Statistical Analysis Plan for TDTS

INTERVENTION	Thinking, Doing, Talking Science (TDTS)
DEVELOPER	Science Oxford (the public brand for The Oxford Trust) and Oxford Brookes University
EVALUATOR	American Institutes for Research and NatCen Social Research
TRIAL REGISTRATION NUMBER	ISRCTN22499525
TRIAL STATISTICIAN	Sami Kitmitto, Ph.D.
TRIAL CHIEF INVESTIGATOR	Sami Kitmitto, Ph.D.
SAP AUTHOR	Sami Kitmitto, Ph.D. and Raquel González, Ph.D.
SAP VERSION	1
SAP VERSION DATE	17/08/17
EEF DATE OF APPROVAL	21/08/17
DEVELOPER DATE OF APPROVAL	03/08/17

Table of contents

Introduction

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 2016-17 school 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.

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?*

Our main research question focuses on two outcomes. The primary outcome is science attainment of pupils at the end of Year 5 using an external science assessment. The secondary pupil outcome is pupil science attitudes at the end of Year 5 using a pupil survey.

Study 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 primary science specialists and are knowledgeable of schools and systems in their respective regions. 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 Years Science in Dorset. During the spring of 2016, the trainers in turn recruited schools to the study within their regions (as listed above). Trainers were instructed to recruit schools in their general geographic regions with an emphasis on recruiting schools that were 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. The TDTS model requires that a minimum of two teachers participate per school. Hence, for one form entry schools recruited into the study the two teachers would be the one Year 5 teacher plus one additional teacher who could be a Science Coordinator or head of Key Stage.

All pupils of the selected teachers were eligible to participate in the evaluation. Recruitment of pupils occurred at the end of pupils' Year 4 (May-June 2016) when schools distributed an opt-out form to the parents of Year 4 pupils. Pupil data will only be collected for pupils whose family did not complete an opt-out form. Data collection occurs once at the end of Year 5 for

the 2016-17 school year with a pupil science assessment and pupil survey about science attitudes. In addition, data from the National Pupil Database will be used to supplement additional information about the pupils.

In order to participate in the programme and evaluation, schools agreed 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, and
- Allow the administration of a science assessment and pupil survey (jointly administered) by the evaluation team to all pupils in the study at the end of the 2016–17 school year.

This is a two-arm design with a "waitlist" control group: in schools assigned to the intervention, Year 5 teachers participate in the TDTS CPD programme during the 2016-17 school year; 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.

Protocol changes

There have not been any changes to the protocol.

Randomisation

Assignment to treatment and control groups occurred at the school level and was conducted separately within each of the seven study regions using minimisation methods.[1] The minimisation included two school characteristics:

1) Percentage of pupils eligible for FSM in three categories:
    a. Low (9 percent FSM or less)
    b. Medium (between 9 and 24 percent FSM)
    c. High (greater than 24 percent FSM)
2) Number of Year 5 teachers in two categories:
    a. One or two
    b. Three or more

In total, 205 schools were recruited into the study and 106 were assigned into the treatment group and 99 into the control group. The minimisation process can result with unequal

[1] The MimimPY program was used for school assignment. We used the "biased coin" method with a base probability of .75.

numbers assigned to the treatment and control groups because the process assigns each school probabilistically one-at-a-time.

The baseline measure of pupil achievement for this study is the academic attainment at the end of Key Stage 1 for pupils in participating schools as measured by the end of Key Stage 1 assessment. Assignment was conducted by the study team in July 2016. We expect to obtain the baseline academic achievement data from the National Pupil Database in the summer of 2017.

Calculation of sample size

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

- Schools would be recruited within region
- Schools would be the unit of randomisation
- Pupils would be the level of observation

Power calculations were conducted using the PowerUp tool (Version: 22/01/2015).[2] The following are parameters used in power calculations:

- Alpha Level (α) = .05, two-tailed test, with Power (1-β) = .80
- Interclass Correlation (ICC) = .12[3]
- Proportion of Level 2 units randomized to treatment = .50
- Number of school level covariates = 2
- Number of pupils per school: 32 pupils, of which 6 would be FSM pupils
- Proportion of pupil variance explained by covariates (R^2_1) = .30
- Proportion of school variance explained by covariates (R^2_2) = .40

