Thinking, Doing, Talking Science

Science Oxford and Oxford Brookes University

American Institutes for Research

Education Endowment Foundation

Evaluation Summary	
Age range	9 to 10
Number of pupils	c. 5,760
Number of schools	c. 180 schools
Design	Randomised control trial
Primary Outcome	Pupil achievement on science assessment

Intervention

Thinking, Doing, Talking Science (TDTS) is an intervention delivered by a team from Science Oxford (the public brand for The Oxford Trust) and Oxford Brookes University, hereafter referred to the Oxford team, aimed at providing teachers with skills and strategies to develop challenging enquiry-based lessons that incorporate more practical activities; deeper thinking and discussion; and less, but better, recording. The theory is that the TDTS professional development (PD) programme would support teachers by providing strategies and background theory to enable them to develop more engaging and challenging lessons. This professional development would then improve teacher self-efficacy and teaching practices, which in turn is thought to improve pupil engagement, content knowledge, and enquiry skills.

For this evaluation, the programme developers will use a "train the trainer" model to deliver the intervention at scale. That is, a group of trainers will be trained in the intervention, and these trainers will in turn provide Year 5 teachers with four one-day continuing professional development (CPD) trainings throughout the year. The trainings will occur outside of the teachers' schools. The teachers will then incorporate the TDTS skills and strategies into their lesson plans and practices, thereby implementing the intervention within their classroom.

Significance

A recent report on science education found that the best science teaching in primary schools included scientific enquiry as a central focus of teaching and had sustaining pupil interest in the subject as a goal.[1] The report's findings emphasised the importance of continuing professional development (CPD) in primary science, but found that such training was not often provided.

These recent findings echo earlier research. Primary science teaching is thought to be a key to turning around a decline in pupil interest in science. In addition, the research literature recommends CPD for teachers to build skills in the teaching of science—particularly in teaching pupils how to conduct investigations.[2]

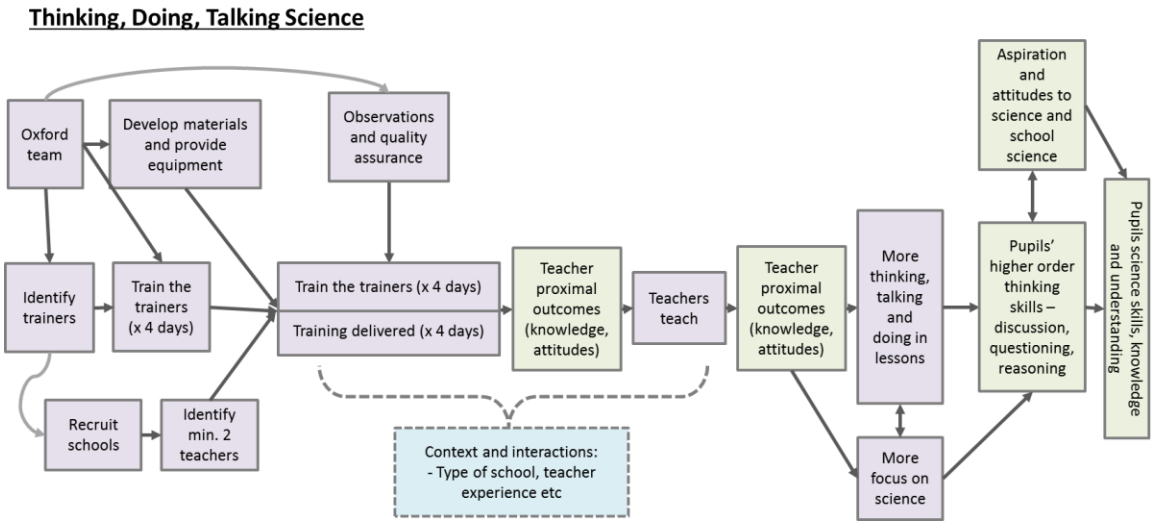
[1] Ofsted. (2013). *Maintaining curiosity: A survey into science education in schools.* Manchester: Ofsted. Retrieved from: https://www.gov.uk/government/publications/maintaining-curiosity-a-survey-into-science-education-in-schools

[2] Murphy, C., & Beggs, J. (2005). *Primary science in the UK: A scoping study. Final report to the Wellcome Trust.* London: Wellcome Trust.

