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

Addendum report

October 2025

Lyn Robinson-Smith, Katie Whiteside, Imogen Fountain, Tom Davill, Alex Mitchell, David Torgerson, Catherine Hewitt, and Caroline Fairhurst

Funded by

The Education Endowment Foundation (EEF) is an independent charity dedicated to breaking the link between family income and education achievement. We support schools, colleges, and early years settings 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.
- **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 from the government, allowing us to continue our work until at least 2032.

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

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

0207 802 1653

info@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Table of contents

About the evaluator

The project was independently evaluated by a team at the University of York. The second year of the trial was conducted by Lyn Robinson-Smith, David Torgerson, Caroline Fairhurst, Katie Whiteside, Imogen Fountain, Alex Mitchell, and Tom Davill. The co-principal investigators were Dr Lyn-Robinson-Smith (maternity leave from February 2024 to December 2024), Professor David Torgerson (until October 2024), and Professor Catherine Hewitt (from October 2024).

Parts of this work were undertaken in the Office for National Statistics (ONS) Secure Research Service using data from ONS and other owners and does not imply the endorsement of the ONS or other data owners.

Contact details:

Dr Lyn Robinson-Smith
York Trials Unit,
Department of Health Sciences,
University of York,
York YO10 5DD

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.

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

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 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 involves four and a half days of face-to-face delivery to Year 5 primary school 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 embed the TDTScience approaches into their normal science lessons for the whole class and are provided with TDTScience course resources (physical and online) and inexpensive science equipment.

The project was a two-armed, cluster randomised controlled trial, involving 180 primary schools from across England. Schools were randomised to either receive TDTScience or to continue with usual teaching practice. The evaluation of TDTScience involved two cohorts of Year 5 pupils in sequential years: Cohort 1 in 2022–2023; and Cohort 2 in 2023–2024. Analysis of Cohort 2 Year 5 pupils permits exploration of the impact of TDTScience once teachers have had more time to embed TDTScience into their teaching practice, since Cohort 1 delivery. Teachers did not receive any further training on TDTScience during Cohort 2 delivery.

Robinson-Smith *et al.* (2025) reported Year 5 pupils' science attainment and attitudes after one year of the intervention. This addendum presents results for Cohort 2 (recruited in Autumn Term 2023), assessed in June 2024 to July 2024 using the Year 5 Science Assessment and an attitudes survey. It also reports a longitudinal analysis of Cohort 1, now in Year 6, using the Year 6 Science Assessment and Key Stage 2 Reading and Maths scores. This evaluation was co-funded by the Education Endowment Foundation (EEF) and the Wellcome Trust, and undertaken by the York Trials Unit, University of York.

Table 1: Key conclusions

Key conclusions	
1.	The impact of TDTScience on Cohort 2 Year 5 pupils' science attainment was consistent with the main trial results (Robinson-Smith *et al.,* 2025). When teachers had an additional year of experience delivering TDTScience, Cohort 2 Year 5 pupils made no additional month's progress in science attainment, on average, compared to pupils in control schools. This finding was also consistent among pupils eligible for free school meals (FSM). These results do not have a security rating.
2.	The impact of TDTScience on Cohort 2 Year 5 pupils' science attitudes was consistent with the main trial results (Robinson-Smith *et al.,* 2025). Cohort 2 Year 5 pupils demonstrated small positive effects in attitudes to science, in TDTScience schools, compared to pupils in control schools.
3.	Considering the longer-term impact of TDTScience, Year 6 pupils from Cohort 1, made no additional month's progress in science attainment, and in Key Stage 2 English and Maths SATs, on average, compared to pupils in control schools.
4.	Year 5 teachers in TDTScience schools reported that they were more confident in teaching science, and reported higher levels of engagement in, and enjoyment of science among their pupils, in comparison to the control group. Year 6 teachers in TDTScience schools also reported that their pupils from Cohort 1 were more engaged, progressive, and confident in science, and enjoyed science lessons more in comparison to pupils in control schools in Year 6.
5.	Teacher surveys indicated that many aspects of the TDTScience approach continued to be implemented frequently within Year 5 science lessons by teachers in many intervention schools. Teaching practice in Year 6 science lessons also differed, with intervention schools adopting aspects of TDTScience techniques and strategies.

EEF security rating

The addendum results do not have a security rating, as the EEF only apply security ratings to the primary outcome of the trial, which is included in the main report by Robinson-Smith *et al.* (2025).

Additional findings

Cohort 2 Year 5 pupils in TDTScience schools, made no additional month's progress, on average, compared to those in the control group. This is our best estimate of impact. As with any study, there is always some uncertainty around the result: the possible impact of this programme also includes negative effects of one month's less progress and positive effects of up to two months' additional progress. Nevertheless, Cohort 2 Year 5 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 small in size but was corroborated within survey data where teachers in intervention schools also considered that their Year 5 pupils enjoyed science more, in comparison to pupils in control schools. This result aligns with the result of the main trial for Cohort 1 Year 5 pupils, in the first year of the trial (Robinson-Smith *et al.,* 2025).

Year 6 pupils made no additional month's progress, on average, in science, maths, or English attainment compared to pupils in control schools. These findings imply that TDTScience does not have a positive or negative longer-term effect on science, maths, or English attainment.

Around a quarter of Year 6 teachers had attended in-person TDTScience training. Anecdotally, we consider that these were Year 5 teachers from the year previously, that had moved into the Year 6 classes. Some respondents to the Year 6 survey also indicated that they had received some cascaded training on TDTScience from trained teachers within their school, although this was not an expectation of teachers who had attended the TDTScience training. Survey data indicated that strategies and techniques used in Year 6 science lessons in intervention schools aligned more to the TDTScience approach—such as more regularly designing/planning and interpreting practical investigations—in comparison to control schools. This is an interesting finding that shows what can happen if the intervention is cascaded.

Teacher survey data indicated that the time spent on Year 5 science lessons was similar across groups, although there was evidence of differences in the frequency of use of some techniques and strategies used during science lessons, with intervention schools using practical work more often than control schools. TDTScience trained Year 5 teachers reported that they continued to implement aspects of the TDTScience approach during science lessons in the second year. The increased confidence in teaching science reported by Year 5 teachers in intervention schools at the end of the first year was sustained at the end of the second year, in comparison to teachers in control schools. Together these data suggest that the continued use of TDTScience approaches by teachers are linked to the positive change in pupils' attitudes towards science.

Cost

The estimated cost of TDTScience is £9,249.64 per school over a three-year period, or £71.71 per pupil 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(s)

Outcome / group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	No. of pupils	P-value	EEF cost rating
Cohort 2 Year 5 Science Assessment	0.03 (-0.07, 0.14)	0	N/A	5,708 (3,048: 2,660)	0.52	£ £ £ £ £
Cohort 2 Year 5 Science Assessment (eligible for FSM)	0.03 (-0.10, 0.16)	0	N/A	1,793 (967: 826)	0.62	N/A
Cohort 1 Year 6 Science Assessment	-0.02 (-0.12, 0.07)	0	N/A	5,717 (2,980: 2,737)	0.64	N/A

N/A, not applicable.

Introduction

Intervention

Thinking, Doing, Talking Science (TDTScience) is a continuing professional development (CPD) programme—and a pedagogical teaching approach—developed by Science Oxford and Oxford Brookes University. TDTScience aims to enable teachers to adapt their pedagogy to plan and confidently teach creative science lessons that overtly encourage their pupils' higher order thinking 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. Following four and a half days of face-to-face training, teachers are provided with course resources (physical and online) and inexpensive science equipment, and encouraged to embed the TDTScience approach into their normal science lessons for the whole class. The original report (Robinson-Smith *et al.,* 2025) provides an in-depth description of the intervention.

A two-armed, cluster randomised controlled trial was conducted to evaluate TDTScience, involving 180 primary schools from across England. This was a two-year, effectiveness trial. Schools were randomised to receive either TDTScience or to continue with usual teaching practice. The first year of the effectiveness trial was conducted in 2022–2023 where the impact of the programme on Year 5 pupils' (Cohort 1) science attainment was measured as a primary outcome, using a Year 5 Science Assessment (Joshi *et al.,* 2022). It found participation in the programme did not result in any additional months' progress in science for Year 5 pupils in the intervention group compared to those in the control group. Science attitudes were assessed as a secondary outcome, using a Science Attitudes Questionnaire. There was improvement in attitudes towards science among Year 5 pupils (Cohort 1) whose schools were allocated to the TDTScience intervention. 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. It found teachers implemented TDTScience as intended, with high compliance and fidelity. Teachers reported that TDTScience improved their lesson planning and that they were more confident in 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 lessons 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. 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 results of the first year of the TDTScience effectiveness trial are fully reported in Robinson-Smith *et al.* (2025).

This addendum report presents the results of the second and final year of the effectiveness trial, which intended to explore the impact of TDTScience once teachers had time to embed TDTScience into their teaching practices for the full year, and to understand if there is a 'soak' effect (e.g. the impact of intervention over a longer period of time) on a second cohort of Year 5 pupils (Cohort 2). It also includes 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.

Evaluation objectives

The second year of the TDTScience trial (2023–2024) aimed to examine the legacy of the TDTScience training and any effects of embedding of the TDTScience practices. The research questions that this addendum report was designed to answer are listed below and align with the evaluation protocol (Hanley *et al.,* 2023) and Statistical Analysis Plan (SAP) (Fairhurst, 2023).

Impact evaluation – research questions

Cohort 1 (2022–2024)

4. What is the long-term impact of the TDTScience programme, in comparison to usual Year 5 provision, on pupils' science attainment at the end of Year 6 and on Key Stage 2 outcomes (Year 6 SATs attainment in Reading and Maths)? *[secondary outcomes]* (Fairhurst, 2023)

Cohort 2 (2023–2024)

5. What is the impact of the TDTScience programme, in comparison to usual Year 5 provision, on the science attainment of Year 5 pupils given the mix of experienced and inexperienced teachers in the intervention group? *[secondary outcome]* (Fairhurst, 2023)

6. What is the impact of the TDTScience programme on pupils' attitudes towards science, in comparison to usual Year 5 provision, given the mix of experienced and inexperienced teachers in the intervention group? *[secondary outcome]* (Fairhurst, 2023)

7. 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 given the mix of experienced and inexperienced teachers in the intervention group? *[subgroup analysis]* (Fairhurst, 2023)

Implementation and process evaluation (IPE) – research questions

1. To what extent was TDTScience implemented as planned? (Hanley *et al.,* 2023)

    b. Classroom practice.

2. What processes are involved for teachers and schools implementing TDTScience—what are the main facilitators and barriers? (Hanley *et al.,* 2023)

3. What are the perceptions of teachers as regards TDTScience? (Hanley *et al.,* 2023)

    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?

5. How does TDTScience compare with practice in business-as-usual science lessons? (Hanley *et al.,* 2023)

    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?

Project team

Evaluation team

The evaluation team were from the York Trials Unit, University of York and included:

- **Dr Lyn Robinson-Smith (Co-Principal Investigator).** 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 Education Endowment Foundation (EEF). 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. On maternity leave from February 2024 to December 2024.

- **Professor Catherine Hewitt (Co-Principal Investigator).** A director of the York Trials Unit and was co-principal investigator on the trial from October 2024.

- **Professor David Torgerson (Co-Principal Investigator).** Was the director of the York Trials Unit. Provided consultancy on methodology and design and was co-principal investigator on the trial from August 2022 to October 2024.

- **Dr Pam Hanley (Co-Principal Investigator).** Has an extensive background in education research, including many randomised controlled trials at the University of York. Their experience with the EEF includes other science-related interventions in addition to the TDTScience efficacy trial. 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).** Jointly responsible for the day-to-day management and coordination of the trial and led on the impact evaluation until August 2022. Has been involved in a large number of trials, including several for the EEF as principal investigator. 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.** An experienced trial support officer and contributed to coordinating schools through recruitment, survey completion, and outcome testing.

- **Caroline Fairhurst.** A senior statistician who has worked on many education trials for the EEF. Oversaw all statistical aspects of the trial, conducted the cost evaluation, and contributed to writing the report. On maternity leave from April 2024 to March 2025.

- **Dr Katie Whiteside.** An experienced trial co-ordinator and has worked on a number of randomised controlled trials evaluating education and health care interventions. Undertook data management and trial coordination oversight and duties for the trial from August 2022 as well as contributed towards report writing.

- **Alex Mitchell.** A trial statistician who has worked on several education trials by the EEF. Carried out the statistical analysis and contributed to writing the report.

- **Tom Davill.** A trial support officer for the evaluation from October 2022 to July 2024 and assisted with recruitment and delivering outcome testing.

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.** The director of Education and Engagement for Science Oxford until November 2023 and has spent their career in the informal science learning sector, with a particular focus on working with primary teachers and pupils in science. Jointly developed and led the TDTScience projects (2013–2023) with Helen Wilson, funded by the EEF and others. 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. Began their career as a secondary physics teacher and then moved into primary teaching. Then went into Initial Teacher Education, eventually leading the primary teacher training programmes at Oxford Brookes University. As a primary science consultant, they researched the links between creative, challenging primary science lessons and pupils' attitudes and attainment. Jointly developed and led the TDTScience projects (2013–2023), funded by the EEF.

- **Andrew Kensley.** 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 STEM [science, technology, engineering, and mathematics] Ambassador) and then a primary school teacher.

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 as well as TDTScience-based twilight sessions. Was a TDTScience-trained trainer and project manager (from 2021) for the TDTScience effectiveness trial 2020–2023.

This core team were joined by a number of trainers:

- **Bryony Turford.** Primary Science Geeks.
- **Wendy Precious.** Precious Learning Ltd.
- **Rachael Webb.** Lancashire County Council.
- **Sarah Earle.** Bath Spa University.
- **Alison Trew.** Primary Science Teaching Trust.
- **Allie Beaumont.** Independent consultant.
- **Mandy Hodgkinson.** East Riding of Yorkshire Council.
- **Nicky Waller, Jane Winter, and Joy Parvin.** Centre for Industry Education Collaboration (CIEC), University of York.
- **Mike Dennis.** Independent consultant.

Methods

Impact evaluation design

Table 3 provides an overview of the evaluation design and includes details of the first year (2022–2023) and second year (2023–2024) of the trial.

This was a two-armed, cluster randomised controlled, effectiveness trial with random allocation at the school level. Schools were randomly allocated (one to one) to one of two groups:

- **Intervention.** Schools allocated to receive the TDTScience programme.
- **Control.** Schools allocated to business-as-usual.

Schools were offered a financial incentive of £500 to help reduce attrition after completing the requirements at the end of the second year of the trial (2023–2024). Financial incentives for the first year (2022–2023) of the trial are outlined in the original report (Robinson-Smith *et al.,* 2025).

Table 3: Evaluation design

Trial design, including number of arms		Two-armed, cluster randomised controlled trial <u>Cohort 1</u> Year 5 (main trial): 2022–2023 (reported in Robinson-Smith *et al.,* 2025) Year 6 (longitudinal follow-up): 2023–2024 (participating Year 5 pupils who moved to Year 6 and continued their participation in the trial) <u>Cohort 2 (longitudinal follow-up)</u> Year 5: 2023–2024
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; 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; reported in Robinson-Smith *et al.,* 2025)
	Measure (instrument, scale, source)	Year 5 Science Assessment, 15-item measure scored 0–45, CIEC and the York Trials Unit, University of York
Secondary outcome(s)	Variable(s)	1. Science attitudes (Cohort 1 – reported in Robinson-Smith *et al.,* 2025) 2. Science attitudes (Cohort 2 – reported in this addendum report) 3. Science attainment at Year 6 (Cohort 1 – reported in this addendum report) 4. Science attainment at Year 5 (Cohort 2 – reported in this addendum report) 5. Key Stage 2 Reading and Maths (Cohort 1 – reported in this addendum report)

	Measure(s) (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	Early Years Foundation Stage Profile (EYFSP)
	Measure (instrument, scale, source)	Average point score from 17 Early Learning Goals (ELGs) that make up the EYFSP, scored 1–3, National Pupil Database (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

Participant selection and sample size

The original report (Robinson-Smith *et al.,* 2025) outlined the recruitment of schools.

School eligibility criteria were as follows:

- 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, whereby science is taught across a two-year cycle, as opposed to one; more common in smaller schools with mixed-age classes).
- 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 co-ordinator) would also need to attend the training.
- Schools within a Multi-Academy Trust (MAT) were 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**' trials 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)**' trials as this is also a science CPD for Key Stage 2.

- The school agreed to all requirements outlined in the Information for Schools and Memorandum of Understanding documents (including commitment to keep same Year 5 teachers across the two years wherever possible).

Pupils

All Year 5 pupils within participating schools during the academic year 2022–2023 were able to participate in Cohort 1 of the trial. Likewise, all Year 5 pupils within participating schools during the academic year 2023–2024 were able to participate in Cohort 2 of the trial. At the beginning of the academic years 2022–2023 and 2023–2024, parents/carers of Year 5 pupils 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 Year 5 pupils with the evaluation team.

Addendum outcome measures

Cohort 2 Year 5 EYFSP and Science Assessment

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 January 2025. This baseline measure was chosen as it was used as the baseline measure for the Cohort 1 Year 5 pupils, as per Robinson-Smith *et al*. (2025)as an alternative to the Key Stage 1 English Reading and Maths scores used in the previous effectiveness trial (Robinson-Smith *et al.,* 2025), as Key Stage 1 results were not available for this cohort of pupils in this trial, who would have been in Year 2 during the academic year 2020–2021 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 2018–2019 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).

The same Year 5 Science Assessment that was used for the primary outcome for Cohort 1 at the end of Year 5 in June 2023 to July 2023 (Robinson-Smith *et al*., 2025) was completed by Cohort 2 at the end of Year 5 in June 2024 to July 2024. This measure was recently developed and validated by the CIEC and the York Trials Unit, University of York (Joshi *et al*., 2022) and was designed to reflect the current curriculum, have a mix of question types, and have an emphasis on 'working scientifically'.