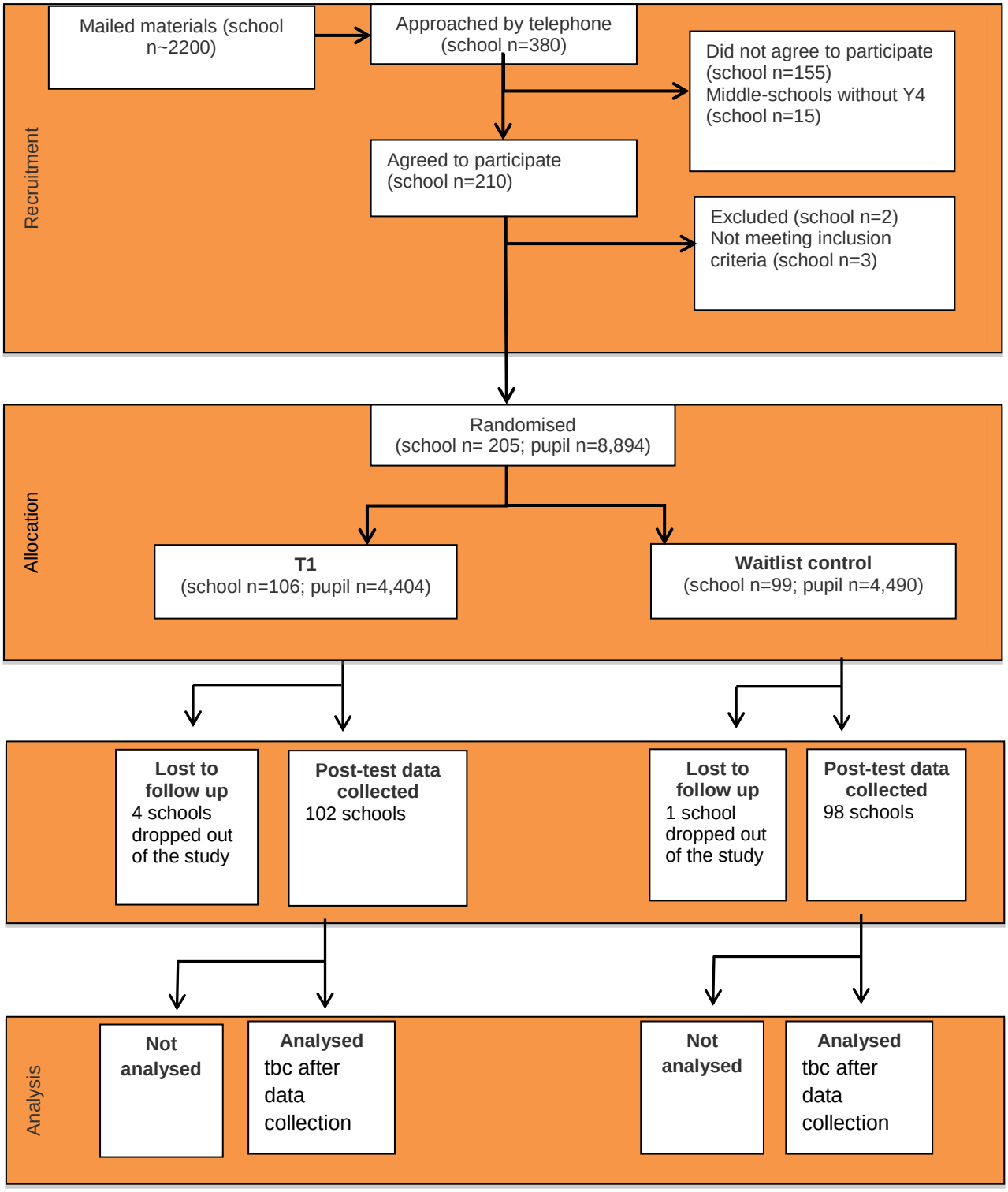
Our goal was to choose a sample size to obtain an MDES of .18 or less. Assuming that 30 schools were recruited within each region and based on the conservative parameters listed above, a total of 6 regions would generate an MDES of .176 for FSM pupils and .127 for all pupils.

[2] Dong, N. and 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

[3] In order to determine the appropriate ICC for the power analysis, we reached out to the study team that led the efficacy trial to ask for their recommendation. The authors of the efficacy trial calculated an ICC of .12 and therefore, this ICC was used in our power analysis. (Hanley, Pam. "Re: Correlations from thinking, talking doing science trial." Message to Camilla Nevill and Sami Kitmitto. 7 October, 2015. Email).