A 2012 synthesis of research on primary school science interventions[3] concluded that interventions focused on supporting teacher instruction showed potential for improving science learning. While enquiry-based interventions that involved the use of specific curriculum, sets of tools, or "kits" for instruction were not found to have significant effects, interventions that supported teachers through CPD that focused on enquiry-based teaching did have significant effects.[4] One of the interventions featured in a widely cited synthesis on science teaching, "Conceptual Challenge in Primary Science", was a direct precursor to TDTS.[5] This CPD intervention, which focused on supporting teachers' ability to provide more dynamic science content, was part of an action research project carried out by Oxford Brookes University's School of Education.

TDTS seeks to improve pupil achievement in and attitudes about science by providing teachers with CPD in primary science that supports their knowledge of and ability to conduct enquiry-based learning, and ability to plan engaging lessons that provide opportunities for deeper thinking. This evaluation will measure the impact of TDTS on pupil outcomes and explore differences in teacher behaviour. Figure 1 depicts the Education Endowment Foundation's (EEF) logic model for this intervention.

Figure 1. EEF's Logic Model for Thinking, Doing, Talking Science Intervention



Methods

Research questions

The main research questions for the evaluation are:

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

Design

This study is a blocked cluster randomised control trial with randomisation at the school level within blocks. The trainers were recruited by the Oxford team, Bridget Holligan from Science Oxford and Helen Wilson from Oxford Brookes University. The trainers are from Centre for Industry Education Collaboration (CIEC) in Teesside and Lincolnshire, Science Made Simple (SMS) in Lancashire, Institute of Education at University College London (UCL) in London, Bath Spa University in Somerset, the Mathematics and Science Learning Centre at the University of Southampton in Hampshire and Early

[3] Slavin, R. E., Lake, C., Hanley, P., & Thurston, A. (2012). *Effective programs for elementary science: A best-evidence synthesis.* Baltimore, MD: Center for Research and Reform in Education.
[4] Interventions that helped teachers incorporate technology were also found to have significant benefits.
[5] Mant, J., Wilson, H., & Coates, D. (2007). The effect of increasing conceptual challenge in primary science lessons on pupils' achievement and engagement. *International Journal of Science Education, 29*(14), 1707–1719.

Years Science in Dorset. The trainers in turn recruit schools to the study within their areas. This is a two-arm design with a "waitlist" control group: in schools assigned to the intervention, Year 5 teachers will participate in the TDTS CPD programme; control schools will be on a "waitlist" and continue with business as usual in the evaluation year and then be offered the TDTS programme for Year 5 teachers during the 2017–18 school year, after data collection is complete.

Participants

The Oxford Team began recruiting the TDTS trainers from different areas of the country in October 2015. The trainers are primary science specialists and are knowledgeable of schools and systems in their respective areas. The trainers will be responsible for the recruitment of schools in their area. Recruitment will take place in April through June 2016. Trainers will be instructed to recruit a collection of schools that are representative of their areas, with targeting of schools that are higher than average in the proportion of pupils receiving free school meals (FSM).

For both intervention and control schools, all Year 5 teachers are required to participate in the study. Trainers have a maximum load that they can handle (75 teachers across 30 schools); in areas that have large schools the trainers will need to monitor recruitment and recruit small schools to balance large ones so that the maximum number of teachers is not exceeded. If a one form entry school is recruited into the study then a minimum of two teachers will be included, which could include a Science Coordinator or head of Key Stage. All pupils of the selected teachers will be eligible to participate in the evaluation. Schools will distribute an opt-out form to the parents of Year 5 pupils. Pupil assessment data will only be collected for pupils whose family does not complete an opt-out form.

