

# Supporting Teachers and ChildRen in Schools (STARS): Incredible Years® Teacher Classroom Management

## Statistical Analysis Plan

Evaluator: National Foundation for Educational Research

Principal investigator: Ben Styles



<b>PROJECT TITLE</b>	Supporting Teachers And childRen in Schools (STARS): Incredible Years® Teacher Classroom Management Evaluation
<b>DEVELOPER (INSTITUTION)</b>	University of Exeter
<b>EVALUATOR (INSTITUTION)</b>	National Foundation for Educational Research
<b>PRINCIPAL INVESTIGATOR(S)</b>	Dr. Ben Styles
<b>SAP AUTHOR(S)</b>	Connie Rennie
<b>TRIAL DESIGN</b>	Two arm randomised controlled trial. Random allocation at school level
<b>TRIAL TYPE</b>	Efficacy
<b>PUPIL AGE RANGE AND KEY STAGE</b>	Ages 5-6, Key Stage 1
<b>NUMBER OF SCHOOLS</b>	140
<b>NUMBER OF PUPILS</b>	5800 (based on a mean cohort size of 42 pupils per school)
<b>PRIMARY OUTCOME MEASURE AND SOURCE</b>	Maths attainment: Maths Key Stage 1 raw scores provided by the school
<b>SECONDARY OUTCOME MEASURE AND SOURCE</b>	Pupil emotional and social well-being: Total Difficulties Score (SDQ) Pupil concentration: Hyperactivity Scale (SDQ) Pupil prosocial behaviour: Prosocial Scale (SDQ) Pupil classroom behaviour: Pupil Behaviour Questionnaire (PBQ) Student-teacher relationship: Student Teacher Relationship Scale (revised version) (STRS)

### SAP version history

VERSION	DATE	REASON FOR REVISION
1.0	06 February 2020	N/A

## Table of Contents

SAP version history .....	1
Introduction.....	3
Design overview.....	4
Sample size calculations overview .....	4
Analysis.....	5
Baseline Measurements .....	5
Outcome Measurements.....	5
Primary outcome analysis.....	7
Secondary outcome analysis.....	8
Subgroup analyses .....	9
Imbalance at baseline .....	10
Missing data.....	10
Compliance .....	11
Intra-cluster correlations (ICCs).....	12
Effect size calculation .....	12
References .....	12

## Introduction

The Incredible Years Teacher Classroom Management (IY-TCM) programme is a teacher training programme designed to support teachers in improving classroom behaviour through the use of specific classroom management techniques and behaviours. The programme focuses on improving teacher-student relationships by reinforcing positive behaviour, and creating a positive learning environment in the classroom.

Research suggests that between three and six percent of children in each school class will have significant social, emotional; or behavioural difficulties that impair their learning, as well as that of their peers, and further negatively affect their future life prospects; children from disadvantaged backgrounds are even more likely to be affected by these issues (Ford *et al.*, 2018). The IY-TCM programme aims to support teachers to deal with these issues by introducing techniques such as behaviour plans, incentive systems and relationship-building tools.

The purpose of this evaluation is to investigate the outcomes of pupils aged five to seven who are taught by IY-TCM trained teachers. The trial aims to explore whether the IY-TCM programme gives teachers the tools to reduce emotional, social, and concentration problems, as well as anti-social problems within the classroom, and if working with these problems in the classroom ultimately leads to improved attainment.

The training for teachers is delivered over six whole days across a period of six months. For this evaluation, Year 1 and Year 2 teachers will receive six training days at monthly intervals during the academic year 2019/20. The trial pupils will be in Year 1 in the Year 2019/20, and will be taught by Year 1 teachers who will be receiving training throughout the year, and will be taught in Year 2 in the year 2020/21 by Year 2 teachers who will have received the training in the previous year. The trial cohort will be exposed to two years of teaching: in the first year by teachers whilst they are being trained, and in the second, by teachers fully trained on the programme. The outcomes of the pupils will be compared to students who have been taught business as usual (BAU) for the same period.

The primary outcome is maths attainment at Key Stage 1. The secondary outcome measures include pupil emotional and social well-being, concentration, prosocial behaviour, classroom behaviour, and the student-teacher relationship. Outcomes are further detailed below in the analysis section. The primary outcome will further be analysed separately for disadvantaged children, as measured by eligibility for free school meals, as subgroup analysis.