Follow-up

Participant flow diagram[4]



[4] Recruitment was led by the Oxford Team. Where we do not have detail regarding recruitment we have left the field blank.

Outcome measures

Primary outcome

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. This study uses the science assessment and scoring guide developed by the team from the Institute for Effective Education (IEE) at the University of York, who conducted the efficacy trial of TDTS.[5] The items in the assessment were drawn from a larger bank of items developed by Terry Russell and Linda McGuigan.[6] The assessment includes items that address the science curriculum content appropriate for the year group and represents a range of topics and item types (e.g., open-ended, closed item).[7] In analysis, we will use raw scores which have a possible range of 0 to 41.

Secondary outcomes

The secondary outcome will be pupil self-efficacy and attitudes towards science, which will be measured at one time point by a pupil-level survey administered at the same time as the science assessment. As with the science assessment, this study uses the pupil survey developed for the efficacy trial of TDTS.[8] That survey was adapted from a questionnaire developed by Kind, Jones and Barmby.[9] The items on the survey are on a five point scale from "Agree a lot" to "Disagree a lot."

We recognise the importance of summarising the items into meaningful constructs. Preliminarily, we interpret these items to fall into the domains of self-efficacy towards science and attitudes towards science. We will use exploratory and then confirmatory factor analysis to determine an underlying factor structure that supports the creation of reliable indices of pupil attitudes and self-efficacy towards science.

Analysis

Primary intention-to-treat (ITT) analysis

The primary pupil outcome measure, pupil science knowledge attainment, will 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. 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. School-level random effects will be included in the model by allowing the intercept to vary randomly across schools. There are too few regions in the study to reliably estimate variation in effects across them, but differences will be accounted for by including fixed-effect region indicators.

---

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

[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 and efficacy trial authors.

[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.

Level 1 – Pupil level

Primary pupil outcome measure

Description	Source
Primary: Pupil score on the study science assessment	Study's science assessment

Pupil-level control measures

Description	Source
KS1_MATHS	NPD
KS1_READWRIT	NPD

Level 2 – School level (level of randomisation)

School-level control measures

Description	Source
School assigned to the treatment group	Study assignment
LEA[yy]_Pct_Pupils_FSM_Eligible	NPD
Number of Year 5 teachers in the school	Study recruitment data
Region indicator	Study recruitment data

No interactions with the treatment indicator are planned for this analysis, except for the sub-group analysis for FSM pupils, which is described below.

Interim analyses

No interim analyses are planned for this trial.

Imbalance at baseline

At the time of school assignment, only school-level information on the percent eligible for FSM and the number of year 5 teachers was available. We will check balance in baseline student characteristics, specifically student KS1 scores which are used as the pre-test control in this trial. Differences in test scores will be reported as effect sizes. While there may be imbalance due to the random nature of assignment, we do not expect imbalance at baseline due to the low attrition of schools (less than three percent of recruited schools were lost to attrition).

Missing data

When data are missing, the power of the analysis to detect statistically significant effects is reduced and, depending on the mechanism by which data are missing, the estimated effects and standard errors can potentially be biased.

If 5 percent or less of pupils have incomplete outcome information, we will conduct analysis omitting these pupils. In other words, we will conduct analysis using listwise deletion of any pupil with incomplete information. Previous research has found that when 5 percent of the data are missing, bias is low across the various approaches to handling missing data in analysis including listwise deletion.[10]

[10] Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price (2009). *What to Do When Data Are Missing in Group Randomized Controlled Trials* (NCEE 2009-0049). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

If greater than 5 percent of the pupils have incomplete outcome information, we will investigate the mechanism by which data are missing and consider multilevel multiple imputation as an alternative to listwise deletion.

A priori, potential reasons why pupils might be missing outcomes in this study include:

- School dropped out after recruitment and assignment to treatment and control groups
- Pupil left the study school between the end of year 4 (i.e., at the time of recruitment when opt-out forms were sent and school assignment to treatment or control groups was conducted) and testing at the end of year 5
- Pupil was absent on the day of testing and not reached in mop-up efforts
- Pupil was not tested due to having special educational needs or disability

To investigate these mechanisms, we will first look at the extent of the missing data for each variable to be included in the analysis and the patterns of missingness across treatment and control groups. Second, we will estimate a model predicting missingness to examine whether the covariates in our primary analysis model jointly (using an F-test) predict the absence of pupil outcome data.