The Year 5 Science Assessment has 15 sections, each of which is comprised of one to five questions (i.e. a to 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). 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 assessment administration procedure used by Robinson-Smith *et al*. (2025) was followed. 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 completed relevant safeguarding and data protection training.

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 arranged for the science assessment to be completed in the pupils' usual classroom, with no special seating arrangements. The evaluation team explained that a teacher or teaching assistant should be present during the assessment to assist pupils who may require pastoral care (as the invigilators would be specifically monitoring the ongoing assessment within the class), 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 support staff should not help the pupils to answer the assessment questions.

The evaluation team recruited and trained a team of markers to score the assessments according to a detailed mark scheme. Approximately 5% of the assessments for each school were second marked by a different marker 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 calculated and entered and, to confirm input accuracy, were cross-checked against total marks calculated using software.

Research invigilators and markers were blind to group allocation for each school. Research invigilators were asked to inform the evaluation team if they became unblinded (e.g. the school mentioned that they had received the TDTScience training). There were 13 known instances of unblinding of research invigilators (eight in control schools and five in intervention schools). Please note that unblinding of research invigilators is reported at a school level, not at a class or year-group level, though there may have been cases where unblinding may not have occurred for every class or year group at the school.

Cohort 2 Year 5 Science Attitudes

The same self-reported Science Attitudes Questionnaire that was used as the secondary outcome for Cohort 1 at the end of Year 5 in June 2023 to July 2023 (Robinson-Smith *et al*., 2025) was completed by Cohort 2 at the end of Year 5 in June 2024 to July 2024. This instrument measures interest, self-efficacy, and activity in science lessons and was also used in both the efficacy trial (Hanley, Slavin, and Elliott, 2015) and the previous effectiveness trial of TDTScience (Kitmitto *et al.,* 2018), though in the TDTScience pre-trial four new items were added to strengthen the self-efficacy scale (Standley *et al.,* 2023).

The questionnaire is comprised of 28 items and 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 questionnaire administration procedure used by Robinson-Smith *et al*. (2025) was followed. The 28-item paper questionnaire was completed in class, supervised by class teachers, who were provided with administration guidance by the evaluation team. The questionnaire was teacher-administered, rather than administered by the trained invigilators during visits to complete the science assessments, 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 requested to administer the questionnaire on a day before the science assessment (to prevent recent science assessment completion from potentially biasing pupil responses), where possible, and advised that the completion of the questionnaire should not immediately follow assessment completion. This questionnaire was also teacher-administered within the previous two TDTScience trials. In total 78% (124/160) of schools completed pupil surveys in all Year 5 classes prior to assessment. In the remaining schools they had either not completed them yet, had not completed them yet in some classes (meaning other classes had completed them prior to assessment), or the completed surveys were not given to the invigilator at the time of their visit (as the process for this data collection was that invigilators returned completed pupil surveys with assessment papers).

Responses to the 20 items identified by factor analysis by Standley *et al*. (2023) were summed to generate a total score from 20–100, 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.

Cohort 1 Year 6 Science Assessment

A new Year 6 Science Assessment measure (Baird *et al.,* in press) was completed by Cohort 1 at the end of Year 6 in June 2024 to July 2024 as a secondary outcome. This measure was developed and validated by the York Trials Unit and Manchester Metropolitan University and was designed to reflect the current curriculum, be age-appropriate, have a mix of question types, and an emphasis on 'working scientifically'.

The Year 6 Science Assessment has 12 sections, each of which is comprised of two to five questions (i.e. a to e) and is worth between 2 and 5 marks (two sections are worth 2 marks, three sections are worth 3 marks, five sections are worth 4 marks, and two sections are worth 5 marks). Incomplete sections are given a score of 0. Section scores are summed to produce a total score from 0 to 43, where a higher score indicates a better outcome.

The assessment was administered as a 45-minute written assessment and was invigilated at a class level by independent, blinded research invigilators who were recruited and trained by the evaluation team. The same assessment administration procedure that was used for the Year 5 Science Assessment was followed and schools and teachers were given the same advice regarding assessment completion and who the assessment would be appropriate for.

As with the Year 5 Science Assessment, the evaluation team recruited and trained a team of markers to score the assessments according to a detailed mark scheme. Approximately 5% of the assessments for each school were second marked by a different marker 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 calculated and entered and, to confirm input accuracy, were cross-checked against total marks calculated using software.

Research invigilators and markers were blind to group allocation for each school. Research invigilators were asked to inform the evaluation team if they became unblinded (e.g. the school mentioned that they had received the TDTScience training). There were 13 known instances of unblinding of research invigilators (eight in control schools and five in intervention schools). Please note that unblinding of research invigilators is reported at a school level, not at a class or year-group level, though there may have been cases where unblinding may not have occurred for every class or year group at the school.

Cohort 1 Year 6 English Reading and Maths

The impact on maths and English attainment was assessed for Cohort 1 at the end of Year 6 by linking to pupils' Key Stage 2 results (English Reading and Maths), via the NPD. These are measured via scaled assessment scores, using the variables KS2_READSCORE and KS2_MATSCORE, both scored on a scale from 0–120, where a higher score indicates greater attainment.

For these outcomes, the objective was to check for 'no negative impact', as it would not be necessarily expected that the intervention would have a positive impact. The rationale for this is based on results of analysis of longer-term outcomes carried out by researchers at Durham University at the end of the first effectiveness trial for TDTScience. They investigated the impact on Key Stage 2 Maths and English (average score for two subjects) among the pupils at the end of Year 6 and found a non-significant, but negative effect (-0.07, 95% confidence interval [CI]: -0.23 to 0.09) (unpublished – personal correspondence). Therefore, we interpret the results of the analysis of maths and English in a non-inferiority, rather than a superiority, framework. This means that focus is on the magnitude of the lower limit of the 95% CI for the treatment effect size, rather than the p-value. We want to ensure that the intervention does not have an overly negative impact on these outcomes.

Statistical analysis

Analysis followed the EEF's most recent guidance and is detailed in full in the published SAP (EEF, 2022; Fairhurst, 2023). 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 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.

Characteristics and the EYFSP measure of prior attainment for pupils recruited in Cohort 2 are summarised descriptively by randomised group both as randomised and as included in the analysis of the Year 5 post-test. Also presented is a comparison of baseline characteristics for Cohort 1 pupils included in the analysis for the Year 6 post-test. No formal comparison of the baseline data are undertaken, but the difference in prior attainment (average EYFSP score) between the groups is reported as a Hedges' g effect size with a 95% 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 with their associated 95% CI and p-value. Treatment effects are also presented as (estimated) Hedges' g effect sizes.

Cohort 2 Primary addendum outcome analysis

A linear mixed effects regression model at the pupil level was used to estimate the adjusted mean difference in science assessment scores at the end of Year 5 for Cohort 2. This was a two-level model with pupils nested within schools. 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%) 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.

The distribution of the standardised residuals was checked using a quantile-quantile (Q-Q) plot and shown to be normal.

Subgroup analyses

A subgroup analysis was conducted for the outcome of Year 5 Science Assessment for Cohort 2 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 at the class level, as opposed to the school level. The delivery team kept registers of teachers' 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 was. Teacher changes throughout the year were recorded by the delivery team and passed to the evaluation team.

For Cohort 2, compliance is defined as: The pupil was in a class that was predominantly taught by a teacher who attended at least three out of the four full days of training in the first year of the trial. The predominant teacher, if there was more than one, was defined as the teacher who taught the class for the largest proportion of the academic year based on termly updates from each school.

A Complier Average Causal Effect (CACE) analysis (Dunn, Maracy, and Tomenson, 2005) was conducted for the Year 5 Science Assessment outcome using the dichotomous compliance measure described above. This analysis used a two-stage least squares (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 for the science assessment outcome at the end of Year 5 for Cohort 2, and reasons for missing data is 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 using chained equations including all available pupil- and school-level baseline variables (school: allocation; location; whether rural or urban; type of school; latest Ofsted [Office for Standards in Education, Children's Services and Skills] rating; pupil: gender; EVER6FSM status; EYFSP average point score).

A ‘burn-in’ of ten was used, which means that the first ten iterations of the multiple 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 multiple 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 outcome analysis

The outcomes of science attainment, Key Stage 2 Maths and English at the end of Year 6 for Cohort 1 and standardised scores from the 20-item ‘interest and self-efficacy’ scale of the Science Attitudes Questionnaire at the end of Year 5 for Cohort 2 were analysed as described for the primary outcome. The remaining items of the Science Attitudes Questionnaire, not used in the ‘interest and self-efficacy’ scale, are summarised separately.

Additional analysis

As part of the protocol for the development of the Year 6 Science Assessment test (Hanley *et al.,* 2022), it was proposed to investigate the correlation between test scores from the Year 5 and Year 6 tests. It was decided that this should be conducted as part of the TDTScience trial. Total science assessment test scores from Year 5 pupils in Summer Term 2023 are correlated with the total test scores from Year 6 pupils in Summer Term 2024, presented using Pearson’s Correlation Coefficient and a 95% CI (using the control arm only).

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, $(\overline{Y}_T - \overline{Y}_C)_{adjusted}$ 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, Kasim, and Higgins, 2016).

Estimation of intracluster correlation coefficient (ICC)

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 methods

All teachers of Cohort 1 (Year 6) and Cohort 2 (Year 5) were asked to complete an online survey during June 2024 to July 2024, to gauge usual practice in science lessons, with additional questions, where appropriate, 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 (Table 4).

Table 4: IPE methods overview (example – please adapt as necessary)

Research methods	Data collection methods	Participants / data sources	Data analysis methods	Research questions addressed	Implementation / logic model relevance
Follow-up at the end of the legacy year	Survey	All participating teachers in intervention and control schools; Year 6 teachers (end of second year only; c360)	Frequency counts; descriptive/thematic analysis	1. b 2. 3. a–e 5. a–d	Comparison of pre- and post-intervention; TDTScience and control

Analysis (IPE)

Descriptive statistics of survey data is presented. Qualitative responses were analysed thematically following the stages outlined by Braun and Clarke (2006): detailed familiarisation; generating initial codes; searching for themes; reviewing themes; defining and naming themes; and data 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 and Year 6 science teachers during June 2024 to July 2024 at the end of the second year of the trial, 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 teacher had spent time, during the 2023–2024 school year: i) sharing TDTScience training with other members of staff within their school who did not receive TDTScience training during the school year 2022–2023; ii) receiving TDTScience training from another member of staff at their school; iii) completing other activities related to the TDTScience programme, for example, reviewing resources, accessing the TDTScience website, etc.; and iv) completing any TDTScience training or activities outside of normal working hours, for example, in their own time, and whether the school had arranged cover for the teacher to undertake any of these activities. 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.

There was no external training in the intervention to any new or existing Year 5 teachers in the second year of the trial (2023–2024), so there are no associated costs of teachers attending training to consider.

Analysis

Following the first year of the trial, a cost evaluation was conducted, which took data from surveys administered in June 2023 to July 2023 to Year 5 teachers to estimate the cost to schools of taking part in the trial, including costs to attend training, for staff cover, and to purchase additional resources (Robinson-Smith *et al.,* 2025). 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. This analysis forms an update to the previous cost evaluation as we are able to determine exactly what resources were bought/replenished by schools in the second year, rather than assuming the same recurring cost from the first year. Additionally, while staff cover was assumed to be a one-off, start-up cost in the original cost evaluation, the survey indicated that some schools provided cover in the second year for teachers to share/receive cascade training; therefore, this has been treated as a recurring cost in this updated analysis.

The total time devoted by school personnel to provide or receive cascade training and for intervention delivery/engagement are summarised descriptively. The average cost to the school of the second year of intervention implementation was calculated by summing the constituent parts (cost of teacher cover and purchase of additional resources) and dividing by the number of schools that responded to the survey.

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).

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 provided data. However, for the purposes of these analyses, we assume 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.

- The TDTScience intervention was delivered to Year 5 teachers in 2022–2023 with the understanding that schools would keep the trained teacher in Year 5 in the 2023–2024 school year wherever possible. However, this was not always the case, and the trained teacher may have left or moved to a different year group in 2023–2024. The survey was administered to all Year 5 and Year 6 science teachers at the end of the school year 2023–2024 (we were interested in survey responses from Year 6 teachers to complement the longitudinal outcomes we were collecting from Year 6 pupils who were in Year 5 in the first year of the trial). In order to gain a more comprehensive understanding of the costs to the *school* of taking part for a second year, we have included data from respondents who reported that they were Year 6 teachers as they may have also taught a Year 5 class at some point in the school year and purchased resources to deliver TDTScience activities, or attended face-to-face training in 2022–2023 and shared that knowledge with a Year 5 teacher in 2023–2024, etc. These data are relevant to the cost evaluation. It is possible there are teachers in other school years who engaged in TDTScience-related activities in 2023–2024 who were not invited to complete the survey so their data are not captured; therefore, the total cost may be a slight underestimation.

- For the purpose of calculating the costs over a three-year period, costs for the first year of taking part in the trial are taken from Robinson-Smith *et al*. 2025, the original costs for the second year have been updated based on the more recent survey data, and we assumed that staff cover and the purchase of additional resources would be recurring in the third year. These costs are considered as optional extras given the inconsistent uptake (i.e. use, if at all, will be different) across schools.

- For the purposes of calculating the average cost of paid cover for teachers to complete TDTScience-related activities, we assumed that the cost of staff cover is £18.33 per hour, which is the average hourly rate of a primary school teacher in the UK.

We estimated the average cost per pupil per year for schools receiving the TDTScience programme following the EEF costing guidance issued in 2023 (EEF, 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 (in-person or cascade) 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.

Impact evaluation

Cohort 2 School recruitment and attrition

Of the 168 schools who were recruited for Cohort 1 and returned pupil details for the collection of post-test data in Year 5, 160 schools agreed to take part in Cohort 2 testing (intervention n=86; control n=74). Seven schools agreed to post-test Cohort 1 but not recruit a Cohort 2 (three schools cited reasons related to staff changes, two schools did not give a specific reason, and two schools cited implementing rolling two-year curriculum), while one school refused to both follow-up Cohort 1 in Year 6 and recruit a Cohort 2.

Cohort 2 Pupil recruitment and attrition

For Cohort 2, at the start of the academic year 2023–2024, 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 160 schools (n=6,863 pupils, intervention, n=3,682; control, n=3,181; mean per school 42.9, SD 19.5, median 37.5, range 9 to 116). These pupils form our Cohort 2 sample.

A post-test Year 5 Science Assessment score was available for 6,087 pupils from 160 schools (intervention, n=3,269/3,682, 88.8%; control n=2,818/3,181, 88.6%; mean per school 38.0, SD 18.1, median 34, range 7 to 103). Reasons for data being unavailable for the remaining 776 include: pupil was absent on day of testing (n=482); pupil had left school (n=187); pupil had/was withdrawn (includes pupils could not access the assessment, e.g. due to Special Educational Needs and Disabilities [SEND] or limited English) (n=95); and other/unknown reason (n=12).

Cohort 1 School attrition for Year 6 post-test

Of the 168 schools who were recruited for Cohort 1 and returned pupil details for the collection of post-test data in Year 5, all but one school agreed to provide post-test data in Year 6 (this school withdrew from the Year 6 post-test as they felt they did not have the capacity to give the trial the time or commitment required). One school that returned data was excluded from the analysis 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 ineligible for the trial. Therefore, valid post-test Year 6 data was received from 166 schools (intervention, n=88; control, n=78).

Cohort 1 Pupil attrition for Year 6 post-test

Post-testing with the Year 6 Science Assessment was due to take place for 6,810 pupils across 167 schools (intervention, n=3,568/3,790, 94.1%; control, n=3,242/3,471, 93.4%). Post-testing was not due to take place with 451 pupils as they either left the school (n=291), withdrew from the trial before this was due (n=139), or their school withdrew (n=21).

A valid post-test Year 6 Science Assessment score was available for 5,941 pupils from 166 schools (intervention, n=3,118/3,790, 82.3%; control n=2,823/3,471, 81.3%; mean per school 35.8, SD 17.4, median 29.5, range 12 to 106). Reasons for data being unavailable for the remaining 869 include: pupil was absent on day of testing (n=591); pupil had left school (n=173); pupil had withdrawn (n=41); other/unknown reason (n=34); and school not eligible so data not included (n=30).

Attrition

A total of 5,709 of the 6,863 recruited Cohort 2 pupils (intervention, n=3,049/3,682, 82.8%; control n=2,660/3,181, 83.6%) from 160 schools had valid baseline EYFSP average point score and post-test Year 5 Science Assessment data; one did not have FSM status data and so 5,708 were included in the primary analysis. This equates to an overall attrition rate of 16.8% (Table 5).

Table 5: Cohort 2 pupil-level attrition from the trial

		Intervention	Control	Total
Number of pupils	Recruited	3,682	3,181	6,863
	Analysed	3,048	2,660	5,708
Pupil attrition (from randomisation to analysis)	Number	634	521	1,155
	Percentage	17.2	16.4	16.8

A total of 5,717 of the 7,261 randomised Cohort 1 pupils (intervention, n=2,980/3,790, 78.6%; control n=2,737/3,471, 78.9%) from 166 schools had valid baseline EYFSP average point score, post-test Year 6 Science Assessment data and FSM status and were included in the longitudinal analysis. This equates to an overall attrition rate of 21.3% (Table 6).

Table 6: Cohort 1 pupil-level attrition from the trial (longitudinal analysis)

		Intervention	Control	Total
Number of pupils	Randomised	3,790	3,471	7,261
	Analysed	2,980	2,737	5,717
Pupil attrition (from randomisation to analysis)	Number	810	734	1,544
	Percentage	21.4	21.1	21.3

Pupil and characteristics

Characteristics of the 6,863 pupils recruited in Cohort 2 are presented in Table 7 and are reasonably similar between the randomised groups. Approximately half the pupils were male, with the control group having a slightly higher proportion of males than the intervention group (intervention, n=1,874/3,682, 50.9%; control, n=1,758/3,181, 55.3%). Just under a third of pupils (30.3%) were eligible for FSM (intervention, n=1,105/3,682, 30.0%; control, n=975/3,181, 30.6%), though there was some missing data in this variable (8.0%) so the proportion eligible for FSM out of those with valid data was 32.9% (intervention, n=1,105/(3,682-321), 32.9%; control, n=975/(3,181-228), 33.0%). The average school-level %FSM was 27.0%.