Eligible schools are mainstream primary schools i.e. all schools other than independent, selective, special, alternative provision and schools in special measures, within the target areas (University of Exeter delivery hubs), or within reasonable travelling distance of the areas of Reading, Cornwall, Bristol, Southampton, Dorset and Liverpool. Eligible teachers are all Year 1 and Year 2 teachers who have at least four days of classroom responsibility per week, or both partners of a job share. The intervention is delivered to whole classes of Year 1 and Year 2 in intervention schools, hence no pupil selection is necessary.

## Design overview

<b>Trial design, including number of arms</b>		Two arm cluster randomised controlled trial
<b>Unit of randomisation</b>		School
<b>Stratification variables (if applicable)</b>		Geographic area; School size proxy (number of reception teachers)
<b>Primary outcome</b>	variable	Maths attainment
	measure (instrument, scale, source)	KS1 total of the two raw scores, (0-25/0-35) provided by the teachers
<b>Secondary outcome(s)</b>	variable(s)	Pupil emotional and social well-being Pupil concentration Pupil prosocial behaviour Pupil classroom behaviour Student Teacher relationship
	measure(s) (instrument, scale, source)	Total Difficulties Score (SDQ) Hyperactivity Scale (SDQ) Prosocial Scale (SDQ) Pupil Behaviour Questionnaire (PBQ) Student Teacher Relationship Scale (revised version) (STRS)
<b>Baseline for primary outcome</b>	variable	Maths attainment
	measure (instrument, scale, source)	The Early Years Foundation Stage Profile (EYFSP) (see analysis section for further details).
<b>Baseline for secondary outcome</b>	variable	Pupil emotional and social well-being Pupil concentration Pupil prosocial behaviour Pupil classroom behaviour
	measure (instrument, scale, source)	Total Difficulties Score (SDQ) Hyperactivity Scale (SDQ) Prosocial Scale (SDQ) Pupil Behaviour Questionnaire (PBQ)

## Sample size calculations overview

	Protocol		Randomisation	
	OVERALL	FSM	OVERALL	FSM
<b>Minimum Detectable Effect Size (MDES)</b>	0.17	0.20	0.17	0.20
<b>Pre-test/ post-test correlations</b>	-	-	-	-
	level 1 (student)	0.51	0.51	0.51
	level 2 (school)	-	-	-

		Protocol		Randomisation	
		OVERALL	FSM	OVERALL	FSM
Intracluster correlations (ICCs)	level 1 (student)	-	-	-	-
	level 2 (school)	0.15	0.15	0.15	0.15
Alpha		0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8
One-sided or two-sided?		Two	Two	Two	Two
Average cluster size		42	10.5	40.6	10.2
Number of schools	intervention	70	70	70	70
	control	70	70	69	69
	<b>total</b>	140	140	139	139
Number of pupils	intervention	2940	735	2718	680
	control	2940	735	2924	731
	<b>total</b>	5880	1470	5642	1411

The ICC and pre and post correlation assumptions were drawn from a previous NFER trial using KS1 maths scores. The recruitment target of 140 schools was only just missed (139), and the assumption of an average cluster size was close to the actual average cluster size at randomisation (40.6). As such the sample size at randomisation it still powered to detect an MDES of 0.17 in the whole sample, and 0.2 in the FSM eligible only sample.

## Analysis

### Baseline Measurements

#### 1. The Early Years Foundation Stage Profile (EYFSP)

The EYFSP is a measure of a child's attainment in relation to 17 early learning goals<sup>1</sup> prior to the age of five. It is intended to provide a reliable, valid and accurate assessment of a child at the end of EYFS. The scale has two Maths items. Due to concerns that two items with a scale of three points each may not be a sufficiently informative covariate, we will conduct exploratory analysis to decide how this profile will be used as a baseline covariate. Prior to conducting the main analysis, we will use three scales, and correlate each with the outcome. The scale that correlates highest with the outcome responses will be used as the baseline covariate. The three scales will be as follows:

- a. Sum of the two Maths elements (G11, G12; 2-6)
- b. Sum of the 'cognitive' elements (G01, G02, G03, G09, G10, G11, G12, G13, G14, G15; 10-30)
- c. Sum of all elements (17-51)

### Outcome Measurements

#### 1. Maths attainment

<sup>1</sup>[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/790580/EYFSP\\_Handbook\\_2019.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/790580/EYFSP_Handbook_2019.pdf)

The primary outcome is maths attainment in KS1 as measured by the KS1 arithmetic and reasoning papers. The raw scores of the two papers will be provided by the teachers, and a total of the two papers will be the outcome score. This measure was chosen as it is a valid and reliable measure of academic achievement, which aligns well with the purpose of the programme's Theory of Change of improving attainment through improved classroom management. Furthermore, this is a summative assessment that will be used regardless of the trial, and therefore reduces additional assessment burden on pupils and teachers.