If any differences are found, significant at $p<.05$, multilevel multiple imputation will be used for analysis and reporting. Results will be reported alongside the completers analysis.

On-treatment analysis

We will conduct a Complier Average Causal Effect (CACE) analysis of TDTS. The Oxford Team has defined an appropriate minimum level of exposure for "compliers" to use in the CACE analysis. Compliers will be schools sending at least one teacher to at least three of the four training sessions. The rationale for this exposure level is that if at least one teacher attended a session, he/she would be able to share knowledge from the training session with colleagues.

Secondary outcome analyses

The indices of pupil self-efficacy and attitudes towards science will be analysed using the same model as described above in the primary ITT analysis, but without pupil level covariates.

Secondary pupil outcome measure

Description	Source
Secondary: Pupil self-efficacy and attitudes towards science	Study pupil survey

Additional analyses

No additional outcomes analyses are expected for this study. As exploratory analysis, we will estimate our primary and secondary analyses models including additional pupil level covariates (see table below).

Pupil-level control measures

Description	Source
KS1_MATHS	NPD
KS1_READWRIT	NPD
KS2_GENDER	NPD
KS2_YEAROFBIRTH and KS2_MONTHOFBIRTH (to calculate age)	NPD
KSF_FSM6	NPD
IDACIScore_[term][yy]	NPD

Subgroup analyses

Subgroup analysis will be conducted for the population of FSM pupils using the "everFSM" variable ("KS2_FSM6"). For this analysis, the primary and secondary outcome analysis models will be re-estimated using data limited to this sample. Furthermore, an interaction model will be run on the entire data. This will mirror the main primary outcome model but also include everFSM and the interaction between everFSM and group as covariates.

Effect size calculation

Primary outcome results will be reported as raw scores as well as effect sizes that standardise the estimated impacts. The numerator will be the regression adjusted estimate of the impact of TDTS from the multi-level model and the denominator will be the standard deviation of the outcome for the full sample.[11]

Report tables

Table 1: Summary of impact on primary outcome

Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	EEF cost rating
Treatment vs. control (science assessment)				
Treatment FSM vs. control FSM (science assessment)				

Table 2: Minimum detectable effect size at different stages

Stage	N [schools/pupils] (n=intervention; n=control)	Proportion of variance in outcome explained by pre-test + other covariates	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
Protocol	[180 schools/5760 pupils] (90 treatment schools, 90	Pupil level: .30 School level: .40	.12	6 regions	.80	.05	.13

[11] Per EEF guidance, see https://educationendowmentfoundation.org.uk/public/files/Evaluation/Analysis_for_EEF_evaluations_REVISED_Dec_2015.pdf.

	control schools; 2880 pupils in treatment schools, 2880 pupils in control schools)						
Randomisation	[205 schools] (106 treatment schools, 99 control schools)			7 regions	.80	.05	
Analysis							

Pupil characteristics

Table 3: Baseline comparison

Variable	Intervention group		Control group	
School-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Region 1 **Region 2** **Region 3** **Region 4** **Region 5** **Region 6** **Region 7**	12/103 (0) 7/103 (0) 13/103 (0) 13/103 (0) 11/103 (0) 12/103 (0) 17/103 (0)		11/98 (0) 7/98 (0) 12/98 (0) 15/98 (0) 10/98 (0) 9/98 (0) 17/98 (0)	
School-level (continuous)	**n (missing)**	**[Mean or median]**	**n (missing)**	**[Mean or median]**
Percent eligible for FSM				
Number of Y5 teachers in the school				
IDACI of pupils in the school				
Percent pupils classified as white British ethnic origin				
Percent of pupils who are EAL				
Pupil-level (categorical)	**n/N (missing)**	**Percentage**	**n/N (missing)**	**Percentage**
Eligible for FSM				
Gender				
Pupil-level (continuous)	**n (missing)**	**[Mean or median]**	**n (missing)**	**[Mean or median]**
Score on KS1 maths				
Score on KS1 reading				
IDACI				
Pupil age				

Outcomes and analysis

Table 4: Primary analysis

	Raw means				Effect size		
	Intervention group		Control group				
Outcome	**n (missing)**	**Mean (95% CI)**	**n (missing)**	**Mean (95% CI)**	**n in model (intervention; control)**	**Hedges g (95% CI)**	**p-value**
Pupil science assessment							
Pupil science self efficacy and attitude survey							