In order to participate in the programme and evaluation, schools must agree to:

- Be randomly assigned to the have the TDTS continuing professional development programme in either a) the 2016–17 school year or b) the 2017–18 school year
- Complete a brief school survey describing professional development opportunities available to Year 5 teachers before randomisation
- Allow all Year 5 teachers (a minimum of two teachers) to participate in the study, to participate in the programme when it is offered, and to utilise skills from that programme in their classrooms
- Provide contact information (i.e., teacher email addresses) for all participating Year 5 teachers to allow the evaluation team to send a survey link directly to the teachers
- Allow teachers in the study to participate in a brief teacher survey about their practices
- Send an opt-out consent form to the parents of all eligible pupils in the study (i.e., all of the pupils in the classes of Year 5 teachers chosen for the study) and record any opt-outs received so that the data for these pupils are not passed on to the evaluation team
- Provide pupil identification information (name, UPN, date of birth) for all pupils in the study (i.e., pupils in the Year 5 classes of teachers chosen for the study) to the evaluation team
- Allow the administration of a science assessment and pupil survey (jointly administered) by the evaluation team to all pupils in the study at the end of the 2016–17 school year

Assignment to intervention and control groups

Once recruitment of schools into the programme and evaluation is complete, schools will be assigned to control or intervention using minimisation methods. The assignment will be conducted separately for each block or area. The minimisation will include the following school characteristics:

- Percentage of pupils eligible for FSM
- Number of Year 5 teachers

Outcome measures

The primary outcome will be pupil science knowledge attainment, which will be measured by a score on a science assessment administered as part of the evaluation. The science assessment was

developed by Terry Russell and Linda McGuigan[6] and includes items that address the science curriculum content appropriate for the year group, and which represents a range of topics and item types (e.g., open-ended, closed item).[7]

The secondary outcome will be pupil attitudes towards science, which will be measured by a pupil-level survey administered at the same time as the assessment. This study uses the pupil survey developed by the team from the Institute for Effective Education (IEE) at the University of York, who conducted the efficacy trial of TDTS.[8] That survey was adapted from a questionnaire developed by Kind, Jones and Barmby.[9]

As noted, these pupil outcomes will be measured using the same instruments that were used in the efficacy trial; therefore, the programme developer will be aware of the outcomes, but the trainers will not. To ensure tests are blind, the evaluator will confirm with the programme developer that there will not be any changes to the delivery of the intervention as a result of knowing the test. The evaluator will ensure that no one else involved in the delivery of the intervention has access to the test. In addition, the evaluator will observe an initial training to ensure that the sessions are not being adapted to the test.

The tests will be administered by a third party, NatCen, and not teachers, to further promote blind administration. In addition, the NatCen staff that administer and score the assessment will not have access to information on whether the school is part of the intervention or control group, to ensure no subtle bias in administration or scoring of the assessment.

Sample size calculations

The sample size for this effectiveness evaluation was determined in part by the capacity of the developers and in part to achieve a minimum detectible effect size of at most .18 (ideally around .15). Power calculations were made using the PowerUp! tool (version: 22 January 2015)[10] specifying 80 percent power and assuming a blocked cluster randomised assignment design with randomisation at the second level (school) and fixed effects at the blocking level (areas). Based on results from the efficacy trial, the power calculations assumed that the current evaluation will have 32 Year 5 pupils per school, 6 FSM pupils per school, and an ICC of .12. The level 1 (pupil) and level 2 (school) R-squared statistics were conservatively set to .3 and .4 respectively.

The analysis model will use pupil KS1 scores as covariates in the model as a substitute for a pre-test in science. Results from the previous evaluation indicated that the correlation between KS1 and KS2 math scores was about .77, meaning that about 60 percent of the variation in KS2 math test could be explained by KS1 test scores. Additionally, the "Variance Almanac of Academic Achievement" (https://arcdata.uchicago.edu/) reports that when using U.S. grade 5 data with a pre-test and demographic controls (5 school-level and 5 pupil-level variables), 58 percent of school-level and 52 percent of pupil-level variation is explained. Hence, the assumptions used in the power analysis were realistic but conservative.

Using these assumptions, a sample including 6 trainers and 30 schools per trainer (180 schools) yielded an MDES of .13 for all pupils and .18 for FSM pupils.