## 2. Pupil emotional and social well-being

Emotional and social wellbeing will be measured using the Total Difficulties score of the Strengths and Difficulties Questionnaire (SDQ, Goodman, 2001). The SDQ was chosen as it is a reliable measure of pupils' emotional and social wellbeing (Goodman, 1997). It consists of 25 items, split into five subscales with five items each (emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems, prosocial behaviour, each scored 0-2). The Total Difficulties score (TDS) is the sum of emotional symptoms, conduct problems, hyperactivity/inattention and peer relationship problems subscale scores. It has a possible range of between 0 and 40, with a higher score representing a higher level of difficulty. The score will be coded using the syntax published by Youth in Mind<sup>2</sup>. The TDS will be analysed as a binary outcome, (in line with categorization used in the previous trial, Ford, *et al.*, 2018), as well as a continuous outcome. A score of 12 or above represents the 80<sup>th</sup> percentile of the British school-age population, and those above the score of 12 have been classified as 'strugglers' in previous analysis of the IY-TCM programme (Ford *et al.*, 2018). This cut-off point will be used for the purpose of comparing to previous research.

## 3. Pupil concentration

Pupil concentration will be measured using the hyperactivity subscale of the SDQ. This is a five item scale that reliably measures levels of restlessness and distractibility. Each item is scored between 0-2 giving a range of scores between 0-10. Higher scores reflect lower levels of concentration. This subscale was chosen as it fits closely with the theory of change and improvements were found in previous investigations of the programme (Ford *et al.*, 2018).

## 4. Pupil prosocial behaviour

Pupil prosocial behaviour will be measured using the prosocial subscale of the SDQ. This is a five item scale that reliably measures levels of helpfulness and kindness. Each item is scored between 0-2 giving a range of scores between 0-10. A higher score reflects higher levels of prosocial behaviour. This subscale was chosen as it fits closely with the theory of change and improvements were found in previous investigations of the program (Ford *et al.*, 2018).

## 5. Pupil Behaviour Questionnaire (PBQ)

The PBQ is a tool developed by Exeter University, and has been previously used as an outcome measure in an IY-TCM evaluation; it was validated in 2018 (Allwood *et al.*, 2018). Whereas the Strengths and Difficulties Questionnaire focuses upon pupil mental health and individual behaviours, this scale captures behaviours that have a negative impact on the classroom. As such, it relates to an important aspect of the programme; managing behaviour within the classroom. It is a six-item scale, each rated on a three point scale; (never

---

<sup>2</sup> <http://www.sdqinfo.org/py/sdqinfo/c0.py>

happens, occasionally happens, frequently happens). The scores range from 0-12 with a higher score reflecting worse behaviour.

## 6. Student Teacher Relationship Scale (STRS)

The STRS is an instrument designed for teachers of children aged between 3 and 12 which measures a teacher's perception of conflict, closeness and dependency with a specific child (Pianta, 2001). The short version of the instrument was suggested for this trial as it allows measurement of pupil-teacher relationship, a focus of the intervention that is not captured in other outcome measures. Following internal discussion, NFER proposed that some questions, through emphasis and phrasing, were inappropriate for a UK context. Following discussion with EEF, a revised version was produced jointly by NFER and the developer team. The revised version is currently being trialled by Exeter University, and depending on the trial results, will be used at the end point of the trial only. Details on the finalised version will be included in the report appendices.

The following table displays the time points at which each baseline and follow-up measure is being measured.

Timepoint	Baseline	Midpoint	Endpoint
EYFSP	x	-	-
KS1	-	-	x
SDQ	x	x	x
PBQ	x	x	x
STRS (UK version)	-	-	x

### *Primary outcome analysis*

The primary outcome analysis of maths attainment will be 'intention-to-treat' (ITT). The primary model will be a multilevel model with two levels (school and pupil). Pupils who have measurements at baseline and follow-up will be included in the model, regardless of whether their teacher attended the IY-TCM training.

The dependent variable for the model will be the KS1 maths total raw score at follow-up with the following covariates:

- Prior attainment as measured by the EYFSP (maths score)
- A series of dummy geographical variables (randomisation stratifiers)
- A series of dummy school size variables (proxy measure: number of reception teachers per school (randomisation stratifiers))
- Intervention allocation dummy variable

This model will determine whether the IY-TCM programme has an impact on pupils' maths attainment. The analysis for all multi-level models in this investigation will be run in R using the lme4 package.

### *Secondary outcome analysis*

The secondary outcome analysis of pupil emotional and social well-being will be 'intention-to-treat' (ITT). Whereas the primary outcome is only measured at one follow-up point, the secondary outcomes are measured at midpoint and endpoint of the trial. Therefore, the model of pupil emotional and social well-being will be a multilevel model with three levels (school, pupil and time point). Using three levels, we will be able to analyse if any potential effects have changed over time. Pupils who have measurements at baseline and follow-up point one and/or follow-up point two will be included in the model, regardless of whether their teacher attended the IY-TCM programme or not.

The dependent variable for the model will be the TDS of the SDQ at follow-up point one and follow-up point two with the following covariates:

- Baseline score of the TDS
- A series of dummy geographical variables
- A series of dummy school size variables (proxy measure: number of reception teachers per school)
- Intervention allocation dummy variable
- A dummy time variable indicating 2<sup>nd</sup> follow-up
- An interaction variable time\*intervention

This model will determine whether the IY-TCM programme has an overall impact on pupils' emotional and social wellbeing, and if any impact has enhanced or attenuated over time through the use of the interaction term.

The same models will be run on the following secondary outcome measures: pupil concentration, prosocial behaviour and the pupil behaviour questionnaire. Each model will be assessed using a three level multilevel model (school, pupil, time point) with the following covariates:

- Baseline measurements of the respective scales
- A series of dummy geographical variables
- A series of dummy school size variables (proxy measure: number of reception teachers per school)
- Intervention allocation dummy variable
- A dummy time variable indicating 2<sup>nd</sup> follow-up
- An interaction variable time\*intervention

These models will determine whether the IY-TCM programme has an overall impact on pupils' concentration, prosocial behaviour and classroom behaviour respectively, and if any impact has enhanced or attenuated over time through the use of the interaction term.

The final secondary outcome, the STRS, will be measured at the end of Year 2 only. Therefore, it will be assessed using a two level multilevel model (school and pupil). It will be an ITT model, therefore pupils who have a measurement at follow-up will be included in the model, regardless of whether their teacher attended IT-TCM programme or not.

The dependent variable for the model will be the raw total score of the teacher-pupil relationship scale at follow-up with the following covariates:



- Baseline measurement of the Pupil Behaviour Questionnaire
- A series of dummy geographical variables
- A series of dummy school size variables (proxy measure: number of reception teachers per school)
- Intervention allocation dummy variable

This model will determine whether the IY-TCM programme has an impact on the quality and effectiveness of teacher-pupil relationships.

### *Subgroup analyses*

As FSM-eligible pupils represent a particularly important subgroup, a separate analysis of FSM-eligible pupils will be carried out as per standard EEF practice. Sample size calculations indicate that a sample of 1470 FSM eligible pupils will ensure enough power to detect an MDES of 0.2. The randomised sample size of 1459 pupils is still powered to detect this MDES, so if minimal attrition is achieved, this model will be sufficiently powered. The model will mimic the model used to assess the primary outcome (maths attainment), however with FSM pupils only.

Three further subgroup analyses will be run on the whole sample, identifying subgroups using interaction terms.

A two level multilevel model (school and pupil) will be run on maths attainment with the following covariates:

- Prior attainment as measured by the EYFSP (maths score)
- A series of dummy geographical variables
- A series of dummy school size variables (proxy measure: number of reception teachers per school)
- Intervention allocation dummy variable
- A dummy variable indicating FSM eligibility
- An interaction variable FSM\*intervention

This model will determine whether the IY-TCM programme has a differential impact on FSM as compared to non-FSM pupils' maths attainment.

A two level multilevel model (school and pupil) will also be run on maths attainment with the following covariates:

- Prior attainment as measured by the EYFSP (maths score)
- A series of dummy geographical variables
- A series of dummy school size variables (proxy measure: number of reception teachers per school)
- Intervention allocation dummy variable
- A dummy variable indicating 'struggling' pupils
- An interaction variable strugglers\*intervention

This model will determine whether the IY-TCM programme has a differential impact on the maths attainment of 'struggling' pupils whose TDSof the SDQ is greater than or equal to 12

(a score which represents above the 80th percentile for the British school-age population), compared to 'non-struggling' pupils<sup>3</sup>.

Finally, a two-level multilevel model (school and pupil) will be run on maths attainment with the following covariates:

- Prior attainment as measured by the EYFSP (maths score)
- A series of dummy geographical variables
- A series of dummy school size variables (proxy measure: number of reception teachers per school)
- Intervention allocation dummy variable
- Baseline measurement of the TDS (continuous)
- An interaction variable TDS (continuous)\*intervention

This model will determine whether the IY-TCM programme has a differential impact on maths attainment, and if this relationship is different for pupils with different levels of social and emotional difficulties.

### *Imbalance at baseline*

We will create a baseline comparison table comparing proportion of baseline characteristics between the intervention and control group, both for the randomised group, and the analysed group. The table will include the following variables:

School level

- School level KS2 maths attainment
- School level FSM quintile
- Region
- School Type
- Ofsted rating
- School size

Pupil level

- Baseline EYFSP numeracy attainment mean and SD
- TDS mean and SD
- FSM eligibility

Due to randomisation, we would not expect any significant imbalance in baseline characteristics between the intervention and control pupils. Furthermore, maths attainment and the TDSscore will be included as covariates in the respective models, and therefore any potential imbalance will be taken into consideration by the model, and will not influence the effect estimate. However, we will produce a baseline comparison table to show if balance was achieved, and to inspect whether attrition may have introduced bias.

### *Missing data*

We will assess the level and pattern of missing data from the primary model. We will report the number and proportion of complete and missing cases included/not included in the

---

<sup>3</sup> <https://www.sdqinfo.com/py/sdqinfo/c0.py> (Instructions in English for scoring by hand **SDQs for 4-17 year olds**, as completed by parents, teachers or youths. These instructions also cover scoring the SDQ for those aged **18+**. Instructions in many other languages are also available, accessed through the page for that [language](#).)

primary model, including extent of missing covariates. In order to assess the missingness mechanism, we will run a multilevel logistic regression model on whether a case had follow-up data for the primary outcome, regressed on the covariates of the primary outcome model plus other school level and available pupil level variables. In conjunction with qualitative judgement on the reasons for missing, if a covariate is found to significantly predict missingness, we will run the primary model with the extra variables included. As per EEF guidance, if the substantive model estimates with and without the covariates are similar, we may conclude that the completers analysis is unbiased. If they differ, it is likely the outcome is missing not at random and sensitivity analysis will need to be carried out.

Should sensitivity analysis be required, this would be preceded by multilevel multiple imputation. Multiple imputation of chained equations is a flexible method that can handle different types of variables (Hughes *et al.*, 2014). (The number of datasets is dependent on the amount of missing data but a minimum would be five datasets, with a minimum of ten iterations. The number of iterations will increase proportional to the level of missing data (White *et al.*, 2011). These iterations are necessary as with only one dataset, the parameter estimates have more sampling variability. Multiple iterations also help in generating the estimates of the standard errors to accurately reflect the uncertainty about the missing values (Allison, 2012). Once the substantive model is run on the multiply imputed data, sensitivity analysis will be carried out using different values of delta to reflect a 'missing not at random' scenario. Resulting intervention coefficients and their standard errors will be assessed and compared to the substantive model. All results will be reported.

In the primary outcome model, the only covariate that might be missing is baseline EYFSP. Since this was collected before randomisation, by definition, any missing cannot bias our estimate of the intervention effect when making conditional inference.

### **Compliance**

The University of Exeter will collect attendance data of teachers at each training session, through the use of attendance logs. We will produce descriptive statistics on this data. Compliance will be measured as a dichotomous variable indicating whether a teacher attended four out of the six sessions or above. Therefore compliance is a teacher level variable. In the case where the students are taught by the same teacher across the two years, the compliance will apply to the one teacher only. If the students are taught by different teachers between Y1 and Y2, both teachers will have had to meet the four out of six criteria.

As per EEF guidance a two-stage least squares multi-level model will be used to calculate the Complier Average Causal Effect (CACE) estimate (Angrist and Imbens, 1995). The first stage of the model will be compliance regressed on all covariates that are used in the main primary outcome model and the group allocation variable. The second stage of the model will regress the primary outcome on the covariates used in the main model and will also include a covariate representing the teacher's estimated level of compliance from the first stage of the model. No interaction term between intervention and compliance will be included as the design of the investigation means contamination is impossible, i.e. pupils will either be in an intervention school, or a control school, and no control pupils can 'comply' and all intervention pupils will have a compliance value. The coefficient of the estimated compliance measure is the CACE estimate of the compliance effect. We will use the R package ivpack to perform the CACE analysis on the primary outcome only.

### Intra-cluster correlations (ICCs)

A two-level multilevel model of the intervention regressed on maths attainment will be run without any covariates, to estimate the ICC, in this case, the proportion of the total variance accounted for by the school level.

Furthermore, the ICC for the main model will be calculated from the output of the primary model, (including the covariates). It will again, be the proportion of the total variance that is accounted for by school level variance.

### Effect size calculation

Effect sizes and confidence intervals will be calculated for all outcome models.

For the primary outcome model, and the student teacher relationship outcome, (i.e. the two level models) the numerator for the effect size calculation will be the coefficient of the intervention group from the multilevel model. The denominator will be the total variance from a multilevel model without covariates, i.e. equivalent to Hedges' g.

$$ES = \frac{(\bar{Y}_T - \bar{Y}_C)_{\text{adjusted}}}{\sqrt{\sigma_S^2 + \sigma_{\text{error}}^2}}$$

For the secondary outcome measures using three level multilevel models with interactions, the numerator will be the model coefficient representing the overall mean difference between the intervention and control groups, while adjusting for the model covariates. The denominator will be the school and pupil level variance (not the time point level variance) of a multilevel model without covariates.

Confidence intervals for each effect size will be derived by multiplying the standard error of the intervention group model coefficient by 1.96. These will be converted to effect size confidence intervals using the same formula as the effect size itself.

## References

- Allison, P. (2012). 'Why maximum likelihood is better than multiple imputation', *Statistical Horizons*, 9 July [online]. Available: <https://statisticalhorizons.com/ml-better-than-mi> [8 January 2020].
- Allwood, M., Allen, K., Price, A., Hayes, R., Edwards, V., Ball, S., Ukoumunne, O. C. and Ford, T. (2018). 'The reliability and validity of the pupil behaviour questionnaire: a child classroom behaviour assessment tool', *Emotional and Behavioural Difficulties*, **23**, 4, 361- 371 [online]. DOI 10.1080/13632752.2018.1478945.
- Angrist, J. D. and Imbens, G. W. (1995). 'Two-stage least squares estimation of average causal effects in models with variable treatment intensity', *Journal of the American statistical Association*, **90**, 430, 431-442 [online]. DOI 10.1080/01621459.1995.10476535.
- Ford, T., Hayes, R., Byford, S., Edwards, V., Fletcher, M., Logan, S., Norwich, B., Pritchard, W., Allen, K., Allwood, M., Ganguli, P., Grimes, K., Hansford, L., Longdon, B., Norman, S., Price, A. and Ukoumunne, O. C. (2018). 'The effectiveness and cost-effectiveness of the Incredible Years® Teacher Classroom Management programme in primary school children: results of the STARS cluster randomised controlled trial', *Psychological Medicine*, **49**, 5, 828-842 [online]. DOI 10.1017/S0033291718001484.

Goodman, R. (1997). 'The Strengths and Difficulties Questionnaire: A Research Note', *Journal of Child Psychology and Psychiatry*, **38**, 5, 581-586 [online]. DOI 10.1111/j.1469-7610.1997.tb01545.x.

Goodman, R. (2001). 'Psychometric properties of the strengths and difficulties questionnaire', *Journal of the American Academy of Child & Adolescent Psychiatry*, **40**, 11, 1337-1345 [online]. DOI 10.1097/00004583-200111000-00015.

Hughes, R. A., White, I. R., Seaman, S. R., Carpenter, J.R., Tilling, K. and Sterne, J. A. C. (2014) 'Joint modelling rationale for chained equations', *BMC Medical Research Methodology*, **14**, 28 [online]. DOI 10.1186/1471-2288-14-28.

Pianta, R. (2001). *Student-Teacher Relationship Scale-Short Form*. Lutz, FL: Psychological Assessment Resources, Inc. Pianta, R. C. and Steinberg, M. (1992). 'Teacher-child relationships and the process of adjusting to school', *New Directions for Child and Adolescent Development*, 1992, 57, 61-80 [online]. DOI 10.1002/cd.23219925706.

White, I. R., Royston, P. and Wood, A. M. (2011) 'Multiple imputation using chained equations: Issues and guidance for practice', *Statistics in Medicine*, **30**, 4, 377-399 [online]. DOI 10.1002/sim.4067.