EYFSP data, used as a measure of prior attainment, was obtained from the NPD. NPD data were requested for 6,768 randomised pupils (98.6%; intervention, n=3,627/3,682, 98.5%; control, n=3,141/3,181, 98.7%) since 95 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 6,328 pupils of which 6,321 had valid EYFSP data. This equates to 92.1% of the randomised population (intervention, n=3,369/3,682, 91.5%; control, n=2,952/3,181, 92.8%). The mean EYFSP average point score was slightly higher in the intervention group than the control group (2.06, SD 0.46, compared with 2.04, SD 0.43; Hedges' g effect size between the groups 0.04; 95% CI: -0.01 to 0.09) (Table 7). This represents a small, non-statistically significant difference (i.e. not different from zero).

Characteristics for the 5,708 pupils included in the primary addendum analysis are presented in Table 8 and are similar to the randomised population.

Table 7: Baseline characteristics of recruited Cohort 2 pupils, by group, as randomised

Pupil level (categorical)	Intervention group (n=3,682)	Control group (n=3,181)	
	Count (%)	Count (%)	
Gender			
Male	1,874 (50.9)	1,758 (55.3)	
Female	1,808 (49.1)	1,423 (44.7)	
Eligible for FSM			
Yes	1,105 (30.0)	975 (30.6)	
No	2,256 (61.3)	1,978 (62.2)	
Missing	321 (8.7)	228 (7.2)	
Pupil level (continuous)	Mean (SD)	Mean (SD)	Effect size
EYFSP average point score	2.06 (0.46)	2.04 (0.43)	0.04 (-0.01, 0.09)

Table 8: Baseline characteristics of recruited Cohort 2 pupils, by group, as analysed

Pupil level (categorical)	Intervention group (n=3,048)	Control group (n=2,660)	
	Count (%)	Count (%)	
Gender			
Male	1,556 (51.0)	1,456 (54.7)	
Female	1,492 (49.0)	1,204 (45.3)	
Eligible for FSM			
Yes	966 (31.7)	823 (30.9)	
No	2,082 (68.3)	1,837 (69.1)	
Pupil level (continuous)	Mean (SD)	Mean (SD)	Effect size
EYFSP average point score	2.07 (0.44)	2.05 (0.43)	0.03 (-0.03, 0.08)

Characteristics of the 5,717 Cohort 1 pupils included in the longitudinal analysis of Year 6 Science Assessment score are presented in Table 9 and are reasonably similar between the randomised groups. Approximately half the pupils were male (intervention, n=1,470/2,980, 49.3%; control, n=1,393/2,737, 50.9%) and 29.6% were eligible for FSM (intervention, n=850/2,980, 28.5%; control, n=841/2,737, 30.7%). The mean EYFSP average point score was slightly higher in the intervention group than the control group (2.09, SD 0.45, compared with 2.05, SD 0.43; Hedges' g effect size between the groups 0.07; 95% CI: 0.02 to 0.13).

Table 9: Baseline characteristics of Cohort 1 pupils, by group, as analysed for Year 6 Science Assessment outcome

Pupil level (categorical)	Intervention group (n=2,980)	Control group (n=2,737)	
	Count (%)	Count (%)	
Gender			
Male	1,470 (49.3)	1,393 (50.9)	
Female	1,510 (50.7)	1,344 (49.1)	
Eligible for FSM			
Yes	850 (28.5)	841 (30.7)	
No	2,130 (71.5)	1,896 (69.3)	
Pupil level (continuous)	Mean (SD)	Mean (SD)	Effect size
EYFSP average point score	2.09 (0.45)	2.05 (0.43)	0.07 (0.02, 0.13)

Outcomes and analysis

Cohort 2 Year 5 Science Assessment score

A valid baseline average EYFSP point score was available for 6,321 randomised pupils (intervention, n=3,369; control, n=2,952), from 160 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 Year 5 Science Assessment score was obtained for 6,087 randomised pupils (88.7%; intervention, n=3,269/3,682, 88.8%; control, n=2,818/3,181, 88.6%) from 160 schools (86 intervention, 74 control). A mean of 22.7 (95% CI: 22.4 to 23.0) was observed in the intervention arm and 22.7 (95% CI: 22.4 to 23.0) in the control arm. The unadjusted mean difference is -0.03 (95% CI: -0.43 to 0.37). 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 (**Error! Reference source not found.**). Baseline and post-test scores were available for 5,709 (83.2%) pupils (intervention, n=3,049/3,682, 82.8%; control, n=2,660/3,181, 83.6%). The correlation between the baseline and post-test scores was 0.51 (95% CI: 0.49 to 0.53). As a check of the analysis model assumptions, the normality of the standardised residuals was checked using a Q-Q plot and were shown to be normally distributed (Figure 2).

The adjusted mean difference in post-test score between the intervention and control groups was 0.27 (95% CI: -0.55 to 1.09, Appendix A Table 1). The estimated Hedges' g effect size was 0.03 (95% CI: -0.07 to 0.14, p = 0.52, Table 10), which does not relate to any months' additional progress in the intervention group. The total variance used to calculate the effect size was 63.24; the sum of 55.40 (random variation between pupils, within-cluster variance) and 7.84 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.12 (95% CI: 0.09 to 0.15). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.12 (0.10 to 0.16).

Given the number of pupils included in this analysis (n=5,708), the observed ICC of 0.12 and the correlation between the pre- and post-test scores of 0.51, the estimated MDES at analysis for the Cohort 2 Year 5 Science Assessment outcome was 0.15, with 80% power.

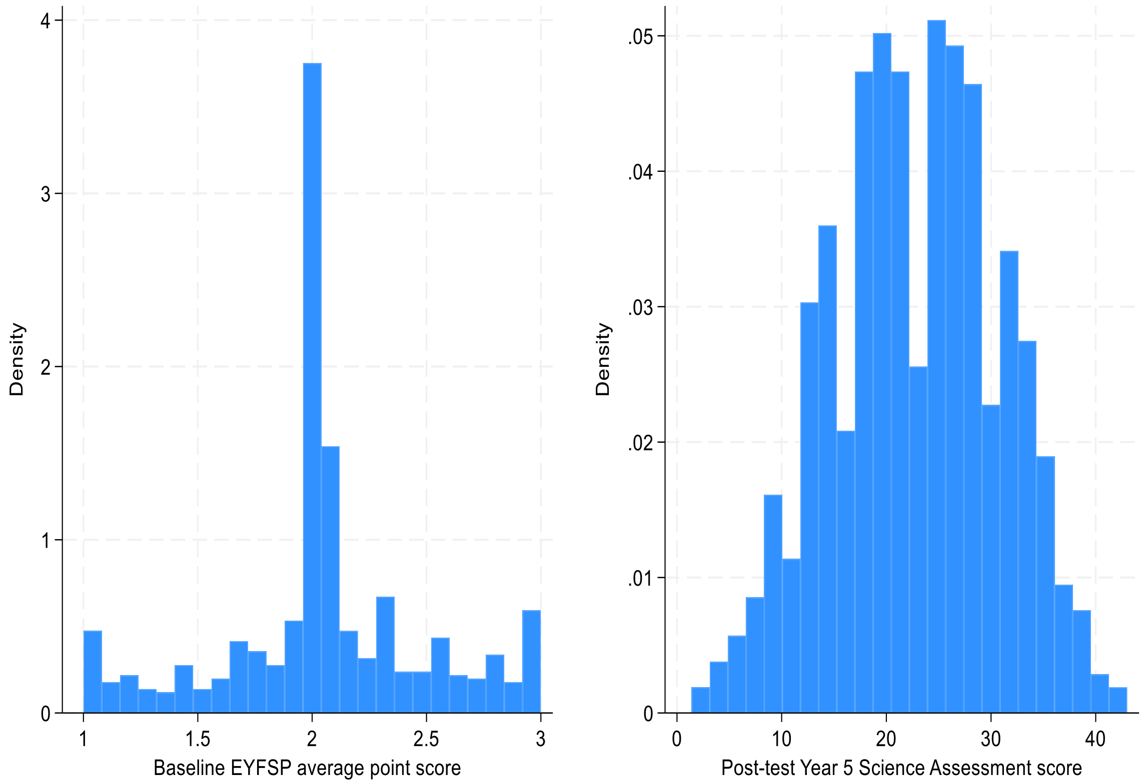

Figure 1: Histograms of baseline EYFSP average point and post-test Year 5 Science Assessment scores (scores have been aggregated to present the mean score for every ten pupils to prevent statistical disclosure)

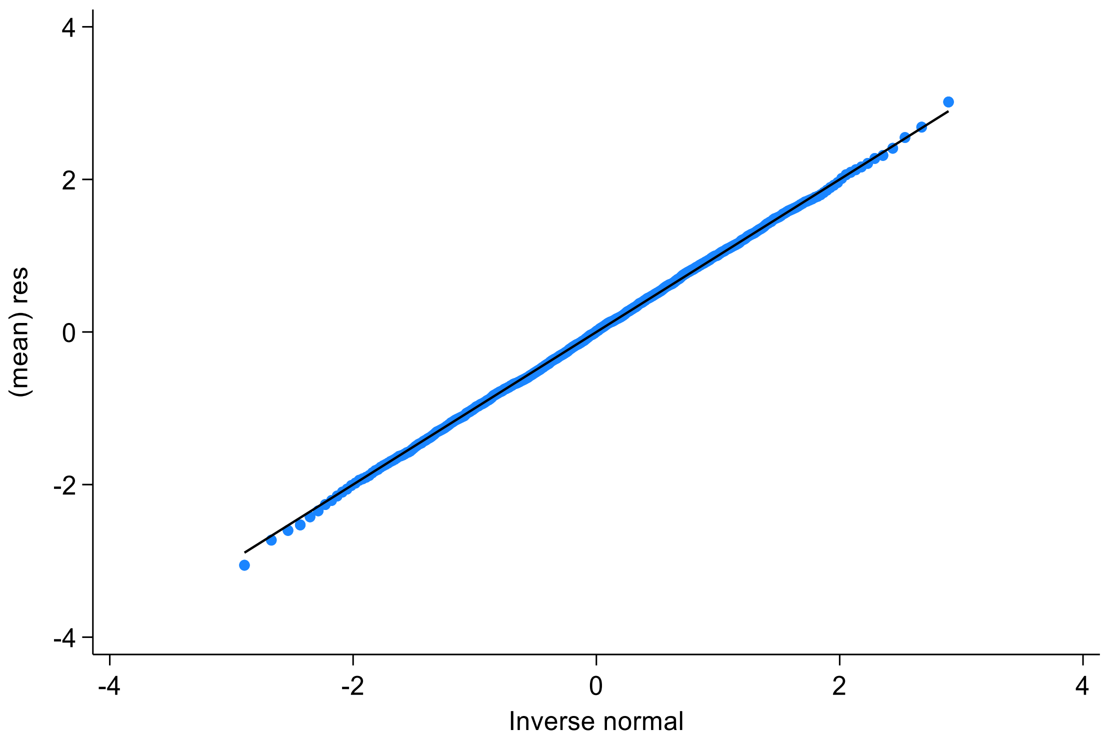

Figure 2: Q-Q plot of the standardised residuals from the primary analysis model to assess the normality assumption (residuals have been aggregated to present the mean residual for every ten pupils to prevent statistical disclosure)

Table 10: Primary addendum analysis (primary outcome)

	Unadjusted means				Effect size		
	Intervention group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)	Total n (intervention; control)	Hedges' g (95% CI)	P-value
Cohort 2 Year 5 Science Assessment (primary addendum analysis)	3,269 (413)	22.7 (22.4, 23.0)	2,818 (363)	22.7 (22.4, 23.0)	5,708 (3,048; 2,660)	0.03 (-0.07, 0.14)	0.52
Cohort 2 Year 5 Science Assessment (eligible for FSM)	967 (138)	19.8 (19.4, 20.3)	826 (149)	20.0 (19.5, 20.5)	1,793 (967; 826)	0.03 (-0.10, 0.16)	0.62

Subgroup analyses

Summary statistics for the post-test Year 5 Science Assessment score are presented by FSM eligibility (Table 11). 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 included an interaction effect, there was no evidence of an interaction between trial allocation and FSM eligibility (-0.12, 95% CI: -0.95 to 0.72, p = 0.79).

Within the recruited sample, there was 2,080 pupils from 160 schools who were eligible for FSM (median per school 10, range 0 to 51 in both groups). Within the restricted sample of pupils eligible for FSM, the correlation between the baseline and post-test scores was 0.44 (95% CI: 0.40 to 0.47, n=1,789 from 155 schools). The adjusted mean difference in post-test score between the intervention and control groups within this sample was 0.24 (95% CI: -0.72 to 1.21, Appendix A Table 1) and the estimated Hedges' g effect size was 0.03 (95% CI: -0.10 to 0.16, p = 0.62, Table 10), which does not relate to any additional months' progress in the intervention group. The total variance used to calculate the effect size (from a model without covariates) was 55.29—the sum of 51.33 (random variation between pupils, within-cluster variance) and 3.96 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.10 (95% CI: 0.07 to 0.16). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.07 (0.05 to 0.11).

Given the number of pupils included in this analysis (n=1,789), the observed ICC of 0.10 and the correlation between the pre- and post-test scores of 0.44, the estimated MDES at analysis for the Cohort 2 Year 5 Science Assessment outcome for the FSM subgroup was 0.17, with 80% power.

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 0.19 (95% CI: -0.79 to 1.17, p = 0.70) and the estimated Hedges' g effect size was 0.02 (95% CI: -0.10 to 0.15), which does not relate to any additional months' progress in the intervention group.

Table 11: 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	967	19.8 (7.5)	826	20.0 (7.3)
No	2,083	24.4 (7.6)	1,838	24.3 (7.7)

Analysis in the presence of non-compliance

A total of 90 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 training sessions, 15 (7.5%) teachers attended one training session, 23 (11.6%) teachers attended two training sessions, 36 (18.1%) teachers attended three training sessions, and 123 (61.8%) teachers attended all four training 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 2023–2024. There was one class from a single school for which a participating teacher was not provided (1.3% of the randomised sample), 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 recruited intervention pupils, 46.4% were deemed compliant (i.e. the pupil was in a class that was predominantly taught by a teacher who attended at least three out of the four full days of training in the first year of the trial).

The CACE estimate of the treatment effect on the pupils' attainment was a predicted increase of 0.10 points in Year 5 Science Assessment score (95% CI: -1.61 to 1.81, p = 0.91; effect size 0.01, 95% CI: -0.20 to 0.23). The partial R^2 from the first stage of the CACE analysis was 0.31 and the F-statistic was $F(8, 5{,}699 = 16.31, p<0.001)$; these indicate a high correlation between the instrumental variable (random allocation) and the endogenous variable, and that the inference of the CACE estimate is most likely reliable.

Missing data analysis

A mixed effect logistic regression model suggested that a 'missing at random' assumption was reasonable, meaning that multiple imputation by chained equations was an appropriate approach. Among the randomised population, whether the school was urban or rural, Ofsted rating, gender, FSM status, and EYFSP score were associated with missingness of the primary outcome.

To investigate the impact of missing data, the primary analysis was repeated using multiple imputation by chained equations. The adjusted mean difference in Year 5 Science Assessment score following multiple imputation was 0.22 (95% CI: -0.54 to 0.99, p = 0.57) and the Hedges' g effect size was 0.03 (95% CI: -0.07 to 0.12). This is a virtually identical effect to the primary analysis.

Secondary analysis

Science attitudes

In total, a valid post-test Year 5 Science Attitudes score was obtained for 5,931 randomised pupils (86.4%; intervention, n=3,202/3,682, 87.0%; control, n=2,729/3,181, 85.8%) from 157 schools (85 intervention, 72 control). A mean of 0.03 (95% CI: 0.00 to 0.07) was observed in the intervention arm and -0.04 (95% CI: -0.08 to 0.00) in the control arm. The unadjusted mean difference is 0.07 (95% CI: 0.02 to 0.12). Histograms of the Year 5 Science Attitudes scores shows they have a slightly negatively (left) skewed distribution (Figure 3). Baseline and post-test scores were available for 5,528 (80.5%) pupils (intervention, n=2,968/3,682, 80.6%; control, n=2,560/3,181, 80.5%). The correlation between the baseline and post-test scores was 0.07 (95% CI: 0.04 to 0.09). As a check of the analysis model assumptions, the normality of the standardised residuals was checked using a Q-Q plot and showed only a slight deviation from a normal distribution (Figure 4).

The adjusted mean difference in post-test score between the intervention and control groups was 0.08 (95% CI: -0.03 to 0.20, Appendix A Table 1). The estimated Hedges' g effect size was 0.08 (95% CI: -0.03 to 0.20, p = 0.16, Table 13), which relates to approximately one month's additional progress in the intervention group. The total variance used to calculate the effect size was 1.01; the sum of 0.90 (random variation between pupils, within-cluster variance) and 0.11 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.11 (95% CI: 0.08 to 0.14). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.11 (0.08 to 0.14).

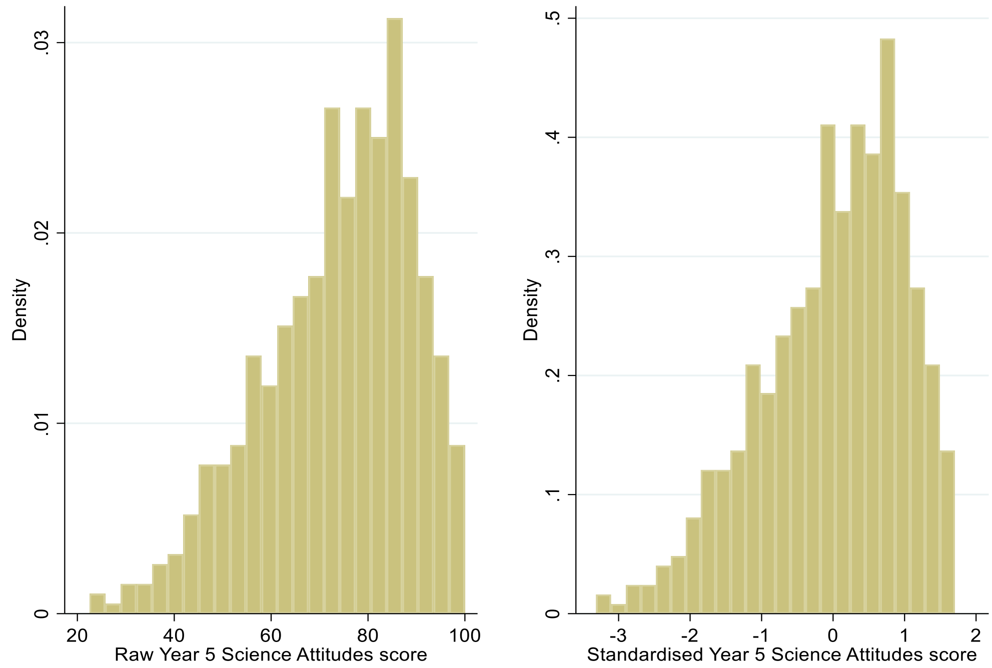

Figure 3: Histograms of raw and standardised post-test Year 5 Science Attitudes scores (scores have been aggregated to present the mean score for every ten pupils to prevent statistical disclosure)

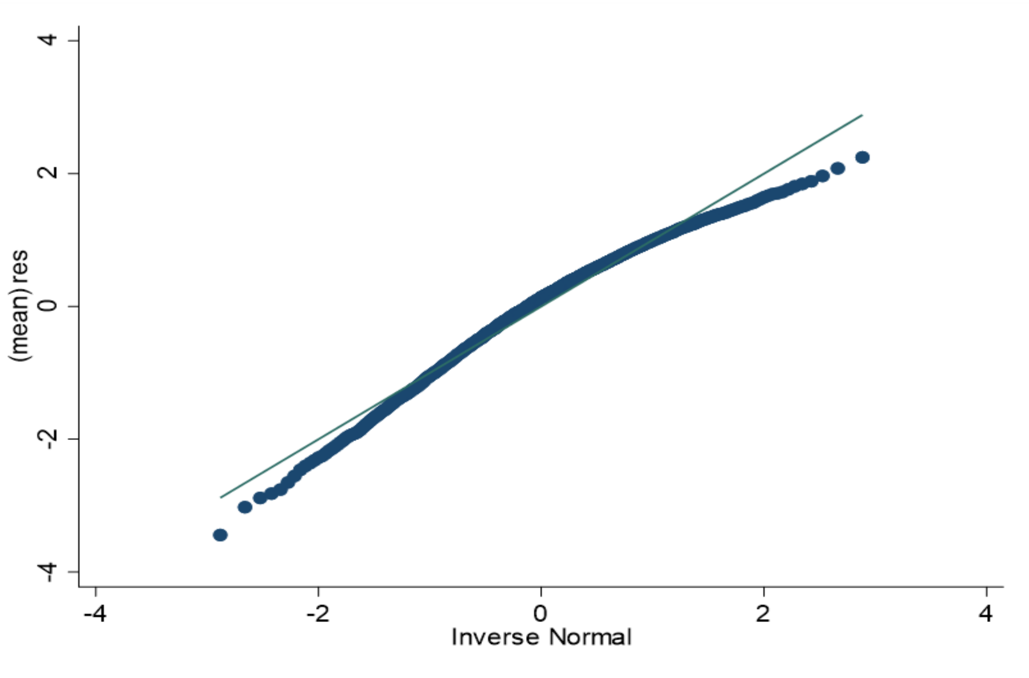

Figure 4: Q-Q plot of the standardised residuals from the secondary analysis model to assess the normality assumption (residuals have been aggregated to present the mean residual for every ten pupils to prevent statistical disclosure)

The remaining items, not used in the scale, are summarised separately (Table 12). Here, pupils' responses in both groups were broadly similar for the majority of Science Attitudes Questionnaire 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, 24.6%; intervention 19.6%).

Table 12: Summary of remaining Science Attitudes Questionnaire items not used in scale

Science Attitudes Questionnaire	Intervention (n=3,682) n (%)	Control (n=3,181) n (%)	Overall (n=6,863) n (%)
We spend a lot of time in science copying from the board			
Agree a lot	388 (9.7)	395 (12.4)	783 (11.4)
Agree a bit	657 (16.5)	599 (18.8)	1,256 (18.3)
Not sure	800 (20.1)	675 (21.1)	1,475 (21.5)
Disagree a bit	736 (18.5)	677 (21.3)	1,413 (20.6)
Disagree a lot	649 (16.3)	401 (12.6)	1,050 (15.3)
Missing	452 (11.4)	434 (13.6)	886 (12.9)
I think science is more for boys			
Agree a lot	98 (2.5)	86 (2.7)	184 (2.7)
Agree a bit	84 (2.1)	89 (2.8)	173 (2.5)
Not sure	424 (10.6)	354 (11.1)	778 (11.3)
Disagree a bit	251 (6.3)	225 (7.1)	476 (6.9)
Disagree a lot	2,384 (59.9)	1,994 (62.7)	4,378 (63.8)
Missing	441 (11.1)	433 (13.6)	874 (12.7)
We often have discussions in science lessons			
Agree a lot	1,804 (45.3)	1,472 (46.3)	3,276 (47.7)
Agree a bit	964 (24.2)	831 (26.1)	1,795 (26.2)
Not sure	343 (8.6)	317 (10.0)	660 (9.6)
Disagree a bit	89 (2.2)	92 (2.9)	181 (2.6)
Disagree a lot	27 (0.7)	35 (1.1)	62 (0.9)
Missing	455 (11.4)	434 (13.6)	889 (13.0)
We do a lot of writing in science lessons			
Agree a lot	781 (19.6)	784 (24.6)	1,565 (22.8)
Agree a bit	1,077 (27.0)	985 (31.0)	2,062 (30.0)
Not sure	689 (18.7)	515 (16.2)	1,204 (17.5)
Disagree a bit	520 (14.1)	365 (11.5)	885 (12.9)
Disagree a lot	175 (4.8)	107 (3.4)	281 (4.1)
Missing	440 (12.0)	425 (13.4)	866 (12.6)
I have made good progress in science this year			
Agree a lot	1,047 (28.4)	868 (27.3)	1,915 (27.9)
Agree a bit	939 (25.5)	777 (24.4)	1,716 (25.0)
Not sure	922 (25.0)	805 (25.3)	1,727 (25.2)
Disagree a bit	168 (4.6)	153 (4.8)	321 (4.7)
Disagree a lot	173 (4.7)	153 (4.8)	326 (4.8)
Missing	433 (11.8)	425 (13.4)	858 (12.5)
We already know what will happen when we do science practical work			
Agree a lot	258 (7.0)	222 (7.0)	480 (7.0)
Agree a bit	572 (15.5)	506 (15.9)	1,078 (15.7)
Not sure	1,059 (28.8)	897 (28.2)	1,956 (28.5)
Disagree a bit	722 (19.6)	588 (18.5)	1,310 (19.1)
Disagree a lot	636 (17.3)	536 (16.9)	1,172 (17.1)
Missing	435 (11.8)	432 (13.6)	867 (12.6)
We can decide for ourselves how to do science practical work			
Agree a lot	420 (11.4)	303 (9.5)	723 (10.5)
Agree a bit	758 (20.6)	582 (18.3)	1,340 (19.5)
Not sure	1,146 (31.1)	1,043 (32.8)	2,189 (31.9)
Disagree a bit	486 (13.2)	425 (13.4)	911 (13.3)
Disagree a lot	402 (10.9)	375 (11.8)	777 (11.3)
Missing	470 (12.8)	453 (14.2)	923 (13.4)
We do practical work in most science lessons			
Agree a lot	685 (18.6)	466 (14.6)	1,151 (16.8)

Science Attitudes Questionnaire	Intervention (n=3,682) n (%)	Control (n=3,181) n (%)	Overall (n=6,863) n (%)
Agree a bit	1,002 (27.2)	836 (26.3)	1,838 (26.8)
Not sure	808 (21.9)	709 (22.3)	1,517 (22.1)
Disagree a bit	497 (13.5)	513 (16.1)	1,010 (14.7)
Disagree a lot	223 (6.1)	197 (6.2)	420 (6.1)
Missing	467 (12.7)	460 (14.5)	927 (13.5)

Cohort 1 Longitudinal analysis

Year 6 Science Assessment score

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 Year 6 Science Assessment score was obtained for 5,941 randomised pupils (81.8%; intervention, n=3,118/3,790, 82.3%; control, n=2,823/3,471, 81.3%) from 166 schools (88 intervention, 78 control). A mean of 22.5 (95% CI: 22.3 to 22.8) was observed in the intervention arm and 22.6 (95% CI: 22.4 to 22.9) in the control arm. The unadjusted mean difference is -0.10 (95% CI: -0.46 to 0.26). Histograms of the baseline and post-test scores show the Year 6 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 5). Baseline and post-test scores were available for 5,717 (78.7%) pupils (intervention, n=2,980/3,790, 78.6%; control, n=2,737/3,471, 78.9%). The correlation between the baseline and post-test scores was 0.46 (95% CI: 0.44 to 0.48). As a check of the analysis model assumptions, the normality of the standardised residuals was checked using a Q-Q plot and were shown to be normally distributed (Figure 6).

The adjusted mean difference in post-test score between the intervention and control groups was -0.16 (95% CI: -0.84 to 0.52, Appendix A Table 1). The estimated Hedges' g effect size was -0.02 (95% CI: -0.12 to 0.07, p = 0.64, Table 13), which does not relate to any months' additional progress in the intervention group. The total variance used to calculate the effect size was 50.09; the sum of 44.01 (random variation between pupils, within-cluster variance) and 6.08 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.10 (95% CI: 0.08 to 0.12). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.12 (0.10 to 0.15).

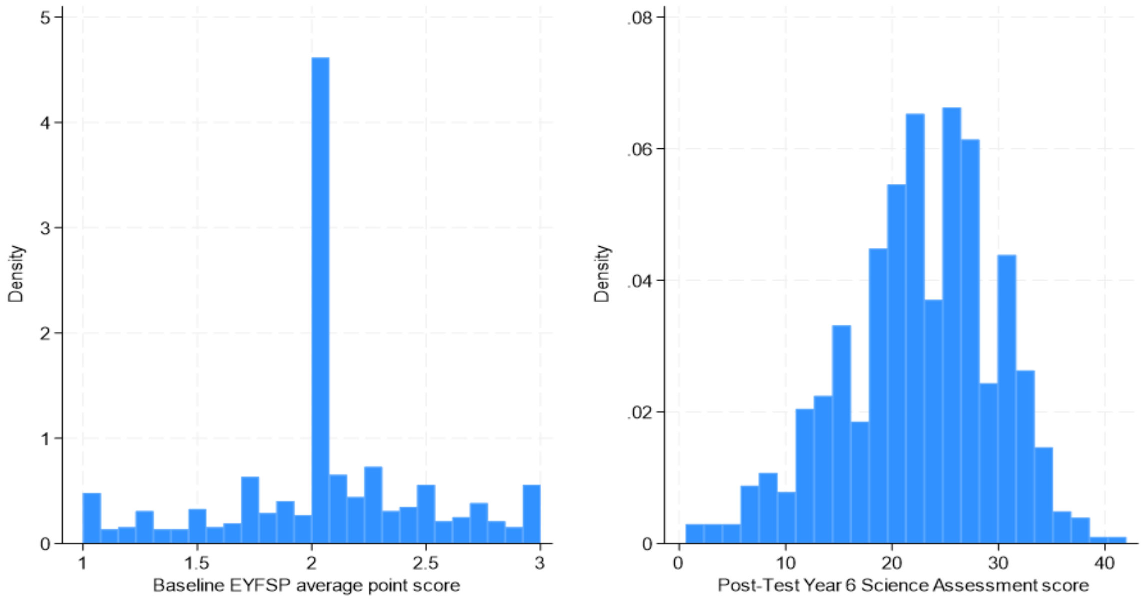

Figure 5: Histograms of baseline EYFSP average point and post-test Year 6 Science Assessment scores (scores have been aggregated to present the mean score for every ten pupils to prevent statistical disclosure)

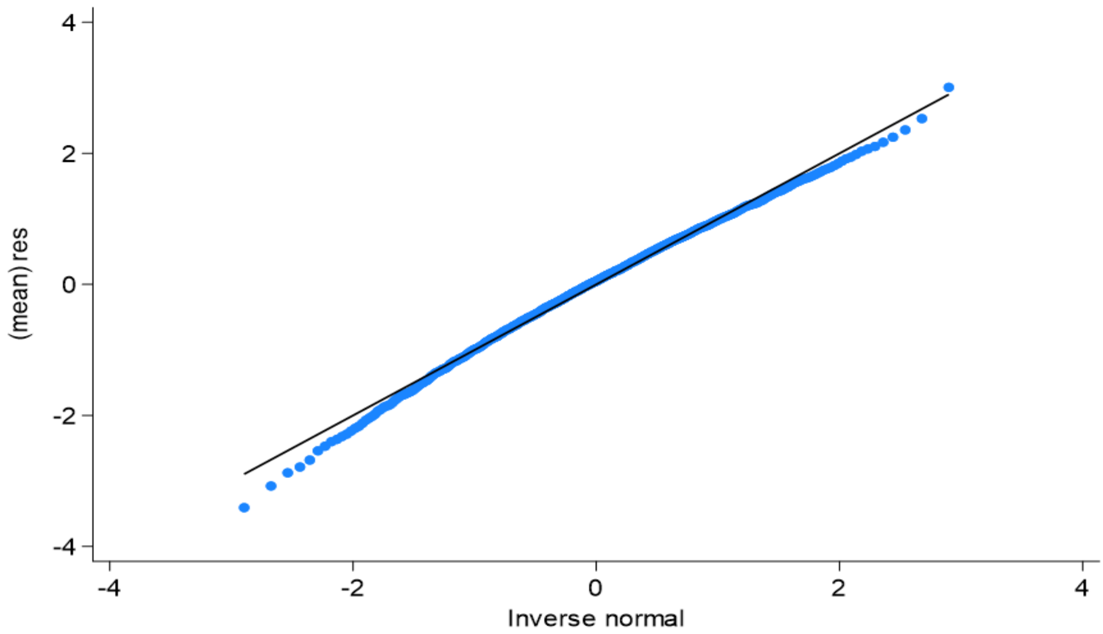

Figure 6: Q-Q plot of the standardised residuals from the analysis model to assess the normality assumption (residuals have been aggregated to present the mean residual for every ten pupils to prevent statistical disclosure)

Key Stage 2 Maths score

In total, a valid Key Stage 2 Maths score was obtained for 6,749 randomised pupils (92.9%; intervention, n=3,536/3,790, 93.3%; control, n=3,213/3,471, 92.6%) from 167 schools (88 intervention, 79 control). A mean of 104.3 (95% CI: 104.1 to 104.6) was observed in the intervention arm and 104.2 (95% CI: 103.9 to 104.5) in the control arm. The unadjusted mean difference is 0.14 (95% CI: -0.22 to 0.50). A histogram of the scores is presented in Figure 7. Baseline and post-test scores were available for 6,463 (89.0%) pupils (intervention, n=3,373/3,790, 89.0%; control, n=3,090/3,471, 89.0%). The correlation between the baseline and post-test scores was 0.47 (95% CI: 0.45 to 0.49). As a check of the analysis model assumptions, the normality of the standardised residuals was checked using a Q-Q plot and were shown to be normally distributed (Figure 8).

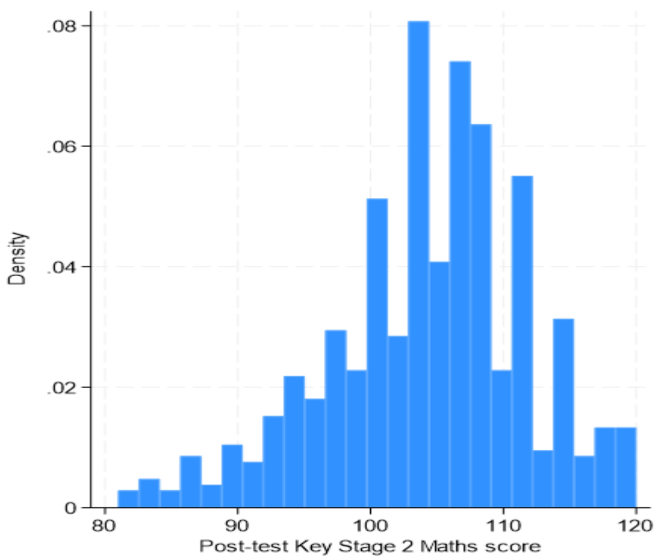

Figure 7: Histogram of Key Stage 2 Maths scores (scores have been aggregated to present the mean score for every ten pupils to prevent statistical disclosure)

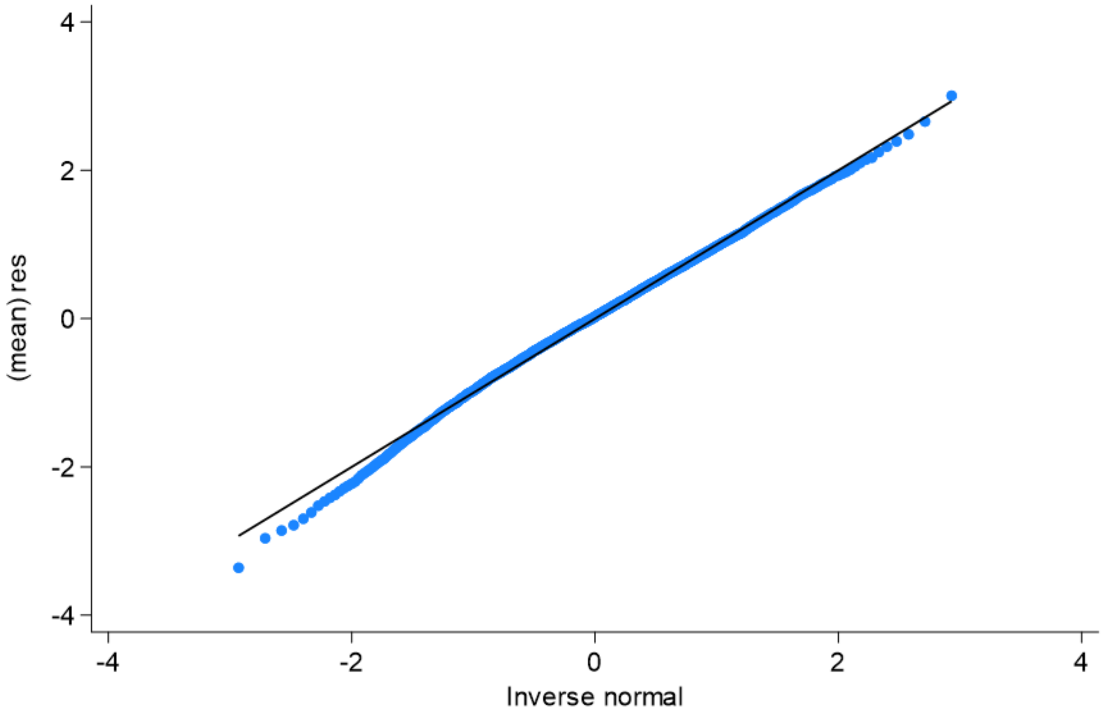

Figure 8: Q-Q plot of the standardised residuals from the analysis model to assess the normality assumption (residuals have been aggregated to present the mean residual for every ten pupils to prevent statistical disclosure).

The adjusted mean difference in post-test score between the intervention and control groups was -0.14 (95% CI: -0.96 to 0.67, Appendix A Table 1). The estimated Hedges' g effect size was -0.02 (95% CI: -0.13 to 0.09, p = 0.74, Table 13), which does not relate to any months' additional progress in the intervention group. The lower limit of the CI equates to a small, negative effect (loss of two months' progress relative to control). The total variance used to calculate the effect size was 56.08; the sum of 49.52 (random variation between pupils, within-cluster variance) and 6.56 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.14 (95% CI: 0.11 to 0.18). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.12 (0.09 to 0.15).

Key Stage 2 English score

In total, a valid post-test Key Stage 2 English score was obtained for 6,751 randomised pupils (93.0%; intervention, n=3,539/3,790, 93.4%; control, n=3,212/3,471, 92.5%) from 167 schools (88 intervention, 79 control). A mean of 105.1 (95% CI: 104.8 to 105.3) was observed in the intervention arm and 104.9 (95% CI: 104.6 to 105.1) in the control arm. The unadjusted mean difference is 0.23 (95% CI: -0.14 to 0.60). A histogram of the scores is presented in Figure 9. Baseline and post-test scores were available for 6,464 (89.0%) pupils (intervention, n=3,375/3,790, 89.0%; control, n=3,089/3,471, 89.0%). The correlation between the baseline and post-test scores was 0.48 (95% CI: 0.46 to 0.50). As a check of the analysis model assumptions, the normality of the standardised residuals was checked using a Q-Q plot and were shown to be normally distributed (Figure 10).

The adjusted mean difference in post-test score between the intervention and control groups was 0.05 (95% CI: -0.65 to 0.74, Appendix A Table 1). The estimated Hedges' g effect size was 0.01 (95% CI: -0.08 to 0.10, p = 0.90, Table 13), which does not relate to any months' additional progress in the intervention group. The lower limit of the CI equates to a small, negative effect (loss of one month's progress relative to control). The total variance used to calculate the effect size was 59.51; the sum of 54.72 (random variation between pupils, within-cluster variance) and 4.80 (heterogeneity between schools, between-cluster variance). The ICC associated with school from the adjusted model was 0.09 (95% CI: 0.07 to 0.11). The ICC (95% CI) for the empty model (i.e. without covariates) was 0.08 (0.06 to 0.10).

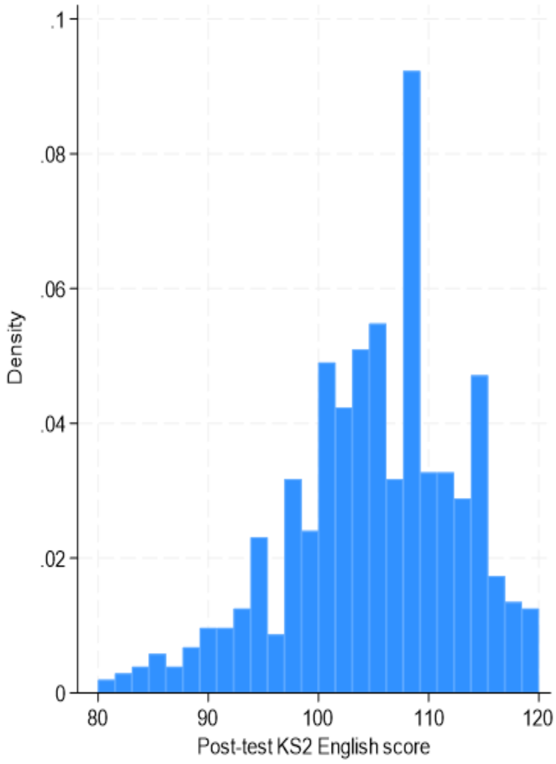

Figure 9: Histogram of Key Stage 2 English scores (scores have been aggregated to present the mean score for every ten pupils to prevent statistical disclosure)

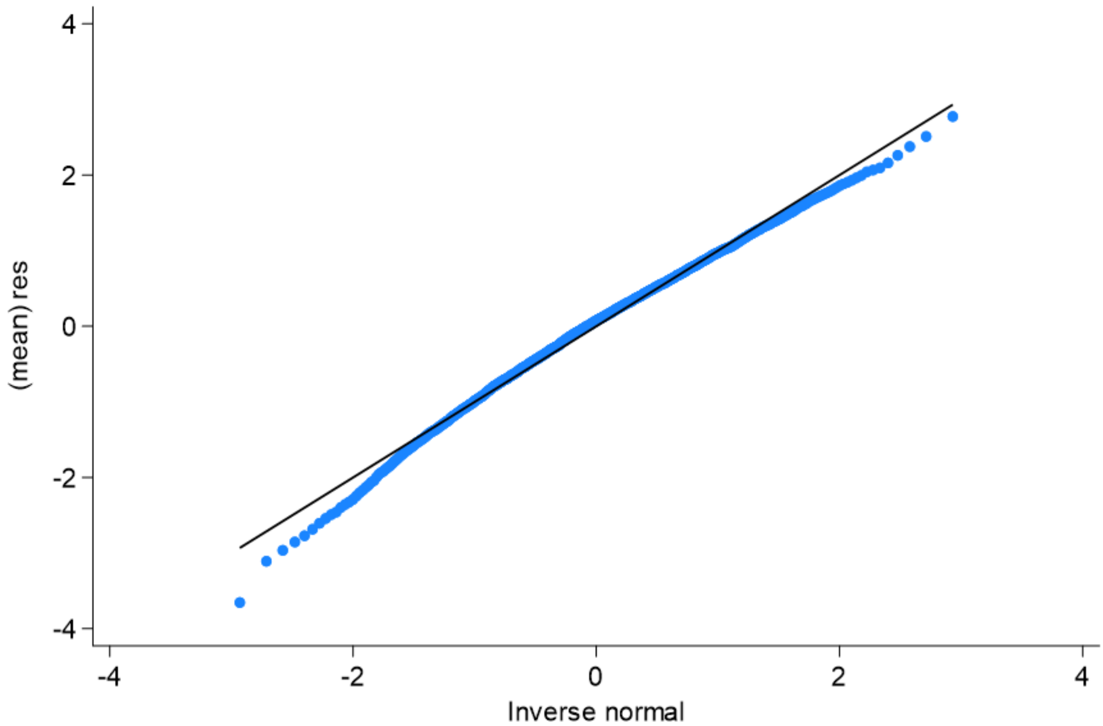

Figure 10: Q-Q plot of the standardised residuals from the analysis model to assess the normality assumption (residuals have been aggregated to present the mean residual for every ten pupils to prevent statistical disclosure).

Table 13: 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 2 Year 5 Science Attitudes Questionnaire	3,202 (480)	0.03 (0.00, 0.07)	2,729 (452)	-0.04 (-0.08, 0.00)	5,528 (2,968: 2,560)	0.08 (-0.03, 0.20)	0.16
Cohort 1 Year 6 Science Assessment	3,118 (672)	22.5 (22.3, 22.8)	2,823 (648)	22.6 (22.4, 22.9)	5,717 (2,980: 2,737)	-0.02 (-0.12, 0.07)	0.64
Cohort 1 Key Stage 2 Maths	3,536 (254)	104.3 (104.1, 104.6)	3,213 (258)	104.2 (103.9, 104.5)	6,463 (3,373: 3,090)	-0.02 (-0.13, 0.09)	0.74
Cohort 1 Key Stage 2 English	3,539 (251)	105.1 (104.8, 105.3)	3,212 (259)	104.9 (104.6, 105.1)	6,464 (3,375: 3,089)	0.01 (-0.08, 0.10)	0.90

Additional analysis

As part of the protocol for the development of the Year 6 Science Assessment test (Hanley *et al.,* 2022), it was proposed to investigate the correlation between test scores from the Year 5 and Year 6 tests. It was decided that this should be conducted as part of the TDTScience trial.

The correlation between Year 5 and Year 6 Science Assessment test score, based on 2,655 observations (control arm only) was 0.73 (95% CI: 0.71 to 0.75).

Implementation and Process Evaluation

Summary of data collected

Table 14 summarises the 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, for example, a staff member who was not the relevant teacher identified using name/role/class, 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 included in the descriptive analysis.

Table 14: Summary of valid responses to IPE surveys at each timepoint by allocation

Survey type / intended respondent	Survey date	Allocation	Intervention n responses / n participating (response rate %)	Control n responses / n participating (response rate %)	Total n responses / n participating (response rate %)
Pre-randomisation baseline survey / Year 5 teacher[a]	October 2022 to December 2022	Teachers	115/[b]	83/[b]	198/[b]
		Schools	80/90 (89%)	62/90 (69%)	142/180 (79%)
Follow-up survey / Year 5 teacher[a]	June 2023 to July 2023	Teachers	76/143 (53%)	75/136 (55%)	151/279 (54%)
		Schools	61/88 (69%)	57/79 (72%)	118/167 (71%)
Follow-up survey / Year 5 teacher	June 2024 to July 2024	Teachers	103/137 (75%)	95/125 (76%)	198/262 (76%)
		Schools	76/88 (86%)	69/79 (87%)	145/166 (87%)
Follow-up survey / Year 6 teacher		Teachers	88/143 (62%)	97/137 (70%)	185/280 (66%)
		Schools	72/88 (82%)	71/74 (96%)	143/160 (89%)

[a]Data presented in Robinson-Smith et al. 2025.
[b]The total number of Year 5 teachers in the study at this point was unknown. The 180 recruited schools were requested to circulate the survey for completion by their Year 5 teachers in Summer Term 2022; however, details of participating teachers were not sent to the evaluation team until the pupil recruitment stage in Autumn Term 2022, by which point school attrition (n=12) had occurred. 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. See Robinson-Smith et al. 2025 for further details.

Of the 198 responses to the Year 5 follow-up survey in 2023–2024, 81 (41 intervention, 40 control) of these respondents also completed the follow-up survey in 2022–2023.

Survey data from Year 5 and Year 6 teachers are presented and discussed separately in this section to align with the purpose of the longitudinal follow-up, which was to examine the legacy of the TDTScience training and any effects of embedding of the TDTScience practices. To summarise the two groups of teachers:

- Year 5 teachers were offered TDTScience training and taught Year 5 pupils (Cohort 1) during academic year 2022–2023. In academic year 2023–2024, Year 5 teachers taught Cohort 2.[1]

[1] At the point of recruitment, schools were advised that Year 5 teachers should remain teaching Year 5 pupils for the duration of the trial (from September 2022 to July 2024), unless circumstances made this impossible. As although the TDTScience training was only delivered in the first year, the expectation was that teachers would continue to use TDTScience in their science lessons in the second year.

- Year 6 teachers were not offered the TDTScience training in 2022–2023. They taught Year 6 pupils (Cohort 1) during academic year 2023–2024.

Year 5 teachers

This section presents data from Year 5 teachers who completed the follow-up survey at the end of 2023–2024. Usual practice for Year 5 science lessons across all schools is presented. We then explore fidelity in terms of classroom practice and whether the intervention was embedded by Year 5 teachers following their training in 2022–2023, their perception of TDTScience and any impact on their own practice and their pupils' (Cohort 2) learning in 2023–2024. In doing so, it addresses the research questions below. A comparison to the IPE data from the first year of the trial are provided where appropriate. For context, within the 2022–2023 sample of responding Year 5 teachers, two-thirds reported that GCSE or equivalent was their highest qualification in science (70% control; 67% intervention) and two-thirds also noted that they had not been involved in any science training or initiatives in 2022–2023 during intervention delivery (70.7% control; 68.4% intervention). Around 80% of survey respondents reported it was usual practice to teach science to a Year 5 class at least once a week (frequency similar between groups) with the average lesson at follow-up lasting 84 minutes (83 minutes, control; 85 minutes, intervention) The 'Usual practice' section within Robinson-Smith *et al*. 2025 provides a detailed summary of Year 5 teachers who responded to the 2022–2023 usual practice follow-up survey.

Four main research questions were addressed in this part of the evaluation:

1. To what extent was TDTScience implemented as planned? (Hanley *et al.,* 2023)

    b. Classroom practice.

2. What processes are involved for teachers and schools implementing TDTScience—what are the main facilitators and barriers? (Hanley *et al.,* 2023)

3. What are the perceptions of teachers as regards TDTScience? (Hanley *et al.,* 2023)

    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?

5. How does TDTScience compare with practice in business-as-usual science lessons? (Hanley *et al.,* 2023)

    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?

Usual practice

Two-thirds of respondents to the Year 5 teacher 2023–2024 survey noted that their highest qualification in science was GCSE or equivalent (58.9%, n=56/95 control; 65.0%, n=67/103 intervention), see Appendix B Table 1. More Year 5 teachers in the control group indicated they had an A Level or equivalent in science than teachers in the intervention group (17.9%, n=17/95 control; 10.7%, n=11/103 intervention). Slightly more teachers in the intervention group (15.5%, n=16/103) had a science postgraduate degree in comparison to the control group (11.6%, n=11/95).

The majority (75.3%, n=149/198) of Year 5 teachers in all participating schools indicated that they had not been involved in any other science-related training or initiatives during 2023–2024. Among those teachers who had, they commonly reported STEM Learning/Science Learning Centre courses, Primary Science Quality Mark and Teacher Assessment in Primary

Assessments (such responses were balanced between allocated groups, see Appendix B Table 2). Despite low levels of participation in science-related training/initiatives in 2023–2024, 79.6% (n=82/103) of teachers in intervention schools reported that they changed the way they taught science in 2023–2024 (presumably impacted by TDTScience) in comparison to 49.4% (n=47/95) of teachers in control schools (see Appendix B Table 3).

As presented in Appendix B Table 4, it was usual practice in Year 5 science lessons in all schools to regularly include science demonstrations, provide pupils time to think and discuss, teach scientific terminology, and to report and interpret results from practical investigations. In comparison to the control group, it was more common for Year 5 science lessons in intervention schools to have pupils plan, design, and conduct practical investigations, and ask them to solve scientific problems during science lessons. The latter was similarly reported in the 2022–2023 survey and could be indicative of Year 5 teachers continuing to implement the TDTScience approach of Bright Ideas Time in their science lessons in 2023–2024. Teachers in intervention schools taught ‘scientific facts’ and required pupils to ‘write up whole investigations’ less frequently during their Year 5 science lessons in comparison to the control group. This was also apparent in the 2022–2023 survey data. Here, it could be indicative of teachers continuing to implement focused recording within their practice in 2023–2024. Teachers discussed the use of focused recording within interviews conducted in 2022–2023. For 80% of Year 5 classes, science lessons took place weekly and lasted on average 79 minutes (range 45–160), with little difference between allocated groups. The duration of Year 5 science lessons at the end of 2023–2024 was five minutes less than the duration reported by Year 5 teachers at the end of 2022–2023, see Table 15 below.

Table 15: Year 5 science lessons duration at each survey timepoint by allocated group

Allocated group	Baseline (2022–2023) minutes (range)	Follow-up (2022–2023) minutes (range)	Long-term follow-up (2023–2024) minutes (range)
Control	79 (28–130)	83 (39–130)	78 (45–121)
Intervention	83 (18–140)	85 (45–120)	79 (45–160)
All schools	81 (18–140)	84 (39–130)	79 (45–160)

Continued implementation

The majority of Year 5 teachers who responded to the survey at the end of 2023–2024 taught science to Year 5 for the duration of the trial (75%, n=77/103)—that is, to Cohort 1 in 2022–2023 and to Cohort 2 in 2023–2024—and 63.1% (n=65/103) confirmed that they received some of the in-person TDTScience training that took place in 2022–2023 (see Table 16 below).

Table 16: Proportion of Year 5 teachers who responded to 2023–2024 follow-up survey who also attended TDTScience training in 2022–2023

In-person attendance at TDTScience training during 2022–2023?	n (%)
Yes	65 (63.1)
No – but a trainer briefed me during the school year	2 (1.9)
No – but a colleague briefed me during the school year	19 (18.4)
No – I received nothing during the school year 2022–2023	15 (14.6)
Missing	2 (1.9)
Total n	**103**

We asked the Year 5 teachers who received TDTScience training a series of statements to gauge their continued understanding of TDTScience and their use of it within the classroom (these data are presented in Appendix B Table 5). The majority of teachers ‘agreed’ (either slightly or strongly) that the training they received in 2022–2023 prepared them well for implementing TDTScience within the classroom (86.1%, n=74/86); it provided them with an understanding of how to construct a TDTScience science lesson, including implementing the core TDTScience teaching strategies such as Bright

Ideas Time, Practical Prompts for Thinking, and focused recording. Additionally, most teachers considered the TDTScience strategies useful and continued to use them regularly and effectively when teaching science to their Year 5 classes in 2023–2024. Qualitative responses also indicated that TDTScience continued to be viewed positively in 2023–2024; and increased teachers' confidence in teaching science, changed how they taught science, and increased the use of practical work, discussion, and pupil enjoyment during science lessons.

> *I have loved adopting the TDTScience approach to teaching science and I feel that the training was delivered very well. It provided me with lots of ideas and made me think about what I wanted children to get out of it and how I can achieve that. My lessons are more engaging, focussed on the science and as a result, the children have grown to love science again.* (Year 5 teacher, End of 2023–2024 survey)

We asked the Year 5 teachers who received the in-person training in 2022–2023 if they commonly shared what they learned from TDTScience training with other teachers in their school during 2023–2024; 71% (n=46/65) of teachers stated that they shared what they learned with others 'a great deal' or 'quite a lot'. Around a third of these teachers (n=19) described that they shared TDTScience techniques with others in their school during staff meetings or during in-school CPD training days (n=4):

> *I led a staff meeting showing all other teachers how to use higher order thinking questions within lessons and provided them with resources to do so. I showed them an example lesson using [B]right [I]deas [T]ime and showed them the website '[E]xplorify' and where to find the [B]right [I]deas [T]ime ideas. I have spoken with the science lead and showed them our practical investigations and focused recording on our online platform 'Seesaw'.* (Year 5 teacher, End of 2023–2024 survey)

From this, some schools stated they adopted a whole-school approach (n=6) to delivering TDTScience by adapting the programme for use in all year groups by encouraging a change to the standard structure of science lessons and promoting an uptake of the techniques and resources.

Other teachers stated that they specifically shared TDTScience training with other Key Stage 2 teachers; for some teachers this was to ensure continuity for their Year 5 2022–2023 cohorts as they moved into Year 6 (n=7).

> *KS2 [Key Stage 2] staff now follow the TDTS [TDTScience] lesson structure with a big question to start a lesson to prompt talking and then a vocabulary slide to support and strengthen vocabulary knowledge. More focus on the practical lesson than [writing] in the book. A greater focus on precise and higher quality recording in shorter amounts.* (Year 5 teacher, End of 2023–2024 survey)

From open-ended qualitative survey responses, there was evidence that cascaded training was not always viewed as being wholly sufficient for underpinning understanding of the TDTScience approach among teachers who did not attend the in-person training. Some schools suggested further training from official TDTScience trainers would have been beneficial for a second consecutive year.

> *Our Science lead used the TDTS [TDTScience] scheme to shape our new planning, which was helpful, however, the strategies used have not been explained in detail, aside to the information placed on the planning documents. I would benefit from information regarding all of the TDTS [TDTScience] strategies to ensure that I am using, planning and facilitating them appropriately to have the desired impact and outcomes.* (Year 5 teacher, End of 2023–2024 survey)

While there was appetite in some schools to implement TDTScience approach in other year groups—as discussed above—it was also mentioned as a barrier to continued implementation as teachers sometimes struggled to apply and adapt the training, resources, and practical lesson plans to pupils within different year groups.

Impact of TDTScience

One of the major changes Year 5 teachers reported in 2022–2023 as a result of TDTScience was increased confidence in teaching science. At the end of 2023–2024, most teachers in the intervention group again reported being more confident in teaching science than those in the control group at follow-up across different domains, see Appendix B Table 6. For example, 95.1% (n=98/103) in the intervention group reported being 'quite confident' or 'very confident' in helping pupils discuss scientific ideas compared to 85.2% (n=81/95) in the control group. This is despite the control group being more confident in teaching science in 2023–2024 than at the end of 2022–2023.

At the end of 2023–2024, all Year 5 teachers were asked a series of statements to gauge how interested and engaged they considered their pupils to be in learning science (see Appendix B Table 7). More teachers in the intervention group than the control group agreed that their pupils came up with their own ideas (74.8%, n=77/103 vs 66.3%, n=63/95) and spent less time writing during their science lessons (61.2%, n=63/103 vs 31.6%, n=30/95). Pupils in intervention schools were also considered to be more engaged during science lessons (97.1%, n=100/103 vs 91.6%, n=87/95) and enjoyed them more (96.1%, n=99/103 vs 94.7%, n=90/95) in comparison to pupils in control schools.

Year 6 teachers

This section presents data from Year 6 teachers who completed the follow-up survey at the end of 2023–2024. Year 6 teachers in intervention schools were not offered TDTScience training in 2022–2023 and taught Cohort 2 pupils in 2023–2024. Usual practice for Year 6 science lessons across all schools is presented. We explore if any Year 6 respondents received any cascaded training and support from their colleagues who attended in-person training and their perception of their Year 6 pupils (Cohort 1) engagement and enjoyment with science to understand if there is any continued impact of TDTScience. In doing so, it addresses the research question below:

5. How does TDTScience compare with practice in business-as-usual science lessons? (Hanley *et al.,* 2023)
    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?

A comparison to IPE data from 2022–2023 is provided where appropriate.

Usual practice

For 80.0% (n=147/185) of Year 6 classes, science lessons took place weekly and lasted on average 79 minutes (range 40–180). Intervention schools reported that their Year 6 science lessons lasted, on average, seven minutes longer than control schools (intervention, 83 minutes; control, 76 minutes). A greater proportion of control schools reported that Year 6 were taught science twice a week in comparison to the intervention group (control 12.4%, n=12/97; intervention, 4.5%, n=4/88).

As presented in Appendix B Table 8, it was usual practice in Year 6 science lessons for teachers in all schools to teach scientific facts. In comparison to science lessons in control schools, Year 6 science lessons in intervention schools more regularly encouraged pupil discussion in the whole class (very often 79.5%, n=70/88 vs 72.2%, n=70/97), gave pupils more opportunities to think (79.5%, n=70/88 vs 64.9%, n=63/97), solve scientific problems (23.9%, n=21/88 vs 16.5%, n=16/97), the design/planning (14.8%, n=13/88 vs 5.2%, n=5/97), and interpretation of practical investigations (11.4%, n=10/88 vs 5.2%, n=5/97). Year 6 science lessons in control schools more regularly involved writing up whole investigations ('very' or 'quite often', intervention 15.9%, n=14/88; control 29.9%, n=29/97) and memorising facts and principles in comparison to intervention schools (intervention 51.1%, n=45/88; control 65.0%, n=63/97).

Such differences in Year 6 science lessons between the control and intervention schools could be a result of Year 6 teachers in intervention schools receiving TDTScience training, see Table 17 below. Among the respondents to the Year 6 teacher survey, 23.9% (n=21/88) indicated that they attended (at least some) in-person TDTScience training in 2022–2023. Although the exact reason for this is unknown, this likely reflects the fact that teachers who had taught Year 5 in 2022–2023 were moved to Year 6 in 2023–2024; teacher movement between year groups each academic year is a commonplace in primary schools.

Table 17: TDTScience training among Year 6 survey respondents to the 2023–2024 end of year survey

In-person attendance at TDTScience training during 2022–2023?	n (%)
Yes	21 (23.9)
No – but a trainer briefed me during the school year	1 (1.1)
No – but a colleague briefed me during the school year	33 (37.5)
No – I received nothing during the school year 2022–2023	29 (33.0)
Missing	4 (4.5)
Total n	**88**

Impact of TDTScience

Year 6 teachers were asked a series of statements to gauge how interested and engaged they considered their pupils to be in learning science (see Appendix B Table 9). In comparison to the control group, Year 6 teachers in TDTScience schools were more likely to report that their pupils enjoyed their science lessons (strongly agree 54.5%, n=48/88 vs 40.2%, n=39/97), made good progress in their understanding of science (51.1%, n=45/88 vs 37.1%, n=36/97), were engaged with (50.0%, n=44/88 vs 39.2%, n=38/97) and confident in science (35.2%, n=31/88 vs 22.7%, n=22/97), could work independently (29.5%, n=26/88 vs 20.6%, n=20/97), and came up with their own scientific ideas (21.6%, n=19/88 vs 10.3%, n=10/97). Such differences between the intervention and control group were also apparent at the end of 2022–2023 when these pupils (Cohort 1) were in Year 5, as reported in Robinson-Smith *et al*. (2025).

Cost evaluation results

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 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. In this evaluation, formal intervention training by Science Oxford was delivered to Year 5 teachers in the 2022–2023 school year. Costs associated with these activities and with schools delivering the intervention in the first year of the trial are reported in Robinson-Smith *et al*. (2025).

There was no external training in the intervention to any new or existing Year 5 teachers in the second year of the trial (2023–2024). They were reliant on the teachers trained in the first year (2022–2023) sharing their knowledge and on accompanying physical/online materials for any learning about TDTScience.

The main costs associated with implementing the programme in the second year are for staff cover and purchase of additional materials, both of which are categorised as optional rather than mandatory as only a proportion of schools utilised these.

Time

The cost and time resource questions from the end of year intervention survey were completed by 190 teachers from 85 schools, of which 103 (from 76 schools) were Year 5 teachers, and 88 (from 72 schools) were Year 6 teachers (one teacher taught both year groups).

Teachers were asked whether they attended any of the in-person TDTScience training during the school year 2022–2023. If they responded ‘Yes’, ‘No – but a trainer briefed me during the school year 2022–2023’, or ‘No – but a colleague briefed me during the school year 2022–2023’ (as opposed to ‘No – I received nothing during the school year 2022–2023’), they were then asked whether, during the school year 2023–2024, they had spent time sharing TDTScience training with other members of staff within their school who did not receive TDTScience training during the school year 2022–2023. A total of 53 out of 140 (37.9%), from 44 schools, said they had, and had spent an average of 3.2 hours (SD 2.1) on this activity. Of these, less than ten[2] teachers, all from different schools, reported that their school arranged for paid cover for them to share the training to other teachers in their school (average of four hours (SD 1.8) cover per teacher). These data are summarised in Table 18.

A total of 45 of 140 teachers (23.7%) from 31 schools reported receiving TDTScience training from another staff member within their school during 2023–2024, averaging 2.7 hours (SD 2.0). Fewer than ten of these teachers, all from different schools, reported that their school arranged for paid cover for them to receive this training (average of 3.3 hours (SD 1.4) cover per teacher).

A total of 77 of 190 teachers (40.5%) from 56 schools reported spending time completing other activities related to the TDTScience programme during 2023–2024, averaging 5.2 hours (SD 9.1). Fewer than ten of these teachers, all from different schools, reported that their school arranged for paid cover for them to receive this training (average of 6.1 hours (SD 9.7) cover per teacher).

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

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

Table 18: Total time devoted by personnel for training, related activities, and staff cover

2023–2024	Year 5 teachers	Year 6 teachers	Overall
	Mean (SD) number of hours per teacher	Mean (SD) number of hours per teacher	Mean (SD) number of hours per teacher
Time spent			
Sharing TDTScience training with other members of staff within their school who missed the in-person training sessions	3.2 (2.3)	2.8 (1.6)	3.2 (2.1)
Receiving TDTScience training from another staff member within their school because they missed one or more in-person training sessions	3.5 (2.7)	2.2 (1.2)	2.7 (2.0)
Completing activities related to the TDTScience programme, excluding the TDTScience training and support	4.2 (5.5)	6.7 (12.8)	5.2 (9.1)
Hours of paid cover for teacher			
Sharing TDTScience training with other members of staff within their school who missed the in-person training sessions	4.3 (2.1)	–[a]	4.0 (1.8)
Receiving TDTScience training from another staff member within their school because they missed one or more in-person training sessions	3.3 (1.5)	–[a]	3.3 (1.4)
Completing activities related to the TDTScience programme, excluding the TDTScience training and support	8.0 (12.3)	3.0 (1.0)	6.1 (9.7)

[a]*Data suppressed under statistical disclosure rules.*

A total of 34 (29.6%) of 115 teachers who said they had shared or received TDTScience training and/or spent time completing other TDTScience-related activities, from 29 schools, reported completing at least some TDTScience training or activities outside of normal working hours/in their own time during 2023–2024, averaging 2.7 hours (SD 1.7) for sharing training, 3.2 hours (SD 3.0) for receiving training, and 4.2 hours (SD 4.2) for other activities. Table 19 summarises the number of hours spent on each activity, on average, per teacher.

Table 19: Total time devoted by personnel for training and related activities outside working hours

2023–2024	Year 5 teachers	Year 6 teachers	Overall
	Mean (SD) number of hours per teacher	Mean (SD) number of hours per teacher	Mean (SD) number of hours per teacher
Time spent			
Sharing TDTScience training with other members of staff within their school who missed the in-person training sessions	2.7 (1.9)	2.7 (1.5)	2.7 (1.7)
Receiving TDTScience training from another staff member within their school because they missed one or more in-person training sessions	3.8 (3.4)	1.7 (0.6)	3.2 (3.0)
Completing activities related to the TDTScience programme, excluding the TDTScience training and support	3.9 (3.6)	4.7 (5.2)	4.2 (4.2)

Financial costs

Staff cover

As detailed above, paid staff cover was used variably by schools in the second year of the trial 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 0.46 hours per teacher (1.04 hours per school) in the academic year 2023–2024 (0.71 hours for Year 5 teachers, and 0.17 hours for Year 6 teachers). To calculate the average cost per school, we multiplied these figures by £18.33 (see 'Methods' subsection 'Assumptions' above). Therefore, on average, per school, £19.00 was spent on cover during the academic year 2023–2024 (Table 20). This is considered as a recurring cost.

Materials

Additional materials

A total of 23 teachers (16.2% of 142 who reported that they had implemented the TDTScience programme with a Year 5 and or Year 6 class during the 2023–2024 school year) from 20 schools, 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 190 teachers, teachers reported spending £7.20 on additional resources (range £0 to £400, median £0) in the year 2023–2024 (average £16.10 per school) (Table 20). On average, Year 5 teachers spent twice as much as Year 6 teachers (£9.37 vs £4.58, median £0 for both), average £12.70 and £5.60, respectively, per school.

A range of items were purchased, and no two teachers reported purchasing the same items. Examples include force meters, plants, balloons, fruit, glitter, and magnets (not an exhaustive list).

Since not all schools reported purchasing additional materials this is considered an optional rather than mandatory cost for implementation.

Table 20: 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 (year one only), 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	£19.00	£19.00	£541.32
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	£16.10	£16.10	£32.20
Total cost of items per school			**£543.45**	**£35.10**	**£35.10**	**£613.65**
Total cost per pupil (43 pupils per school)						**£14.27**
Total cost per pupil per school year (43 pupils per school over three years)						**£4.76**

Overall costs

Table 21 presents the updated total cost per pupil per year over three years based on the trial costs of £71.71.

Table 21: 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	£541.32	£4.20
Staff travel	Start-up	£24.40	£0.19
Optional materials	Start-up / recurring cost per school	£47.93	£0.37
Total		**£9,249.64**	**£71.71**

Conclusion

Table 22: Key conclusions

Key conclusions
1. The impact of TDTScience on Cohort 2 Year 5 pupils' science attainment was consistent with the main trial results (Robinson-Smith *et al.,* 2025). When teachers had an additional year of experience delivering TDTScience, Cohort 2 Year 5 pupils made no additional month's progress in science attainment, on average, compared to pupils in control schools. This finding was also consistent among pupils eligible for FSM. These results do not have a security rating.
2. The impact of TDTScience on Cohort 2 Year 5 pupils' science attitudes was consistent with the main trial results (Robinson-Smith *et al.,* 2025). Cohort 2 Year 5 pupils demonstrated small positive effects in attitudes to science, in TDTScience schools, compared to pupils in control schools.
3. Considering the longer-term impact of TDTScience, Year 6 pupils from Cohort 1, made no additional month's progress in science attainment, and in Key Stage 2 English and Maths SATs, on average, compared to pupils in control schools.
4. Year 5 teachers in TDTScience schools reported that they were more confident in teaching science, and reported higher levels of engagement in, and enjoyment of science among their pupils, in comparison to the control group. Year 6 teachers in TDTScience schools also reported that their pupils from Cohort 1 were more engaged, progressive, and confident in science, and enjoyed science lessons more in comparison to pupils in control schools in Year 6.
5. Teacher surveys indicated that many aspects of the TDTScience approach continued to be implemented frequently within Year 5 science lessons by teachers in many intervention schools. Teaching practice in Year 6 science lessons also differed, with intervention schools adopting aspects of TDTScience techniques and strategies.

Interpretation

Research investigating the impact of TDTScience on Year 5 pupils' science attainment and attitudes has taken place over the past decade, and the results relating to attainment have been mixed. Hanley, Slavin, and Elliott (2015) reported the findings of a small-scale efficacy trial, which showed that Year 5 pupils of teachers trained in TDTScience made three months' additional progress in science, with a particularly positive effect among girls, pupils with lower prior attainment, and pupils eligible for FSM. There was also a positive impact on attitudes towards science. A subsequent effectiveness trial including 205 schools was commissioned and reported in 2018. This trial reported no evidence of additional attainment 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 (Kitmitto *et al.,* 2018). Following this, the developers of TDTScience —Science Oxford and Oxford Brooks University—worked to strengthen the train-the-trainers model used to deliver the intervention to teachers in schools, and this process was reported by Standley *et al*. (2023). This work took place ahead of this second large effectiveness trial, which was conducted over two academic years, 2022–2023 and 2023–2024. The results relating to the first year of the trial are reported in Robinson-Smith *et al*. (2025), and this addendum report presents the results of the second year (2023–2024). The purpose of the second year of the trial was to examine the legacy of the TDTScience training and any effects of embedding the TDTScience practices in the academic year following training.

The trial followed a cohort of pupils who were recruited at the start of Year 5 in 2022–2023 until the end of Year 6 in 2023–2024, known as Cohort 1. Teachers in schools allocated to the intervention were offered in-person TDTScience training during 2022–2023. As presented in Robinson-Smith *et al*. (2025), at the end of 2022–2023 Year 5 pupils in TDTScience schools did not make any additional months' progress in science in comparison to pupils in schools in the control group. The second year of the trial followed the same cohort of pupils until the end of Year 6. When assessed using the Year 6 Science Assessment (Baird *et al.,* in press), Year 6 pupils in intervention schools did not make any additional months' progress in science in comparison to schools in the control group. These findings imply that TDTScience does not have a longer-term effect on science attainment. The longer-term impact on non-science subjects, namely Key Stage 2 English and Maths, was evaluated using a non-inferiority framework to assess that the intervention did not have a substantially negative impact on these subjects. The lower confidence limit for these results equate to one or two months' less progress in these subjects, respectively. While a non-inferiority margin was not formally pre-specified for these outcomes to compare these results against, these are very small effects, which would suggest any potentially negative impact on these subjects would be minimal.

There was evidence from the Year 6 teachers' survey at the end of 2023–2024 that around a quarter had attended in-person TDTScience training in 2022–2023; anecdotally we consider that these teachers were Year 5 teachers in 2022–2023 who had moved into the Year 6 class in 2023–2024. Some respondents to the Year 6 survey also indicated that they had received some cascaded training on TDTScience from trained teachers within their school. Survey data indicated that strategies and techniques used in Year 6 science lessons in intervention schools aligned more to the TDTScience approach—such as more regularly designing/planning and interpreting practical investigations—in comparison to control schools. When in Year 5, Cohort 1 pupils completed a pupil attitudes survey in Summer Term 2023, and analyses suggested that pupils from schools that were offered TDTScience had a small improvement in their attitudes to science (Robinson-Smith *et al.,* 2025). Year 6 teachers in intervention schools also reported that their pupils (Cohort 1) were more engaged, progressive, and confident in science, and enjoyed science lessons more in comparison to pupils in control schools.

In order to understand if the Year 5 teachers embedded the TDTScience approach into their teaching practices in 2023–2024, a subsequent cohort of Year 5 pupils (Cohort 2) were recruited and their science attainment assessed in Summer Term 2024 using the Year 5 Science Assessment (Joshi *et al.,* 2022). These data indicated that Year 5 pupils in intervention schools did not make any additional months' progress in science in comparison to the control schools. From this we can summarise that there was no evidence of a 'soak effect' of the intervention in relation to attainment. This may be because the compliance data indicated that only 46% of Year 5 pupils (Cohort 2) in 2023–2024 were taught by a teacher who attended in-person TDTScience training during the first year of the trial (this is discussed as a limitation in the 'Limitations' section below). There was also no evidence that the intervention was more or less beneficial among pupils eligible for FSM than those not eligible.

The number of pupils included in the analysis of the Cohort 2 Year 5 Science Assessment outcome means this analysis would have had 80% power to detect an effect size of 0.15 if such a difference existed. This is a fairly small effect size, while still representing a meaningful difference, so the trial was well-powered. Ultimately, the effect size observed was much smaller than this (0.03), so while the trial was not powered for this small effect size, such an effect size would not represent a meaningful improvement in attainment.

There was evidence that Year 5 pupils in 2023–2024 (Cohort 2) from schools that were offered TDTScience had a small improvement in their attitudes to science. This is supported from the survey data, where teachers from intervention schools reported greater enjoyment and engagement with science lessons among their pupils in comparison to those in control schools. Such gains were present for Cohort 1 (Robinson-Smith *et al.,* 2025), and consistent with the findings from previous evaluations of TDTScience (Kitmitto *et al*., 2018; Hanley, Slavin, and Elliott, 2015).

Teacher survey data indicated that the time spent on Year 5 science lessons was similar across groups, although there was evidence of differences in the frequency of use of some usual techniques and strategies used during science lessons, with intervention schools using practical work more often than control schools. TDTScience trained Year 5 teachers reported that they continued to implement aspects of the TDTScience approach during science lessons in 2023–2024. The increased confidence in teaching science reported by Year 5 teachers in intervention schools at the end of 2022–2023 was sustained at the end of 2023–2024 in comparison teachers in control schools. Primary teachers are often reported to lack confidence in teaching science, and improving teacher confidence and primary pupils science experience can influence subsequent subject attitudes (e.g. see Markwick and Reiss 2023; Ofsted 2021; Slavin *et al.,* 2014; Harlen and Qualter, 2018). A meta-analysis found a positive and moderate relationship between science attitudes and attainment (Mao *et al.,* 2021) suggesting that enhancing pupils' positive attitudes towards science could be conducive to students' learning. Although the increase in positive attitudes towards science did not translate into greater attainment for pupils in either cohort within this trial, it remains to be seen whether this increased favourable attitude to science influences educational decisions and attainment further into the future.

The estimated cost of TDTScience is £71.71 per pupil 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. It is therefore, a reasonably low-cost intervention. However, this is an average cost and the costs to school were reasonably variable so the actual cost per pupil for a school could be higher or lower than this figure.

Limitations

Overall, this was a well-powered, two-armed, cluster randomised controlled, effectiveness trial, conducted over two years; however, there were a few limitations relating to the design, which are already detailed in Robinson-Smith *et al*. (2025). Those specific to the second year of the trial are detailed below.

1. Of the 168 schools who were recruited for Cohort 1 and returned pupil details for the collection of post-test data in Year 5 in 2022–2023, 160 schools agreed to take part in Cohort 2 testing. Seven schools agreed to post-test Cohort 1 but not recruit a Cohort 2, while one school refused to both follow-up Cohort 1 in Year 6 and recruit a Cohort 2. In total, 591 Cohort 1 pupils and 482 Cohort 2 pupils were absent on the day of attainment outcome testing, this figure could have been decreased with mop-up testing. However, it was not considered financially or practically viable to resend invigilators to schools for mop-up testing for fewer than 15 pupils; and there were no instances where 15 or more pupils missed the assessment during the initial school visit.
2. Both the Year 5 and Year 6 Science Attainment outcomes were assessed and marked blinded to intervention, with limited recorded instances of unblinding. At school level, unblinding occurred eight times in the control group and five times in the intervention group. Even where unblinding occurred, it is likely that the invigilator(s) was not unblinded to all assessments at the school; although we have assumed that at such schools all assessments were invigilated unblinded for simplicity and transparency. Nevertheless, 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.
3. The pupil attitudes survey was administered by teachers as opposed to independent research invigilators due to practical and financial constraints, i.e. 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.
4. At the point of recruitment, schools were advised that Year 5 teachers should remain teaching Year 5 pupils for the duration of the trial (September 2022 to July 2024), unless circumstances made this impossible. As although the TDTScience training was only delivered in the first year, the expectation was that teachers would continue to use TDTScience in their science lessons in the second year. Compliance data indicated that only 46% of pupils were taught by a teacher who attended at least three out of four of the TDTScience in-person training days. These figures suggest it is not usual practice or practical for schools to retain teachers in the same class from year to year. This has implications for the sustainability of interventions and initiatives that are tailored specifically for use in certain year groups and delivered within an academic year—such as TDTScience—and the evaluation of them within longitudinal research.
5. The CACE analysis did not suggest that increased compliance with the intervention resulted in greater gains in attainment. However, the definition of compliance was limited as it only considered teacher attendance at training, not whether, or how often/well, teachers incorporated what they had learned into their teaching.
6. Usual practice surveys were administered to the teacher(s) of Cohort 1 (Year 6) and Cohort 2 (Year 5) at the end of 2023–2024. When administering the end of 2023–2024 surveys, we requested that the Year 6 (science) teacher of Cohort 1 complete the Year 6 survey. Similarly, we requested that the Year 5 (science) teacher(s) of Cohort 2 complete the survey; in intervention schools this teacher may or may not have received the TDTScience training in 2022. Survey data indicated that 63% of respondents to the Year 5 survey received in-person TDTScience training in 2022–2023. We acknowledge that not every teacher will have completed all surveys, which may result in response bias and is a limitation of the comparisons made between survey data.
7. 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. New data presented here also demonstrated that there is no benefit on science, maths, or reading attainment at the end of Year 6. We would therefore, not recommend any further evaluation of attainment at this timepoint.

A further piece of research could investigate if the increased attitude towards science observed within this trial in both Cohort 1 and Cohort 2 are sustained. This could involve following pupils and evaluating their further education choices.

The evaluation team are exploring the possibility of publishing a paper relating to Key Stage 2 teacher confidence in teaching science.

References

Baird, K., Crompton, Z., Elliott, L., Fountain, I., Hanley, P., Davill, T., Hugill-Jones, J., Robinson-Smith, L., Whiteside, K., Keding, A. (in press) *'Development of a Year 6 Science Assessment for Use as an Evaluation Outcome Measure. Measure Development Report'*. London: Education Endowment Foundation.

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, 443–50. https://doi.org/10.1080/01621459.1995.10476536

Braun, V. and Clarke, V. (2006) 'Using Thematic Analysis in Psychology'. *Qualitative Research in Psychology*, 3: 2, 77–101. https://doi.org/10.1191/1478088706qp063oa

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, 369–95. https://doi.org/10.1191/0962280205sm403oa

Education Endowment Foundation (EEF). (2023) '*Cost Evaluation Guidance for EEF Evaluations'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/documents/evaluation/evaluation-design/Cost-Evaluation-Guidance-Feb_2023.pdf?v=1684358910 (accessed 01/04/2024).

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 01/04/2023).

Fairhurst, C. (2023) '*Thinking, Doing, Talking Science (TDTS): Statistical Analysis Plan'*. London: EEF. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/TDTS_SAP-for-publication.pdf?v=1737043051 (accessed 01/04/2024).

Hanley, P., Elliott, L., Fountain, I., Kalpita, B., Keding, A. and Crompton, Z. (2022) *'Development of Year 6 Science Assessment: Protocol'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/outcome-measures-and-databases/Y6-science-measure-development-protocol.pdf?v=1746559325 (accessed 01/04/2025).

Hanley, P., Elliott, L., Fountain, I., Roche, J., Mandefield, L., Fairhurst, C. and Robinson-Smith, L. (2023) *'Trial Evaluation Protocol: Thinking, Doing, Talking Science'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/TDTS-protocol-v1.1-update-for-publication.pdf?v=1737043051 (accessed 01/04/2024).

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).

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

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, 871–93. 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)'*. Thousand Oaks: SAGE Publications, Inc.

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: 784068. https://doi.org/10.3389/fpsyg.2021.784068

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, 1339–59. https://doi.org/10.1080/09500693.2023.2288660

Office for Standards in Education, Children's Services and Skills (Ofsted). (2021) *'Research Review Series: Science'*. GOV.UK. www.gov.uk/government/publications/research-review-series-science/research-review-series-science (accessed 15/09/2025).

Robinson-Smith, L., Fairhurst, C., Tracey, L., Fountain, I., Whiteside, K., Davill, T., Hugill-Jones, J., Turkenburg van-Diepen, M., Brakovic-Thomas, M., Bottomley-Wise, R., Hanley, P., Elliott, L. and Torgerson, D. (2025) *'Thinking, Doing, Talking Science (Second Re-Grant – A Two-Armed, Cluster Randomised Trial). Evaluation Report'*. London: Education Endowment Foundation. Available at: https://d2tic4wvo1iusb.cloudfront.net/production/documents/projects/tdts_evaluation_report_2025-07-02-150320_ixre.pdf?v=1751810751 (accessed 02/05/2025).

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

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

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, 518–29. https://doi.org/10.1198/073500102288618658

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: Effect size estimation

Table 1: Effect size estimation

			Intervention group		Control group		
Outcome	Unadjusted differences in means (95% CI)	Adjusted differences in means (95% CI)	n (missing)	Variance of outcome	n (missing)	Variance of outcome	Pooled variance
Cohort 2 Year 5 Science Assessment (primary addendum analysis)	-0.03 (-0.43 to 0.37)	0.27 (-0.55 to 1.09)	3,269 (413)	63.02	2,818 (363)	62.95	63.24
Cohort 2 Year 5 Science Assessment (eligible for FSM)	-0.20 (-0.89 to 0.49)	0.24 (-0.72 to 1.21)	967 (138)	56.68	826 (149)	53.86	55.29
Cohort 2 Year 5 Science Attitudes Questionnaire	0.07 (0.02 to 0.12)	0.08 (-0.03 to 0.20)	3,202 (480)	0.96	2,729 (452)	1.04	1.01
Cohort 1 Year 6 Science Assessment	-0.10 (-0.46 to 0.26)	-0.16 (-0.84 to 0.52)	3,118 (672)	49.02	2,823 (648)	51.15	50.09
Cohort 1 Key Stage 2 Maths	0.14 (-0.22 to 0.50)	-0.14 (-0.96 to 0.67)	3,536 (254)	56.93	3,213 (258)	54.88	56.08
Cohort 1 Key Stage 2 English	0.23 (-0.14 to 0.60)	0.05 (-0.65 to 0.74)	3,539 (251)	57.77	3,212 (259)	60.96	59.51

Appendix B: IPE data tables

Table 1: Year 5 teachers' highest qualification in science at 2023–2024 follow-up, by allocated group

Qualification	Intervention n (%)	Control n (%)	Total n (%)
GCSE or equivalent	67 (65.0)	56 (58.9)	123 (62.1)
A Level or equivalent	11 (10.7)	17 (17.9)	28 (14.1)
Undergraduate degree or equivalent	8 (7.8)	7 (7.4)	15 (7.6)
Postgraduate degree	16 (15.5)	11 (11.6)	27 (13.6)
Other	0 (0.0)	2 (2.1)	2 (1.0)
Missing	1 (1.0)	2 (2.1)	3 (1.5)
Total n	**103**	**95**	**198**

Table 2: Other science training or initiatives undertaken by Year 5 teachers during 2023–2024, by allocated group

Training or initiative	Intervention n (%)	Control n (%)	Total n (%)
Primary Science Quality Mark	8 (7.8)	4 (4.2)	12 (6.1)
Primary Science Teaching Trust	1 (1.0)	1 (1.1)	2 (1.0)
Teacher Assessment in Primary Science	7 (6.8)	6 (6.3)	13 (6.6)
STEM Learning/Science Learning Centre courses	11 (10.7)	11 (11.6)	22 (11.1)
Association for Science Education (ASE) courses	1 (1.0)	0 (0.0)	1 (0.5)
Secondary school collaboration	2 (1.9)	0 (0.0)	2 (1.0)
Other	6 (5.8)	4 (4.2)	10 (5.1)
No	75 (72.8)	74 (77.9)	149 (75.3)
Total n	**103**	**95**	**198**

Table 3: Year 5 teachers changes to teaching science in 2023–2024, by allocated group

'I have changed the way I teach science'	Intervention n (%)	Control n (%)	Total n (%)
Agree strongly	35 (34.0)	10 (10.5)	45 (22.7)
Agree slightly	47 (45.6)	37 (38.9)	84 (42.4)
Neither agree nor disagree	17 (16.5)	31 (32.6)	48 (24.2)
Disagree slightly	2 (1.9)	10 (10.5)	12 (6.1)
Disagree strongly	0 (0.0)	4 (4.2)	4 (2.0)
Missing	2 (1.9)	3 (3.2)	5 (2.5)
Total n	**103**	**95**	**198**

Table 4: Strategies and techniques used by teachers in Year 5 science lessons at the end of 2023–2024, by allocated group

Strategy or technique	Intervention n (%)	Control n (%)	Total n (%)
Teacher demonstration			
Very often	24 (23.3)	19 (20.0)	43 (21.7)
Quite often	57 (55.3)	54 (56.8)	111 (56.1)
Not very often	18 (17.5)	17 (17.9)	35 (17.7)
Rarely	2 (1.9)	3 (3.2)	5 (2.5)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Pupil discussion in whole class			
Very often	82 (79.6)	78 (82.1)	160 (80.8)
Quite often	19 (18.4)	14 (14.7)	33 (16.7)
Not very often	0 (0.0)	1 (1.1)	1 (0.5)
Rarely	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Pair or small-group discussion			
Very often	79 (76.7)	72 (75.8)	151 (76.3)
Quite often	22 (21.4)	21 (22.1)	43 (21.7)
Not very often	0 (0.0)	0 (0.0)	0 (0.0)
Rarely	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Give pupils time to think			
Very often	79 (76.7)	68 (71.6)	147 (74.2)
Quite often	20 (19.4)	24 (25.3)	44 (22.2)
Not very often	2 (1.9)	1 (1.1)	3 (1.5)
Rarely	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Teach scientific facts			
Very often	55 (53.4)	70 (73.7)	125 (63.1)
Quite often	46 (44.7)	20 (21.1)	66 (33.3)
Not very often	0 (0.0)	2 (2.1)	2 (1.0)
Rarely	0 (0.0)	1 (1.1)	1 (0.5)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Teach scientific terminology			
Very often	71 (68.9)	70 (73.7)	141 (71.2)
Quite often	28 (27.2)	22 (23.2)	50 (25.3)
Not very often	2 (1.9)	1 (1.1)	3 (1.5)
Rarely	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Ask pupils to solve scientific problems			
Very often	22 (21.4)	13 (13.7)	35 (17.7)
Quite often	64 (62.1)	54 (56.8)	118 (59.6)
Not very often	12 (11.7)	23 (24.2)	35 (17.7)
Rarely	3 (2.9)	3 (3.2)	6 (3.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Require pupils to write up the whole investigation			
Very often	3 (2.9)	4 (4.2)	7 (3.5)