[6] Russell, T., & McGuigan, L. (2001). Science Assessment Series 1. Teachers' Guides and Assessment Units KS1 and KS2. NFER-Nelson.
Russell, T., & McGuigan, L. (2001). Science Assessment Series 2. Teachers' Guides and Assessment Units. NFER-Nelson.

[7] AIR has obtained permission to use this pupil science assessment from the developers.

[8] AIR has obtained permission to use this pupil science survey from IEE.

[9] Kind, P., Jones, K., & Barmby, P. (2007). Developing attitudes towards science measures. *International Journal of Science Education 29*(7), 871–893.

[10] Dong, N., & Maynard, R. A. (2013). *PowerUp!*: A tool for calculating minimum detectable effect sizes and sample size requirements for experimental and quasi-experimental designs. *Journal of Research on Educational Effectiveness*, *6*(1), 24–67. doi: 10.1080/19345747.2012.673143

Analysis plan

The primary and secondary pupil outcome measures will both be analysed using a mixed multilevel regression model to reflect the nested nature of the data and the method of assignment, with pupils nested within schools nested within areas. The model will include pupil and school covariates, including all variables used in the assignment process as well as individual pupil's prior KS1 scores. Intervention and control groups will be compared by including an intervention indicator at the school level. There are too few areas in the model to reliably estimate variation in effects across areas, but differences across areas will be accounted for by including area fixed effects.

For science attainment, the primary outcome, the measure analysed will be the raw score given to each pupil's assessment per the scoring rubric. For science attitudes, the measure analysed will be the average of pupil responses over a number of related Likert-type items designed to measure the construct.

Results will be reported as Hedge's g effect sizes that standardise the estimated impact using the unadjusted pooled within-group standard deviations of the outcome.[11]

Subgroup analysis will be conducted for the population of FSM pupils using the "everFSM" variable. To conduct this analysis, the model will be re-estimated using only FSM pupils.

Implementation and process evaluation methods

There are two research questions designed to address implementation and process methods. These research questions are:

- How is Thinking, Doing, Talking Science implemented at scale?
- Were there any differences in teacher practice between those whose school was assigned to the control group and those whose school was assigned to the intervention group?

To determine how well TDTS is implemented at scale and to assess implementation fidelity to the programme model, we will collect information from schools, administrative data from the developers, and data collected from evaluation forms completed by CPD trainers and teachers.

First, prior to intervention/control assignment and at the time of recruitment, schools will be asked a few questions about plans for CPD, and specifically CPD in science, for Year 5 teachers. This information will help describe what "business as usual" is in study schools.

Second, during the project's inception phase we will work with the TDTS developers to set up processes for recording, retaining, and delivering to the study team key administrative data on the implementation of the training of CPD providers and the training of teachers. These data will include process information on how potential trainers were identified, recruited, and selected for participation; the dosage and intensity of the training of the CPD providers; lists of any materials used for training; and attendance records of CPD training of teachers from the intervention schools.

Third, an evaluation form about the effectiveness of the training will be created and administered to both the CPD providers and the teachers receiving the training. The evaluation forms will obtain information that is relevant for each party. For example, to gauge the effectiveness of the training of trainers, the form will obtain information on the CPD providers' understanding of the TDTS model and confidence in their ability to deliver the TDTS professional development programme. This evaluation form will also include background questions on previous experience in delivering professional development. Teacher evaluation forms will gauge the effectiveness of the training, the usefulness of materials provided, and the teacher's confidence in his/her ability to implement new lesson plans in the classroom.

[11] Per both EEF and IES guidance, we will not use the level 1 within-group variance. (For EEF guidance see https://educationendowmentfoundation.org.uk/public/files/Evaluation/Analysis_for_EEF_evaluations_REVISED_Dec_2015.pdf; for IES guidance see http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf)

In addition, members of the study team will observe trainings of the CPD providers and subsequent teacher training.

We also will gather information on teacher practice. Though teachers may come out of PD sessions feeling they have learned useful skills, it will be important to know whether this translates into the actual use of the teaching practices and strategies on which teachers were trained. We therefore propose to use a teacher survey as the main data source for these teacher-level outcomes. This teacher survey will be administered to both intervention and control teachers to obtain information on teacher behaviours and attitudes towards science and to establish business as usual in the control schools. Teacher survey data will be analysed using a regression model with teacher and school characteristics as covariates as well as area fixed effects.

