miss [teacher] told me not to write, one day. She said, 'Make a poster but don't write'.* (S2–FG)

For one pupil, the changes in the teacher's approach to focused recording were seen as a facilitator for better learning and understanding.

> *We all have different mind-sets and we all think differently so if we have different options then I think it would just be a lot better for us to understand it…It gives us options and it makes us…what we want to do…say if we do it one way and we don't understand it then say if you didn't have any other options you just won't understand it, but if you have different options of what you're doing it makes you understand it more.* (S10–FGV1)

Other pupils also picked up on the change in practice regarding recording:

> *So in Year four in science we did lots of writing and not much creativity and stuff in science because all we did was writing down paragraphs. But in Year five the teachers have said we should…instead of writing, it doesn't help very much, you're just writing a paragraph. Instead, we should do more talking and experiments.* (S8–FG)

In terms of their response to different TDTScience strategies, there was a general sense that pupils enjoyed the thinking activities, like Odd One Out, primarily delivered as lesson starters. Many talked about those strategies as 'brain starters'.

> *It's like it's normally for me it's an exercise for our brains and then we begin the real lesson. So it's like turning on the brain.* (S12–FGV2)

> *I think they help you because say if you're jumping straight into a lesson you might not be ready for it. Your brain might not be ready to do it. Where if you start it off nice and calmly just slowly, gradually, doing stuff it jogs your brain so you're ready to learn about the science and you might have a better understanding.* (S10–FGV1)

Some found it more difficult to make that connection and it is not always clear that pupils understood the purpose of those starters beyond it being fun or 'getting the brain to work':

> *They're fun, but then when he says that there wasn't an odd one out, I'm just like, 'What is the point in this?'* (S7–FGV1)

However, some pupils were able to make a clear connection between the TDTScience starter strategy and the learning objectives.

> *At first we did it at the start and it was all like 'Oh the duck has webbed feet' or like the 'Tiger's a mammal' but then when we finished our work we put the same slide back up and then thought about it again but linked it to their learning cycles instead of actually like just the first thing that comes to mind.* (S7–FGV2)

> *The teacher gives us questions and gets us thinking, thinking and wondering about how this is doing this and how it is doing that. Like for example, one time we did, which is the same, which is the odd one out, but just today we did like, she did this thing with books and it got us thinking why is airplanes shaped like this. So, it gets the brain thinking about these things.* (S12–FGV1)

> *So, the one before it was like a streamlined shape and it's not about what's in there but it's about what type of shape, because they all link to your topic. …that's not what you're meant to be thinking about like, 'Penguin,*

> *person, and boat', but it's more like a streamlined shape. So, that's the scientific and that's what we're doing about Forces.* (S10–FGV1)

The pupils had picked up on the increased amount of thinking they are explicitly required to do, with comments such as *'I think we think more'* (S3–FGV2), referring to *'opening thinkers'* (S7–FGV2), and

> *[P1] It was the fact that the Earth wasn't flat lesson. But that's just like a starter question. Sometimes the starter questions aren't actually linked to what you're doing. It's just to get your brains ready for the science that you're going to do.*
>
> *[P2] Yeah, I feel like it's helping us learn because it's giving us more ideas in our head instead of teachers telling us the ideas that we have to do.* (Pupil interaction, S9–FG)

<u>TDTScience practicals</u>

According to some pupils, the change in teacher's practice was particularly evident in their approach to practical work, which was both more frequent and independent.

> *This year, mostly because we've done a lot more experiments and most of the experiments that we've done haven't been like, 'Do this'. They've been like, 'Here's some materials. Try and make something out of this'.* (S9–FG)
>
> *There's a lot more independence, for instance if we were doing the Newton meters last year they would have done it for us, they would have held up the Newton meter and go, has anyone got a pencil case?* (S8–FG)

For some pupils, there was a clear link between the practical work and their ability to better learn and understand the topics:

> *[Teacher] was telling us questions about it and we were just writing it down. We didn't do as many experiments as we do now. I think with actually doing the experiments it gives us more of an understanding of what we're doing and it helps us explain a bit so you actually know what you're doing.* (S10–FGV1)

<u>Pupil engagement</u>

As suggested above, pupils indicated high levels of engagement in science lessons as a result of the TDTScience strategies and practicals. For example:

> *But now in these we get to engage in more conversations and it's really interesting and it's very fun and I like talking to people that way when I get to engage in a conversation with them.* (S5–FG)

Pupils commented on the joy of building scientific knowledge together and of listening to each other:

> *Like me and [classmate] sit next to each other and we make two good ideas and we like add them together to make it like a real good idea.* (S7–FGV1)
>
> *Sometimes we say our own idea and then we listen to the rest of the table's ideas, and then we listen to Miss [teacher] first of all and see which one is the right answer. And then think why.* (S3-FGV2)

Overall, there was a general sense that most pupils enjoyed their science lessons, including those who had struggled with the subject in the past:

> *I think science is actually my favourite because it's not the only time we get to actually experiment, but it's like it's not just writing down in a book. It's like actually doing something.* (S7–FGV1)
>
> *Not that but in Year four I always used to be, no, I hate science. Now in Year five when the teacher says, we're going to do science tomorrow, I'm just like, when I'm sleeping, I'm just like, I've got science tomorrow!* (S8–FG)

Summary of impacts of TDTScience key findings

A summary of the key 'impacts of TDTScience' findings are presented below linked to the relevant research questions.

3. What are the perceptions of teachers as regards TDTScience?

d. How has it affected their engagement with and confidence in teaching science?

Teachers reported increased confidence in teaching science, in line with the logic model.

e. How do they think it has impacted on pupils?

Teachers reported positive impacts of the programme on their pupils in terms of pupil engagement, confidence, and knowledge. Teachers commented on the inclusivity of the programme in terms of impacting all pupils within the classroom regardless of socio-economic status or prior attainment in science.

4. How do pupils respond to TDTScience?

a. What is their experience of, and reaction to, the different TDTScience strategies?

They reported a noticeable change in their Year 5 science classes, particularly an increase in science-based discussions and some pupils noted a move to focused recording (i.e. 'less writing').

b. What is their experience of practical work in the science classroom?

Pupils reported a noticeable increase of practical work in the science, as well as these being more independent.

c. What is their engagement with science lessons?

Pupils reported high levels of engagement with science as a result of the programme.

Usual practice

This section will explore the IPE results in relation to usual practice for Year 5 science across all participating schools at baseline and explores what usual practice was for all participating schools during the trial. The IPE research questions that will be addressed in this section are:

5. How does TDTScience compare with practice in business as usual science lessons?
    a. What strategies and techniques are used in science lessons?
    b. How interested and engaged are teachers and pupils in science teaching and learning?
    c. What is the frequency and length of science lessons?
    d. What practical science takes place?
    e. How much training have Year 5 teachers received in science?

In order to address these research questions this section draws on data collected from the following sources:

- teacher surveys; and
- samples of pupil's work.

As reported in Table 22, for approximately two-thirds of responding Year 5 teachers the highest qualification in science was GCSE (General Certificate of Secondary Education) or equivalent, at both baseline and follow-up. There were a slightly smaller proportion of teachers with an A level (Advanced level) or equivalent in science that responded to the follow-up survey in comparison to the baseline survey, and slightly more teachers reporting to have a postgraduate degree in science at follow-up in comparison to baseline.

Table 22: Teachers' highest qualification in science at baseline and follow-up, by allocated group

Highest qualification in science at baseline	Control		Intervention	
	Baseline *n* (%)	Follow-up *n* (%)	Baseline *n* (%)	Follow-up *n* (%)
Postgraduate degree	6 (7.2)	8 (10.7)	6 (5.2)	7 (9.2)
Undergraduate degree or equivalent	5 (6.0)	4 (5.3)	9 (7.8)	6 (7.9)
A Level or equivalent	14 (16.9)	7 (9.3)	22 (19.1)	11 (14.5)
GCSE or equivalent	57 (68.7)	53 (70.7)	76 (66.1)	51 (67.1)

Missing	1 (1.2)	3 (4.0)	2 (1.7)	1 (1.3)
Total *n*	83	75	115	76

Year 5 teachers in all schools were asked if they had been involved in any other science-related training or initiatives during academic year 2022–2023. As detailed in Table 22, most had not but for those who had, the most common was STEM Learning/Science Learning Centre courses; involvement in such courses was generally balanced between allocated groups.

Table 22: Other science training or initiatives undertaken by teachers during intervention delivery at baseline and follow-up, by allocated group

Science-related training or initiatives	Control		Intervention	
	Baseline *n* (%)	Follow-up *n* (%)	Baseline *n* (%)	Follow-up *n* (%)
PSQM	5 (6.0)	5 (6.7)	10 (8.7)	9 (11.8)
Primary Science Teaching Trust (PSTT)	3 (3.6)	1 (1.3)	2 (1.7)	5 (6.6)
Teacher Assessment in Primary Science (TAPS)	6 (7.2)	4 (5.3)	8 (7.0)	5 (6.6)
STEM Learning/Science Learning Centre courses	8 (9.6)	10 (13.3)	16 (13.9)	12 (15.8)
ASE courses	1 (1.2)	0 (0.0)	2 (1.7)	3 (3.9)
Secondary school collaboration	1 (1.2)	0 (0.0)	2 (1.7)	2 (2.6)
None	66 (79.5)	53 (70.7)	83 (72.2)	52 (68.4)
Other[a]	0 (0.0)	3 (4.0)	2 (1.7)	3 (3.9)
Total *n*	83	75	115	76

[a] *Other baseline: Intervention (n=2): Science leader network meetings (n=1); Science week: planning with other teachers a week of science activities (n=1). Follow-up: Intervention (n=3): Science leader network update meetings (n=1); PSQM Silver previously (n=1); the part of a local Science Cluster (n=1). Control (n=3): inset training from Rachel Webb Lancashire CC (n=1); spent a day with WEST science leads and planned a curriculum for WEST to use (n=1); Phizzi Earth & Space—Dudley Partnership (n=1).*

Almost 80% of survey respondents reported that it was usual practice to teach science to a Year 5 class at least once a week; this frequency was similar between allocated groups at both baseline and follow-up.

Respondents reported science lessons lasted:

- At baseline in all schools: average 81 minutes:
    - control group: average 79 minutes (range 28–130); and
    - intervention group: average 83 minutes (range 18–140).
- At follow-up in all schools: average 84 minutes:
    - control group: average 83 minutes (range 39–130); and
    - intervention group: average 85 minutes (range 45–120).

In summary, from baseline to follow-up, intervention schools reported that the duration of their science lessons increased by two minutes (from 83 to 85 minutes) and control schools reported a higher increase of four minutes (from 79 to 83 minutes).

Table 23: Frequency of science lessons at baseline and follow-up, by allocated group

How often do you teach science to your own Year 5 class?	Control		Intervention	
	Baseline *n* (%)	Follow-up *n* (%)	Baseline *n* (%)	Follow-up *n* (%)
Every day	1 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)
Three or four times a week	1 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)
Twice a week	9 (10.8)	3 (4.0)	10 (8.7)	6 (7.9)
Once a week	64 (77.1)	58 (77.3)	92 (80.0)	60 (78.9)
Once a fortnight	1 (1.2)	6 (8.0)	2 (1.7)	2 (2.6)

Never	1 (1.2)	2 (2.7)	1 (0.9)	1 (1.3)
Blocked classes for two to three weeks every half-term	3 (3.6)	3 (4.0)	3 (2.6)	2 (2.6)
Missing	3 (3.6)	3 (4.0)	7 (6.1)	5 (6.6)
Total n	83	75	115	75

Figure 8: Strategies and techniques used in science lessons at baseline and follow-up, by allocated group

Figure 8 details the strategies and techniques used in Year 5 science lessons as reported by responding teachers within both the baseline and follow-up surveys. At baseline, the majority of teachers in all schools indicated that their usual Year 5 science lessons often ('very often' or 'quite often') involved teacher demonstration, whole-class discussion, pair/small group discussion, thinking time for pupils, the teaching of scientific facts, and science terminology. Additionally:

- approximately 65% of all teachers in all schools indicated they often asked pupils to solve scientific problems;
- approximately 50% of all teachers indicated that they often required pupils to memorise facts or principles; and
- approximately 25% of all teachers required pupils to write up the whole investigation.

At baseline, there are notable differences between allocated groups in the frequency of how often the science techniques and strategies are implemented in most domains, with the intervention group reporting an observable lower frequency than the control group. Anecdotally, we consider that this could be because the intervention group completed the survey while being aware of their group allocation and in some cases having already started to receive the TDTScience intervention. These factors may have influenced the intervention group to think more in-depth, and more critically, about the content and delivery of their science lessons.

Among the intervention group, there was an observable shift in self-reported practice between baseline and follow-up in two domains:

- Pupils were required to write up the whole investigations less often. This could be indicative of teachers implementing focused recording within their practice. Teachers discussed the use of focused recording within interviews described above.
- Pupils were asked to solve scientific problems more often. This could be indicative of teachers implementing the TDTScience approach of Bright Ideas Time.

It is important to note when interpreting these results that not all survey respondents will have completed both surveys, namely, respondents who completed a baseline survey, may not have completed a follow-up and vice versa. This should be considered when making baseline and follow-up survey comparisons. There is the potential for response bias, particularly among the intervention group where non-response was higher at follow-up.

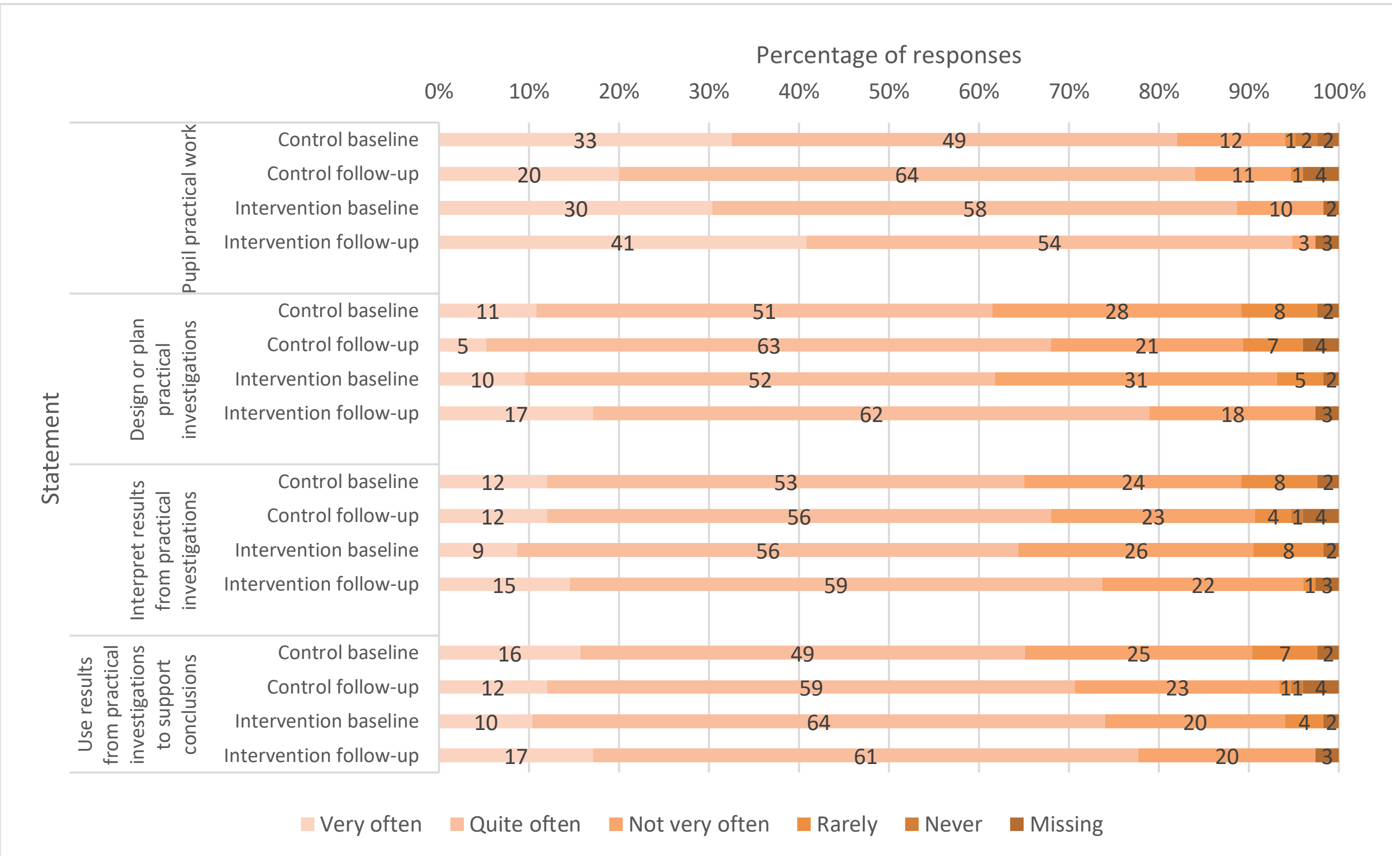


Figure 9: Frequency of practical sciences at baseline and follow-up, by allocated group

Respondents were asked a series of questions to gauge the frequency in which practical science was incorporated into their Year 5 science lessons. As detailed in Figure 9, at baseline the majority of teachers across all schools reported that they often (‘very often’ or ‘quite often’) had pupils complete practical work, design or plan practical investigations, interpret the results from investigations, and use the results from practical investigations to support conclusions.

The proportion of teachers reporting that they incorporated the ‘design or planning of practical investigations’ into their science lessons increased in the intervention group from 62% at baseline (10% ‘very often’; 52% ‘quite often’) to 79% at follow-up (17% ‘very often’; 62% ‘quite often’). Such differences in practices were not observed among the control group at follow-up. There were no other observable difference across the other domains.

As indicated in the section on focused recording in intervention classes, the expansive recording, which was noted as the norm by teachers before TDTScience was recognisable in some of the pupils’ books from control schools, where in one lesson several aspects of an experiment were recorded, namely the different variables, a prediction, a table of results, and a discussion/conclusions section (see example on the left, Image 11 below). In comparison, in a TDTScience school, the focus of recording from an experiment was on one aspect, for example, the setting out of results in a graph, with the subsequent lesson being dedicated to evaluating the results (see example on the right, Image 12 below).

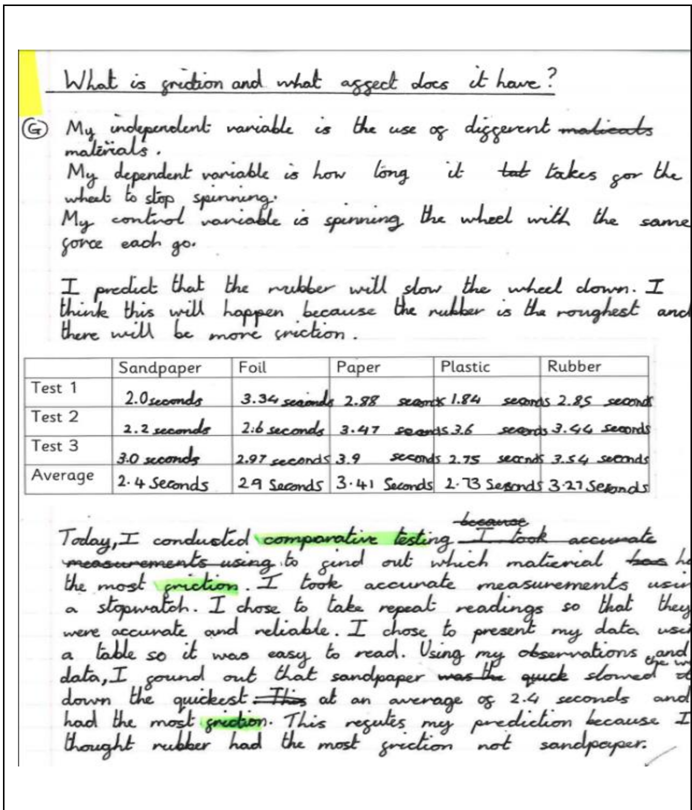


Image 11: Example of an experiment write up on Materials (control group) (S13–C–B)

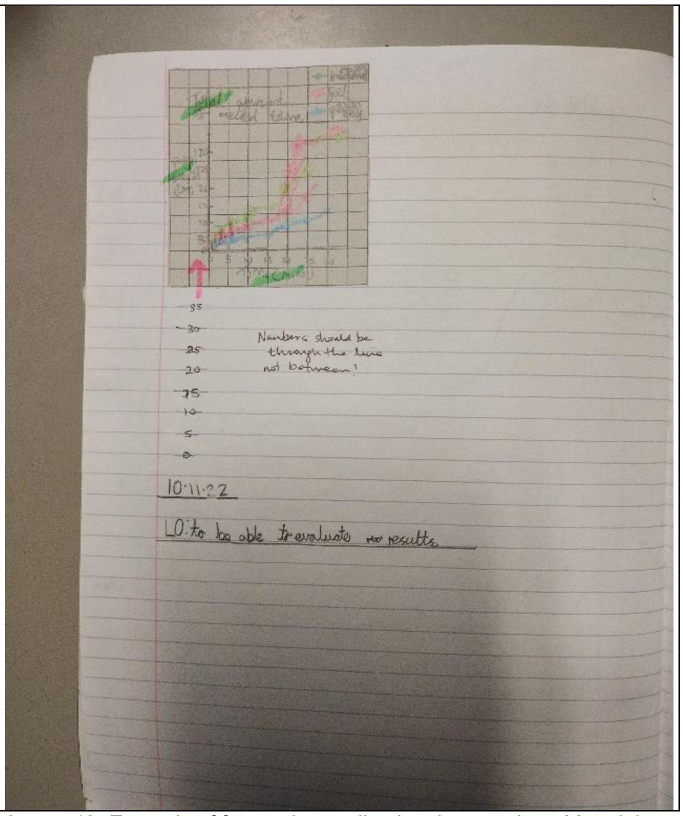


Image 12: Example of focused recording in a lesson about Materials (intervention group) S12–I–B–V1

Respondents were asked a series of questions at baseline and follow-up to gauge how interested and engaged they considered their pupils to be in learning science during 2022–2023. These data are presented in Figure 1010. Generally, the majority of teachers reported that their pupils had relatively high levels of interest and engagement. In comparison to the control group at follow-up, a greater proportion of teachers in the intervention group considered their pupils to have enjoyed their science lessons, made good progress in their understanding of science, have confidence in science, ability to work independently in science, come up with their own scientific ideas, and engage with science more at follow-up in comparison to baseline.

Figure 10: How interested and engaged are pupils in learning science at baseline and follow-up, by allocated group.

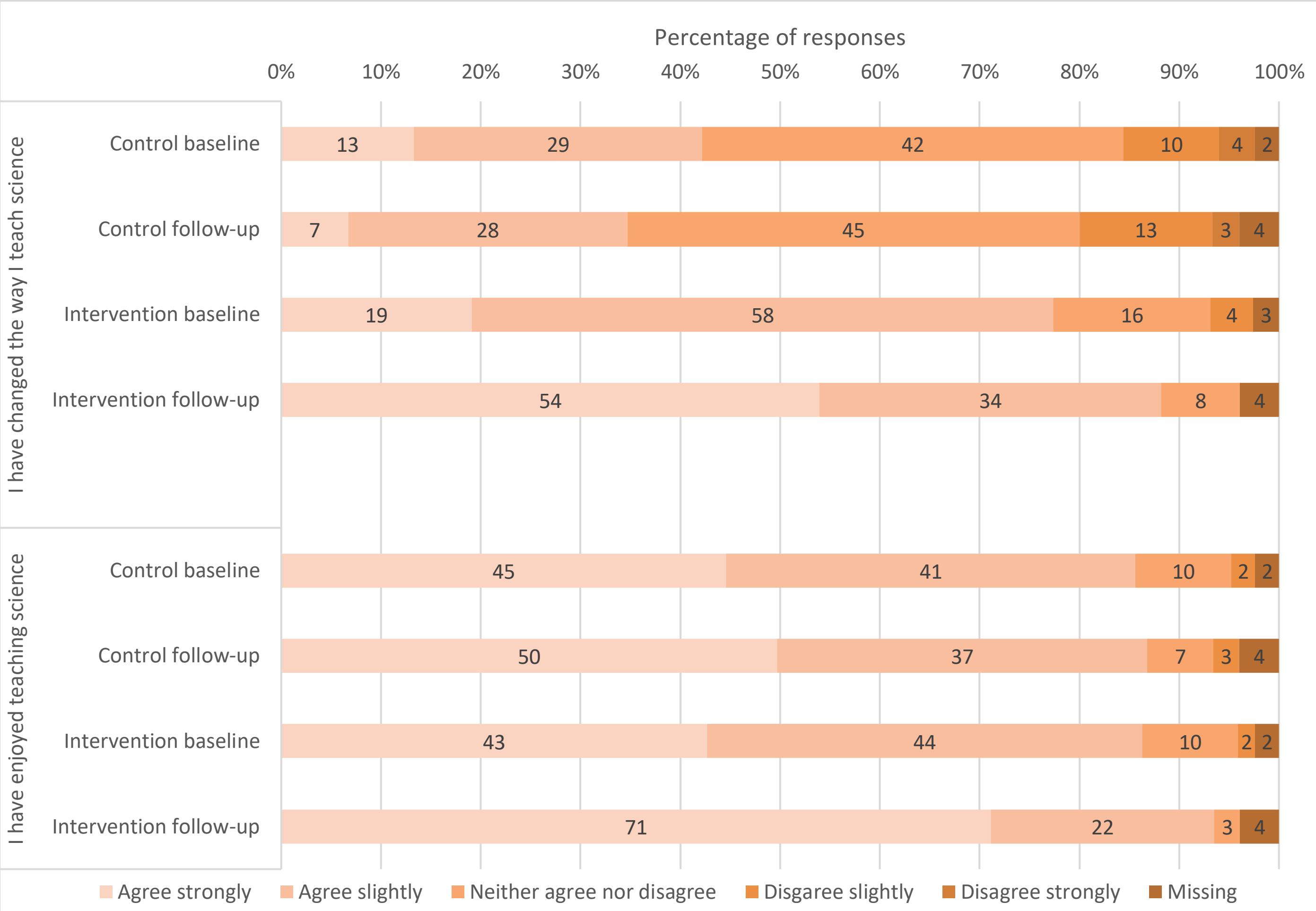
Teachers in all schools were asked if they had changed the way they taught science so far that year (Figure 11).

- In total, 77% of respondents in the intervention group at baseline agreed (19% 'strongly'; 58% 'slightly') that they had changed the way they taught science, in comparison to 42% (13% 'strongly'; 29% 'slightly') within the control group. As discussed above, anecdotally, the observed difference in baseline figures between the allocated groups could be because the intervention group completed the survey while being aware of their group allocation and in some cases having already started to receive the TDTScience intervention.
- Nevertheless, the proportion of respondents in the intervention group selecting that they 'strongly' agreed that they had changed the way they had taught science increased from 19% at baseline to 54% at follow-up. This

increase is noteworthy, and as supported by interview data in the sections above, attributable to participating in TDTScience.

Only 7% of teachers in the control group ‘strongly’ agreed with this statement at follow-up, and 28% selected the next most positive answer (‘slightly’ agree). In total, 34% of control teachers agreed that they had changed practice when asked at follow-up, this may be in response to any science CPD that they had undertaken; as noted above, 30% of control schools indicted at follow-up that they had completed some form of science CPD that academic year. Teachers were also asked at baseline and follow-up to indicate on a Likert scale how strongly they agreed with the following statement: ‘I have enjoyed teaching science’. At baseline, responses were similar between allocated groups with 43% of all teachers reporting they ‘strongly’ agreed, see Figure 111. However, this increased from 43% at baseline to 71% at follow-up in the intervention group and is supported by interview data; there were no large observable differences among the control group (45% at baseline and 50% at follow-up, Figure 11).

Figure 11: How interested and engaged are teachers in teaching science at baseline and follow-up, by allocated group



Summary of usual practice key findings

A summary of the key ‘usual practice’ findings is linked to the relevant research questions and presented below.

5. How does TDTScience compare with practice in business as usual science lessons?

a. What strategies and techniques are used in science lessons?

The majority of teachers in all schools indicated that their usual Year 5 science lessons often (‘very often’ or ‘quite often’) involved teacher demonstration, whole-class discussion, pair/small group discussion, gave pupils time to think, and taught scientific facts and science terminology. Less often, strategies including asking pupils to solve scientific problems, requiring pupils to memorise facts or principles, and write up the whole investigation were used.

b. How interested and engaged are teachers and pupils in science teaching and learning?

A greater proportion of teachers in the intervention group reported that they had changed the way they taught science in 2022–2023, which aligns to interview data in the ‘Impacts of TDTScience’ section above.

There was an increase in the proportion of intervention respondents that had enjoyed teaching science in comparison to the control group at follow-up. Among the control group, 34% agreed that they had changed their science teaching practice in 2022–2023. This may be due to them completing other (non-TDTScience) science CPD that year, which was reported in 30% of control schools.

More teachers in the intervention group considered their pupils to have been more highly engaged and interested in science in comparison to the control group.

c. What is the frequency and length of science lessons?

Science was typically taught once a week in 80% of schools for approximately 81 minutes, a baseline, which increased very slightly at follow-up; by two minutes in intervention schools and four minutes in control schools.

d. What practical science takes place?

In relation to the use of practical science in usual science lessons, the majority of teachers across all schools at baseline reported that they often (‘very often’ or ‘quite often’) had pupils complete practical work, design or plan practical investigations, interpret the results from investigations, and use the results from practical investigations to support conclusions.

The proportion of teachers reporting that they incorporated the ‘design or planning of practical investigations’ into their science lessons increased in the intervention group at follow-up. Such differences in practice were not observed among the control group at follow-up.

e. How much training have Year 5 teachers received in science?

Around a third of participating schools (30%) acknowledged that they had been involved in any other science-related training or initiatives during academic year 2022–2023 (other than TDTScience for the intervention group), and this was evenly balanced between the intervention and control group. The majority of teachers’ highest qualification in science was GCSE or equivalent, with a third having an A level or higher qualification.

Cost evaluation results

For the purposes of the trial, schools allocated to receive the TDTScience intervention were not required to pay for training.

Table 24 details the resources needed to implement the programme as per the ingredients method (Levin *et al.,* 2017). The main cost for implementing the programme is the cost of training. Other costs identified include staff cover, staff travel, and purchase of additional materials; however, staff cover and purchase of additional resources are categorised as optional rather than mandatory as only a proportion of schools utilised these.

Table 24: List of resources, 'ingredients'

Category		Item
Personnel for preparation and delivery		Schools: All Year 5 teachers
		Schools: Another class teacher, ideally science subject lead (optional, if there is only one Year 5 class/teacher)
		Trainers
Personnel for training	Trainers	Science Oxford team
	Trainees	Schools: All Year 5 teachers
		Schools: Another class teacher, ideally science subject lead (optional, if there is only one Year 5 class/teacher)
		Trainers
	Project management and admin	Science Oxford team
Facilities, equipment, and materials (prerequisites)		Information and communications technology (ICT) equipment, e.g., laptop, computer, or iPad, and internet connection
Extras		Staff travel to attend in-person training
Optional extras[a]		Staff cover (as needed)
		Additional learning resources to support science learning (as needed)

[a] Some schools may need to provide staff cover to release teachers for training and intervention activities. Some schools may need to purchase additional resources if they do not already have them in place as part of their usual provisions.

Prerequisites

To take part in TDTScience, schools needed to have access to ICT equipment such as a laptop or a personal computer, and an internet connection in order to access the dedicated website, which provided online versions of TDTScience course resources. To our knowledge, no school purchased such equipment specifically for the purpose of taking part in the evaluation and we anticipate that most schools would already have this equipment should the programme be rolled-out.

Time

The TDTScience programme is intended to be delivered to Year 5 teachers but can then be cascaded by them to other members of staff at the school. Given this, the time costs are largely front loaded and associated with attending in-person training sessions, and schools potentially having to provide cover for teachers to do this and to continue with development activities undertaken during the course of the implementation period.

Data relating to the number of in-person training sessions attended by each teacher was provided by the Science Oxford team and indicated that 199 teachers from 87 intervention schools were nominated by their school to attend training (average 2.3 per school). These teachers attended an average of 3.3 out of 4 sessions (range 0 to 4), equating to 21.6 hours per teacher. Some teachers took over attendance at the sessions for other teachers in their school who may have left or were unable to attend. On average, 7.6 sessions were attended per school (approximately equating to two teachers each attending the full four days of training). This is 49.4 hours per school on average.

The cost and time resource questions from the end of year intervention survey were completed by 73 teachers from 59 schools.

Teachers were asked whether their school arranged for paid cover for them to attend the TDTScience in-person training, and if so to estimate the total number of hours of cover the school arranged. These data are summarised in Table 25. Fifty-eight teachers (79.5% of 73) from 49 schools (78.0% of 59) said their school had paid for cover for them to attend training in the 2022–2023 academic year; the number of hours averaged 21.3 per teacher across all 59 schools (range 0 to 50). Less than ten[9] teachers (<13.7% of 73) from less than ten schools (16.9% of 59) said they had received other support from the TDTScience team but none of these required paid cover. Twenty teachers (27.4% of 73) from 20 schools (33.9% of 59) spent time sharing TDTScience training with other members of staff within their school who missed the in-person training sessions, of which less than ten reported that this required their school to arrange paid cover (average 0.09 hours across all teachers). Eleven teachers (27.4% of 73) from 11 schools (18.6% of 59) received TDTScience training from another staff member within their school because they missed one or more in-person training sessions, of which less than ten reported that this required their school to arrange paid cover (average 0.07 hours across all teachers). Seventy-one teachers (97.3% of 73) from 58 schools (98.3% of 59), spent time completing activities related to the TDTScience programme (excluding the TDTScience training and support), but less than ten reported that this required their school to arrange paid cover (average 0.84 hours across all teachers).

Given the variable uptake of staff cover, this is considered an optional cost that would be utilised on a school-by-school basis.

Table 25: Total time devoted by personnel for training, related activities, and staff cover

	Year one (2022–2023)
	Mean (SD) number of hours per teacher
Time spent	
Attending in-person training	21.59 (6.59)[a]
Receiving other support from the TDTScience team	0.16 (0.47)[b]
Sharing TDTScience training with other members of staff within their school who missed the in-person training sessions	0.51 (1.00)[b]
Receiving TDTScience training from another staff member within their school because they missed one or more in-person training sessions	0.27 (0.77)[b]
Completing activities related to the TDTScience programme, excluding the TDTScience training and support	8.79 (10.8)[b]
Hours of paid cover for teacher	
Attending in-person training	21.35 (14.00)[b]
Receiving other support from the TDTScience team	0.00 (–)[b]
Sharing TDTScience training with other members of staff within their school who missed the in-person training sessions	0.09 (0.47)[b]
Receiving TDTScience training from another staff member within their school because they missed one or more in-person training sessions	0.07 (0.42)[b]
Completing activities related to the TDTScience programme, excluding the TDTScience training and support	0.84 (4.14)[b]

[a] *199 teachers from 87 schools.*
[b] *73 teachers from 59 schools.*

Twenty-three (31.5% of 73) teachers from 21 schools (35.6% of 59) reported that they attended/completed at least some TDTScience training or activities outside of normal working hours/in their own time. Table 26 summarises the number of hours spent on each activity on average per teacher.

Table 26: Total time devoted by personnel for training and related activities outside working hours

	Year one (2022–2023)
	Mean (SD) number of hours per teacher[a]
Attending in-person training	3.92 (10.42)
Receiving other support from the TDTScience team	0.47 (3.52)
Sharing TDTScience training with other members of staff within their school who missed the in-person training sessions	0.36 (0.93)
Receiving TDTScience training from another staff member within their school because they missed one or more in-person training sessions	0.05 (0.37)

[9] Counts less than ten for individuals (teachers) suppressed to avoid statistical disclosure.

Completing activities related to the TDTScience programme, excluding the TDTScience training and support	2.54 (6.47)

[a] *73 teachers from 59 schools.*

Financial costs

Travel to training (cost to schools)

Teachers were asked to estimate how much it cost for them to travel to the in-person sessions overall. The average was £19.72 (SD 20.19) based on data from 73 teachers from 59 schools (£24.40 per school) (Table 27).

Staff cover

As detailed above, paid staff cover was used variably by schools and so it is considered an optional cost that schools considering implementing TDTScience should factor into their considerations. On average, across all schools providing data, cover was required for 22.34 hours per teacher (27.64 hours per school) in the academic year 2022–2023. To calculate the average cost per school, we multiplied these figures by £18.21. Therefore, on average, per school, £503.32 was spent on cover during the academic year 2022–2023 (Table 27).

Materials

Additional materials

Sixteen teachers (21.9% of 73) from 16 schools (27.1% of 59) reported that they had needed to buy additional resources as a result of the TDTScience programme in their school. They were asked what items/resources they bought, the quantity and the cost. On average, across all 73 teachers, teachers reported spending £12.71 on additional resources (range £0 to £500) in the year 2022–2023 (average £15.73 per school) (Table 27).

A range of items were purchased, and no two teachers reported purchasing the same item. Examples include a ball of string, batteries, candles, washing-up liquid, force meters, funnels, and sieves (not an exhaustive list).

Since not all schools reported purchasing additional materials this is considered an optional rather than mandatory cost for implementation. Costs associated with additional resources could represent both a start-up or recurring cost depending on the frequency that resources need to be replaced. Costs are likely to be lower at start-up for schools that already have a wealth of relevant learning resources and higher for those that do not.

Training (cost to Science Oxford team)

Training of ‘trainers’ was delivered by colleagues from the Science Oxford team. Trainers in turn trained teachers during in-person sessions. Teachers were trained in groups. Programme delivery costs were provided by the Science Oxford team to the evaluation team in two forms: i) according to the actual costs incurred during this trial (Table 28) including intervention set-up and development, setting recruitment, and intervention delivery; and ii) costs of delivering the intervention in the ‘real-world’ (non-trial) setting (Table 29), which include training and intervention delivery costs for 30 teachers. During the trial, the estimated total cost of delivering the intervention was £8,635.99 per school over three years, which equates to £66.95 per pupil per school year (assuming 43 pupils per school) (Table 29). Going forward, the estimated cost of delivering the intervention to a school would be £2,049.33 over three years (assuming two teachers per school were trained), or £11.39 per pupil per school year (assuming 30 pupils per teacher) (Table 30).

Table 27: Cost of the implementation of the programme to schools, per ingredient

Category	Cost ingredient	Start-up or recurring?	Nominal values			
			£ Year one	£ Year two	£ Year three	Total
Staff cover (start-up) (optional)	Paid cover for teachers to attend training, receive support from the TDTScience team, share TDTScience training with other members of staff, receive TDTScience training from another staff member, complete other activities related to the TDTScience programme	Start-up	£503.32	£0	£0	£503.32
Staff travel (start-up) (mandatory)	Staff travel to attend in-person training	Start-up	£24.40	£0	£0	£24.40
Materials (start-up) (optional)	Additional resources purchased	Start-up	£15.73	£0	£0	£15.73
Materials (ongoing) (optional)	Additional resources purchased	Recurring	£0	£15.73	£15.73	£31.46
Total cost of items per school			£543.45	£15.73	£15.73	£574.91
Total cost per pupil (43 pupils per school)						£13.37
Total cost per pupil per school year (43 pupils per school over three years)						£4.46

Table 28: Cost of the implementation of the programme to Science Oxford, per mandatory ingredient – actual costs during the trial

Cost ingredient	Start-up or recurring?	Nominal values			
		£ Year one	£ Year two	£ Year three	Total
Intervention development[a]	Start-up	£82,684	£0	£0	£82,684
Setting recruitment	Start-up	£13,395	£0	£0	£13,395
Intervention delivery[b]	Start-up	£681,160	£0	£0	£681,160
Total cost of programme delivery for all schools		£777,239	£0	£0	£777,239
Total cost of programme delivery per school (90 schools)					£8,635.99
Total cost per pupil per school year (43 pupils per school over three years)					£66.95

[a] Programme set-up and intervention development (e.g. staffing costs, management and admin costs, recruitment of trainers, updating the TDTScience course for Year 5 focus for the trial).
[b] Developer and training costs, train-the-trainer, project management and admin, teacher resource grants for intervention schools.

Table 29: Cost of the implementation of the programme to Science Oxford, per mandatory ingredient – non-trial ('real-world') cost of intervention (covers cost for 30 teachers, assumes 30 pupils per teacher – equates to 15 schools assuming two teachers per school)

Cost ingredient	Start-up or recurring?	Nominal values			
		£ Year one	£ Year two	£ Year three	Total
Intervention delivery[a]	Start-up	£30,740	£0	£0	£30,740
Total cost of programme delivery for all schools		£30,740	£0	£0	£30,740
Total cost of programme delivery per school (15 schools)					£2,049.33
Total cost per pupil per school year (60 pupils per school over three years)					£11.39

[a] Training (travel, venue hire, catering), intervention delivery (marketing, travel, accommodation, catering, venues, other resources).

Overall costs

Table 30 presents the total cost per pupil per year over three years based on the trial costs at £71.40.

Table 30: Total combined costs of training and extras (actual trial costs)

Item	Type of cost	Total cost per school over three years	Total cost per pupil per year over three years (43 pupils per school per year)
Programme delivery	Start-up cost per school	£8,635.99	£66.95
Optional cover	Start-up	£503.32	£3.90
Staff travel	Start-up	£24.40	£0.19
Optional materials	Start-up recurring cost per school	£47.19	£0.37
Total		£9,210.90	£71.40

Conclusion

Table 31: Key conclusions

Key conclusions	
1.	Pupils in TDTScience schools, made the equivalent of zero months' additional progress in science attainment, on average, compared to pupils in other schools. This result has a high security rating.
2.	Pupils' attitudes to science in TDTScience schools were positively impacted, on average, compared to pupils in other schools. This positive change in attitudes was of a small size.
3.	Pupils receiving free school meals (FSM) in TDTScience schools made the equivalent of zero months' additional progress in science attainment, on average, compared to pupils receiving FSM in other schools. These results may have lower security than the overall findings because of the smaller number of pupils.
4.	Teachers changed their practice in accordance with TDTScience's theory of change. This was evidenced through lesson observation and noted by pupils in focus group discussions who indicated more time for discussion, practical work, and less writing. Teachers observed, and pupils reported, high levels of engagement and self-efficacy in science, which translated into improvements in pupils' attitudes to science, but was not reflected at the end of the academic year in science attainment.
5.	TDTScience was implemented as intended, with high compliance and fidelity. Interviews and surveys reported that teachers felt TDTScience improved their lesson planning, and they were more confident teaching science, with clear strategies, techniques, and practicals embedded in the programme.

Impact evaluation and IPE integration

Below we detail the relevant items of the logic model from Figure 1 (in italics) and discuss the evidence base from the impact evaluation and the IPE in support (or not) of each logic model item. Overall, the results of the impact evaluation and the IPE provide mixed evidence to support or refute the logic model. The majority of outputs supported most of the short-term and medium-term outcomes as per the IPE. However, these factors did not translate into long-term improvements in pupils' science attainment at the end of Year 5 (primary outcome) for Cohort 1, or for pupils eligible for FSM although improvements were made to pupils' attitudes towards science (secondary outcome).

Inputs and activities

Inputs: TDTScience programme; Training materials; Access to TDTScience website; Trainers; Teachers.

Activities: Trainers deliver the course to teachers in trial schools, 4x one-day sessions (two Autumn Term, and two Spring Term) plus a half-day (Summer Term)

Inputs and activities detailed in the logic model related to the Science Oxford team and trainers delivering the training and providing schools full access to the TDTScience programme; all training sessions were delivered in all regions as intended. In addition to the CPD, teachers received hard copies of all TDTScience course resources in a ring binder, some low-value science equipment at the point of course delivery and access to online versions of TDTScience course resources via a dedicated website. All of the teachers interviewed as part of the IPE felt further supported by the resources provided via the training, both physical (for experiments/often referred to as 'goodies' or 'freebies') and the paper-based and online resources (TDTScience folders and links to relevant websites).

Outputs

Delivered as planned with fidelity; Session goals met; Attendance at training.

Outputs detailed in the logic model related to training activities being delivered as planned and being sufficiently attended by the trainers and teachers alike.

Overall, the train-the-trainer model worked well. In the post-training day surveys, all trainers (12/12) reported that they 'strongly' agreed that they felt confident delivering four of the training days and only one feeling slightly less so ('slightly' agreed) for the remaining training day (Day 3). The majority (over 75%) also reported feeling 'very well prepared'.

Additionally, in the post-training day surveys trainers agreed that they were able to effectively train teachers in the use of TDTScience strategies. The extent of the agreement relating to training teachers in specific TDTScience strategies (e.g. 'Practical prompts for thinking', 'Practical problem-solving') varied depending on the training day.

Implementation fidelity during the training was judged to be high following the training observations by researchers, and trainers reported high levels of fidelity to the tutor file in the post-training day survey. In the trainer interviews, they discussed the importance of replicating the training they had previously received and only reported making minor 'tweaks', primarily based on their own previous experiences and the need to be responsive to the specific needs of the group of teachers they were training.

Teachers were overall positive about the training they received and felt that the trainers were both knowledgeable and supportive. Teachers felt that some days were more successful than others in addressing different strategies. Teachers were particularly receptive to learning new ideas around teaching science during the training days as well as the opportunities to experience practical sessions as they could be delivered in the classroom and to share ideas with other teachers.

Training attendance data for the four full-day training sessions was provided by the delivery team for 199 teachers from 87 intervention schools (median 2, range 1 to 4, per school). Two (1.0%) of the teachers did not attend any sessions, 15 (7.5%) of the teachers attended one session, 23 (11.6%) of the teachers attended two sessions, 36 (18.1%) of the teachers attended three sessions, and 123 (61.8%) of the teachers attended all four sessions.

Moderators

Teacher pedagogical experience; Amount of classroom time for science; Pupils prior attainment; Trainer experience; Dosage; Fidelity; FSM.

The majority of teachers' highest qualification in science was GCSE or equivalent, with no notable differences by allocated group. Science was typically taught once a week in 80% of schools for approximately 81 minutes, and this did not alter substantially between baseline and follow-up or by allocated groups. This finding suggests that the intervention did not influence the amount of time schools allocated to teaching science in intervention schools. During IPE interviews teachers expressed that they did not have enough time to implement the TDTScience fully as they were often 'losing' science time to extracurricular activities; as science is not considered to be as important as maths and English, it is often scheduled for afternoons where extracurricular activities take place.

The baseline measure was the average point score from the 17 ELGs that make up the EYFSP (Key Stage 1 results were unavailable for this cohort due to the COVID-19 pandemic). EYFSP data, used as a measure of prior attainment, was obtained from the NPD. NPD data could be matched for 93.4% of the randomised population (intervention, n=3,527/3,790, 93.1%; control, n=3,258/3,471, 93.9%); missing data may be caused if a record could not be found in the NPD that could be matched to the pupil details (e.g. forename, surname, date of birth, and UPN) provided by the evaluation team with a high degree of confidence, for instance if there is an error in the spelling of the name. The mean EYFSP average point score was slightly higher in the intervention group than the control group (2.07, SD 0.45, compared with 2.03, SD 0.44; Hedges' g effect size between the groups 0.09; 95% CI: 0.04 to 0.13).

As reported in Standley *et al*. (2023), Science Oxford considered trainers to be highly experienced. All 12 trainers had previously been teachers (11 primary, one secondary) as well as being science CPD providers—either as independent consultants or as part of the work of their organisation. Of these, seven had previous experience of training teachers in TDTScience (either in the first effectiveness trial or in TDTScience training offered by the developers). The trainers were all previously known to the developers and were selected based on prior experience in delivering primary science teacher training and interest in the TDTScience approach and previous evaluation findings.

Compliance was high with over 80% of pupils taught in a class that was taught by a teacher who attended at least three out of the four full days of training. However, the CACE estimate of the treatment effect on the pupils' attainment was similar to the ITT estimate (effect size 0.03, 95% CI: -0.09 to 0.14) in that there was no impact to science attainment. There was no evidence of contamination in any teachers from the control school attending training sessions.

There was a statistically significant interaction between intervention and FSM status, which indicates that there was evidence that the intervention effect for science attainment differs between FSM and non-FSM pupils. The analysis restricting to pupils eligible for FSM produced a very small, negative effect of the intervention within this subgroup, which was not statistically significant. The IPE noted that teachers regarded the TDTScience to be inclusive to all groups of pupils due to the emphasis on practical work and discussion as well as focused recording. Teachers mentioned this in relation to pupils from disadvantaged backgrounds, as well as pupils with low prior attainment, those from disadvantaged socio-economic backgrounds, SEND, or who had EAL.

The IPE identified ‘school culture’ and ‘resources’ to be both facilitators and barriers to implementation and therefore, these themes have been included as ‘moderators’ within a revised logic model.

Mediators

Pupils’ engagement with science; HOT; Teacher confidence in teaching science.

IPE survey data at follow-up indicated that more teachers in the intervention group considered their pupils to have been more highly engaged and interested in science in comparison to the control group. Pupils themselves, in the focus groups, reported high levels of engagement with science in Year 5.

Teachers delivering TDTScience in their classrooms expressed mixed views as to pupils’ adoption and development of HOT, and this is discussed further in the ‘Outcomes’ section below.

During training, some teachers felt that their own lack of experience sometimes prevented them from participating in discussions as much as they would have and made some concepts difficult for them to grasp. One of the major changes teachers reported as a result of TDTScience was increased confidence in teaching science. During interviews, some mentioned that they felt more confident with doing practical activities because they did them in training, they understood how they work (or don’t) and felt they were better equipped to deal with possible misconceptions or what might not go to plan. Teachers’ increased confidence was also observed when comparing the baseline and follow-up survey responses from the intervention group. Teachers in the intervention group tended to report being more confident in teaching science than those in the control group at follow-up,

Outcomes

Short- and medium-term outcomes

Teachers: TDTScience strategies are embedded in all Year 5 science lessons; Familiarity with TDTScience strategies; Preparedness and confidence to use in the classroom.

Pupils: Pupils thinking encouraged; More dedicated discussion time; Challenging practical work; Clearly focused pupil recording; Increased engagement with science; Improved attitudes towards science.

Data from the follow-up usual practice survey indicated that a greater proportion of teachers in the intervention group reported they had changed the way they taught science and had enjoyed teaching science in comparison to the control group. Data from focus groups reported that pupils noticed that their teacher was ‘doing something different’ in science this academic year, particularly in terms of their ‘thinking, talking, and doing’ science. As discussed in the ’Mediators’ section above, teachers reported increased confidence in teaching and delivering science lessons; this was considered a major change resulting from the intervention.

Overall, IPE classroom observation data concluded that teachers were prepared and implemented learning from the TDTScience programme well within their lessons. Observation data highlighted that teachers’ did not always teach topics in the order in which they were delivered in training. Some teachers did not feel all of the activities could be transferred directly into the classroom. Teachers did express that the free resources were particularly valued and helped them to deliver TDTScience in the classroom.

Teachers reported improvements in pupil’s confidence, knowledge, and thinking skills. These impacts were felt to be beneficial to all pupils.

Teachers frequently praised the strategies designed to promote HOT among pupils thinking, for example, Bright Ideas Time, such as Odd One Out, the Big Question, and PMI. However, some teachers queried the practicality of some of the openness of the lines of enquiry encouraged by the programme in their own classrooms, feeling that their own pupils needed a little more scaffolding than they understood the programme to require. Some teachers sensed that pupils prefer ‘the right answer’, which meant that TDTScience techniques needed to be chosen carefully and strategically at every stage of a topic.

As anticipated by the delivery team, in their interview, the concept of focused recording was the area that teachers reported struggling with the most given schools’ reliance on work in pupils’ exercise books to track learning and this was also reflected in the analysis of pupils’ textbooks. However, teachers’ also perceived benefits from the approach. The data from the usual practice surveys indicated that the use of ‘practical science’ within science lessons was frequently used by teachers across both groups, with no observable differences at follow-up. However, it was more common for the control group to report that they taught ‘scientific facts’ during lessons at follow-up in comparison to the intervention group.

Long-term outcomes

Primary outcome (impact): Pupils science attainment improves (measured by science measure for Year 5).

Secondary outcomes (impact): Pupils’ attitudes towards science improves (measured by Science Attitude Questionnaire).

Within the IPE, the predominant theme from teachers was that TDTScience had improved student engagement with science and science lessons, and consequently their thinking skills. However, as per the results of the impact evaluation, participation in the programme did not result in any additional progress for pupils in the intervention group compared to those in the control group in science attainment as measured using the Year 5 Science Assessment.

There was improvement in attitudes towards science among pupils whose schools were allocated to the TDTScience intervention. During IPE focus groups, pupils made connections between the increase in discussion and practical work and changes in focused recording. In most cases, pupils felt the new approach resulted in less writing, and there was a general sense that most pupils enjoyed their science lessons. A revised logic model describing the evidenced-based pathways to improved pupil attitudes is provided in Appendix E.

Interpretation

Research has shown that the primary science experience heavily influences subsequent subject attitudes but is often low priority and teachers may lack confidence teaching it (e.g. see Markwick and Reiss 2023; Ofsted 2021; Slavin *et al.,* 2014; Harlen and Qualter, 2018). This evaluation was delivered following the COVID-19 pandemic, which saw unprecedented challenges for the education sector. In 2021, Ofsted reported that there had been a decline in science teaching in primary schools during the COVID-19 pandemic that resulted in a narrowing of the science topics being taught as schools. The Parliamentary Office for Science and Technology (2003) identified that many primary school teachers tended to concentrate too much on teaching factual content in science and not enough on core scientific ideas.

As evaluated in Year 5, the TDTScience programme was designed to develop science teaching to improve pupils HOT and science attainment and attitudes towards the subject. The TDTScience intervention is a low-moderate intensity CPD intervention, which invites all Year 5 teachers within a participating school to attend at least four and a half-days of face-to-face training over the course of one academic year. Hardcopies of the TDTScience resources, low-value science equipment, and access to the online version of the TDTScience course resources were also provided to schools. This was the second effectiveness trial evaluating TDTScience. The main difference between this effectiveness trial and the previous one (Kitmitto *et al.,* 2018) was the train-the-trainers model adopted by Science Oxford to deliver TDTScience to teachers in trial schools.

This evaluation took the form of a large (180 schools), one-year, cluster randomised controlled trial aiming primarily to improve science attainment among Year 5 pupils. The trial recruited to target in terms of the number of schools randomised (n=180). Randomisation used minimisation and the school characteristics were well balanced between trial arms at baseline. However, for logistical reasons, schools were randomised before pupils were recruited and 12 schools (predominantly control schools, n=11) were lost at this stage. The number of randomised pupils was therefore, based on the remaining 168 schools, rather than the 180 randomised. Randomised pupils were similar in terms of gender and FSM status between the intervention and control groups, but those in the intervention group had a very slightly higher EYFSP average point score, which was used as our measure of prior attainment. We allowed for 15% pupil-level attrition in the analysis (from the randomised population), and the amount we saw was just 16% and this was balanced across the two arms. The observed correlation between the chosen measure of prior attainment (EYFSP average point score) from the end of Reception year and follow-up Year 5 Science Assessment score was as anticipated (0.50). However, the ICC at analysis was lower (0.10 as opposed to 0.15) as was the average cluster (number of pupils per school) size (36 as opposed to 45), which meant the trial was well-powered to detect an estimated effect size that was slightly smaller than the one we had originally powered for (0.13 as opposed to 0.15).

The intervention was very well received by teachers in schools delivering the intervention with teachers reporting positive changes to their confidence in teaching science as a result. However, the results of the primary outcome—pupil science attainment at the end of Year 5—did not show benefit. The observed effect size for the primary outcome was smaller than 0.13, at 0.02, and it did not reach statistical significance. Observed effect sizes, whether statistically significant or not, need to be interpreted in terms of whether the difference would be deemed meaningful in practice. The 95% CI for the effect size for the primary result indicated that the intervention could result in an effect of two months' additional progress.

The impact evaluation suggests there is little to no benefit of the TDTScience intervention on Year 5 Science attainment, for all pupils and for those eligible for FSM. This is in contrast with the previous small-scale efficacy trial, which showed indications that the intervention might be especially beneficial for pupils eligible for FSM (Hanley, Slavin, and Elliott, 2015) and the previous effectiveness trial where pupils eligible for FSM made a small amount of additional progress (Kitmitto *et al.,* 2018).

Among control schools 34% of teachers who responded to the follow-up survey agreed that they had changed their practice. It is possible that this finding can be explained by data relating to the completion of other (non-TDTScience) training or initiatives whereby 30% of control schools reported completing such initiates/training during the intervention delivery period. As a similar proportion of intervention schools also reported being involved in such science initiative/training (non-TDTScience) during the same period, there is balance between allocated groups and therefore, is not considered to compromise the design of this evaluation. The low proportion of science CPD reported here aligns with findings from Ofsted (2021) who found most primary schools were focusing on catch-up work in maths and English in response to the COVID-19 pandemic. There is no suggestion within national data that science attainment has improved over recent years, which could have diluted the effect of the intervention. In fact, Ofsted (2023) recently reported concerns over the low status of science in some primary schools, particularly since national tests in science were removed in 2009, which coincided with a relative decline in the performance in science of ten-year-old pupils in England. Within the IPE, teachers also noted several barriers to implementation including a lack of science resources, having to implement TDTScience alongside other existing schemes of science work, aligning TDTScience with the content and volume of the national curriculum requirements, general school culture, and a lack of time for science within the curriculum.

Our findings relating to pupils science attainment are in contrast with the previous small-scale efficacy trial involving Year 5 pupils in 41 schools (Hanley, Slavin and Elliott, 2015), where pupils of teachers trained in TDTScience made three months' additional progress in science. The findings of this evaluation are consistent with those found within the first effectiveness trial of TDTScience in 205 schools (Kitmitto *et al.,* 2018), which also failed to show evidence of additional progress for most pupils at the end of Year 5. As within Kitmitto *et al.* (2018) this evaluation also saw teachers been trained and attainment measures collected from pupils within the space of one academic year. Whether there is a 'soak' effect of the TDTScience intervention is being explored within the longitudinal follow-up to be reported in Spring Term 2025, namely, Year 5 teachers received training during 2022–2023 preparing them to fully implement the CPD into their teaching practices for the following cohort in the academic year 2023–2024.

Secondary analyses from this evaluation suggest that Cohort 1 (Year 5) pupils from schools that were offered TDTScience had a small improvement in their attitudes to science. There was also some positive impact on attitudes towards science in the efficacy trial (Hanley, Slavin and Elliott, 2015) and in the previous effectiveness trial (Kitmitto *et al.,* 2018). This is an important finding given the reported worldwide decline in pupils' favourable attitudes toward science (e.g. Kennedy et al., 2016). A meta-analysis found a positive and moderate relationship between science attitudes and attainment (Mao *et al.,* 2021), suggesting that enhancing students' positive attitudes toward science could be conductive to students' learning in science. While the increase in positive attitudes towards science did not translate into greater attainment at the end of Year 5 within this evaluation, the longer-term impact of the intervention will be explored when pupils reach the end of Year 6 and reported as part of the longitudinal follow-up. Improving attitudes towards science among school pupils is considered to be very important to enhance their interest in pursuing science-based subjects in later education and careers in later life.

As with most evaluations, the results of the study are only applicable to the sample of schools that were enrolled into the study. Only a relatively small proportion of schools expressed an interest in taking part and were recruited to the study out of those approached (approximately 4%, i.e. 182/4,402). This could potentially introduce selection bias in that the participating schools may not be representative of all those approached or of the 'average' state primary school. When school- and pupil-level characteristics of the randomised sample were compared with national data we saw some small differences. In our sample, relative to the national picture, we had a smaller proportion of community schools and a higher proportion of academy converter schools; a higher proportion of schools in urban areas, as opposed to rural areas; and, of the schools for which the latest Ofsted inspection rating was available, we have a lower proportion of schools rated 'Outstanding' and a higher proportion rated 'Good'. Additionally, we saw that randomised pupils were more likely to be eligible for FSM than pupils in primary schools in England in general; this reflects the fact that schools with higher than the national average percentage of pupils eligible for FSM were targeted for enrolment in this trial.

Limitations and lessons learned

There were a few limitations to this study, including that, given the nature of the intervention, schools had to be randomised at the end of the academic year before the year in which the evaluation was taking place. This was to enable schools to participate in the trial, for example, knowing planned training dates in advance so that necessary arrangements could be made to enable them to attend. This meant that the baseline survey was completed by schools once group allocation was known, and in many cases, the intervention had begun. This may have impacted how teachers, particularly in intervention schools, responded. Additionally, the recruitment of pupils to the trial could only be identified after randomisation, at the start of the evaluation year when it was confirmed, which pupils would be in the Year 5 class at the school. This resulted in the loss of 12 randomised schools from the evaluation who withdrew at this point. The loss was not evenly distributed across the intervention and control groups, as 11 of these schools were in the control group. Another school, from the intervention group, did not provide outcome data as it was discovered following randomisation that they met one of the exclusion criteria for the trial. However, a comparison of school and pupil characteristics between intervention and control groups among those included in the analysis indicate reasonable balance. Therefore, this differential loss-to-follow-up does not appear to have introduced substantial selection bias. The overall attrition from randomisation to analysis was very similar, at 16%, to the amount we had allowed for in the power calculation (15%).

The baseline measure was the average point score from the 17 ELGs that make up the EYFSP. This was chosen as an alternative to the Key Stage 1 English (Reading) and Mathematics scores used in the previous effectiveness trial as Key Stage 1 results are not available for this cohort of pupils, who would have been in Year 2 during the academic year 2019–2020 when national Key Stage 1 assessments were cancelled due to the COVID-19 pandemic. Key Stage 1 results would have been the preferred choice for the baseline measure as this would have allowed a direct comparison of results with the previous effectiveness trial, and it is likely that the correlation between Key Stage 1 results and the outcomes in the trial would have been higher than with EYFSP results as these were assessed longer ago. This may have resulted in a more precise result but would have been unlikely to change the magnitude or direction of the effect.

There was a small, statistically significant difference between the EYFSP average point score, used as the baseline measure, between the intervention and control group for pupils included in the primary analysis (0.08 of an effect size). However, EYFSP average point score was included as a covariate in the analysis, which would control for this imbalance.

Overall, this was a well-powered, cluster randomised controlled, effectiveness trial. The primary outcome was assessed and marked blinded to intervention, with limited recorded instances of unblinding. Even where unblinding occurred, which occurred six times in the control group and 12 times in the intervention group), the risk of any bias being introduced is extremely low as assessments were not administered one-to-one between the invigilators and pupils, but rather at a class level.

In total, 505 pupils were absent on the day of primary outcome testing, this figure could have been decreased with mop-up testing; however, it was not considered finanaically or practically viable to resent invigilators to schools for mop-up testing for <15 pupils. This scenario did not occur.

The pupil attitudes survey was administered by teachers as opposed to independent research invigilators due to practical and financial constraints, namely, the evaluation deemed it was not appropriate for the survey to be completed on the same day as the science assessment, and it would have been too costly to resend invigilators to schools for a secondary outcome measure. Although unlikely, teachers may have influenced pupil's survey responses.

It is possible that schools that participated in IPE interviews are systematically different from schools that did not and, consequently, the findings from this aspect of the evaluation should be treated with a measure of caution. These were a self-selecting population and, as such, bias could have been introduced. For example, teachers who were extremely enthusiastic about the TDTScience programme may have been more likely to volunteer to be interviewed than those who experienced issues or challenges with the programme.

Usual practice surveys were administered at baseline and follow-up. All teachers in the school who taught Year 5 science were asked to complete the usual practice surveys at each timepoint. Baseline response rates were high. Not all respondents who completed a baseline survey also completed a follow-up survey, which may result in response bias. There were some very slight differences in the proportion of teachers with an A level or equivalent in science that responded to the follow-up survey in comparison to the baseline survey, and in the proportion of teachers reporting to have a postgraduate degree in science at follow-up in comparison to baseline. Such small differences are unlikely to impact on the outcomes, however, we acknowledge here that this is a limitation of these survey data.

The data for the cost evaluation were based on information provided by teachers/schools and so is only as accurate and reliable as the information they could recall. Not all schools responded to the cost survey, so the schools that did not respond may have had different costs, which may or may not have changed the results.

Future research and publications

This is the second effectiveness trial to suggest no benefit of the TDTScience intervention specifically on science attainment at the end of Year 5. We would therefore, not recommend any further evaluation of attainment at this timepoint.

Future research includes assessing the impact of the TDTScience intervention on attainment and attitudes to science among the next cohort of Year 5 pupils (Cohort 2) in schools that participated in the main trial. This will investigate whether the effect of TDTScience appears to be modified in any way after teachers have received the entire training package (which is delivered across the academic year) and had a greater opportunity to incorporate TDTScience in their science teaching. Cohort 1 will also be followed into the second year and requested to complete a science assessment at the end of Year 6 (June 2024–July 2024). Pupils will also be followed up based on their attainment in Mathematics and Reading in the Year 6 SATs, using data from the NPD. The results of the longitudinal follow-up will be reported in an addendum report in 2025.

References

Abrahams, I., Bennett, J., Cheung, A., Elliott, L., Hanley, P., Oberio, Z., Rudd, P., Slavin, R., Turkenburg, M. (2014) 'Evaluation of the Impact of a Continuing Professional Development (CPD) *'Course for Primary Science Specialists: Final Report to the Wellcome Trust'.* York: University of York, Department of Education. Available at: https://pure.york.ac.uk/portal/en/publications/evaluation-of-the-impact-of-a-continuing-professional-development (accessed 15 July 2024).

Aksela, M. 2005. *'Supporting Meaningful Chemistry Learning and Higher-Order Thinking Through Computer-Assisted Inquiry: A Design Research Approach'.* Doctoral Thesis. Helsinki: University of Helsinki. Available at: https://core.ac.uk/download/pdf/14916675.pdf accessed 19 July 2024).

Altman, D.G. and Bland, J.M. (2005) 'Treatment Allocation by Minimisation'. *BMJ*, 330: 7495, 843. https://doi.org/10.1136/bmj.330.7495.843

Bax, S. (2002) 'The Social and Cultural Dimensions of Trainer Training'. *Journal of Education for Teaching*, 28: 2, 165–178. https://doi.org/10.1080/0260747021000005592

Bianchi, L., Whittaker, C. and Poole, A. (2021) *'The 10 Key Issues With Children's Learning in Primary Science in England'.* Manchester and London: The University of Manchester and The Ogden Trust. Available at https://documents.manchester.ac.uk/display.aspx?DocID=57599

(accessed day month year).

Bound, J., Jaeger, D.A. and Baker, R.M. (1995) 'Problems With Instrumental Variables Estimation When the Correlation Between the Instruments and the Endogenous Explanatory Variable is Weak'. *Journal of the American Statistical Association*, 90: 430, 443–450. https://doi.org/10.1080/01621459.1995.10476536

Data Protection Act.2018a. c12., art.6 Available at: https://gdpr-info.eu/art-6-gdpr/

Data Protection Act. 2018b. c.12., art.9. Available at: https://gdpr-info.eu/art-9-gdpr/

Department for Education. (2013) '*The National Curriculum in England: Key Stages 1 and 2 Framework Document*'. London: Department for Education. Available at: https://assets.publishing.service.gov.uk/media/5a81a9abe5274a2e8ab55319/PRIMARY_national_curriculum.pdf (accessed 7 March 2024).

Dichaba, M.M. and Mokhele, M.L. (2012) 'Does the Cascade Model Work For Teacher Training? Analysis of Teachers' Experiences'. *International Journal of Educational Sciences*, 4: 3, 249–254. https://doi.org/10.1080/09751122.2012.11890049

Dunn, G., Maracy, M. and Tomenson, B. (2005) 'Estimating Treatment Effects From Randomized Clinical Trials With Noncompliance and Loss to Follow-Up: The Role of Instrumental Variable Methods'. *Statistical Methods in Medical Research*, 14: 4, 369–395. https://doi.org/10.1191/0962280205sm403oa

Education Endowment Foundation (EEF). (2022) *'Statistical Analysis Guidance For EEF Evaluations'.* London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/evaluation-guidance-and-resources/evaluation-design (accessed 15 February 2024).

Education Endowment Foundation (EEF). 2023. *Cost Evaluation Guidance for EEF Evaluations*. Education Endownment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1684358910

Fairhurst, C. 2023. *Thinking, Doing, Talking Science (Second Re-Grant – A Two Armed, Cluster Randomised trial): Statistical Analysis Plan.* Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/TDTS_SAP-for-publication.pdf?v=1737043051

FitzPatrick, B. and Schultz, H. (2015) 'Do Curriculum Outcomes and Assessment Activities in Science Encourage Higher Order Thinking?' *Canadian Journal of Science, Mathematics and Technology Education,* 15: 2, 136–154 https://doi.org/10.1080/14926156.2015.1014074

Gask, L., Coupe, N. and Green, G. (2019) 'An Evaluation of the Implementation of Cascade Training For Suicide Prevention During the 'Choose Life' Initiative in Scotland – Utilizing Normalization Process Theory'. *BMC Health Services Research*, 19: 1, 588. https://doi.org/10.1186/s12913-019-4398-1

Hanley, P., Slavin, R. and Elliott, L. (2015) '*Thinking, Doing, Talking Science: Evaluation Report and Executive Summary'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/thinking-doing-talking-science (accessed 7 March 2024).

Hanley, P., Elliott, L., Fountain, I., Roche, J., Mandefield, L., Fairhurst, C. and Robinson-Smitj, L. 2023. *Thinking, Doing, Talking Science (Second Re-Grant – A Two Armed, Cluster Randomised trial): Protocol*. Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/TDTS-protocol-v1.1-update-for-publication.pdf?v=1737043051

Harlen, W. and Qualter, A. (2018) '*The Teaching of Science in Primary Schools (7th Edition)*'. London: David Fulton Publishers.

Hayes, D. (2000) 'Cascade Training and Teachers' Professional Development'. *ELT Journal*, 54: 2, 135–145. https://doi.org/10.1093/elt/54.2.135

Jarvis, T. and Pell, A. (2004) Primary Teachers' Changing Attitudes and Cognition During a Two-Year Science In-Service Programme and Their Effect on Pupils'. *International Journal of Science Education*, *26*: 14, 1787–1811. https://doi.org/10.1080/0950069042000243763

Jensen, J. L., McDaniel, M.A., Woodard, S.M., and Kummer, T.A. (2014) 'Teaching to the Test…Or Testing to Teach: Exams Requiring Higher Order Thinking Skills Encourage Greater Conceptual Understanding'. *Educ Psychol Rev,* 26: 307–329. https://doi.org/10.1007/s10648-013-9248-9

Joshi, K., Burke, L., Elliott, L., Fountain, I., Hanley, P., Haynes, A., Keding, A., Parvin, J. and Waller, N. (2022) '*Development of a Year 5 Science Assessment For the EEF'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluation/eef-outcome-measures-and-databases/year-5-science-assessment (accessed 3 July 2024).

Kind, P., Jones, K. and Barmby, P. (2007). Developing Attitudes Towards Science Measures. *International Journal of Science Education*, 29 (7), 871-893. https://doi.org/10.1080/09500690600909091

Kitmitto, S., González, R., Mezzanote, J. and Chen, Y. (2018) '*Thinking, Doing, Talking Science: Evaluation Report and Executive Summary'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/thinking-doing-talking-science-effectiveness-trial (accessed 7 March 2024).

Levin, H.M., McEwan, P.J., Belfield, C., Bowden, A.B. and Shand, R. (2017) '*Economic Evaluation in Education: Cost-Effectiveness and Benefit-Cost Analysis (Third Edition)'*. London: SAGE Publications.

Kennedy, J.P., Quinn, F., and Taylor, N. (2016) 'The School Science Attitude Survey: A New Instrument For Measuring Attitudes Towards School Science'. *International Journal of Research & Method in Education*, 39: 4, 422–445. https://doi.org/10.1080/1743727X.2016.1160046

Markwick, A. and Reiss, M.J. (2023) 'Professional Learning in Primary Science: Developing Teacher Confidence to Improve the Leadership of Teaching and Learning'. *International Journal of Science Education*, *46*: 13, 1339–1359. https://doi.org/10.1080/09500693.2023.2288660

Mao,P., Zhihui, C., Jinbo, H., Xinjie, C., and Xitao, F. (2021) The Relationship Between Attitude Toward Science and Academic Achievement in Science: A Three-Level Meta-Analysis'. *Frontiers in Psychology,* 12. https://doi.org/10.3389/fpsyg.2021.784068

Parliamentary Office for Science and Technology (2003). Primary Science. Postnote 202. Available at: https://post.parliament.uk/research-briefings/post-pn-202/

Ofsted. (2021) *'Research Review Series: Science'.* GOV.UK. https://www.gov.uk/government/publications/research-review-series-science/research-review-series-science#print-or-save-to-pdf (accessed 20 July 2024).

Ofsted. (2023) *'Finding the Optinmum: The Science Subject Report'.* GOV.UK. Available at: https://www.gov.uk/government/publications/subject-report-series-science/finding-the-optimum-the-science-subject-report--2#:~:text=In%202019%2C%2083%25%20of%20Year,the%20expected%20standard%20in%20science.&text=However%2C%20we%20are%20concerned%20that,af (accessed 18 July 2024).

Ray, M.L., Wilson, M.M., Wandersman, A., Meyers, D.C. and Katz, J. (2012) 'Using a Training-of-Trainers Approach and Proactive Technical Assistance to Bring Evidence Based Programs to Scale: An Operationalization of the

Interactive Systems Framework's Support System'. *American Journal of Community Psychology*, 50: 3–4, 415–427. https://doi.org/10.1007/s10464-012-9526-6

Roy, P., Rutt, S., Easton, C., Sims, D., Bradshaw, S. and McNamara, S. (2019) '*Stop and Think: Learning Counterintuitive Concepts. Evaluation Report'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/learning-counterintuitive-concepts (accessed 8 March 2024).

Rubin, D.B. (1987) '*Multiple Imputation for Nonresponse in Surveys'*. New York: John Wiley & Sons, Inc. https://doi.org/10.1002/9780470316696.fmatter

Saghaei, M. and Saghaei, S. (2011) 'Implementation of an Open-Source Customizable Minimization Program for Allocation of Patients to Parallel Groups in Clinical Trials'. *Journal of Biomedical Science and Engineering*, 4: 11, 734–739. https://doi.org/10.4236/jbise.2011.411090

Senn, S. (1994) 'Testing For Baseline Balance in Clinical Trials'. *Statistics in Medicine*, 13: 17, 1715–1726. https://doi.org/10.1002/sim.4780131703

Slavin, R.E., Lake, C., Hanley, P. and Thurston, A. (2014) 'Experimental Evaluations of Elementary Science Programs: A Best-Evidence Synthesis'. *Journal of Research in Science Teaching*, 51: 7, 870–901. https://doi.org/10.1002/tea.21139

Standard and Testing Agency. (2019) *'Key Stage 2 Science Sampling 2018: Methodology Note and Outcomes'.* Coventry: Standards & Testing Agency. https://assets.publishing.service.gov.uk/media/5d307b6ee5274a14f6df1b0c/Key_stage_2_science_sampling_2018_methodology_note_and_outcomes.pdf (accessed 20 July 2024).

Standley, E., Carr, R.M., Hanley, P., Leggett, H., Fountain, I., Parkinson, G., Fairhurst, C. and Robinson-Smith, L. (2023) '*Thinking, Doing, Talking Science: Pre-Trial Report Train-the-Trainers Phase'*. London: Education Endowment Foundation. Available at: https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/thinking-doing-talking-science-effectiveness-trial-2 (accessed 7 March 2024).

Stock, J.H., Wright, J.H. and Yogo, M. (2002) 'A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments'. *Journal of Business & Economic Statistics*, 20: 4, 518–529. https://doi.org/10.1198/073500102288618658

Turner, F., Brownhill, S. and Wilson, E. (2017) 'The Transfer of Content Knowledge in a Cascade Model of Professional Development'. *Teacher Development*, 21: 2, 175–191. https://doi.org/10.1080/13664530.2016.1205508

Wedell, M. (2005) 'Cascading Training Down into the Classroom: The Need For Parallel Planning'. *International Journal of Educational Development*, 25: 6, 637–651. https://doi.org/10.1016/j.ijedudev.2005.03.004

Xiao Z., Kasim, A. and Higgins, S.E. (2016) Same Difference? Understanding Variation in the Estimation of Effect Sizes from Educational Trials. *International Journal of Educational Research,* 77: 1-14 http://dx.doi.org/10.1016/j.ijer.2016.02.001

Appendix A: EEF cost rating

Figure 12: Cost rating

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

OUTCOME: *SCIENCE ATTAINMENT AT THE END OF YEAR 5*

Please use this template to assign a separate security rating for each primary outcome. Secondary outcome analysis and/or subgroup analyses are NOT included in the security ratings unless otherwise stated.

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	MDES	Attrition					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	4				4
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity -0		
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Risk rating	Comments
Threat 1: Confounding	Low	This was designed as an RCT. Imbalance of 0.09 SD observed in baseline scores and controlled for in regression model. The statistical analysis plan had pre-specified that the baseline EYFSP score would be adjusted for as a covariate in the primary analysis.

Threat 2: Concurrent Interventions	Low	Concurrent interventions investigated in both control and treatment groups and no evidence of differential impacts.
Threat 3: Experimental effects	Low	The proportion of teachers reporting that they had changed their practice in the control group was similar at baseline (13% agreed strongly; 29% agreed slightly) and follow-up (7% agreed strongly; 28% agreed slightly), suggesting this is 'usual practice' for primary schools in relation to science teaching. This is supported by the fact that around 30% of all schools reported that they had been involved in other science related training or initiatives during academic year 2022-2023; this figure was similar between randomised groups. The change in lesson time from baseline to follow-up for groups is considered insignificant (2 minutes intervention; 4 minutes control – 2 minutes difference) especially given that the length of a science lesson is not part of TDTS.
Threat 4: Implementation fidelity	Low	High implementation fidelity at training and delivery.
Threat 5: Missing Data	Low	Missing data is moderate but identical percentages between groups and accounted for in MI models and MI shows similar results to original primary models.
Threat 6: Measurement of Outcomes	Moderate	Some limitations discussed with unavailability of key stage assessment (KS1 English-Reading and Mathematics) and the need to use an alternate measure (EYFSP results). To score as "low", the descriptor is "thoroughly justified in relation to reliability, validity, utility and acceptably with target population" whereas it appears here, a compromise has had to be accepted for this trial, and it was not the optimal outcome.
Threat 7: Selective reporting	Low	This study was registered, and the analytical approach was identified before outcomes were observed.

- **Initial padlock score:** 4 padlocks
  - Design: Randomised design
  - MDES at randomisation: 0.16
  - Attrition (pupil level) at analysis = 15.9% → -1 padlock
- **Reason for adjustment for threats to validity**:
  - One moderate threat → -0 padlock (No adjustment made)
- **Final padlock score (initial score adjusted for threats to validity):** 4 padlocks

Appendix C: Effect size estimation

Appendix Table 1: 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**
Primary: Cohort1 Year 5 Science Assessment	0.45 (0.06, 0.84)	0.17 (-0.58, 0.92	3,527 (263)	64.60	3,258 (186)	61.26	63.27
Secondary: Cohort 1 Year 5 Science Attitudes questionnaire	0.15 (0.10, 0.20)	0.12 (0.01, 0.24)	3,320 (470)	0.96	3,052 (419)	1.03	1.00

Appendix D: Usual practice survey response tables

Appendix D, Table 1: Teachers confidence in teaching science at baseline and follow-up, by allocated group.

	Baseline			Follow-up		
How confident do you feel doing the following when teaching science to you Yr5 class?	**Intervention N=115 n (%)**	**Control N=83 n (%)**	**Total N=198 n (%)**	**Intervention N=76 n (%)**	**Control N=75 n (%)**	**Total N=151 n (%)**
Inspiring pupils to learn science						
Very confident	26 (22.6)	27 (32.5)	53 (26.8)	37 (48.7)	14 (18.7)	51 (33.8)
Quite confident	79 (68.7)	49 (59.0)	128 (64.6)	35 (46.1)	51 (68.0)	86 (57.0)
Not very confident	8 (7.0)	5 (6.0)	13 (6.6)	2 (2.6)	7 (9.3)	9 (6.0)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Planning science lessons						
Very confident	14 (12.2)	16 (19.3)	30 (15.2)	31 (40.8)	15 (20.0)	46 (30.5)
Quite confident	84 (73.0)	57 (68.7)	141 (71.2)	40 (52.6)	53 (70.7)	93 (61.6)
Not very confident	15 (13.0)	8 (9.6)	23 (11.6)	3 (3.9)	4 (5.3)	7 (4.6)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Explaining science concepts or principles by doing science practicals						
Very confident	15 (13.0)	12 (14.5)	27 (13.6)	30 (39.5)	15 (20.0)	45 (29.8)
Quite confident	81 (70.4)	61 (73.5)	142 (71.7)	44 (57.9)	49 (65.3)	93 (61.6)
Not very confident	17 (14.8)	8 (9.6)	25 (12.6)	0 (0.0)	8 (10.7)	8 (5.3)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Providing challenging tasks for the highest achieving pupils						
Very confident	9 (7.8)	7 (8.4)	16 (8.1)	11 (14.5)	8 (10.7)	19 (12.6)
Quite confident	41 (35.7)	42 (50.6)	83 (41.9)	47 (61.8)	37 (49.3)	84 (55.6)
Not very confident	60 (52.2)	28 (33.7)	88 (44.4)	16 (21.1)	26 (34.7)	42 (27.8)
Not at all confident	3 (2.6)	4 (4.8)	7 (3.5)	0 (0.0)	1 (1.3)	1 (0.7)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Adapting my teaching to engage pupils' interest						
Very confident	14 (12.2)	13 (15.7)	27 (13.6)	25 (32.9)	13 (17.3)	38 (25.2)
Quite confident	83 (72.2)	62 (74.7)	145 (73.2)	48 (63.2)	51 (68.0)	99 (65.6)
Not very confident	16 (13.9)	6 (7.2)	22 (11.1)	1 (1.3)	8 (10.7)	9 (6.0)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Assessing pupil learning in science						
Very confident	7 (6.1)	14 (16.9)	21 (10.6)	16 (21.1)	13 (17.3)	29 (19.2)
Quite confident	64 (55.7)	39 (47.0)	103 (52.0)	47 (61.8)	43 (57.3)	90 (59.6)
Not very confident	38 (33.0)	27 (32.5)	65 (32.8)	11 (14.5)	15 (20.0)	26 (17.2)
Not at all confident	4 (3.5)	1 (1.2)	5 (2.5)	0 (0.0)	1 (1.3)	1 (0.7)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Improving the understanding of struggling pupils						
Very confident	10 (8.7)	9 (10.8)	19 (9.6)	18 (23.7)	11 (14.7)	29 (19.2)
Quite confident	65 (56.5)	48 (57.8)	113 (57.1)	44 (57.9)	47 (62.7)	91 (60.3)
Not very confident	37 (32.2)	24 (28.9)	61 (30.8)	12 (15.8)	13 (17.3)	25 (16.6)
Not at all confident	1 (0.9)	0 (0.0)	1 (0.5)	0 (0.0)	1 (1.3)	1 (0.7)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Helping pupils discuss scientific ideas						
Very confident	17 (14.8)	16 (19.3)	33 (16.7)	28 (36.8)	18 (24.0)	46 (30.5)

Quite confident	77 (67.0)	60 (72.3)	137 (69.2)	43 (56.6)	47 (62.7)	90 (59.6)
Not very confident	19 (16.5)	5 (6.0)	24 (12.1)	3 (3.9)	7 (9.3)	10 (6.6)

Appendix D, Table 2: Strategies and techniques used in science lessons at baseline and follow-up, by allocated group.

	Baseline			**Follow-up**		
How often do you do the following in Yr5 science lessons?	**Intervention N=115**	**Control N=83**	**Total N=198**	**Intervention N=76**	**Control N=75**	**Total N=151**
Teacher demonstration						
Very often	22 (19.1)	22 (26.5)	44 (22.2)	11 (14.5)	10 (13.3)	21 (13.9)
Quite often	63 (54.8)	50 (60.2)	113 (57.1)	44 (57.9)	45 (60.0)	89 (58.9)
Not very often	22 (19.1)	6 (7.2)	28 (14.1)	12 (15.8)	14 (18.7)	26 (17.2)
Rarely	6 (5.2)	3 (3.6)	9 (4.5)	7 (9.2)	3 (4.0)	10 (6.6)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Pupil discussion in whole class						
Very often	78 (67.8)	65 (78.3)	143 (72.2)	64 (84.2)	61 (81.3)	125 (82.8)
Quite often	34 (29.6)	16 (19.3)	50 (25.3)	10 (13.2)	11 (14.7)	21 (13.9)
Not very often	1 (0.9)	0 (0.0)	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Pair or small group discussion						
Very often	65 (56.5)	63 (75.9)	128 (64.6)	60 (78.9)	53 (70.7)	113 (74.8)
Quite often	44 (38.3)	16 (19.3)	60 (30.3)	14 (18.4)	17 (22.7)	31 (20.5)
Not very often	3 (2.6)	2 (2.4)	5 (2.5)	0 (0.0)	2 (2.7)	2 (1.3)
Rarely	1 (0.9)	0 (0.0)	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Give pupils time to think						
Very often	58 (50.4)	60 (72.3)	118 (59.6)	57 (75.0)	52 (69.3)	109 (72.2)
Quite often	44 (38.3)	21 (25.3)	65 (32.8)	17 (22.4)	19 (25.3)	36 (23.8)
Not very often	9 (7.8)	0 (0.0)	9 (4.5)	0 (0.0)	1 (1.3)	1 (0.7)
Rarely	2 (1.7)	0 (0.0)	2 (1.0)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Teach scientific facts						
Very often	62 (53.9)	52 (62.7)	114 (57.6)	38 (50.0)	53 (70.7)	91 (60.3)
Quite often	47 (40.9)	25 (30.1)	72 (36.4)	34 (44.7)	19 (25.3)	53 (35.1)
Not very often	3 (2.6)	4 (4.8)	7 (3.5)	2 (2.6)	0 (0.0)	2 (1.3)
Rarely	1 (0.9)	0 (0.0)	1 (0.5)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Teach scientific terminology						
Very often	65 (56.5)	56 (67.5)	121 (61.1)	47 (61.8)	50 (66.7)	97 (64.2)
Quite often	45 (39.1)	24 (28.9)	69 (34.8)	26 (34.2)	22 (29.3)	48 (31.8)
Not very often	3 (2.6)	1 (1.2)	4 (2.0)	1 (1.3)	0 (0.0)	1 (0.7)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Ask pupils to solve scientific problems						
Very often	13 (11.3)	10 (12.0)	23 (11.6)	12 (15.8)	9 (12.0)	21 (13.9)
Quite often	58 (50.4)	43 (51.8)	101 (51.0)	48 (63.2)	38 (50.7)	86 (57.0)
Not very often	38 (33.0)	23 (27.7)	61 (30.8)	14 (18.4)	23 (30.7)	37 (24.5)
Rarely	4 (3.5)	5 (6.0)	9 (4.5)	0 (0.0)	2 (2.7)	2 (1.3)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Require pupils to write up the whole investigation						

Very often	5 (4.3)	3 (3.6)	8 (4.0)	0 (0.0)	3 (4.0)	3 (2.0)
Quite often	22 (19.1)	13 (15.7)	35 (17.7)	2 (2.6)	16 (21.3)	18 (11.9)
Not very often	39 (33.9)	41 (49.4)	80 (40.4)	19 (25.0)	31 (41.3)	50 (33.1)
Rarely	24 (20.9)	18 (21.7)	42 (21.2)	15 (19.7)	12 (16.0)	27 (17.9)
Never	23 (20.0)	6 (7.2)	29 (14.6)	38 (50.0)	10 (13.3)	48 (31.8)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Require pupils to memorise facts and principles						
Very often	10 (8.7)	10 (12.0)	20 (10.1)	4 (5.3)	8 (10.7)	12 (7.9)
Quite often	42 (36.5)	34 (41.0)	76 (38.4)	24 (31.6)	31 (41.3)	55 (36.4)
Not very often	44 (38.3)	30 (36.1)	74 (37.4)	27 (35.5)	25 (33.3)	52 (34.4)
Rarely	9 (7.8)	6 (7.2)	15 (7.6)	11 (14.5)	7 (9.3)	18 (11.9)
Never	8 (7.0)	1 (1.2)	9 (4.5)	8 (10.5)	1 (1.3)	9 (6.0)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)

Appendix D, Table 3: Frequency of practical sciences at baseline and follow-up, by allocated group.

	Baseline			**Follow-up**		
How often do you do the following in Yr5 science lessons?	**Intervention N=115 n (%)**	**Control N=83 n (%)**	**Total N=198 n (%)**	**Intervention N=76 n (%)**	**Control N=75 n (%)**	**Total N=151 n (%)**
Pupil practical work						
Very often	35 (30.4)	27 (32.5)	62 (31.3)	31 (40.8)	15 (20.0)	46 (30.5)
Quite often	67 (58.3)	41 (49.4)	108 (54.5)	41 (53.9)	48 (64.0)	89 (58.9)
Not very often	11 (9.6)	10 (12.0)	21 (10.6)	2 (2.6)	8 (10.7)	10 (6.6)
Rarely	0 (0.0)	1 (1.2)	1 (0.5)	0 (0.0)	1 (1.3)	1 (0.7)
Never	0 (0.0)	2 (2.4)	2 (1.0)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Design or plan practical investigations						
Very often	11 (9.6)	9 (10.8)	20 (10.1)	13 (17.1)	4 (5.3)	17 (11.3)
Quite often	60 (52.2)	42 (50.6)	102 (51.5)	47 (61.8)	47 (62.7)	94 (62.3)
Not very often	36 (31.3)	23 (27.7)	59 (29.8)	14 (18.4)	16 (21.3)	30 (19.9)
Rarely	6 (5.2)	7 (8.4)	13 (6.6)	0 (0.0)	5 (6.7)	5 (3.3)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Interpret results from practical investigations						
Very often	10 (8.7)	10 (12.0)	20 (10.1)	11 (14.5)	9 (12.0)	20 (13.2)
Quite often	64 (55.7)	44 (53.0)	108 (54.5)	45 (59.2)	42 (56.0)	87 (57.6)
Not very often	30 (26.1)	20 (24.1)	50 (25.3)	17 (22.4)	17 (22.7)	34 (22.5)
Rarely	9 (7.8)	7 (8.4)	16 (8.1)	1 (1.3)	3 (4.0)	4 (2.6)
Never	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.3)	1 (0.7)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)
Use results from practical investigations to support conclusions						
Very often	12 (10.4)	13 (15.7)	25 (12.6)	13 (17.1)	9 (12.0)	22 (14.6)
Quite often	73 (63.5)	41 (49.4)	114 (57.6)	46 (60.5)	44 (58.7)	90 (59.6)
Not very often	23 (20.0)	21 (25.3)	44 (22.2)	15 (19.7)	17 (22.7)	32 (21.2)
Rarely	5 (4.3)	6 (7.2)	11 (5.6)	0 (0.0)	1 (1.3)	1 (0.7)
Never	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.3)	1 (0.7)
Missing	2 (1.7)	2 (2.4)	4 (2.0)	2 (2.6)	3 (4.0)	5 (3.3)

Appendix D, Table 4: How interested and engaged are pupils in learning science at baseline and follow-up, by allocated group.

	Baseline			Follow-up		
Reflecting on your own experience of teaching your Y5 class so far this year, how much would you agree or disagree with the following statements?	Intervention N=115 n (%)	Control N=83 n (%)	Total N=198 n (%)	Intervention N=76 n (%)	Control N=75 n (%)	Total N=151 n (%)
My pupils have enjoyed their science lessons						
Agree strongly	56 (48.7)	48 (57.8)	104 (52.5)	57 (75.0)	33 (44.0)	90 (59.6)
Agree slightly	54 (47.0)	33 (39.8)	87 (43.9)	16 (21.1)	35 (46.7)	51 (33.8)
Neither agree nor disagree	2 (1.7)	0 (0.0)	2 (1.0)	0 (0.0)	4 (5.3)	4 (2.6)
Missing	3 (2.6)	2 (2.4)	5 (2.5)	3 (3.9)	3 (4.0)	6 (4.0)
My pupils have made good progress in their understanding of science						
Agree strongly	29 (25.2)	29 (34.9)	58 (29.3)	38 (50.0)	24 (32.0)	62 (41.1)
Agree slightly	63 (54.8)	49 (59.0)	112 (56.6)	34 (44.7)	47 (62.7)	81 (53.6)
Neither agree nor disagree	20 (17.4)	3 (3.6)	23 (11.6)	1 (1.3)	1 (1.3)	2 (1.3)
Missing	3 (2.6)	2 (2.4)	5 (2.5)	3 (3.9)	3 (4.0)	6 (4.0)
My pupils are confident in science						
Agree strongly	13 (11.3)	11 (13.3)	24 (12.1)	28 (36.8)	11 (14.7)	39 (25.8)
Agree slightly	59 (51.3)	57 (68.7)	116 (58.6)	38 (50.0)	49 (65.3)	87 (57.6)
Neither agree nor disagree	36 (31.3)	11 (13.3)	47 (23.7)	7 (9.2)	12 (16.0)	19 (12.6)
Disagree slightly	4 (3.5)	2 (2.4)	6 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	3 (2.6)	2 (2.4)	5 (2.5)	3 (3.9)	3 (4.0)	6 (4.0)
My pupils can work independently in science						
Agree strongly	10 (8.7)	7 (8.4)	17 (8.6)	22 (28.9)	13 (17.3)	35 (23.2)
Agree slightly	49 (42.6)	47 (56.6)	96 (48.5)	45 (59.2)	40 (53.3)	85 (56.3)
Neither agree nor disagree	32 (27.8)	16 (19.3)	48 (24.2)	5 (6.6)	14 (18.7)	19 (12.6)
Disagree slightly	21 (18.3)	11 (13.3)	32 (16.2)	1 (1.3)	5 (6.7)	6 (4.0)
Missing	3 (2.6)	2 (2.4)	5 (2.5)	3 (3.9)	3 (4.0)	6 (4.0)
My pupils come up with their own scientific ideas						
Agree strongly	6 (5.2)	5 (6.0)	11 (5.6)	17 (22.4)	9 (12.0)	26 (17.2)
Agree slightly	48 (41.7)	39 (47.0)	87 (43.9)	43 (56.6)	36 (48.0)	79 (52.3)
Neither agree nor disagree	33 (28.7)	18 (21.7)	51 (25.8)	11 (14.5)	17 (22.7)	28 (18.5)
Disagree slightly	23 (20.0)	15 (18.1)	38 (19.2)	2 (2.6)	10 (13.3)	12 (7.9)
Disagree strongly	2 (1.7)	4 (4.8)	6 (3.0)	0 (0.0)	0 (0.0)	0 (0.0)
Missing	3 (2.6)	2 (2.4)	5 (2.5)	3 (3.9)	3 (4.0)	6 (4.0)
My pupils do a lot of writing in science						
Agree strongly	6 (5.2)	6 (7.2)	12 (6.1)	1 (1.3)	4 (5.3)	5 (3.3)
Agree slightly	22 (19.1)	18 (21.7)	40 (20.2)	3 (3.9)	19 (25.3)	22 (14.6)
Neither agree nor disagree	29 (25.2)	31 (37.3)	60 (30.3)	20 (26.3)	27 (36.0)	47 (31.1)
Disagree slightly	38 (33.0)	22 (26.5)	60 (30.3)	28 (36.8)	18 (24.0)	46 (30.5)
Disagree strongly	17 (14.8)	4 (4.8)	21 (10.6)	21 (27.6)	4 (5.3)	25 (16.6)
Missing	3 (2.6)	2 (2.4)	5 (2.5)	3 (3.9)	3 (4.0)	6 (4.0)
My pupils have been engaged with science						
Agree strongly	50 (43.5)	42 (50.6)	92 (46.5)	52 (68.4)	25 (33.3)	77 (51.0)
Agree slightly	49 (42.6)	38 (45.8)	87 (43.9)	19 (25.0)	42 (56.0)	61 (40.4)
Neither agree nor disagree	13 (11.3)	1 (1.2)	14 (7.1)	2 (2.6)	5 (6.7)	7 (4.6)
Missing	3 (2.6)	2 (2.4)	5 (2.5)	3 (3.9)	3 (4.0)	6 (4.0)

Appendix D, Table 5: How interested and engaged are teachers in teaching science at baseline and follow-up, by allocated group.

	Baseline			**Follow-up**		
Reflecting on your own experience of teaching your Y5 class so far this year, how much would you agree or disagree with the following statements?	**Intervention N=115 n (%)**	**Control N=83 n (%)**	**Total N=198 n (%)**	**Intervention N=76 n (%)**	**Control N=75 n (%)**	**Total N=151 n (%)**
I have changed the way I teach science						
Agree strongly	22 (19.1)	11 (13.3)	33 (16.7)	41 (53.9)	5 (6.7)	46 (30.5)
Agree slightly	67 (58.3)	24 (28.9)	91 (46.0)	26 (34.2)	21 (28.0)	47 (31.1)
Neither agree nor disagree	18 (15.7)	35 (42.2)	53 (26.8)	6 (7.9)	34 (45.3)	40 (26.5)
Disagree slightly	5 (4.3)	8 (9.6)	13 (6.6)	0 (0.0)	10 (13.3)	10 (6.6)
Disagree strongly	0 (0.0)	3 (3.6)	3 (1.5)	0 (0.0)	2 (2.7)	2 (1.3)
Missing	3 (2.6)	2 (2.4)	5 (2.5)	3 (3.9)	3 (4.0)	6 (4.0)
I have enjoyed teaching science						
Agree strongly	49 (42.6)	37 (44.6)	86 (43.4)	54 (71.1)	21 (28.0)	75 (49.7)
Agree slightly	50 (43.5)	34 (41.0)	84 (42.4)	17 (22.4)	39 (52.0)	56 (37.1)
Neither agree nor disagree	11 (9.6)	8 (9.6)	19 (9.6)	2 (2.6)	8 (10.7)	10 (6.6)
Disagree slightly	2 (1.7)	2 (2.4)	4 (2.0)	0 (0.0)	4 (5.3)	4 (2.6)
Missing	3 (2.6)	2 (2.4)	5 (2.5)	3 (3.9)	3 (4.0)	6 (4.0)

Appendix E: Revised logic model

Goals: Change teachers' pedagogy in primary science lessons, to maximise pupils' higher order thinking. Teachers will actively encourage their pupils' higher order thinking, by enabling their pupils to think and talk about scientific concepts through dedicated discussion times and providing pupils with a wide range of inclusive opportunities for creative investigations and problem solving. They will focus pupils' recording so that is there is always time for practical science.

Inputs

- TDTS programme, training materials and access to TDTS website
- Trainers
- Teachers

Activities

Trial Teacher training: Trainers deliver the course to teachers in trial schools. (See * Programme Activities below).4x 1 day sessions (2 Autumn Term & 2 Spring term) + 0.5 day in Summer Term

Programme activities: TDTS strategies are introduced systematically throughout the 4-day programme, exemplified through specific curriculum areas. Teachers are given gap tasks to implement the strategies in their own classrooms between the training days and then reflect upon the outcomes and impact on their

Outputs

- Delivered as planned with fidelity, session goals met
- Attendance at training

Short-term

Medium-term outcomes

TEACHERS
TDTS strategies are embedded in all Year 5 science lessons
- Familiarity with TDTS strategies
- Preparedness and confidence to use in the classroom

PUPILS
- pupils' thinking encouraged
- more dedicated discussion times
- challenging practical work
- clearly focused pupil recording
- increased pupil engagement with science
- improved attitude towards science

Explored through session observations, trainer post-session interviews and surveys; attendance registers; school visits; teacher and pupil surveys

Long-term outcomes

Secondary outcomes (impact):
- Pupils' attitude towards science improves (measured by science attitude questionnaire)

Moderator - Dosage; Teacher pedagogical experience; FSM; Fidelity; Amount of classroom time for science; Trainer experience; Pupils prior attainment; school culture; resources.

Mediator - Pupils engagement with science; Higher Order Thinking; Teacher confidence in teaching science

Further appendices

Please see accompanying document '*Further Appendices'*.

Education Endowment Foundation

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn