Thinking, Doing, Talking Science: Pre-trial Report Train-the-Trainers Phase

September 2023

Emma Standley
Rachel Margaret Carr
Pam Hanley
Heather Leggett
Imogen Fountain
Gill Parkinson
Caroline Fairhurst
Lyn Robinson-Smith

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

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

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

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus Trust (now part of Impetus - Private Equity Foundation) and received a founding £125m grant from the Department for Education. Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

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

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

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from the York Trials Unit, University of York: Dr Pam Hanley, Emma Standley, Dr Rachel Margaret Carr, Dr Heather Leggett, Imogen Fountain, Caroline Fairhurst, Gill Parkinson, Louise Elliott, and Dr Lyn Robinson-Smith.

The principal investigator was Dr Pam Hanley until December 2022, and Dr Lyn Robinson-Smith from December 2022.

Contact details

Dr Lyn Robinson-Smith
Research Fellow
York Trials Unit
University of York
YO10 5DD

Tel: 01904 321692
Email: lyn.robinson-smith@york.ac.uk

Acknowledgements

Thanks to all the colleges and schools that participated in this trial, particularly the many teachers and trainers who continued to contribute despite the difficulties caused by the COVID-19 pandemic. We also recognise the support and encouragement we received from the EEF team: Guillermo Romero, Katie Luxton, and (formerly) Celeste Cheung.

Executive summary

This report focuses on the research conducted before the main trial of Thinking, Doing, Talking Science (TDTScience). Its purpose was to evaluate the train-the-trainers model and draw any lessons for future scale-ups from developer-delivered interventions to trainer-delivered interventions. The TDTScience programme taught and supported teachers to embed opportunities for higher order thinking into their teaching through discussion strategies and approaches to practical work. TDTScience involves primary school staff attending four and a half days of training over a school year; the training is designed to be delivered to primary teachers by two trainers.

The study focused on the extent to which the training model was implemented as planned, the effectiveness and necessity of each element of the model, and how the model might be improved to benefit the TDTScience intervention and training design more widely.

The train-the-trainers phase of the study was divided into three consecutive stages with trainers fulfilling a different role in each. In Stage 1, trainers, acting as if they were teachers, received training in TDTScience from the developers. In Stage 2, trainers were trained by the developers to deliver TDTScience training to teachers. In Stage 3, trainers trained pre-trial teachers in TDTScience. The pre-trial occurred prior to the full-scale trial to test the feasibility of the research instruments (surveys, topic guides, and observation schedules) and to conduct the first three stages of the trial, primarily including training the trainers in TDTScience in preparation for the full trial. The pre-trial also allowed Science Oxford (SO)time to test the final stage of the training model.

This year-long process intended to create more accurately the real-life conditions in which the TDTScience intervention would be delivered: not by trainers who were novices at delivering the programme but by trainers with a mix of experience none of whom would be delivering it unsupported for the very first time.

Implementation generally went as planned. Despite various delays and some absences due to COVID-19, the pandemic did not seem to have a negative effect on the process. All three stages broadly met the expectations of developers and trainers.

The three pre-trial stages were designed to give trainers more understanding of TDTScience philosophy and practice, and in this they proved effective. Stages 1 and 2 succeeded in providing a solid basis for trainers to be confident about delivering the programme to teachers and during Stage 3, trainers were regularly considering how delivery could be enhanced by de-briefing with their partner and reflecting on their own practice. It was clear from trainer and teacher feedback and lesson and training observations that most teachers were engaged in TDTScience and that the approach had been put into practice in the classroom. Trainers were aware of which concepts and approaches teachers found more difficult and, where possible, adjusted their timings and emphases to accommodate this. Teachers being reflective and sharing good practice were judged particularly effective with trainers often surprised by the extent to which classroom practice seemed to have changed. Moreover, there was evidence from teachers and observations that the expected changes in pupil abilities and behaviour—such as more discussion and practical activities in lessons—were starting to take place.

It was clear that there was a need for all three stages of the pre-trial process to produce confident and competent trainers for the main trial. Being trained in TDTScience as if they were teachers (Stage 1) meant trainers could familiarise themselves with the content and approach and make suggestions for possible improvements. The train-the-trainers element (Stage 2) was valued particularly for providing the opportunity to plan and practise with their delivery partners. The real-life delivery of TDTScience to teachers (Stage 3) acted as a rehearsal for the main trial. They could refine how to run the days, gain confidence, and identify where more practice and support was needed. Stage 3 also gave them an opportunity to explore the extent to which they could make the delivery their own (by inserting personal anecdotes, for instance) while maintaining fidelity to the TDTScience programme.

Although the fundamentals of the programme were already firmly fixed, the developers were willing to consider trainers' suggested improvements at every stage of the pre-trial, and to adopt them if they were considered appropriate and easy to incorporate. This helped give trainers a sense of ownership. During Stage 3, some trainers had ideas for amendments for delivery that did not affect the content of the training and could be adopted in their own practice in the main trial.

There was little feeling among trainers that the training model could have been improved by eliminating or compressing any elements, even among those who had delivered TDTScience before, and the developers also expressed satisfaction with the process. Moreover, trainers' prior concerns about keeping to time in Stage 3 proved unfounded, with tweaks to the schedule being made throughout the days as appropriate for their groups. Preparing and delivering with a partner worked well with benefits in terms of complementary knowledge and skills and providing mutual support. Their main challenge, which they seemed to cope with successfully, was balancing fidelity to the programme with a desire to establish ownership of delivery.

Introduction

Thinking, Doing, Talking Science (TDTScience) is a continuing professional development programme for primary school teachers, evaluated via trials with Year 5 teachers, to develop their pedagogy to deliver science lessons that enhance pupils' higher order thinking (HOT) and subsequently their attainment in science. The logic model predicts that, by encouraging these HOT skills, pupils will engage more deeply and actively, developing their understanding along with an increased interest and self-efficacy in science.

Pupils are encouraged to think and talk about scientific concepts in every science lesson through dedicated discussion slots (the Bright Ideas Time) linked to the topic being taught. For example, a teacher may show three pictures and ask which one is the ''Odd One Out'' inviting pupils to offer an answer along with their reasoning. Teachers are encouraged to allow pupils to contribute until all ideas have been saturated in a bid to promote thinking and creativity. Teachers will also facilitate their pupils' thinking through doing purposeful practical science, providing them with frequent opportunities for creative investigations, problem solving, and other types of enquiry activity. Pupils will not record everything they do in a practical as the teacher will focus the recording on the lesson's learning objectives so that the time for thinking, doing, and talking is prioritised.

TDTScience evaluations

A small-scale efficacy trial (Hanley et al., 2015) suggested pupils of teachers trained in TDTSscience made three months' additional progress in science and had more positive attitudes towards science than those following 'business as usual'. However, a subsequent effectiveness trial in 205 schools (Kitmitto et al., 2018) failed to show evidence of additional progress for most pupils—although pupils eligible for free school meals (FSM) made a small amount of additional progress and pupils' attitudes to science showed a small improvement.

The main changes between the two trials related to teacher training. In contrast to the efficacy trial, the effectiveness trial used a 'train-the-trainers' model rather than the developers training the teachers directly. Train-the-trainer is the theory that a group of individuals can be given training in a new concept and then go on to train a large group in this newly acquired skill (Ray et al., 2012). A further difference was that the continuing professional development (CPD) days were reduced from five to four and the funding to cover two in-school preparation days per teacher was eliminated.

After the first effectiveness trial, the Science Oxford team recommended adapting the train-the-trainers model to improve the impact of TDTScience in the event of a retrial (internal report, 2019) in four main ways:

- before trainers start delivering any training to teachers they receive the full TDTScience course—as if they were teachers—as well as a complete package of training in how to deliver the training to teachers (previously, trainers received training throughout the intervention year in tandem with delivering to teachers);
- trainers deliver to teachers in pre-trial schools before the main trial begins;
- quality assurance was improved, for instance, by having developers observe trainers delivering the training to teachers in pre-trial schools and providing further support as required in consultation with the trainers;
- improved trainer resources.

This pre-trial study has been designed not only to re-evaluate TDTScience but also to inform scale-ups more generally, with particular reference to those using a train-the-trainer model. As such, the study has a substantial initial component that evaluates the train-the-trainers model for efficacy and fidelity of delivery. The main trial will then focus on the training of the main trial teachers and the experience, attitudes, and performance of their Year 5 pupils. There will also be a second year of the trial to follow the next cohort of Year 5 pupils and the original cohort (now in Year 6) to measure legacy effects. More detail can be found in the protocol (Hanley et al., 2021).

This report focuses on the research conducted before the main trial to evaluate the train-the-trainers model.

The TDTScience programme is led by Science Oxford and independently evaluated by the York Trials Unit, University of York. The study is funded by the Education Endowment Foundation (EEF) and the Wellcome Trust.

Research questions and methodology

Pre-trial study research questions

There were four main research questions:

- **RQ1** To what extent is the training model implemented as planned?
- **RQ2** How effective is each element within the model at achieving its aim(s)?
- **RQ3** How necessary is each element to the training model overall?
- **RQ4** What improvements could be made to the model to benefit the TDTScience intervention and training design more widely?

This stage of the evaluation provided an important opportunity to assess the implementation and acceptability of the train-the-trainer model. On completion of the three stages of the train-the-trainers model, a final team of trainers will be selected to train the teachers participating in the intervention arm of the main trial.

Training and evaluation structure

The train-the-trainers phase was divided into three consecutive stages, see Table 1.

Table 1: Timeline and stages of the TDTScience train-the-trainer model

	2021						**2022**						
	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul
Stage 1: Developers deliver TDTScience to trainers													
Stage 2: Developers deliver train-the-trainers sessions to trainers													
Stage 3: Trainers train pre-trial teachers													

A summary of the data collection methods for the study are detailed in Table 2. Further information about each of the stages are detailed below along with a description of the relevant evaluation approach adopted at each stage.

Table 2: Data collection methods overview

Research focus	**Data collection methods**	**Participants/data sources**	**Data analysis**	**Research questions addressed**
Developer feedback (after Stages 2 and 3)	Paired semi-structured interview	Two interviews	Descriptive/ thematic analysis	1, 4
Developer-run TDTScience course	Observation schedule		Combination of inductive and deductive analysis	1

Trainer feedback (after observed sessions)	Paired semi-structured interview with 'new' and 'experienced' trainers	Six interviews (12 trainers)	Combination of inductive and deductive analysis	1, 4
Trainer feedback (after each delivery phase)	Survey	Four timepoints (12 trainers per occasion)	Descriptive/ thematic analysis	1, 4
Trainer-run TDTScience course	Observation schedule	Six observations (12 trainers)	Combination of inductive and deductive analysis	1
Teacher feedback (after Stage 3)	Focus group	Two focus groups (nine teachers)	Descriptive/ thematic analysis	2, 3
Instrument (teacher)	New survey		Descriptive/ thematic analysis	2, 3, 4
Instrument (lesson)	New observation schedule	Four observations (four teachers)	Combination of inductive and deductive analysis	2, 3
Instrument (pupil)	New survey		Descriptive/ thematic analysis	2, 3, 4
Technique (lesson)	Audio recording	Four observations (four teachers)	Assessment of data utility	2, 3

Stage 1 (July 2021) - developers deliver TDTScience to trainers

The aim

Trainers, as if they were participant teachers, received training in TDTScience from the developers. This provided immersion in TDTScience and an appreciation of the teacher experience. It is worth noting that the content was condensed from the usual model of four separate days into a three-day residential. Table 3 summarises the evaluation approach for Stage 1.

Table 3: Stage 1 evaluation approach

Focus	Method	Why?	Research Question	Number
Developer-run TDTScience course	Observation schedule and field notes	To gain more insight into TDTScience and how the developers run the training.	RQ1	All three days
Developer	Post-session interview	To establish whether sessions ran as planned	RQ1	At the end of each day
Trainers	Survey (devised by Oxford team in consultation with Evaluation team)	To obtain feedback from sessions	RQ2	At the end of each day

Trainers	Paired interview with 'new' trainers	To gain deeper insight into training experience	RQ2, RQ3, RQ4	One (two participants)

Stage 2 (September–November 2021) - developers deliver train-the-trainer sessions to trainers

The aim

Trainers were trained by the developers to deliver TDTScience training to teachers. This deepened their knowledge of how the course was designed to deepen teachers' understanding of the approach across the four and a half days, taught them how to deliver the content, and allowed them to explore and plan delivery with their delivery partner. Table 4 summarises the evaluation approach for Stage 2.

Table 4: Stage 2 evaluation approach

Focus	Method	Why?	Number
Developer	Pre-session communication	To determine training goals for each session	One (two to three participants)
Developer-run train-the-trainer course	Observation schedule and fieldnotes	To understand the expected model of TDTScience training delivery	All (four days)
Developer	Post-session	To establish whether sessions ran to plan and goals were met	Two (after second and fourth days)
Trainers	Inter-event communication	To check learning, recall, issues from the first session, and expectations of the final session	One (all trainers)
Trainers	Post-course survey	To check learning, recall, issues from the final session, and reflect on the course overall	One (all trainers)
Trainers	Paired interview with 'new' trainers	To gain deeper insight into training experience (first pair); to explore preparations and confidence (second pair)	Two (one post-training, one predelivery)

Stage 3 (January–July 2022)—trainers train pre-trial teachers

The aim

Trainers trained pre-trial teachers (who will not be involved in the main trial) across four full days and a fifth half-day round-up across a six-month period.[1] This gave them chance to develop experience and expertise in training teachers in TDTScience so, in a more accurate representation of real life, they would not enter the main trial with no delivery experience. It also allowed developers to observe the trainers (each of the two developers visited one training day run by each pair) and select the final team for the trial.[2] Table 5 summarises the evaluation approach for Stage 3.

[1] See Appendix 3 for summary agenda.
[2] In the event, all trainers were retained for the trial.

Table 5: Stage evaluation approach

Focus	Method	Why?	Number
Trainer-run TDTScience course	Observation	To assess fidelity of delivery	Six sessions (one per region and covering each of the four days)
Trainers	Post-session interview	To establish whether sessions ran to plan and goals were met	Six (as above sessions)
Teachers	Focus group	To explore feedback from the sessions	Two groups (different regions) with four to six teachers
Trainers	Survey	To gather over-arching feedback on training and how useful they find the pre-trial phase	All (four sessions)

The developers had a further over-arching aim for these three stages. They felt that a weakness of the previous effectiveness trial was finding appropriate trainers—those willing to adopt the approaches and strategies of someone else's training programme as their own. This model allowed developers to invest more time training, observing, screening, and supporting potential trainers to ensure they were appropriate to deliver the main trial.

The developers issued feedback forms to the trainers after Stage 1 and the two two-day parts of Stage 2. This was analysed and used in a formative manner if appropriate, leading to adjustments in future stages. The evaluation team shared early, very brief findings from the research reported here to the developers. Developers read the feedback from the trainers and used it to inform the ongoing development of the training in line with the 'collaboration with trainers' sub-theme as it demonstrates the developers welcoming feedback from the trainers.

Recruitment of pre-trial schools to Stage 3

The target was for the lead trainer within each of the six regions (with oversight from the developers) to recruit 60 schools, ten in each of the six geographical areas—South West, Lancashire, Staffordshire/West Midlands, North East, Yorkshire (Leeds), Lincolnshire/East Midlands)—during autumn term 2021/2022. All state primary schools in England,[3] including academies, could take part in the pre-trial as long as the following eligibility criteria were met:

- nominated teachers have not been TDTScience-trained;
- school will not take part in the main trial;
- if the school is part of a multi-academy trust (MAT) then no school in the MAT will take part in the main trial;[4] and
- school is aware the training is Year 5-focused.

These criteria were a lot more relaxed than those for the main trial. For instance, in the main trial, any schools (rather than individuals) trained in TDTScience would be excluded and all Year 5 teachers would be required to participate. Although it was recommended that two Year 5 teachers from each school take part in the pre-trial, this was not an eligibility requirement.

Sixty-five schools were recruited in total across the regions.

[3] Or middle schools if they include both Year 5 and Year 6.
[4] This was relaxed in the main trial to not sharing or collaborating with any school from the MAT that had participated in the pre-trial due to difficulties recruiting.

Context

Background of trainers

All 12 trainers had previously been teachers (11 primary, one secondary) as well as being science CPD providers—either as independent consultants or as part of the work of their organisation. Of these, seven had previous experience of training teachers in TDTScience (either in the first effectiveness trial or in TDTScience training offered by the developers). The trainers were all previously known to the developers and were selected based on prior experience in delivering primary science teacher training and interest in the TDTScience approach and previous evaluation findings. Initially 14 trainers were recruited but the year-long delay to the trial due to COVID-19 led to two of them having to withdraw from participation.

Analysis of data

Data from the interviews and focus groups were imported into Nvivo and analysed inductively using thematic analysis to explore the perspectives of developers, trainers, and teachers on developing, delivering, and receiving the TDTScience training as well as teacher's experiences of putting the training into practice in their classrooms. Transcripts were initially coded line by line by researchers from the evaluation team. Following this, the codes were grouped into those with similar meanings and this led to the generation of themes. Themes were discussed by the evaluation team as they developed and were refined to ensure they were an accurate representation of participants' experiences.

Descriptive statistics were used to analyse the trainer survey data. Answers to open-ended questions were analysed deductively using the themes that had been developed from the analysis of the interviews and focus groups. However, additional codes were added as necessary.

The teacher survey data was examined to ensure that the questions seemed to be understood in a consistent and unambiguous manner. The survey also asked how the survey might be improved, for example the wording, topic coverage, and length.

Findings

This section reports the findings mapped to the research questions. Each section is preceded by a short summary of the main points.

RQ1. To what extent is the training model implemented as planned?

Summary of findings

- Although COVID-19 caused delays in delivery (to the main effectiveness trial and subsequently to Stage 3) and resulted in some absences among trainers and teachers, it did not seem to have a particularly negative effect.
- Trainers delivering in pairs worked well with benefits in terms of complementary knowledge and skills and providing mutual support.
- All three stages broadly met the expectations of developers and trainers.
- Trainers found it a challenge to balance fidelity to the TDTScience programme with a desire to establish ownership of delivery, but this seemed to be achieved successfully.
- Despite concerns beforehand about fitting in all the Stage 3 content, trainers managed to run the days to time with minor deviations from the schedule where necessary.
- Trainers spent a considerable amount of time preparing, both independently and with their delivery partner. Their sense of preparedness did not increase directly in line with their experience of delivering but appeared to be more influenced by the content of the individual days.

Impact of COVID-19

The training process was initially due to start in summer 2020 but the lockdowns resulting from the COVID-19 pandemic resulted in the study being delayed by a full year. The alternative would have been to run the TDTScience training process online but from previous experience the developers felt this medium would not be appropriate. As observed in the training sessions, the approach includes practical investigations that require a variety of resources, teamwork, and physical activity impossible to recreate online. Additionally, the developers felt that face to face delivery allowed trainers to engage properly with the content and work together more coherently and facilitated the developers observing the trainers and their engagement. Consequently, the process of training the trainers began in July 2021. The developers felt an additional benefit of the delay was an extra year to develop the materials and recruit high quality trainers, optimising the opportunities for successful delivery of training.

> *'I think Covid in a way… has helped us … it actually gave us a year longer to get the materials ready, so it was a real luxury … we've had longer to recruit appropriate trainers' (developer, Stage 2, day 4).*

However, the developers recognised that the COVID-19 pandemic created some practical issues, particularly for the July 2021 training, including delivering the training in cold rooms due to having the doors open, having to stay in 'bubbles' during breaks, and various staff absences due to illnesses, self-isolation, or shielding as a consequence of the pandemic. Despite the pandemic, trainers felt comfortable and safe attending in person as everyone was doing lateral flow tests, wearing face coverings, and keeping a two-metre distance from one another. Trainers enjoyed being around people for the training and for some the July 2021 training was the first time seeing and working with people in-person since the pandemic.

> *'It was really nice to be face to face. So I feel Covid didn't impact it as much as I thought it was going to … I keep listening to things on the news about cases going up and I was thinking, please don't let us have to cancel and do it via Zoom this date' (trainer interview, Stage 2, new trainer).*

During the January to July 2022 training of pre-trial teachers, 'Covid [was] obviously still affecting schools quite a lot' (trainer interview, Stage 3). This was mostly due to other teachers in the schools having COVID-19 and, as a result, the TDTScience teachers could not be released or were called back to school during the training day to support capacity at school. Once, a teacher was called back to school before the training day even began. Trainers commented that they were grateful that, whenever possible, the teachers were able to attend the training, and that 'they have pulled out all the stops to have people here' (trainer interview, Stage 3).

The trainer surveys showed that on two occasions trainers missed pre-trial training sessions because of COVID-19 and other arrangements had to be made. One of these absences was on day five and the remaining trainer commented that they did not have the opportunity to say goodbye to the teachers together and they missed their training partner. Another trainer was concerned about low attendance due to COVID-19 but was relieved that only one teacher had been unable to attend day one and only two has missed day two.

The pandemic also affected the timings of the Stage 3 training. The postponement of the first day (usually to the intended day two date) concertinaed the training so that day five was much closer to the fourth day than planned—in one instance only a week apart.

There was a concern that poor attendance would affect the fidelity of the project, which reflects commitment to the project as trainers felt a responsibility for the successful completion of the trial:

> *'Possibly my only concern was the absentee rate. Clearly schools are still recovering from Covid but it seems that for one or two schools in particular they were not able to attend many of the sessions. I'm concerned how this will impact the main trial results and whether the evaluation will take poor attendance into account when looking at the impact on learning' (trainer survey, post day five, experienced trainer).*

Trainers delivering in pairs

The model stipulates that the trainers deliver the training in pairs and developers felt it was important for the trainers to be able to 'work as a pair and play to each other's strengths' (developer interview, Stage 3). This was supported by the trainers who felt it was extremely useful for delivery: '[my partner] and I have got different strengths, which is really good because you can see how you can use those best.' (trainer interview, Stage 2, new trainer). Where possible, the developers paired new trainers with those who had previous experience delivering TDTScience.

Trainers enjoyed working with a partner to prepare for the training and then deliver it to the teachers and commented that it was 'quite nice to have that time to reflect with each other and discuss some plan together but then also to share that delivery' (trainer interview, Stage 2, new trainer). They felt that they had sufficient time to plan the training and that this was improved by having a partner to share the planning with.

The observations of the training sessions confirmed that it was useful for one trainer to lead a session while the other prepared resources for the next activity. Moreover, both trainers could contribute their ideas and anecdotes to make the delivery rich and meaningful, and trainers were observed chipping into each other's sessions in a complementary rather than an intrusive manner.

The interviews also demonstrated that it was beneficial to share out the topics that each trainer felt more confident with and felt reassured that their partner could contribute if they had missed anything. During the delivery of training to the teachers, trainers tended to divide up the sections between them that they felt most comfortable in delivering. Sometimes, trainers wanted to challenge themselves and engage in delivering the activities they felt less confident with, particularly if they had done less of these in previous training sessions.

> *'I'm not a performer when I do CPD, you know, some of the things like the practical prompts for thinking, it's not my bag and [name] also admits the same. So when she did it in the first trial, she deferred all of that to her partner at the time and said, "Oh, you do all that, I'm not very good at that and this year [name] has been, like, I really want to do those things because I know that I copped out last time. So that's good." She's been practising all those sorts of things and then I've taken other roles' (trainer interview, Stage 3, new trainer).*

One aspect that helped with confidence was that the trainers had given permission to their partners to interrupt any sections they were presenting and contribute if needed, either with anything they had missed or if they had their own anecdotes to contribute. This allowed trainers to feel more confident that they would not miss any content out. However, occasionally there was some miscommunication with these arrangements.

Trainers were happy that the pairings would be retained for the main trial—although note that the retention of pairings was not assumed from the outset as the developers informed trainers that they reserved the right to change pairings for the main trial should it emerge from observations that different pairings may work better.

Venues

Trainers were keen to provide teachers with venues that were in the vicinity of most of the schools (to minimise teacher travel) as well as having an appropriate room for the training and a good quality lunch and other refreshments to minimise the burden on teachers and maximise learning. In most cases, trainers considered that the venues were of a good standard, met the required needs, and allowed them to focus on delivering the training.

The trainer surveys suggested that last minute changes to the venue may lead to anxiousness.

> *'I was aware the venue had to be changed at short notice so was a bit apprehensive … but in the event the day seemed to run well' (trainer survey, post day two, experienced trainer).*

Expectations of training—Stages 1 and 2

During their interviews at the end of the first two stages, the developers commented on the responsiveness of the trainers during the July to November 2021 training.

Developers felt that some trainers were less engaged than others during this period and that the 'level of effort and engagement' was 'just slightly below the other people' (other trainers; developer, Stage 2). However, largely, developers felt that trainers were highly engaged and this increased from the first day of their training onwards. The trainers interacted well with one another and the atmosphere was very warm and friendly, to an extent that exceeded the developers' expectations. Although some trainers had done the training before, they seemed to approach it without fixed presuppositions and considered it a new opportunity for professional development.

> *'I knew that they would engage well because they are trainers and they understand. But they sort of exceeded my expectations, bearing in mind that half of them, you know, were on the effectiveness trial with us, I thought they did a brilliant job at just kind of coming at something afresh and going, "I'm just going to find a new way of thinking about this, because I've got the opportunity to think about it again and I'm just going to go for it"' (developer, Stage 1, day one).*

Trainers commented that the training met their expectations and was extremely well run. They knew what to expect in Stage 1: 'I was expecting to pretend that I was a teacher and the course was going to be done to me' (trainer, Stage 1, new trainer). Overall, trainers enjoyed the programme and they felt well supported by the developers throughout the process and felt they could ask questions at any point: 'It's a superb programme. Always thought it was a superb programme. I've really loved delivering it. The content is superb and the reactions of the teachers is always really good' (trainer interview, Stage 3, experienced trainer). The trainer survey data also supported this positive view about delivering TDTScience.

Expectations of training—Stage 3

Generally, the course met the expectations of the trainers and teachers and they were happy with the content that was being delivered. Some of the teachers came from schools with prior experience with TDTScience and this helped with knowing what to expect, for example, being aware of the general principles of the approach.

Table 6 summarises the responses of trainers to the question, 'How did the day compare with your expectations?', which was included in each post-training day survey (see the table below). Day five had the greatest number of trainers saying

the day went about the same as expected, with the remainder saying it went much or a little better. This could be a reflection of the growing experience of the training or the slightly different content and format of the final half-day.

Table 6: Post-training day survey responses about expectations

The day compared to their expectations	Post Day 1	Post Day 2	Post Day 3	Post Day 4	Post Day 5
Much better	4	0	4	4	3
A little better	5	4	5	6	4
About the same as expected	3	5	3	2	5
A little worse	0	2	0	0	0
Much worse	0	1	0	0	0

Overall, some trainers were initially apprehensive but then felt the days went better than expected. They were 'worried about timings and also not rushing' (trainer survey, post day one, new trainer), concerned that attendance would be low due to the COVID-19 pandemic, less familiar with the content, or worried about the lack of practical activities.

> *'I was cautious of the perceived lack of practical activities in [day four] compared to the others and given the name of the programme—'doing' science. However, the wide ranging activities that were a part of the day were great: they all worked well and the teachers really enjoyed them' (trainer survey, post day four, new trainer).*

Other trainers had positive expectations and they 'appreciated the training [partner]' and felt that they 'worked well together' and this met their expectations (trainer survey, post day three, experienced trainer).

When discussing a training group that had particularly quiet delegates, the trainers of this group noted how initial concern that it might indicate a lack of engagement turned into a realisation that it might instead mean that delegates were pondering the issue, had a deeper understanding, and were using 'higher order thinking - HOT'. Several trainers gradually realised that they could not tell how much the teachers had engaged without asking them to reflect on their understanding.

> *'There was not as much "buzz" as previous sessions; however, with reflection and discussion [with one of the developers] this was because delegates were thinking hard. That was also the reason why they (and we) were tired. After reflection I realised that the important part was the thinking and discussion on tables—that was high quality. This is a big measure of impact on delegate thinking and ultimately delegate practice' (trainer survey, post day three, new trainer).*

There was a distinct tension between trainers wanting to take more ownership of TDTScience and recognising that they needed to provide consistent delivery and ensure all trainers involved in the trial were receiving the same programme.

> *'We tried to keep pretty much to script because that was kind of the message we were given—that because it's a trial we want to kind of have that consistency as much as possible that everyone is getting the same story and the same approach. But they said to us that it's okay to add the odd anecdote (trainer interview, Stage 2, new trainer).*

Some trainers felt adding in anecdotes may distract them from the script and they were conscious of presenting it according to how the developers had trained them. They were reluctant to discuss anything that was not mentioned in the written materials.

> *'We do try very hard to say it as it is written. It would be very easy to go off. The last little bit with the practical prompts and the string and the bees—I used something with Play-Doh and strips of paper and I kind of want to tell them about that activity as well but it's not in the written materials, so I haven't' (trainer interview, Stage 3, new trainer).*

There were trainers who felt more confident knowing they had notes to hand that they could refer back to: 'I do like having the computer facing me with the slide notes—that does help me flow better; that's something that I haven't done on one session and I started to do a little bit last time' (trainer interview, Stage 3, new trainer). However, although developers were keen for the programme to be delivered accurately, they wanted to see trainers owning the delivery and not using their notes. Developers noted when training the trainers that they could see that 'the newer ones' had 'work to do but the potential is there'. In general, they were confident that the trainers 'knew the course well' (developer, Stage 2).

Pace

During Stage 1 of the pre-trial, when developers delivered the training to the trainers as if they were teachers, they condensed the intended four-day delivery into a three-day residential. Although this meant there was a lot to fit in, the benefits of time-saving prevailed.

> *'But for this group of trainers, it's not fair to ask them to travel this distance and not fill the days. So no, I wouldn't change it' (developer interview, Stage 1, day two).*

Trainers felt that Stage 3 (training pre-trial teachers) was extremely valuable and they were motivated to deliver all the content in a timely and effective fashion. While the developers were confident in the pace and amount of content in the schedule, the trainers felt it was sometimes a struggle and felt slightly rushed.

> *'Yeah, but what I think is because we're so pushed for time, the time is so tight. We were just constantly looking at that clock and knowing, oh my gosh, I've got 23 slides here in ten minutes!' (trainer interview, Stage 3, experienced trainer).*

To support themselves, trainers used detailed schedules with annotations of timings they had prepared prior to the training day. Often, despite the challenge of sticking to the timetable, trainers still felt confident that they would finish on time. They slightly adjusted the timings depending on personal preference of delivery and borrowed time from other sections. For instance in one observation, two sessions ran five to ten minutes late (these were on the second day: 'Which Shoes Have the Best Grip' and 'The Science of Forces'—see Appendix 3 for more detail) but the session prior to lunch that day ('Higher Order Questioning') was shortened slightly to make up the time and the day was back on schedule by the start of the afternoon.

> *'I think I might tweak some of the timings … I think we knew that some bits, for example, the levers [Day 2, Force] had 20 minutes, so we kind of knew that as we were heading towards that, that even though we were over time we would still catch up (trainer interview, Stage 3, experienced trainer).*

The importance of pace was also supported by the trainer surveys, where the trainers felt the timings were quite strict and they had to amend the timings slightly (for example, depending on familiarity with the materials). They discussed needing more time in the sessions on areas of discomfort for teachers. Trainers suggested that it may be useful within the main trial to ask delegates what they are unsure about and focus on giving them time to reflect on this. The trainer surveys also showed that trainers felt apprehensive about the length of time allocated to participant sharing and whether they would have enough to say. There were also some comments about changes to the suggested agenda timings.

> *'I felt we spent too long looking at candles. I also felt we could have opened up the paper towel activity more ['Which Paper Towel?' practical, Materials session], in particular giving them a larger range of paper towels to choose from ... (trainer survey, post day one, experienced trainer).*

On the first day in particular, some trainers expected to feel rushed in their delivery but in the end the timings worked well, which contributed to them feeling that the day went better than expected.

> *'I was worried about timings and also not rushing—to ensure that delegates understood the main principles. We amazingly stuck to the time and worked really well together' (trainer survey, post day one, new trainer).*

The trainers showed ownership by amending the timings of the sections, depending on what their group was more familiar with. In some cases, more time was needed because activities were unfamiliar.

> *'For day five we amended the timings slightly to give more time for the Action planning for impact, as few delegates have ever done any before' (trainer survey, post day five, new trainer).*

Preparation

Trainers felt they had a wealth of time to prepare for the delivery in between receiving the training themselves and the first day of teacher training. They preferred to use their own methods of preparing (for example, checklists, flipcharts, spreadsheets) in order to consolidate and process the information they had learnt, and this was often done independently. Trainers then met with their partner, either on Zoom or in-person, to plan out which sections would be delivered by whom and check with one another they were happy with the content. Although these meetings included a lot of planning, trainers found this a relatively straightforward process.

Although trainers felt it was necessary to commit a significant amount of time to the preparation, they felt that this was key for feeling comfortable delivering the TDTScience sessions, even if they felt generally confident as a trainer. Minor elements such as downloading videos and arranging the resources into the right kits also took some additional time but it all contributed to the trainers feeling prepared and organised for the day.

> *'And I think for me, no matter how many times I've presented this before, you still always have to put in the leg work. So even though I do know this stuff, I haven't presented it for a while. So I did still spend quite a lot of time going through it all, going through the PowerPoint, going through my notes and the training' (trainer interview, Stage 3, experienced trainer)*

In the surveys, the trainers were asked how many hours of preparation they did ahead of each day, both on their own and with their training partner. The responses for preparing with their partner ranged from 30 minutes to eight hours, with preparation meetings sometimes being spread across multiple days or half days. Trainers spent a lot more time preparing on their own, with many references to spending more than seven hours or a full day (at least) preparing alone. The responses ranged from one hour to four days, however, it was difficult for respondents to state how long they spent preparing, as illustrated by this quote.

> *'A lot of background reading as well as focused on the file/materials. Probably four days ...? Many smaller chunks as well as the luxury of a couple of days emersion [sic])' (trainer survey, post day one, new trainer).*

This indicates that there was a great deal of variety in the time allotted to preparing and the varying arrangements adopted during preparation.

Within the surveys, trainers were also asked how well-prepared they felt to deliver the sessions. On average across the five days, the responses were overwhelmingly positive, as summarised in Table 7. Post day five responses would suggest the trainers seemed to feel a bit more apprehensive. Interview data suggested that trainers may have felt it was harder to prepare due to this day relying on more teacher 'sharing', which was often seen as daunting. However, some of the reasons for feeling less prepared on that day were personal, for example, 'I usually prepare the day before but because of another commitment had to do this a few days before' (trainer survey, post day five, new trainer).

Table 7: Post-training day survey responses about preparedness

Level of preparedness	Post Day 1	Post Day 2	Post Day 3	Post Day 4	Post Day 5
Very well-prepared	11	7	8	10	4
Quite well-prepared	1	5	4	2	8
Not very well-prepared	0	0	0	0	0
Not at all well-prepared	0	0	0	0	0

Since trainers did not report feeling consistently better prepared as Stage 3 progressed, this suggests that it was the content of individual days rather than growing familiarity with TDTScience and their group of teachers that determined how well-prepared they felt.

A few trainers described other issues that affected their ability to prepare—for example, the session was during a busy time for them, there were ICT issues, or they were sharing equipment with their partner—and some trainers commented that they made a few mistakes as a result. The timing of preparation was seen as important and trainers liked having the day before to prepare.

> *'Spending a day preparing beforehand was worth it: we had all the kit bagged up and ready to go meaning we could focus on getting the room right when we arrived' (trainer survey, post day one, new trainer).*

It seems that working well with their partner was key to feeling prepared and meeting their expectations:

> *My co-trainer and I try to allocate sessions according to our strengths and I wasn't sure whether we had got this right but after delivering the training, I think we had' (trainer survey, post day four, new trainer).*

Resources

Trainers received a comprehensive tutor file along with various other resources to support their delivery. They were grateful for these resources—for example, the manual, additional notes and the resource kit. The 'crafting a lesson' template was particularly useful and was said to be 'the most important piece of paper from the whole three days*'* (new trainer, Stage 1) and helped the trainers to put the TDTScience strategies into perspective However, it was noted by the teachers that the activities taught during the training are often 'resource heavy' and 'being able to replicate that in school' can be difficult, particularly when it relies on the teachers funding it themselves and struggling to find similar equipment.

The trainer surveys highlighted that they felt that the course was 'well resourced' (trainer survey, post day two, experienced trainer), that the 'materials were very supportive' (trainer survey, post day two, new trainer), and that they 'knew it would be fine because of the high quality of the training materials' (trainer survey, post day five, experienced trainer).

Trainers prepared by checking they had the right kit and discussing the equipment with their partner. Some trainers developed new materials such as prompt cards, which reflects a feeling of ownership and wanting to do a good job by being extra-prepared.

> *'[My training partner] and I met a couple of times—I had read/rehearsed and also had extra props—also developed prompt cards for the activities to support and aid memory' (trainer survey, post day two, new trainer).*

The feedback on resources was positive with at least one trainer remarking that the resources worked well or were quick to set up in survey feedback after each of the days.

However, updated notes were received from the developers shortly before day five. A number of trainers commented on the quality of the notes been less informative than others and how their late arrival made them feel less prepared to deliver the session, although it did not impact on session delivery overall. It illustrates the importance of getting both the timing and content of the notes right.

> *'The extra notes needed a little extra prep—and mostly this was fine. I think I had one slide I could have explained better to teachers but I suspect only I noticed that!' (trainer survey, post day five, experienced trainer).*

RQ2. How effective is each element within the model at achieving its aim(s)?

Summary of findings

- The three pre-trial stages were designed to give trainers more understanding of TDTScience philosophy and practice. Stages 1 and 2 had provided a solid basis for trainers to be confident about delivering the programme to teachers in Stage 3.
- During Stage 3, trainers were regularly considering how delivery could be enhanced by de-briefing with their partner and reflecting on their own practice.
- On the whole, trainers felt that they had achieved the engagement they were seeking from the teachers, and this was endorsed by training observations. It was clear from trainer and teacher feedback and lesson and training observations that the TDTScience approach had been put into practice in the classroom.
- Trainers were aware that teachers struggled more with some concepts and approaches (such as HOT or Practical Prompts for Thinking) than others (for example, Bright Ideas Time).
- Teachers being reflective and sharing good practice were judged particularly effective with trainers often surprised by the extent to which classroom practice seemed to have changed.

Comparison with previous effectiveness trial

The developers discussed how this trial differed from the previous TDTScience effectiveness trial (Kitmitto et al., 2018). The current trial (Hanley et al., 2021) was designed to provide more time for the trainers to process and understand the content of the training before they delivered it to teachers—trainers were able to study the whole tutor file before delivering the first day and gain an overview of the full training programme. Previously, timing constraints meant developers delivered one day of training to the trainers who then delivered the same day to teachers immediately without having chance to experience the full course.

> *'There's no comparison, I think, with last time … this time, they've had the whole course and now we're talking about how you train with familiarising themselves with the tutor file, which we didn't have time, at all, to do last time and also they're getting an overview of what the whole thing looks like' (developer, Stage 2).*

Developers felt that the background experience of those trainers who had participated previously was advantageous for the pre-trial, and it was these trainers that were showing high levels of engagement with high effort levels. The developers felt that it was relevant and important to know which trainers had more experience than others to identify whether they retained their interest; they were able to identify who had 'done it before but [they] were obviously still engaged' (developer, Stage 1).

The trainers with prior experience felt that they were more proficient because of delivering the first trial, influencing the way training went during the pre-trial. Trainers seemed to learn from past mistakes and were much more familiar with the materials.

> *Also, knowing that we had an issue with the first practical last time we were prepared for this and delivered it in a way that would support a greater understanding of the underlying principles' (trainer survey, post day three, experienced trainer).*

Trainer understanding of TDTScience and aims for delivery

The aims for delivery from the developers' perspective were that the trainers had received the TDTScience course as if they were teachers (Stage 1) so they could be fully immersed in the programme and understand it from both perspectives. The developers then trained the trainers in how to deliver the TDTScience course so that they understood

how to deliver the content in lines with the developers' design and concepts. As highlighted below, trainers appeared to find this structure beneficial and appreciated receiving the training from both perspectives.

After Stages 1 and 2, most trainers felt confident and left the November training with a readiness to deliver. This was aided by the level of detail in the notes and resources they were provided with and could refer back to, helping trainers to feel confident they were adhering closely to the programme and its intended delivery. There was also some clarity around using one's own experience while remaining faithful to TDTScience. For example, in an interview, one trainer said that they felt able 'to bring in any kind of anecdotes drawing from our own experience of [running training] and using any of those sort of ideas and techniques' (trainer interview, Stage 3, experienced trainer).

This level of confidence was supported by the trainer surveys: responses highlighted that trainers often felt confident in their delivery and some felt that although they were apprehensive that 'once you're in the flow it became easier' (trainer survey, post day two, experienced trainer) or they 'relaxed into it' (trainer survey, post day two, experienced trainer). Trainers felt confident to deliver the materials and felt that by the end of Stage 3 they had 'everything ready to go' (trainer survey, post day four, new trainer).

The data demonstrated that the trainers were reflective in their thinking and were constantly considering what could be improved. Trainers indicated that reflective thinking is important in their line of work and it will be an extremely useful skill in delivering the TDTScience programme.

> *'I found at times I was trying to reflect, do my reflections and listen to the next bit … maybe I needed to think— just what is the main thing I'm taking away?' (trainer interview, Stage 3, new trainer).*

One method of reflective practice that the trainers used was to de-brief with their training partner following each day's training in Stage 3; they found that this was valuable for their development. Trainers reflected on whether they should deliver similar sections of the training next time, on whether their preparation methods were sufficient, and on any constructive criticism they had for one another. This is linked to trainers taking ownership over the training as they are ensuring they are delivering it accurately and to the best of their ability through reflective practice.

It seems that the relationship with their partner was bolstered by the pre-trial and this close connection helped them to feel confident to deliver.

> *'It has really helped to become familiar with the course materials to run the entire course for the pre-trial. I have become more confident when asking teachers questions or for contributions. I've also benefitted from working with an experienced partner trainer who has a different but complementary style. We have got to know each other and worked well together, which will really help us to deliver the main trial course as effectively as possible (trainer survey, post day five, experienced trainer).*

Stage 3 outcomes

Trainers' hopes for the training seemed to centre on increasing teachers' understanding of the bigger picture and key messages and promoting HOT. Most trainers felt this had been achieved and that their delivery of the training had met their high expectations: 'Listening to discussions, delegates had a much better understanding of the key messages of the TDTS project' (trainer survey, post day two, new trainer).

Trainers also wanted to make it interesting and exciting, so they 'worked on developing the awe and wonder' (trainer interview, Stage 3, new trainer). Trainers were keen to deliver what they viewed as exciting content and wanted the teachers to 'go away buzzing about it' (trainer interview, Stage 3, new trainer). From the trainers' perspective, they felt this was achieved with most teachers as there was a high amount of engagement and excitement. Trainers felt that even during the lecture-style activities, interest was maintained and trainers could see teachers were taking notes and processing the information being delivered. The trainers noticed that if teachers liked a particular activity a lot, they would be keen to go back to the classroom the following day and teach it.

> *'I thought the "habitats" activity—where they had to make their own animal in Play-Doh—I thought that was brilliant and I always knew that was a good activity from when we'd done it as trainers. But the teachers really, really went for it and I've actually seen it in two schools on Twitter since as well. They've obviously*

gone straight back to school and gone, "We're going to do that tomorrow!"' (trainer interview, Stage 3, experienced trainer).

Even before they started delivering to teachers, trainers were concerned about whether they would engage and the level of science knowledge that the teachers had. They wanted to provide a level of scientific explanation that was neither patronising nor inadequate. Trainers were conscious of sticking to the TDTScience programme but providing an appropriate level of subject knowledge for the audience and enough to take back to school and deliver effectively.

'There's always the question of how much do you go with that explanation. So I'm always mindful of keeping to the plan … but giving the kind of subject knowledge bit with enough credibility that hopefully I know what I'm talking about and it makes sense without dumbing down the science—does that make sense?' (trainer interview, Stage 3).

Afterwards, they felt that teachers had enjoyed all aspects of the training including sharing good practice, group discussions, practical investigations, and time for reflection. It seemed that the training provided the teachers with many useful skills that they were aware of prior to the course but had not engaged with properly in the past, such as reflective thinking.

'They like hearing from each other how things are going ... They like reflecting on some of the approaches that they might have tried before but haven't gone as well or they might have tried them in another subject area or they might have tried them in a different year group or a different context' (trainer interview, Stage 3, experienced trainer).

In the surveys, trainers commented that 'teachers were engaged and appreciated the training' (trainer survey, post day three, experienced trainer) and felt that the teachers really enjoyed the activities, even on days with fewer practicals.

'I knew that it would be fine because of the high quality of the training materials. However, there was a real buzz in the room and teachers were full of how much they appreciated the approach (trainer survey, post day five, experienced trainer).

It was important that the sessions allowed time for teachers to reflect and communicate what they had learnt. Particularly on day five, trainers commented that the teachers had reflective discussions about how they could apply their practice to the classroom.

'The teachers attending had used the TDTScience approaches and really developed their practice. The discussions were extremely reflective and they identified impact both on their own practice and pupils learning precisely and backed up their views with really good examples' (trainer survey, post day five, new trainer).

Here, the teachers were asked to bring examples of children's work to share; teacher's participation in this was mixed.

'I was a little concerned the teachers might not engage with the 'sharing' part of the session. I needn't have worried, they were brilliant and brought lots of examples of what they had tried in school' (trainer survey, post day five, new trainer).

'Only about a third of teachers brought examples of children's work to share' (trainer survey, post day five, new trainer).

Despite some teachers not bringing examples of children's work, the session achieved its purpose: 'The teachers really engaged with each other, sharing their ideas and experiences of using the TDTScience approach in schools' (trainer survey, post day five, new trainer).

Teacher feedback obtained by the developers showed that over 80% rated their enjoyment and engagement as 'excellent' for the first two training days, falling to around 70% for the other three days (most of the remainder saying 'good'). There was a steady increase in teachers' rating of their understanding of HOT, from 59% after day one to 72% after day four. On completion of day five, all teachers rated their understanding of the TDTS approach as either excellent

(77%) or good (23%) and all said they would recommend the course to other teachers. Trainers provided positive accounts and discussed the benefits of the TDTScience approach for teachers, but how open teachers are to learning can affect this experience.

Anticipated impact on pupils

Trainers identified some aspects of the training that they felt the pupils would benefit from once the teachers had integrated the approach into their teaching: that pupils may demonstrate 'more engagement' and inquisitiveness and improve in areas such as 'scientific vocabulary', 'listening skills' towards the teacher and each other, 'valuing everybody's opinions', justifying their choices and ideas, and being able to fully understand a concept. Trainers felt these skills were valuable life lessons and were transferable to other areas of the curriculum. Additionally, it is evident that the trainers feel passionate about the TDTScience strategies because they can see the impact it will have on pupils at school and their development in terms of HOT and problem solving.

> *'I think that idea about the benefit of discussion in terms of unpicking their thinking and helping them to see—that's what I'm thinking now but then by having that opportunity to reflect and think and pair and share and go a bit deeper with it that they're having that opportunity to address their own thinking, which could be unpicking misconceptions or just filling in gaps from partial understanding' (trainer interview, Stage 3, experienced trainer).*

Additionally, trainers highlighted that TDTScience may be particularly beneficial for pupils who struggle with written work where achievements are heavily based on literacy skills: it was felt that TDTScience strategies, that depend more heavily on oral contributions and practical work, helped pupils to develop and use a range of skills. This allows such pupils to excel in a topic like science in which they may previously have struggled.

In general, these teachers felt that TDTScience strategies inspire curiosity in pupils and help them to believe that they are 'scientists themselves' and to 'think outside the box' (trainer interview, Stage 3, new trainer). They commented that pupils often worry about the fact they have to write a paragraph or extended piece of writing at the end of the lesson and lose focus on the actual topic of the lesson. Teachers felt focused recording will help pupils to consolidate their learning through a method that they do not feel anxious about and therefore will help to engage pupils in the practical work throughout the rest of the lesson, making this a much more effective teaching strategy.

> *'So if they've done it and you've had that discussion, they're more likely to remember it unless, 'Now you must write a paragraph.' Well, they don't remember it, do they, because they're so focused on writing the paragraph rather than talking to you about what you think' (teacher focus group, Stage 3).*

During one training observation, a teacher anecdote supported this point remarking that one of their pupils who struggles with written work had excelled in a scientific investigation task in which they could verbally contribute ideas and learning which they may have had difficulty articulating in writing.

This is important to note for the purpose of understanding the train-the-trainer model as it demonstrates how teachers were able to return to schools quickly after each training session and implement strategies, and the pupils were responding well to this. This also highlights that the trainers were delivering programme content to a high standard that teachers felt able to use with confidence before the training was even complete, demonstrating good pragmatism, adaptability, and acceptability in teachers' and pupils' of the TDTScience methods.

RQ3. How necessary is each element to the training model overall?

Summary of findings

- It was clear that there was a need for all three stages of the pre-trial process to produce confident and competent trainers for the main trial.
- Being trained in TDTScience as if they were teachers (Stage 1) meant trainers could familiarise themselves with the content and approach and make suggestions for possible improvements.
- The train-the-trainers element (Stage 2) was valued particularly for having a chance to plan and practise with their delivery partners.
- The real-life delivery of TDTScience to teachers (Stage 3) acted as a rehearsal for the main trial. They could refine how to run the days, gain confidence, and identify where more practice and support was needed.
- Stage 3 also gave them an opportunity to explore the extent to which they could make the delivery their own—by inserting personal anecdotes, for instance—while maintaining fidelity to the TDTScience programme.

Importance of Stages 1 and 2

The training observations, together with the feedback survey administered by the developer, showed that at Stage 1 some trainers were unclear about the extent to which they could adapt the TDTScience materials and process. There were requests (which were rejected by the developer) to replace major TDTScience activities with ones of their own, although some less fundamental suggestions (possible supplementary material, for instance) were accepted. Experiencing the course as if they were teachers enabled trainers to spot possible improvements such as providing summary sheets for each day to give delegates an overview and help them build across the course and take elements back to the classroom. Trainers also became aware of their own weak areas, for instance, two expressed least confidence in the Forces topic and two had concerns about Earth and Space.

The importance of this first stage of the training process was underlined when one of the experienced trainers was absent for part of it, including when the delivery of day four was covered. This had unexpected consequences that impacted negatively on that pair's delivery of the fourth day in Stage 3.

> *'I certainly didn't pick it up early enough that actually … that they might have needed more support in relation to that day because one of them was missing and day four was the day that was the newest day—completely new content. So nobody had experience of that. Even we didn't have experience of delivering that day' (developer, Aug 2022 interview).*

One trainer—who had delivered TDTScience previously—suggested that Stage 1 could be curtailed to allow more planning time. Another discussed the possibility of shortening Stages 1 and 2.

There were various factors impacting how confident the trainers felt to deliver to the teachers following the Stage 2 training. There were several pleas for more time to talk and plan with their partner after the first two days were delivered in September 2021, and this was already scheduled into the November timetable but consequently got even more emphasis. As a result, trainers felt they were given adequate time to prepare and process the information and this was crucial for building their confidence. Trainers also felt they had ample opportunity to raise issues with the developer.

Additionally, trainers asked for more practice in areas they felt less secure delivering during the Stage 2 training, allowing them to build their confidence prior to delivering the course to the teachers. Trainers felt nervous about the 'practical prompts for thinking', a strategy used in TDTScience which shows how demonstrations can be used to stimulate thinking and talking and to deepen understanding of scientific phenomena. A wide range of demonstrations are used, from those

which focus on everyday items to those with real 'wow' factor and scope for 'performance'. Developers were surprised that these activities caused anxiety for the trainers but were happy to accommodate additional practice opportunities to improve confidence.

> *'It never occurred to me that that would be a source of anxiety for them because they're straightforward but actually, the time to practice—they have really valued, haven't they?' (developer interview, Stage 2, day two).*

Observing the final day of the Stage 2 sessions, it was clear that trainers had a much firmer grasp of the approach.

> *'I've really developed my confidence, I think … I feel confident in doing it delivering it with the teachers … I really like the way we've looked at them over the four days so you can see how those strategies are deepening and embedding, so I feel, you know, I'm really clear now on how we can do that' (new trainer, Stage 2 interview).*

Trainers, however, recognised that more work was needed, for instance to—

> *'go away and read the files again and just almost do like we did today [final November training day], practising it but in my own little head space and going through it' (trainer interview, Stage 2, new trainer).*

Trainers valued the chance in Stage 2 to present the training to each other. In addition, they felt this training provided them with a wealth of preparation time, which was important for feeling confident in delivering the training to the teachers.

> *'Lots of time to have it taught to us and experience it from a delegate point of view and then during the sessions putting us out of our comfort zone so we had to deliver parts of the day to other colleagues in the room, which when you're doing it to people of a calibre presenting it was a bit like, "Oh there's such and such a body"—it made you really engage with it so, yeah, really good prep time' (trainer interview, Stage 2, new trainer).*

Importance of Stage 3

Stage 3 was also a crucial part of the process of ensuring trainers felt ready to deliver the TDTScience training programme, being described as 'invaluable as a rehearsal for the main trial' (trainer interview, Stage 3, new trainer). This stage could alert trainers to considerations they had not realised were important, as with this trainer's concerns about the impact of teacher or pupil background on content knowledge.

> *'Prediction graph time v temperature as you boil water—just do not have last click as too high level for primary and could lead to a lot of misconceptions' (trainer survey, post day two, new trainer).*

> *'Audience has lots of young teachers and some subject knowledge quite a bit lower than expected. Although misconceptions were dealt with, it could have been more effective in future and I would be more prepared' (trainer survey, post day two, new trainer).*

This final stage of the pre-trial allowed trainers' real-life experiences to highlight issues that had not been obvious in Stages 1 or 2—some trainers, for instance, found they had quieter groups and would have welcomed strategies for managing them. However, the four and a half days of delivery to teachers provided them with the opportunity to explore potential solutions for themselves, such as mixing the groups. Trainers recognised the importance of their previous experience with the materials. Those new to TDTScience training felt they needed to spend more time planning and familiarising themselves with the content to boost their confidence. This was exemplified in post day four survey responses. Day four included a lot of rewritten material that had not been delivered by any of the trainers before and several had been worried about delivery because of this, but their expectations were exceeded as illustrated by the following observation.

> *'I felt that day four is a little different from some other days—less practical investigation and with some content I had not presented before—so I was a little apprehensive about this. However, in the event I felt*

the day went well and the teachers continued to be very engaged' (trainer survey, post day four, experienced trainer).

Trainers acknowledged that Stage 3 would put them in good stead for the main trial.

'I am so pleased to have seen the whole cycle through from a trainer's perspective and I feel even better prepared for the main trial' (trainer survey, post day five, new trainer).

'I think having to run it as the pre-trial gives us great practice for the main trial' (trainer interview, Stage 3, experienced trainer).

Importance of the pre-trial phase overall

The evaluation made it clear that each of the three stages made an essential contribution to creating trainers who felt confident and prepared to deliver the TDTScience course in the main trial. There was evidence that trainers' understanding of the TDTScience approach was crystallising throughout the pre-trial process, even during the later parts of Stage 3 delivery.

'I feel I have developed in my understanding about TDTScience strategies and the bigger picture of how each day adds to the teacher's TDTScience toolkit' (trainer survey, post day four, new trainer).

This could be true whether trainers were new to TDTScience or had been involved previously, and meant main trial delivery was continually being re-thought, as this interview extract shows.

T1: 'I think we found more threads, golden threads that pull it all together that isn't in the training manual. So obviously there are links but I think we've found more in that we've constantly referred back to things that we introduced, to things that we've done already. "Do you remember this? Now we're doing this and this links to that …" We've had more penny-drops since we've linked things up from day to day.'

T2: 'We were saying it will be interesting when we do day one again if we do that differently now that we see the overall picture' (T1 experienced trainer, T2 new trainer, Stage 3).

The data highlights that trainers found it difficult to take ownership of the training as they were very conscious of delivering someone else's programme, and this was something that they needed to work on particularly during Stage 2 and Stage 3. Trainers felt that all aspects of the training were predetermined and their natural delivery styles had to be adapted to accommodate this. This was exaggerated when trainers did not feel confident in delivering some of the topics: they were keen to do the programme justice but were conscious that the training belongs to Science Oxford and is not theirs to get wrong.

'Very, very rarely would I include something like that in my own CPD delivery and what I think would be good is almost being aware of when you're training what are the things you wouldn't normally do. So basically these are the ways we want you to train. There are going to be different ways we want you to train. Some of it is leading discussions. Some of it is talking to PowerPoint slides. Some of it is leading practical activities. Some of it is delivering practical prompts for thinking. So, like, a range of training strategies and everybody will have different confidence levels and comfort levels in different ones and for me, all the others are absolutely fine but the practical prompts for thinking' (trainer interview, Stage 3, new trainer).

This was echoed in the trainer surveys where trainers felt conscious of presenting the work of others.

'I still feel it is hard to be as well prepared for something that is someone else's "baby" and we have to deliver it is closely as possible to the original' (trainer survey, post day one, experienced trainer).

Trainers were positive about the training process as a whole, discussing how the developers 'have this beautifully in hand' (trainer survey, post day two, experienced trainer). Trainers appreciated the time this whole process gave them to prepare: 'So by the time we do that for the main trial, I'm going to look really slick!' (trainer interview, Stage 3, new trainer).

RQ4. What improvements could be made to the model to benefit the TDTScience intervention and training design more widely?

Summary of findings

- Although the fundamentals of the TDTScience programme were already firmly fixed, the developer was willing to consider suggestions for improvements from trainers at every stage of the pre-trial and ready to adopt them if they were considered appropriate and easy to incorporate. This helped give trainers a sense of ownership.
- During Stage 3, some trainers had ideas for amendments that could be adopted autonomously because they referred to practicalities of delivery rather than the content of the training.
- There was little feeling among trainers that the training model could have been improved by eliminating or compressing any elements, even among those who had delivered TDTScience before.

Ongoing changes

The developers invited and received feedback from the trainers during the Stage 1 training and in some cases modified the course accordingly. Developers were willing to accommodate suggested improvements provided they were true to the programme and relatively straightforward to incorporate, especially if it gave trainers the confidence and ownership to deliver the programme in Stage 3.

> *'It's working well, because [the trainers have] taken on board that we're not at the point where the course can be fundamentally changed but actually we're still very open to suggestions which are quite easy, relatively easy, for us to implement at this stage. They're not suggesting things that affect the whole structure of the way we do things and the order' (developer interview, Stage 2).*

Developers considered that such contributions have been valuable in the development of the TDTScience programme and that implementing this feedback from the trainer evaluations was something they could not do on the previous effectiveness trial. Additionally, trainers were grateful that the developers took on their feedback and comments and it helped the trainers feel valued.

> *'And they've really taken on everybody's feedback as well, haven't they … they've altered the materials in light of the comments' (trainer interview, Stage 2, new trainer).*

Suggested changes for main trial

Throughout the pre-trial, trainers offered a range of recommendations for changes to the programme and developers also identified changes they wanted to make prior to the main trial.

> *'There are some tweaks that we've written down as we've gone through, as you'll have seen. So the first PMI, when you're writing it, I wrote this bit. I looked up a PMI that was linked to 'earth and space' and the minute I put it up there, I thought, that's far too hard. So what I will change, with talking to [other developers], is I'd like to put in a PMI that's a much more straightforward one from another area of the curriculum. I would put in, door handles are made of chocolate and do one that I know works well. And then say, you can do these in Earth and Space. I think the earth stopping spinning, I'll get rid of. I think it's too difficult for primary children' (developer interview, Stage 1, day two).*

Although planning time was incorporated into Stage 2 training, some trainers would have liked more, especially with their partner.

> *'Probably [I would have liked] just having more time to plan who was doing what for each day of training so that this was not additional work to try and fit in around other work commitments after the training days' (trainer survey, post day one, experienced trainer).*

The pairings were not decided from the beginning, although trainers felt it may have been useful to know who they would be working with from the outset. They believed this may have helped visualise training when taking notes and planning. By contrast, developers felt that the delay in announcing the partnerships had allowed them to observe the individuals and gain more understanding of who might work well together, as well as considering geographic factors. Trainers suggested further documentation that might be helpful, such as a summary of the main TDTScience principles with an overview of where such principles or strategies appear in the training.

> *'As a new person to TDTScience I would have valued a summary of the main principles—the schematic with HOT in the centre, practical investigations, problem solving etc. etc. and a summary of each of the strategies. An overview of where these principles/strategies appeared in the training would have been useful' (trainer survey, post day one, new trainer).*

Others had a self-development focus, as this trainer alludes to:

> *'My main aims are personal—to further develop my questioning skills, which are a bit rusty, so I need to practise these. I feel that I know the main messages, had chance to experience all the activities and practice the practical prompts' (trainer survey, post day one, new trainer).*

For Stage 3, some trainers felt that it may have been useful for the teachers to have more documents that they could annotate throughout the sessions, such as a pro forma listing strategies and one-page summaries for each practical investigation. Alternatively, they could be provided a sheet at the end of the day that highlights all the strategies, principles, and key terms they have covered during that session. Teachers would be able to refer to these to consolidate their knowledge when planning a lesson using TDTScience strategies. Trainers felt this would also help teachers to engage in 'sharing good practice' because they could reflect on how their notes were implemented in the classroom.

> *'I think they can then annotate as they get, "Oh, that's a Bright Ideas! We've done that, we've done that, that's what that means. I'm building that, I think I know what that is, I'm building that in my toolkit", or "We've done that." So we tried to model it on that picture[5] … we've got into the habit of doing that, is quite a good retrieval for them to talk about what did we do last time because they're using that cycle and also freshen their memories of what we've done' (trainer interview, Stage 3, new trainer).*

The trainers suggested that sometimes it was not clear how the practical activities linked to the topic or how the content links to the TDTScience principles.

> *'It would have been good to link the practical investigations more explicitly to the Space topic' (trainer survey, post day three, experienced trainer).*
>
> *'I'd also like to see clear links to HOT for more of the practical prompts for thinking, for example, suncream/UV activity' (trainer survey, post day five, experienced trainer).*
>
> *'Also I think it is strange that we introduce PMI in day three but the gap task does not link to this at all' (trainer survey, post day three, experienced trainer).*

They also wanted more cohesion between the different days and how each day builds on the last.

> *'[It would be helpful to have] more opportunities to make links between different days and reference things covered previously and make more explicit how this builds. Also, give all participants the main objectives for each day at the start of each day' (trainer survey, post day four, experienced trainer).*

[5] Image of TDTScience cycle for trainers to display at training sessions.

Day five covered how TDTScience relates to Ofsted requirements, and there were different views about how engaging this section was.

Some trainers suggested alternative ways of tackling this section.

> *'The section on Ofsted and how TDTScience addresses concerns was quite dry—could we do it as a card matching exercise or something similar?' (trainer survey, post day five, new trainer).*
>
> *'I wonder about the way the Ofsted ideas are presented and wonder if it might be better to encourage the teachers to look at the main points and then justify the TDTScience approach from their perspective?' (trainer survey, post day five, experienced trainer).*

Another priority for teachers was that the training content, such as the practical activities, explicitly linked to curriculum knowledge.

> *'Some of our teachers struggled to see links between some of the investigations and National Curriculum content. Of course, we emphasised the importance of focusing on Working Scientifically as another focus, but there was still a sense of having more investigations/practical activities that directly link to curriculum knowledge. This was true of "glider", "protected egg", and "longest legs"' (trainer survey, post day five, experienced trainer).*

The trainers were anxious that teachers asked for clarification when they were struggling to understand and felt this was an area where they could tighten their practice.

> *'I felt that they probably asked a few more questions around their areas of lack of confidence, which I think is the first time they'd really done that but I guess I pushed them quite a lot. I was really pushing them to say, look if there's anything that you don't understand about any of these strategies, this is pretty much your last chance to ask, you need to ask now ... There was one question around the focused recording and she really didn't get it and I thought it's a shame that you've left it so late but then that's something that I would definitely change in the next training, in the main event, is ask them after every strategy' (trainer interview, Stage 3, experienced trainer).*

Trainers suggested other ideas that they felt would improve the course, which shows a level of ownership as they wanted to make it more engaging and appealing. These included:

- using names badges;
- email teachers between sessions to remind them about gap tasks;
- a physical globe and boat to demonstrate that the earth is spherical (day three);
- writing a song or rap to illustrate the power of creative recording (day four);
- more relevant activities, for instance, when illustrating the difference between discrete data—a finite value that can be counted—and continuous data—an infinite number of possible values that can be measured (day two); and
- a practical element for day four, for example, a trainer suggested, 'Maybe design and test the ultimate seed that can be dispersed by air, water, and animal—to increase problem-solving element and discuss focused recording further' (trainer survey, post day four, experienced trainer).

On completion of the three training stages that preceded the main trial, the developers felt the process could not have been improved upon.

> *D1: 'We've given this our very best shot. We've got the best group of trainers I think we could have and we can do no more. If we don't get the results, I just don't know what we would do differently.'*
>
> *D2: 'I think it would just demonstrate how incredibly difficult it is to replicate impact through a train-the-trainer model, to be honest' (developer interview, August 2022).*

References

Bartlett, M. S. (1954) ‘A Note on the Multiplying Factors for Various χ2 Approximations’, *Journal of the Royal Statistical Society, Series B (Methodological)*, 16 (2), pp. 296–298. http://www.jstor.org/stable/2984057

Browne, M. W. (2001) ‘An Overview of Analytic Rotation in Exploratory Factor Analysis’, *Multivariate Behavioral Research*, 36 (1), pp. 111–150. doi:10.1207/S15327906MBR3601_05

Hanley, P., Elliott, L. (2021) ‘Thinking, Doing, Talking Science Protocol’ (second re-grant: a two-armed, cluster randomised trial), London: Education Endowment Foundation. https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/thinking-doing-talking-science-effectiveness-trial-2

Horn, J. L. (1965) ‘A Rationale and Test for the Number of Factors in Factor Analysis’, *Psychometrika*, 30 (2), pp. 179–185. doi:10.1007/BF02289447

Kaiser, H. F. (1974) ‘,An Index of Factorial Simplicity’, *Psychometrika*, 39 (1), pp. 31–36. doi:10.1007/BF02291575

Kind, P., Jones, K. and Barmby, P. (2007) ‘Developing Attitudes Towards Science Measures’, *International Journal of Science Education*, 29 (7), pp. 871–893. doi: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. https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/TDTS.pdf

Lorenzo-Seva, U. (1999) ‘Promin: A Method for Oblique Factor Rotation’, *Multivariate Behavioral Research*, 34 (3), pp. 347–365. doi:10.1207/S15327906MBR3403_3

Lorenzo-Seva, U. and Ferrando, P. J. (2013) ‘FACTOR 9.2: A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models’, *Applied Psychological Measurement*, 37 (6), pp. 497–498. doi:10.1177/0146621613487794

Appendix 1: Factor analysis (pupil survey)

Introduction

The pre-trial pupil survey was used to collect data on pupils' attitude towards science using a questionnaire adapted from the one used in the first effectiveness trial (Kitmitto, González, Mezzanote, & Chen, 2018) which was itself adapted from a questionnaire developed in 2007 by Kind, Jones, and Barmby (2007). The version used in the pre-trial is shown in Appendix 2 and differed from the one used in the first effectiveness trial in two ways: i) four items intended to strengthen the self-efficacy index (2i, 2j, 2k and 3h) were added, and ii) the wording of item 3c was changed from "*I can decide what to do for myself in science practical work*" to "*We can decide for ourselves how to do science practical work*". Item 2j was accidently repeated in the questionnaire as item 2l; data for this item was removed prior to analysis. The 27 items were measured on a five-point Likert scale ranging from 1 ('disagree a lot') to 5 ('agree a lot'). Reverse scoring was implemented for negatively phrased questions (1f, 1h, 2a, 2b, 2c, 2f, 3f and 3b) to ensure consistency across questions with a high score reflecting a strong positive response and a low score a strong negative response. For evaluation purposes, it was important to summarise the items into meaningful constructs. As in the first effectiveness trial, exploratory factor analysis (EFA) was conducted to support the creation of indices to represent underlying constructs in the questionnaire.

The pre-trial survey also included a series of questions asking about the frequency in which different in-class activities are performed in science lessons and an open-ended question for pupils to write anything else they wanted to say about their science lessons in the last year. Responses to these are summarised separately and were not included in the factor analysis for the attitudes to science questionnaire.

Methods

A total of 103 pupil surveys were collected, of which 25 (24.3%) were from Year 4 students and 78 (75.7%) from Year 5 students, though only those with complete data for all items were included in analyses. The suitability of the data for factor analysis was assessed using the Kaiser-Meyer-Olkin measure (Kaiser, 1974) and Bartlett's test (Bartlett, 1954). Polychoric correlations were used as these can be calculated for ordinal response Likert scales typical on outcome measures, such as was used here. The method of factor extraction was unweighted least squares, with Promin rotation (Browne, 2001; Lorenzo-Seva, 1999). Based upon the findings of a previous TDTScience study, it was hypothesised that there would be three factors contained within the survey, so a three-factor model was fitted initially. Parallel Analysis (Horn, 1965) was then used to determine how many factors were required to explain the variation.

Two sets of analyses were conducted. First, we attempted to reproduce the factor analysis conducted in the first effectiveness trial by excluding the items that were added for this study. Then we conducted the factor analysis for the whole new item set.

FACTOR 9.2 software (Rovira i Virgili University, Tarragona, Spain) was used for the factor analysis (Lorenzo-Seva & Ferrando, 2013). All other analyses and data preparation were conducted in Stata v17 (StataCorp LP, College Station, TX, USA). Significance was assessed at the 5% level.

Results

Reproducing Previous Factor Analysis

The additional questions, 2i, 2j, 2k and 3h, were removed and an EFA conducted on the resulting 74 records with complete data for the remaining 23 items. Bartlett's test was statistically significant ($\chi^2 = 742.6$, degrees of freedom (df) = 253, $p<0.001$), suggesting that the 23 variables were sufficiently correlated with each other. The Kaiser-Meyer-Olkin test yielded a value of 0.75, exceeding the commonly recommended value of 0.6 (Kaiser, 1974), indicating fair partial correlations between the variables after controlling for all other variables. These tests provided evidence that factor analysis could be applied to these data.

Three factors accounted for 49.1% of the variance (32.7%, 8.6% and 7.7%, respectively). Loadings for the three-factor model are presented in Table 1. Two items did not load ≥|0.3| on any of these factors (1f and 2f), eight items loaded ≥|0.3| on factor 1, nine on factor 2 and 11 on factor 3. There was cross-loading of one item on factors 1 and 2, five items

on factors 1 and 3, and one on factors 2 and 3. The inter-factor correlations were 0.42 between factors 1 and 2, 0.26 between factors 1 and 3, and 0.64 between factors 2 and 3.

However, Parallel Analysis indicated that a one-factor solution was sufficient (Table 1). This factor accounted for 32.7% of the variation. Eighteen items loaded on this factor ≥|0.3|. Cronbach's Alpha was employed to estimate the internal consistency of this factor. Cronbach's Alpha was 0.91, indicating a very high level of reliability.

This factor contains all the same items as the items included in the 'interest in science' and 'self-efficacy' indices generated in the first factor analysis (Kitmitto et al., 2018), but with the inclusion of item 2c.

Table 1: Factor loadings of original question set from the EFA

	Three-factor solution Original survey data				**Three-factor solution New survey data**				**One-factor solution**
Pre-Trial TDTScience Survey	**Factor 1: Interest**	**Factor 2: Self-efficacy**	**Factor 3: Activity**		**Factor 1**	**Factor 2**	**Factor 3**		**Factor 1**
1a Science lessons make me think	0.5401						0.597		0.530
1b I look forward to my science lessons	0.7811				0.371		0.776		0.733
1c Science lessons are interesting	0.7871				0.324		0.366		0.782
1d I would like to do more science at school	0.7280						0.656		0.721
1e Science is fun	0.8206					0.365			0.627
1f We spend a lot of time in science lessons copying from the board			0.3508						
1g I enjoy discussions in science lessons	0.6042				0.512		0.346		0.461
1h Science lessons are boring	0.6499					0.340	0.464		0.751
2a I find science difficult to understand		0.6772					0.524		0.472
2b I am just not good at science		0.6068					0.336		0.537
2c I think science is more for boys							0.533		0.362
2d I understand everything in my science lessons		0.5025				0.428			0.673
2e We often have discussions in science lessons	0.3007					0.500			0.440
2f We do a lot of writing in science lessons									
2g It is important that we learn science	0.6259				0.316	0.474			0.470
2h I like thinking about scientific ideas	0.6719				0.454		0.760		0.736
3a Doing practical work in science lessons is fun	0.5878					0.742			0.608
3b We already know what will happen when we do science practical work			0.4269		-0.523		0.301		
3c We can decide for ourselves how to do science practical work					0.563				
3d We do practical work in most science lessons						0.470			
3e I look forward to doing science practicals	0.6451					0.692			0.762

3f Practical work in science is boring	0.4493		0.3510			0.854			0.521
3g Solving science problems is enjoyable	0.6473				0.547				0.675

Factor Loadings < │0.3│ were suppressed.

Pre – Trial TDTScience Factor Analysis

The extended survey included 27 questions, with complete data available for 72 participants. Bartlett's test was statistically significant (χ^2 = 704.2, degrees of freedom (df) = 351, p<0.001) and the Kaiser-Meyer-Olkin Measure score was 0.72, both of which suggested the suitability of the data for factor analysis.

Three factors accounted for 46.0% of the variance (32.2%, 7.6% and 6.2%, respectively). Loadings for the three-factor model are presented in Table 2. Four items did not load ≥|0.3| on any of these factors (2b, 2k, 3d and 3h), 12 items loaded ≥|0.3| on factor 1, eight on factor 2 and 11 on factor 3. There was cross-loading of three items on factors 1 and 2, two items on factors 1 and 3, and three on factors 2 and 3. The inter-factor correlations were 0.61 between factors 1 and 2, 0.71 between factors 1 and 3, and 0.56 between factors 2 and 3.

However, Parallel Analysis recommended a one-factor solution (Table 2). This factor accounted for 32.2% of the variation. Twenty items loaded on this factor ≥|0.3|. Cronbach's Alpha was employed to estimate the internal consistency of this factor. Cronbach's Alpha was 0.92, indicating a very high level of reliability.

This factor contains all the same items as the items included in the 'interest in science' and 'self-efficacy' indices generated in the first factor analysis (Kitmitto et al., 2018), except Question 2e is not included, plus the four additional items.

Table 2: Factor loadings of new question set from the EFA

	Three-factor solution				**One-factor solution**
Pre-Trial TDTScience Survey	**Factor 1**	**Factor 2**	**Factor 3**		**Factor 1**
1a Science lessons make me think			0.721		0.478
1b I look forward to my science lessons			0.932		0.689
1c Science lessons are interesting	0.302		0.419		0.813
1d I would like to do more science at school			0.778		0.725
1e Science is fun	0.582				0.791
1f We spend a lot of time in science lessons copying from the board	0.460				
1g I enjoy discussions in science lessons			0.320		0.307
1h Science lessons are boring	0.377		0.470		0.730
2a I find science difficult to understand			0.450		0.439
2b I am just not good at science					0.340
2c I think science is more for boys		-0.429	0.582		
2d I understand everything in my science lessons	0.545				0.648
2e We often have discussions in science lessons	0.371				
2f We do a lot of writing in science lessons		-0.336			
2g It is important that we learn science	0.308	0.492			0.485
2h I like thinking about scientific ideas			0.924		0.732
2i Science is my strength		0.438	0.328		0.771
2j I have made good progress in science this year	0.510	0.451			0.642
2k I like how my teacher teaches science					0.444
3a Doing practical work in science lessons is fun	0.708				0.612
3b We already know what will happen when we do science practical work		-0.711	0.305		

3c We can decide for ourselves how to do science practical work	-0.333	0.699			
3d We do practical work in most science lessons					
3e I look forward to doing science practicals	0.584				0.785
3f Practical work in science is boring	0.882				0.556
3g Solving science problems is enjoyable		0.542			0.673
3h I enjoy doing practical work with my classmates					0.549

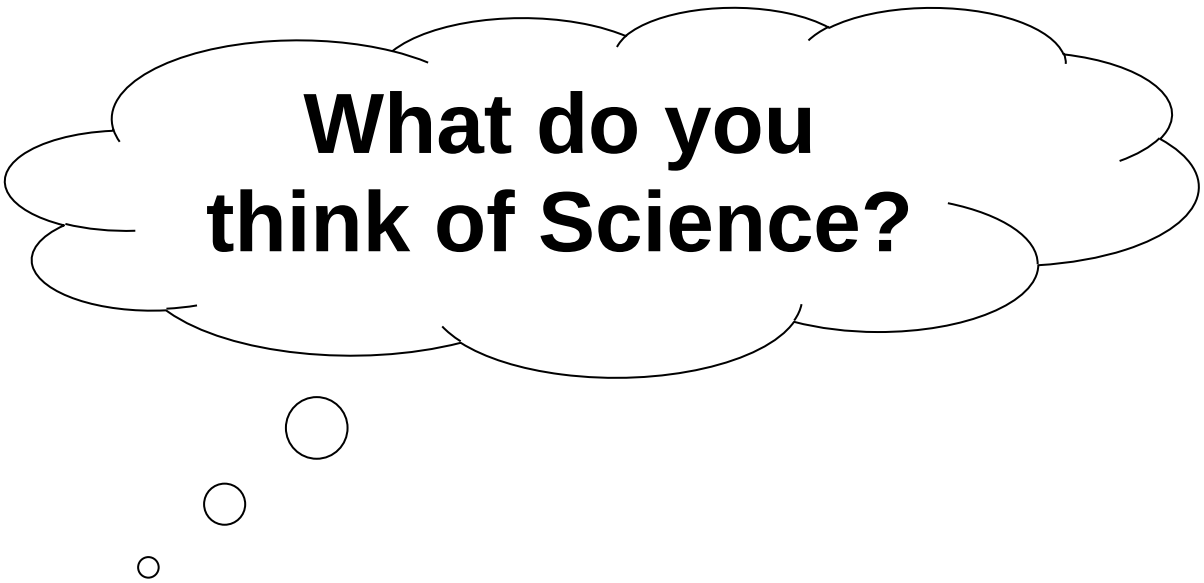
Factor Loadings < |0.3| were suppressed.

Interpretation and limitations

The findings of the EFA suggest the existence of one underlying construct rather than the three hypothesised, with the variables loading onto the singular factor implying the underlying factor is a combination of the previously found interest and self-efficacy factors. Hence, if the research to be conducted within the TDTScience main trial is solely aiming to assess these two aspects of student attitudes towards science, the questionnaire could be reduced to include only the 20 items loaded on Factor 1. Utilising the most recent version of the survey in its complete form, however, would still be useful if there is an interest in understanding pupils' views on in-class activities and exercises, as these components are largely addressed by the questions that failed to load in the one-factor model. Responses to these items could be summarised separately, while the other 20 items could be combined in one index.

A limitation of this factor analysis was that the sample size was very small, and did not meet the general rule of thumb to have at least 10 observations per variable (for example, there was 72 observations for 27 items in the full analysis – a ratio of only 2.7 observations per item). The ratio of 10 observations per item is recommended to allow the correlation matrix used in the factor analysis to stabilise and so increase the likelihood of the results being able to be reproduced with a different sample population.

Appendix 2: Pre-trial questionnaire



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

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

Q1 Learning Science at school

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

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

Q2 Learning Science at school

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

	Agree a lot	Agree a bit	Not sure	Disagree a bit	Disagree a lot
a) I find science difficult to understand	☐	☐	☐	☐	☐
b) I am just not good at science	☐	☐	☐	☐	☐
c) I think science is more for boys	☐	☐	☐	☐	☐
d) I understand everything in my science lessons	☐	☐	☐	☐	☐
e) We often have discussions in science lessons	☐	☐	☐	☐	☐
f) We do a lot of writing in science lessons	☐	☐	☐	☐	☐
g) It is important that we learn science	☐	☐	☐	☐	☐
h) I like thinking about scientific ideas	☐	☐	☐	☐	☐
i) Science is my strength	☐	☐	☐	☐	☐
j) I have made good progress in science this year	☐	☐	☐	☐	☐
k) I like how my teacher teaches science	☐	☐	☐	☐	☐
l) I have made good progress in science this year	☐	☐	☐	☐	☐

Q3 About practical work in school science

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

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

Q4 How often do you do the following in science lessons?

	Very often	Quite often	Not very often	Rarely	Never
a) Discuss things together as a whole class	☐	☐	☐	☐	☐
b) Discuss things in pairs or small groups	☐	☐	☐	☐	☐
c) Do practical work in pairs or small groups	☐	☐	☐	☐	☐
d) Learn scientific facts	☐	☐	☐	☐	☐
e) Solve scientific problems	☐	☐	☐	☐	☐
f) Design or plan practicals	☐	☐	☐	☐	☐
g) Carry out practicals	☐	☐	☐	☐	☐
h) Interpret results from practicals	☐	☐	☐	☐	☐

Q5 Finally, is there anything you would like to say about your science lessons this year?

Appendix 3: Brief overview of TDTScience 4.5 Day training sessions

Training Day	Sessions include	Brief content notes
Day 1 Materials	Introduction to the TDTS project	Overview of Teacher Folder and website; Background and evidence; Ethos of CPD and mapping to curriculum content
	Various practicals	e.g. Paper Flowers & Protect an Egg practicals
	Challenge and Higher Order Thinking (HOT) in science	Definitions and evidence
	Introduction to the Bright Ideas Time (BIT): 'Odd One Out' (OOO)	Examples of pupil responses and teacher feedback
	The Science of Materials	States of matter: role play
	Practical Prompts for Thinking	e.g. Use PPT to go from 'wow' to 'wonder' and HOT
GAP TASK – trying an 'Odd One Out' with pupils in science lessons		
Day 2 Forces	Introduction and Practical Prompts for Thinking (PPT)	TDTS strategies reminder
	Various practicals	e.g. Which shoes have the best grip & The Marble Maze
	Sharing of good practice	Discussion: OOO gap task teacher feedback
	The Science of Forces	Pushes/pulls and getting a 'feel' for Newtons
	Higher Order Questioning (HOQ) BIT: 'Big Question' (BQ)	HOQ and inclusive challenge BQ examples and pupil feedback
	More PPTs for HOT	
GAP TASK – trying a 'Big Question' discussion with pupils		
Day 3 Earth & Space	Introduction BIT: The 'Big Question' (BQ) cont.	TDTS strategies reminder Constructivist view of learning
	Further thinking about Higher Order Questioning (HOQ)	Types and examples of questions teachers ask Planning (& celebrating) HOQ and HOT: OOO, BQ
	Sharing of good practice	Discussion: BQ gap task teacher feedback
	Various practicals	e.g. Strongest legs & Glider Challenge
	PPTs for HOT in Earth and Space Galaxies – seeing history	Scale and use of models
GAP TASK – practicals for Higher Order Thinking with Focussed Recording		
Day 4 Living Things	Introduction The Bright Ideas Time (BIT)	TDTS strategies reminder e.g. of OOO, PMI, BQ and making your own Questioning – value of open and closed questions
	Life Cycles	Researching secondary sources – HOTS Observation over time - HOT and FR
	Sharing of good practice	Discussion: Practical and FR gap task teacher feedback
	Various Practicals	e.g. Create an Animal & Seed dispersal
GAP TASK – crafting a lesson		
Day 5 Leading TDTS in your school	Sharing of TDTS practice	Based on 'crafting a lesson' gap task
	Leading TDTS in Your School Part 1: your classroom practice Part 2: working with others	Link to OFSTED 2019 primary science research Discussion: how TDTS practice addresses issues Small Changes Big Impact: value of TDTS evidence Discussion: effective staff meetings, dissemination

Appendix 4: Overview of the findings relating to the content of TDTScience

Training Day	Sessions include	Main Findings
Day 1 Materials	Introduction to the TDTS project Various practicals Challenge and Higher Order Thinking (HOT) in science Introduction to the Bright Ideas Time (BIT): 'Odd One Out' (OOO) The Science of Materials Practical Prompts for Thinking	• Some teachers were familiar with OOO so there was no need to go into detail about this strategy. • Trainers felt BIT was a particularly useful strategy for teachers; can be implemented straight away; encourages pupils to have scientific discussions. • Teachers commented that 'talking activities' like BIT or OOO are useful at the start of the lesson; encourage reflection on previous learning; promotes exploration and curiosity.
Day 2 Forces	Introduction and Practical Prompts for Thinking (PPT) Various practicals Sharing of good practice The Science of Forces Higher Order Questioning (HOQ) BIT: 'Big Question' (BQ) More PPTs for HOT	• Trainers and teachers felt nervous about delivering PPT and felt they needed a lot of practice before presenting one. • Forces was agreed to be a difficult topic for teachers. • Sharing good practice was a particularly valuable tool and a lot was learnt from speaking to one another about their experiences so far.
Day 3 Earth & Space	Introduction BIT: The 'Big Question' (BQ) cont. Further thinking about Higher Order Questioning (HOQ) Sharing of good practice Various practicals PPTs for HOT in Earth and Space Galaxies – seeing history	• This day was challenging to deliver, and difficult and potentially daunting for teachers to receive due to the complex nature of the topic. • Teachers felt that it often difficult to address misconceptions that children have in this topic. TDTS provided some practical demonstration ideas to help with this.
Day 4 Living Things	Introduction The Bright Ideas Time (BIT) Life Cycles Sharing of good practice Various Practicals	• Trainers felt the topic didn't lend itself well to the TDTS structure and was more difficult to present. • Teachers expressed worry about using focused recordings due to it not being enough evidence of learning for Ofsted, but felt that taking part in TDTS provided a good justification for using this strategy. Recognised that it requires good preparation. • Trainers were concerned that teachers hadn't fully understood the use of focused recording and this was a challenge. • 'Crafting a lesson' practical was extremely useful on this day for returning to the classroom and implementing strategies.
	Focused recording	
Day 5 Leading TDTS in your school	Sharing of TDTS practice Leading TDTS in Your School Part 1: your classroom practice Part 2: working with others	• Useful for sharing good practice and preparing teachers to disseminate the information to other staff members.

Appendix 5: Recommendations for the main trial research methods

Stage 3 was used by the evaluation team to explore the usage of new or amended research instruments to be employed in the subsequent trial: teacher surveys, lesson observation schedules, pupil measures, and audio recordings.

This latter research instrument on recording lessons was conceived as a potential alternative approach to obtain examples of more "typical lessons" for the purposes of the main trial evaluation. Because TDTScience is not a prescriptive programme or pedagogy but uses various techniques and strategies, lessons that were observed in-person might be particularly susceptible to the "measurement effect" where teachers prepare lessons with more TDTScience characteristics than usual. Although this can be triangulated with pupil feedback, it might also be possible to reduce the problem by using remote recording of lessons. The intention was to ask teachers in two schools to audio-record their science lessons over several weeks. After discarding the very first as potentially atypical, the evaluators would randomly sample two from each teacher for analysis. This would minimise the measurement effect as teachers/children become used to the recorder. The effectiveness of this evaluation approach would be assessed and, if judged successful, it would be used alongside in-school observations in the main trial. Success would be judged on several criteria including the lessons being recorded as planned, recordings returned to York Trials Unit, usable audio quality, and the content providing information that would enhance lessons captured by a physically present observer.

Table 8: Stage evaluation approach

Focus	Method	Why?	Number
Instrument (teacher)	New survey	Develop and pilot teacher survey for main trial	15 schools (30+ teachers)
Instrument (lesson)	New observation schedule	Develop and pilot observation schedule and fidelity measures for main trial	4 lessons (2 schools)
Instrument (pupil)	New survey	Develop and pilot pupil survey (including attitude statements) for main trial	2 classes x 4 schools
Technique (lesson)	Audio recording	Triangulate face-to-face observation with data from audio-recorded lessons (two sampled from several). Test usability of recording in main trial. To overcome possible "measurement effect" when observing in-school lessons that may have been more carefully prepared than usual.	2 classes x 2 schools

As outlined in the Introduction, the pre-trial stage represented an opportunity to pilot research instruments and methodology. Where relevant, some of the findings have been included in the main body of this report, but the key purpose was to guide the future approach.

The compressed delivery of Stage 3 (over five months rather than six) had some impact on the fieldwork that could be completed. All elements were covered, however there was some reduction in anticipated sample size as detailed in Table 8 below.

Table 9: Amendments to data collection

Instrument/technique	Aim	Target	Achieved
New teacher survey	Develop and pilot teacher survey for main trial	15 schools (30+ teachers)	9 schools (14 teachers)
New lesson observation schedule	Develop and pilot observation schedule and fidelity measures for main trial	4 lessons (1 per teacher across 2 schools)	4 lessons (1 per teacher across 2 schools)
New/adapted pupil survey	Develop and pilot pupil survey (including attitude statements) for main trial	2 classes per school at 4 schools	2 classes per school at 2 schools (N=103)
Audio recording	Triangulate face-to-face observation with data from audio-recorded lessons (two sampled from several). Test usability of recording in main trial. To overcome possible "measurement effect" when observing in-school lessons that may have been more carefully prepared than usual.	2 teachers' classes at 2 schools (2+ recordings per teacher)	1 teacher's class at 2 schools (2+ recordings per teacher)

Teacher survey

The instrument overall functioned as expected, although there were some areas where it could be improved. These were all incorporated for the baseline teacher survey in the main trial:

- Finer grained pre-selected responses for "How often do you teach science (to your own class)?" (all chose the most frequent category of "once a week or more often"): suggest every day; three or four times a week; twice a week; once a week; once a fortnight; less often; other.
- Finer grained pre-selected responses for "How long, on average, does a science lesson last?" (all chose "1-2 hours"): suggest allow teachers to use a slider with scale from 0 to 180 minutes.
- Add a "no" option to the pre-selected responses for "Have you personally been involved with any science-related training or initiatives so far this school year?" (accidentally omitted).

Only one teacher made a suggestion about the survey when invited to do so: they proposed asking about the different units covered (the survey focused on the TDTS approaches rather than the content). It could be considered whether this line of questioning would be informative for the main trial post-survey.

Lesson observation schedule

Two of the four lesson observations were conducted jointly by two researchers. This showed that the schedules were straightforward to interpret, could be filled in consistently and worked well. Structured rating scales (developed in conjunction with Science Oxford) covered key components of TDTS delivery and allowed an assessment of fidelity of delivery. Free-form fieldnotes allowed the flexibility to capture any other pertinent information.

Pupil survey

The survey comprised the attitude statements as used in the previous efficacy and effectiveness trial, with four additional statements designed to strengthen the self-efficacy index. Another intended new statement ("I can solve difficult problems related to science") was omitted from the survey in error. A new question was added to explore the frequency with which pupils said they did various activities in their science lessons, for triangulation with other elements of the

evaluation. A factor analysis was conducted on the attitude statements and the behavioural question was explored through descriptive statistics. Details of the questionnaire and the factor analysis can be found in Appendix 1.

In brief, with the important caveat that it was based on a small sample, the factor analysis did not support the generation of two separate scales for efficacy and self-interest. We propose that the whole questionnaire is used in the main trial, and we would rerun the factor analysis on the larger sample size. If this produces a similar result, we would only analyse one scale (20 items) and provide a summary of the other items.

Analysis of the new question revealed a possible irregularity. Fewer pupils reported that they often carried out practicals (50% of which 29% stated quite often, 21% stated very often) than said they often did practical work in pairs or small groups (72% of which 41% stated quite often, 31% stated very often). Logically, it might be expected that the former would include, and therefore be a larger percentage than the latter. One possibility is that “practicals” and “practical work” were being interpreted differently. Elsewhere in the questionnaire, the term “practical work” is nearly always employed, so it is recommended that all references to “practicals” are changed to “practical work” to eliminate inconsistency. There may, however, be other reasons for the discrepancy. For instance, pupils may be saying that when they do practical work, it is usually as pairs or small groups. Consequently, the statement might be changed to “watch the teacher do practical work”.

Audio recording

Two teachers at different schools agreed to audio-record some of their science lessons. They were sent the equipment and instructions for use.

In terms of practicality of the approach, it was not a popular thing to volunteer for, with 20 pre-trial schools declining to participate compared with 11 for the visits and just 5 for each survey. The two teachers recruited to record the activity managed the technical side well, although not as many lessons were recorded as anticipated. The equipment (which was expensive) was returned undamaged. It was quite a technical challenge to upload the recordings from the recording units and very time-consuming to clean the data (matching the recordings from the teacher lapel microphone and the one picking up the classroom environment).

It was difficult to get the gist of the classroom environment from the recordings, making it problematic to complete some of the key elements on the lesson observation schedule (e.g., teacher engaging with individual pupils, pupils engaging with each other). One of the main pillars of the TDTScience approach is making lessons more hands-on and practical, and it was difficult to judge how well this was being achieved or, indeed, what was being undertaken simply from listening.

The approach was trialled because it offered two main advantages: less of a “hothouse” effect if selecting one or two recordings from a sequence compared with an external evaluator making an in-person visit; and considerable resourcing/cost saving. For this intervention it was not considered a suitable substitute to in-school visits. However, it would be worth considering for interventions that depend less on activities that need to be seen as well as heard. For example, for practical programmes similar to TDTScience, it may be useful to consider video recordings to capture visual cues which will contribute to a more representative and holistic evaluation.