Costs

The TDTS programme theoretically impacts pupil achievement through teacher training. If proven effective, the programme could potentially be applied to and have impact on teacher education programmes; however, for assessing the costs of the programme, the approach will be to treat it in its current form as a service to schools. The costing approach will seek to assess the direct marginal costs associated with an additional school's participation in the programme.

Costs that will be collected and reported in monetary terms include:

- Salaries to trainers for time spent being trained on the TDTS programme (i.e., time in train-the-trainer sessions). These are fixed costs that, per EEF guidance, we will assume could theoretically be spread over three years.
- Salaries to trainers for time providing the TDTS CPD to a group of teachers. This would include time in any periodic meetings with the Oxford team to provide maintenance in their delivery of the programme.
- Materials needed to provide the TDTS CPD (e.g., printed materials).

In this evaluation, we will need to take care to separate (1) the salaries paid to the trainers that are associated with learning the programme and providing the CPD to teacher from (2) the salaries paid to the trainers for time assisting with the evaluation (e.g., for recruitment of schools). Cost that will be collected and reported in units of time include teacher supply cover.

This information will be collected directly from the Oxford team. Where distinctions need to be made in costs (e.g., distinguishing trainer salaries associated with the CPD from those associated with the evaluation), the evaluation team will work with the Oxford team to make necessary judgements.

Ethics and registration

AIR has submitted an application to AIR's Institutional Review Board (IRB) and has received ethics approval to continue with the study. AIR's IRB (IRB00000436) has conducted expedited and full-board reviews of research involving human subjects for more than 24 years. AIR is registered with the Office of Human Research Protection (OHRP), an office within the U.S. Department of Health and Human Services, as a research institution (IORG0000260) and conducts research under its own Federalwide Assurance (FWA00003952).

For this evaluation, AIR will use opt-out consent, in which families will receive information about the study and instructions on how to opt out of the study. Sensitive personal data will not be collected. Therefore, per EEF's interpretation, the opt-out procedures correspond with Section 2 Conditions 1 and 6 of the Data Protection Act.

The study has been registered, and the International Standard Randomised Controlled Trial Number (ISRCTN) is: ISRCTN22499525.

Personnel

Sami Kitmitto, Ph.D., a Principal Researcher at AIR, will serve as the principal investigator. He will oversee the study, provide theoretical and statistical guidance, and participate in report writing.

Raquel L. González, Ph.D., a Senior Researcher at AIR, will serve as the project director. She will oversee the day-to-day activities of the study, serve as the primary contact, provide guidance and oversight on instrument development, conduct the data analysis, ensure high-quality work, manage the budget, and participate in report writing.

Jarah Blum, a Researcher at AIR, will develop materials for recruitment, respond to evaluation questions from sites, develop instruments, and participate in analysis and report writing.

John Mezzanotte, a Research Associate at AIR, will create the web versions of the data collection instruments, administer the survey, assemble the data, and participate in data analysis and report writing.

Rakhee Patel, Research Director at NatCen, will lead the data collection of the pupil science assessment and survey, including distributing opt-out consent forms, training the data collectors, and scoring the science assessment. In addition, Ms. Patel will oversee the pull of NPD data of participating pupils and the data merge and de-identification of the pupil-level data (i.e., NPD data, pupil science assessment, pupil survey) which will be shared with AIR. Operational staff at NatCen who will be involved in the project include Alessio Fiaccio (programmer), Lesley Mullender (Head of Logistics, responsible for mailings to schools), Bryan Mason (Head of Data, who will oversee the scoring of the science assessment) Michael Lumpkin (Research Administrator) and Claire Jones (Field Project Manager, who will provide oversight of the regional field managers who manage the data collection staff).

Tom Chadwick, Researcher at NatCen, will be responsible for helping to design and test the information form to collect pupil information and will be responsible for briefing the data collectors.

Anticipated roles of programme developer
For this evaluation study, the seven partners selected and supported by the Oxford team will lead the recruitment of schools into the study, provide information on which schools were recruited, work with schools in informing them of intervention/control status, provide programme administrative data (as referenced earlier), and support the process evaluation as requested by the evaluation team. Helen Wilson, from Oxford Brookes University, and Bridget Holligan and Catherine Aldridge, from Science Oxford, will provide the training and oversight of the programme. The trainers are: David Price and Ruth Perkins, from SMS in Lancashire; Stuart Twiss and Caroline Galpin, from Early Years Science in Dorset; Sarah Earle and Allie Beaumont, from Bath Spa University in Somerset; Joy Parvin, Jane Winter, Jenny Harvey, and Nicky Waller, from the Centre for Industry Education Collaboration (CIEC) in Teesside and Lincolnshire; Jillian Trevethan and Esme Glauert, from Institute of Education at UCL in London; and Caroline Whittaker and Julie Reynolds, from the Mathematics and Science Learning Centre at the University of Southampton in Hampshire.

Risks

Risks to the evaluation
The use of assessment items that were previously used in the efficacy trial could potentially compromise a blind assessment of the test and is one risk to the evaluation. As noted above, AIR will ensure the developer does not change the intervention to teach to the test and that the assessment is not shared with programme staff. In addition, AIR will observe a training to ensure that the training is not overly aligned with the assessment.

Another risk to the evaluation is recruitment of sites and pupils. The Oxford team will need to recruit at least 180 schools to power the study as originally planned. It may be difficult to recruit such a

large number of schools. The developer, though, is working with staff who are familiar with school systems and have connections in the schools.

Parents will receive opt-out consent forms, and the evaluation team has assumed that only small numbers of families will return these forms. However, it is possible that parents may not want their child to participate in the study, or a large number of opt-out forms may be received within certain schools. However, we do not anticipate this to be a problem because it was not a challenge in the efficacy trial or other EEF trials.

Data protection

AIR takes the protection of data seriously and has implemented numerous companywide policies to promote the security of data. Staff are required to participate in yearly trainings on data security. All AIR computers are encrypted and password protected with stringent requirements for passwords, including specification on password length, character usage, and password expiration dates. Any login into AIR's servers and systems (including accessing email or shared network) outside of AIR offices require two-factor authentication.

Folders on the shared network have been created with limited access, so only select members of the project team have access the folder. Folders on the shared network also will be encrypted for an additional layer of encryption (beyond encryption of staff laptops). In addition, all IRBs require a data security plan, which must first be approved prior to any approval of the IRB research application. The data security plan outlines ways in which data will be protected.

For this study, data will be protected in compliance with AIR's strict rules and regulations, and in accordance with the Data Protection Act and the Safe Harbour Agreement.

All identifiable pupil data will be merged by NatCen. NatCen will merge the data from multiple sources (i.e., pupil assessment and survey data and NPD data) and de-identify the data before sharing the data with AIR. The identifiable data will not be kept by NatCen any longer than absolutely necessary to complete the merging and transfer process.

The Safe Harbour Agreement only applies to personal data,[12] where personal data are identifiable. Since only de-identified data will be shared outside of the UK, AIR will be in compliance with the Safe Harbour Agreement.

Timeline

Date	Activity
Spring 2016	AIR meet with the Oxford team, observe CPD training, and administer evaluation forms for CPD trainings
June 2016	AIR work with NatCen to distribute opt-out consent forms
July 2016	AIR randomise schools
Autumn 2016 through spring 2017	AIR work with the Oxford Team to administer teacher evaluation forms after each teacher training
February 2017	NatCen request NPD data
April 2017	AIR administer pupil assessment and survey
April 2017	AIR administer teacher survey
July 2017	AIR obtain all data files and conduct analyses
August 2017	AIR write up report and submit to EEF

[12] Article 2 of Safe Harbour defines personal data as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity."
http://curia.europa.eu/juris/document/document.jsf?text=&docid=169195&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=125031

Date	Activity
December 2017	AIR make edits per peer reviewer, submit final report to EEF