Strategy or technique	Intervention n (%)	Control n (%)	Total n (%)
Quite often	8 (7.8)	22 (23.2)	30 (15.2)
Not very often	27 (26.2)	32 (33.7)	59 (29.8)
Rarely	27 (26.2)	22 (23.2)	49 (24.7)
Never	36 (35.0)	13 (13.7)	49 (24.7)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Require pupils to memorise facts and principles			
Very often	7 (6.8)	11 (11.6)	18 (9.1)
Quite often	32 (31.1)	37 (38.9)	69 (34.8)
Not very often	32 (31.1)	29 (30.5)	61 (30.8)
Rarely	22 (21.4)	12 (12.6)	34 (17.2)
Never	8 (7.8)	4 (4.2)	12 (6.1)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Pupil practical work			
Very often	41 (39.8)	18 (18.9)	59 (29.8)
Quite often	58 (56.3)	64 (67.4)	122 (61.6)
Not very often	2 (1.9)	9 (9.5)	11 (5.6)
Rarely	0 (0.0)	2 (2.1)	2 (1.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Design or plan practical investigations			
Very often	15 (14.6)	4 (4.2)	19 (9.6)
Quite often	59 (57.3)	64 (67.4)	123 (62.1)
Not very often	25 (24.3)	23 (24.2)	48 (24.2)
Rarely	2 (1.9)	2 (2.1)	4 (2.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Interpret results from practical investigations			
Very often	15 (14.6)	10 (10.5)	25 (12.6)
Quite often	69 (67.0)	59 (62.1)	128 (64.6)
Not very often	16 (15.5)	23 (24.2)	39 (19.7)
Rarely	1 (1.0)	1 (1.1)	2 (1.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Use results from practical investigations to support conclusions			
Very often	16 (15.5)	15 (15.8)	31 (15.7)
Quite often	66 (64.1)	56 (58.9)	122 (61.6)
Not very often	19 (18.4)	21 (22.1)	40 (20.2)
Rarely	0 (0.0)	1 (1.1)	1 (0.5)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Total n	**103**	**95**	**198**

Table 5: Year 5 teachers' understanding and continued use of TDTScience during 2023–2024

Statement	Intervention n (%)
The training prepared me well for using the TDTScience strategies	
Agree strongly	55 (64.0)
Agree slightly	19 (22.1)
Neither agree nor disagree	8 (9.3)
Disagree slightly	0 (0.0)
Disagree strongly	1 (1.2)
Missing	3 (3.5)
The materials provided increased my understanding of the TDTScience strategies	
Agree strongly	56 (65.1)
Agree slightly	20 (23.3)
Neither agree nor disagree	6 (7.0)
Disagree slightly	1 (1.2)
Disagree strongly	1 (1.2)
Missing	2 (2.3)
I understand how to implement Bright Ideas Time	
Agree strongly	58 (67.4)
Agree slightly	19 (22.1)
Neither agree nor disagree	4 (4.7)
Disagree slightly	1 (1.2)
Disagree strongly	2 (2.3)
Missing	2 (2.3)
I understand how to implement Practical Prompts for Thinking	
Agree strongly	46 (53.5)
Agree slightly	29 (33.7)
Neither agree nor disagree	7 (8.1)
Disagree slightly	0 (0.0)
Disagree strongly	2 (2.3)
Missing	2 (2.3)
I understand how to implement practical investigations	
Agree strongly	45 (52.3)
Agree slightly	33 (38.4)
Neither agree nor disagree	5 (5.8)
Disagree slightly	1 (1.2)
Disagree strongly	0 (0.0)
Missing	2 (2.3)
I understand how to implement practical problem-solving	
Agree strongly	33 (38.4)
Agree slightly	37 (43.0)
Neither agree nor disagree	10 (11.6)
Disagree slightly	2 (2.3)
Disagree strongly	2 (2.3)
Missing	2 (2.3)
I understand how to use focused recording	
Agree strongly	46 (53.5)
Agree slightly	27 (31.4)
Neither agree nor disagree	8 (9.3)
Disagree slightly	1 (1.2)
Disagree strongly	2 (2.3)
Missing	2 (2.3)
I understand how to construct a TDTScience lesson	
Agree strongly	48 (55.8)

Statement	Intervention n (%)
Agree slightly	27 (31.4)
Neither agree nor disagree	6 (7.0)
Disagree slightly	2 (2.3)
Disagree strongly	1 (1.2)
Missing	2 (2.3)
During the current school year 2023–2024, the TDTScience strategies have been useful to me for teaching science	
Agree strongly	54 (62.8)
Agree slightly	20 (23.3)
Neither agree nor disagree	8 (9.3)
Disagree slightly	1 (1.2)
Disagree strongly	0 (0.0)
Missing	3 (3.5)
During the current school year 2023–2024, I've used the TDTScience strategies regularly	
Agree strongly	54 (62.8)
Agree slightly	20 (23.3)
Neither agree nor disagree	9 (10.5)
Disagree slightly	1 (1.2)
Disagree strongly	0 (0.0)
Missing	2 (2.3)
During the current school year 2023–2024, I've been able to implement the TDTScience strategies effectively	
Agree strongly	42 (48.8)
Agree slightly	32 (37.2)
Neither agree nor disagree	8 (9.3)
Disagree slightly	2 (2.3)
Disagree strongly	0 (0.0)
Missing	2 (2.3)
During the current school year 2023–2024, my pupils have enjoyed the TDTScience activities	
Agree strongly	54 (62.8)
Agree slightly	25 (29.1)
Neither agree nor disagree	5 (5.8)
Disagree slightly	0 (0.0)
Disagree strongly	0 (0.0)
Missing	2 (2.3)
Total n	**86**

Table 6: Year 5 teachers' confidence in teaching science at the end of 2023–2024, by allocated group

Statement	Intervention n (%)	Control n (%)	Total n (%)
Inspiring pupils to learn science			
Very confident	35 (34.0)	26 (27.4)	61 (30.8)
Quite confident	64 (62.1)	59 (62.1)	123 (62.1)
Not very confident	2 (1.9)	8 (8.4)	10 (5.1)
Not at all confident	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Planning science lessons			
Very confident	36 (35.0)	30 (31.6)	66 (33.3)
Quite confident	61 (59.2)	55 (57.9)	116 (58.6)
Not very confident	4 (3.9)	8 (8.4)	12 (6.1)
Not at all confident	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Explaining science concepts or principles by doing science practicals			
Very confident	29 (28.2)	28 (29.5)	57 (28.8)
Quite confident	68 (66.0)	52 (54.7)	120 (60.6)
Not very confident	4 (3.9)	13 (13.7)	17 (8.6)
Not at all confident	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Providing challenging tasks for the highest achieving pupils			
Very confident	17 (16.5)	13 (13.7)	30 (15.2)
Quite confident	56 (54.4)	46 (48.4)	102 (51.5)
Not very confident	26 (25.2)	32 (33.7)	58 (29.3)
Not at all confident	2 (1.9)	2 (2.1)	4 (2.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Adapting my teaching to engage pupils' interest			
Very confident	36 (35.0)	23 (24.2)	59 (29.8)
Quite confident	61 (59.2)	64 (67.4)	125 (63.1)
Not very confident	4 (3.9)	5 (5.3)	9 (4.5)
Not at all confident	0 (0.0)	1 (1.1)	1 (0.5)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Assessing pupil learning in science			
Very confident	18 (17.5)	17 (17.9)	35 (17.7)
Quite confident	68 (66.0)	57 (60.0)	125 (63.1)
Not very confident	14 (13.6)	19 (20.0)	33 (16.7)
Not at all confident	1 (1.0)	0 (0.0)	1 (0.5)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Improving the understanding of struggling pupils			
Very confident	22 (21.4)	19 (20.0)	41 (20.7)
Quite confident	67 (65.0)	61 (64.2)	128 (64.6)
Not very confident	12 (11.7)	13 (13.7)	25 (12.6)
Not at all confident	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Helping pupils discuss scientific ideas			
Very confident	37 (35.9)	27 (28.4)	64 (32.3)
Quite confident	61 (59.2)	54 (56.8)	115 (58.1)
Not very confident	3 (2.9)	12 (12.6)	15 (7.6)
Not at all confident	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	2 (2.1)	4 (2.0)
Total n	**103**	**95**	**198**

Table 7: How interested and engaged are Year 5 pupils (Cohort 2) in learning science at the end of 2023–2024, by allocated group

Statement	Intervention n (%)	Control n (%)	Total n (%)
My pupils have enjoyed their science lessons			
Agree strongly	71 (68.9)	56 (58.9)	127 (64.1)
Agree slightly	28 (27.2)	34 (35.8)	62 (31.3)
Neither agree nor disagree	2 (1.9)	2 (2.1)	4 (2.0)
Disagree slightly	0 (0.0)	0 (0.0)	0 (0.0)
Disagree strongly	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	3 (3.2)	5 (2.5)
My pupils have made good progress in their understanding of science			
Agree strongly	50 (48.5)	44 (46.3)	94 (47.5)
Agree slightly	47 (45.6)	42 (44.2)	89 (44.9)
Neither agree nor disagree	4 (3.9)	6 (6.3)	10 (5.1)
Disagree slightly	0 (0.0)	0 (0.0)	0 (0.0)
Disagree strongly	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	3 (3.2)	5 (2.5)
My pupils are confident in science			
Agree strongly	34 (33.0)	28 (29.5)	62 (31.3)
Agree slightly	56 (54.4)	53 (55.8)	109 (55.1)
Neither agree nor disagree	11 (10.7)	10 (10.5)	21 (10.6)
Disagree slightly	0 (0.0)	1 (1.1)	1 (0.5)
Disagree strongly	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	3 (3.2)	5 (2.5)
My pupils can work independently in science			
Agree strongly	23 (22.3)	18 (18.9)	41 (20.7)
Agree slightly	59 (57.3)	57 (60.0)	116 (58.6)
Neither agree nor disagree	14 (13.6)	11 (11.6)	25 (12.6)
Disagree slightly	5 (4.9)	6 (6.3)	11 (5.6)
Disagree strongly	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	3 (3.2)	5 (2.5)
My pupils come up with their own scientific ideas			
Agree strongly	25 (24.3)	10 (10.5)	35 (17.7)
Agree slightly	52 (50.5)	53 (55.8)	105 (53.0)
Neither agree nor disagree	20 (19.4)	18 (18.9)	38 (19.2)
Disagree slightly	4 (3.9)	10 (10.5)	14 (7.1)
Disagree strongly	0 (0.0)	1 (1.1)	1 (0.5)
Missing	2 (1.9)	3 (3.2)	5 (2.5)
My pupils do a lot of writing in science			
Agree strongly	2 (1.9)	5 (5.3)	7 (3.5)
Agree slightly	17 (16.5)	28 (29.5)	45 (22.7)
Neither agree nor disagree	19 (18.4)	29 (30.5)	48 (24.2)
Disagree slightly	34 (33.0)	23 (24.2)	57 (28.8)
Disagree strongly	29 (28.2)	7 (7.4)	36 (18.2)
Missing	2 (1.9)	3 (3.2)	5 (2.5)
My pupils have been engaged with science			
Agree strongly	59 (57.3)	43 (45.3)	102 (51.5)
Agree slightly	41 (39.8)	44 (46.3)	85 (42.9)
Neither agree nor disagree	0 (0.0)	4 (4.2)	4 (2.0)
Disagree slightly	1 (1.0)	1 (1.1)	2 (1.0)
Disagree strongly	0 (0.0)	0 (0.0)	0 (0.0)
Missing	2 (1.9)	3 (3.2)	5 (2.5)
Total n	**103**	**95**	**198**

Table 8: Strategies and techniques used by teachers in Year 6 science lessons at the end of 2023–2024, by allocated group

Strategy or technique	Intervention n (%)	Control n (%)	Total n (%)
Pupil practical work			
Very often	17 (19.3)	14 (14.4)	31 (16.8)
Quite often	59 (67.0)	62 (63.9)	121 (65.4)
Not very often	7 (8.0)	16 (16.5)	23 (12.4)
Rarely	1 (1.1)	1 (1.0)	2 (1.1)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Teacher demonstration			
Very often	18 (20.5)	14 (14.4)	32 (17.3)
Quite often	43 (48.9)	63 (64.9)	106 (57.3)
Not very often	19 (21.6)	14 (14.4)	33 (17.8)
Rarely	3 (3.4)	2 (2.1)	5 (2.7)
Never	1 (1.1)	0 (0.0)	1 (0.5)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Pupil discussion in whole class			
Very often	70 (79.5)	70 (72.2)	140 (75.7)
Quite often	14 (15.9)	21 (21.6)	35 (18.9)
Not very often	0 (0.0)	2 (2.1)	2 (1.1)
Rarely	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Pair or small-group discussion			
Very often	68 (77.3)	64 (66.0)	132 (71.4)
Quite often	16 (18.2)	29 (29.9)	45 (24.3)
Not very often	0 (0.0)	0 (0.0)	0 (0.0)
Rarely	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Give pupils time to think			
Very often	70 (79.5)	63 (64.9)	133 (71.9)
Quite often	14 (15.9)	30 (30.9)	44 (23.8)
Not very often	0 (0.0)	0 (0.0)	0 (0.0)
Rarely	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Teach scientific facts			
Very often	57 (64.8)	65 (67.0)	122 (65.9)
Quite often	26 (29.5)	27 (27.8)	53 (28.6)
Not very often	1 (1.1)	1 (1.0)	2 (1.1)
Rarely	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Teach scientific terminology			
Very often	66 (75.0)	68 (70.1)	134 (72.4)
Quite often	16 (18.2)	24 (24.7)	40 (21.6)
Not very often	2 (2.3)	1 (1.0)	3 (1.6)
Rarely	0 (0.0)	0 (0.0)	0 (0.0)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Ask pupils to solve scientific problems			
Very often	21 (23.9)	16 (16.5)	37 (20.0)
Quite often	45 (51.1)	56 (57.7)	101 (54.6)
Not very often	15 (17.0)	17 (17.5)	32 (17.3)

Strategy or technique	Intervention n (%)	Control n (%)	Total n (%)
Rarely	3 (3.4)	4 (4.1)	7 (3.8)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Design or plan practical investigations			
Very often	13 (14.8)	5 (5.2)	18 (9.7)
Quite often	54 (61.4)	63 (64.9)	117 (63.2)
Not very often	16 (18.2)	19 (19.6)	35 (18.9)
Rarely	1 (1.1)	6 (6.2)	7 (3.8)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Interpret results from practical investigations			
Very often	10 (11.4)	5 (5.2)	15 (8.1)
Quite often	56 (63.6)	61 (62.9)	117 (63.2)
Not very often	16 (18.2)	26 (26.8)	42 (22.7)
Rarely	2 (2.3)	1 (1.0)	3 (1.6)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Use results from practical investigations to support conclusions			
Very often	12 (13.6)	7 (7.2)	19 (10.3)
Quite often	56 (63.6)	59 (60.8)	115 (62.2)
Not very often	14 (15.9)	26 (26.8)	40 (21.6)
Rarely	2 (2.3)	1 (1.0)	3 (1.6)
Never	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Require pupils to write up the whole investigation			
Very often	3 (3.4)	4 (4.1)	7 (3.8)
Quite often	11 (12.5)	25 (25.8)	36 (19.5)
Not very often	32 (36.4)	34 (35.1)	66 (35.7)
Rarely	18 (20.5)	20 (20.6)	38 (20.5)
Never	20 (22.7)	10 (10.3)	30 (16.2)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Require pupils to memorise facts and principles			
Very often	11 (12.5)	15 (15.5)	26 (14.1)
Quite often	34 (38.6)	48 (49.5)	82 (44.3)
Not very often	25 (28.4)	18 (18.6)	43 (23.2)
Rarely	12 (13.6)	10 (10.3)	22 (11.9)
Never	2 (2.3)	2 (2.1)	4 (2.2)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Total n	**88**	**97**	**185**

Table 9: How interested and engaged are Year 6 pupils (Cohort 1) in learning science at the end of 2023–2024, by allocated group

Statement	Intervention n (%)	Control n (%)	Total n (%)
My pupils have enjoyed their science lessons			
Agree strongly	48 (54.5)	39 (40.2)	87 (47.0)
Agree slightly	34 (38.6)	49 (50.5)	83 (44.9)
Neither agree nor disagree	2 (2.3)	3 (3.1)	5 (2.7)
Disagree slightly	0 (0.0)	2 (2.1)	2 (1.1)
Disagree strongly	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
My pupils have made good progress in their understanding of science			
Agree strongly	45 (51.1)	36 (37.1)	81 (43.8)
Agree slightly	37 (42.0)	50 (51.5)	87 (47.0)
Neither agree nor disagree	2 (2.3)	6 (6.2)	8 (4.3)
Disagree slightly	0 (0.0)	1 (1.0)	1 (0.5)
Disagree strongly	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
My pupils are confident in science			
Agree strongly	31 (35.2)	22 (22.7)	53 (28.6)
Agree slightly	41 (46.6)	56 (57.7)	97 (52.4)
Neither agree nor disagree	12 (13.6)	12 (12.4)	24 (13.0)
Disagree slightly	0 (0.0)	3 (3.1)	3 (1.6)
Disagree strongly	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
My pupils can work independently in science			
Agree strongly	26 (29.5)	20 (20.6)	46 (24.9)
Agree slightly	42 (47.7)	54 (55.7)	96 (51.9)
Neither agree nor disagree	12 (13.6)	17 (17.5)	29 (15.7)
Disagree slightly	3 (3.4)	2 (2.1)	5 (2.7)
Disagree strongly	1 (1.1)	0 (0.0)	1 (0.5)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
My pupils come up with their own scientific ideas			
Agree strongly	19 (21.6)	10 (10.3)	29 (15.7)
Agree slightly	42 (47.7)	49 (50.5)	91 (49.2)
Neither agree nor disagree	16 (18.2)	22 (22.7)	38 (20.5)
Disagree slightly	6 (6.8)	11 (11.3)	17 (9.2)
Disagree strongly	1 (1.1)	1 (1.0)	2 (1.1)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
My pupils have been engaged with science			
Agree strongly	44 (50.0)	38 (39.2)	82 (44.3)
Agree slightly	38 (43.2)	45 (46.4)	83 (44.9)
Neither agree nor disagree	1 (1.1)	9 (9.3)	10 (5.4)
Disagree slightly	1 (1.1)	1 (1.0)	2 (1.1)
Disagree strongly	0 (0.0)	0 (0.0)	0 (0.0)
Missing	4 (4.5)	4 (4.1)	8 (4.3)
Total n	**88**	**97**	**185**

Appendix C: EEF cost rating

Figure 10: 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.

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

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn