

<b>INTERVENTION</b>	<b>Zippy's Friends</b>
<b>DEVELOPER</b>	Partnership for Children
<b>EVALUATOR</b>	Queen's University Belfast
<b>TRIAL REGISTRATION NUMBER</b>	ISRCTN82558815
<b>TRIAL STATISTICIAN</b>	Seaneen Sloan and Sarah Miller
<b>TRIAL CHIEF INVESTIGATOR</b>	Seaneen Sloan
<b>SAP AUTHOR</b>	Seaneen Sloan and Sarah Miller
<b>SAP VERSION</b>	1
<b>SAP VERSION DATE</b>	28 February 2017
<b>EEF DATE OF APPROVAL</b>	15 August 2017
<b>DEVELOPER DATE OF APPROVAL</b>	15 August 2017

## Introduction

Zippy's Friends is a universal, classroom-based social and emotional learning programme that aims to develop children's repertoire of coping skills and their ability to adapt those coping skills to various situations. It is designed to be delivered to five to seven year olds by the class teacher over the course of an academic year, through 24 weekly, 45 minute sessions which focus on specific themes: feelings, communication, making and breaking relationships, conflict resolution, dealing with change and loss, and general coping skills. Weekly sessions are centred around a set of illustrated stories about a group of children, their families, friends and Zippy, a pet stick insect. Pupils also actively participate in activities such as role playing, group discussions, drawing and crafts, and playing games. Through listening to the stories and joining in activities, it is proposed that children learn how to find their own solutions to problems and in turn they become better equipped to cope with difficulties in life, and get on better with others.

Teachers undergo one day of training and are provided with a set of teaching and children's materials. Three follow up, twilight support sessions are held after modules two, four and six, in which teachers in the same area are brought together to discuss experiences, address any questions and look at the forthcoming modules.

The impact of Zippy's Friends on children's academic, social and emotional skills is being tested using a cluster-randomised controlled trial design. Specifically, the research questions addressed by the trial are:

1. What is the overall effectiveness of the programme on:
  - pupil reading attainment and emotional self-regulation (primary outcomes)
  - self-regulated learning and social skills (secondary outcomes).
2. Is there a differential impact of the programme for pupils eligible for Free School Meals (FSM)?
3. Is any variability in implementation associated with variability in outcomes?

The evaluation report is due to be submitted to the EEF in November 2017 for publication in Spring 2018.

## Study design

The trial is designed as a randomised controlled trial with two levels (pupils clustered within schools) and with two arms (a control arm and an intervention arm).

Schools in five local authority areas (Nottingham, Lincolnshire, Dudley, Croydon and Cheshire) were invited to participate in the trial. These local authorities were selected as they had indicated that there was sufficient interest from schools in the area. Recruitment of local authorities and schools took place between January and May 2016, and was led by the delivery team. A local coordinator, generally a professional working in a school or with a remit for working in schools (e.g. educational psychologist), was recruited by the delivery team within each of the five areas with a remit to engage with schools in the local area, initially to gauge interest in the programme, and then to monitor fidelity and providing ongoing support to schools.

All state infant and primary schools in these areas were eligible to take part in the trial if met the following criteria:

- They have not previously delivered the programme;
- They are prepared to allow 1 hour per week in the Year 2 timetable for the programme;
- They are prepared to release Year 2 teachers for training and support sessions; and

- They had provided pupil data needed for linking to the National Pupil Database (NPD), i.e. first name, last name, date of birth and unique pupil number, and had administered and returned the baseline measures (described below) to the research team before random allocation.

Within eligible schools, the target cohort was identified as pupils in Year 1 during the 2015/16 school year (and thus in Year 2 the following school year which is the trial period). All pupils were eligible to participate, and parental opt-out consent was sought for pupil participation in data collection.

The intervention arm received the programme, with Year 2 teachers being in trained in September/October 2016 and delivering the programme thereafter with their class. The control arm was asked to continue as normal and did not receive the programme. Schools allocated to the control group will receive a payment of £1000 at the end of the trial, and can then choose to purchase the programme.

There are two measurement points in this trial:

1. Pre-testing, which took place between May and July 2016 (prior to randomisation);
2. Post-testing, which is due to take place in June 2017.

## Protocol changes

There have been no changes to the protocol.

## Randomisation

Randomisation is at the school level to minimize the risk of contamination between trial arms. Schools were randomly allocated to either an intervention arm (which is delivering the programme) or a control arm (which is continuing with standard practice) in July 2016.

Schools were allocated to groups using the program Minim<sup>1</sup> to create groups that are balanced across pre-specified characteristics. This method of allocation is known as minimisation and is a widely accepted alternative to simple or stratified randomisation (Altman & Bland, 2005). For this trial, minimisation was based on two factors: local authority area and school-level attainment. Local authority was used to ensure that equal numbers of schools within each of the five local authority area are allocated to each group, to avoid the scenario whereby one area has very few intervention schools and another has many.

School-level attainment was used to ensure there are a similar proportion of high, medium and low attaining schools in both groups. School-level attainment was based on percentage of pupils obtaining level 4 or above in reading, writing and maths at Key Stage 2 in 2015. These data were collected from DfE school performance tables, published in December 2015<sup>2</sup>, using the variable PTREADWRITTAMATX which was available for 95% of the sample (79/83 schools). Four schools had no available KS2 results; one was a recently established primary school that had not yet had a cohort of pupils sit KS2, and the remaining three were infant schools. KS1 results (percent attaining level 2 or above in reading, writing and maths in 2015) were obtained directly from these schools were used in place of KS2 in these four instances.

School level attainment was split into three equally sized categories using internally-derived cut-offs. This involved taking published figures on the percentage of pupils attaining level 4 or above in reading, writing and maths at Key Stage 2 for each school recruited to the trial, and identifying cut off points such that three equal groups are created representing the top,

<sup>1</sup> <http://www-users.york.ac.uk/~mb55/guide/minim.htm>

<sup>2</sup> <https://www.gov.uk/government/statistics/primary-school-performance-tables-2015> (last accessed 12 May 2017)

middle and bottom third attaining schools within the sample. Cut-offs derived for the current sample were as follows:

1. Low attainment:  $\leq 76\%$
2. Medium attainment:  $\geq 77\%$  and  $< 86\%$
3. High attainment:  $\geq 86\%$

Allocation took place after baseline testing. Schools were made aware gained parental (opt-out) consent, completed and returned the baseline measures and provided the evaluators with data on participating pupils (for linking to the NPD) to be included in the random allocation process.

A total of 83 schools were entered into the random allocation process. Of these, 41 were allocated to the control group and 42 were allocated to the intervention group.

## Calculation of sample size

The sample size calculation in the protocol estimated that 70 schools with an average of 37 pupils per school (approx. 2590 pupils in total) would be sufficient to detect an effect size of 0.20 in an intention-to-treat analysis. This was calculated using Optimal Design (version 3.01) and is based on a two-level cluster design with an estimated pre-post correlation ( $R^2$ ) of 0.60, an intra-cluster correlation coefficient ( $\rho$ ) of 0.16 for the primary outcome (reading), and standard power and alpha thresholds of 0.80 and 0.05 respectively. However, during the course of the school recruitment stage of the project, a higher number of schools than anticipated expressed interest in taking part in the trial. An additional 15-20 schools would power the trial to detect an effect size of 0.18 (using the same assumptions as above), thus it was agreed with the funder that up to 90 schools could be recruited.

Based on recent DfE statistics<sup>3</sup> indicating that the average proportion of pupils eligible for and claiming FSM is 16.5%, it was estimated that there would be an average of 6 pupils eligible for FSM (based on an average year group size of 37 pupils). Using the same assumptions as above, the minimum detectable effect size for FSM pupils is 0.33.

### Updated sample size calculation

Following recruitment of 83 schools, the sample size calculation was recalculated using more precise estimates of the school size, intra-cluster correlation coefficient and pre-post correlation. Average number of pupils per year group increased to 45 (as recruited schools were bigger on average). Intra-cluster correlation ( $\rho$ ) was revised to 0.106 based on figures calculated by EEF<sup>4</sup>, and estimated pre-post correlation was increased to 0.70 (giving a  $R^2$  value of 0.49). Based on these assumptions, the trial has 80% power to detect an effect size of 0.17 (and 0.27 for FSM pupils).

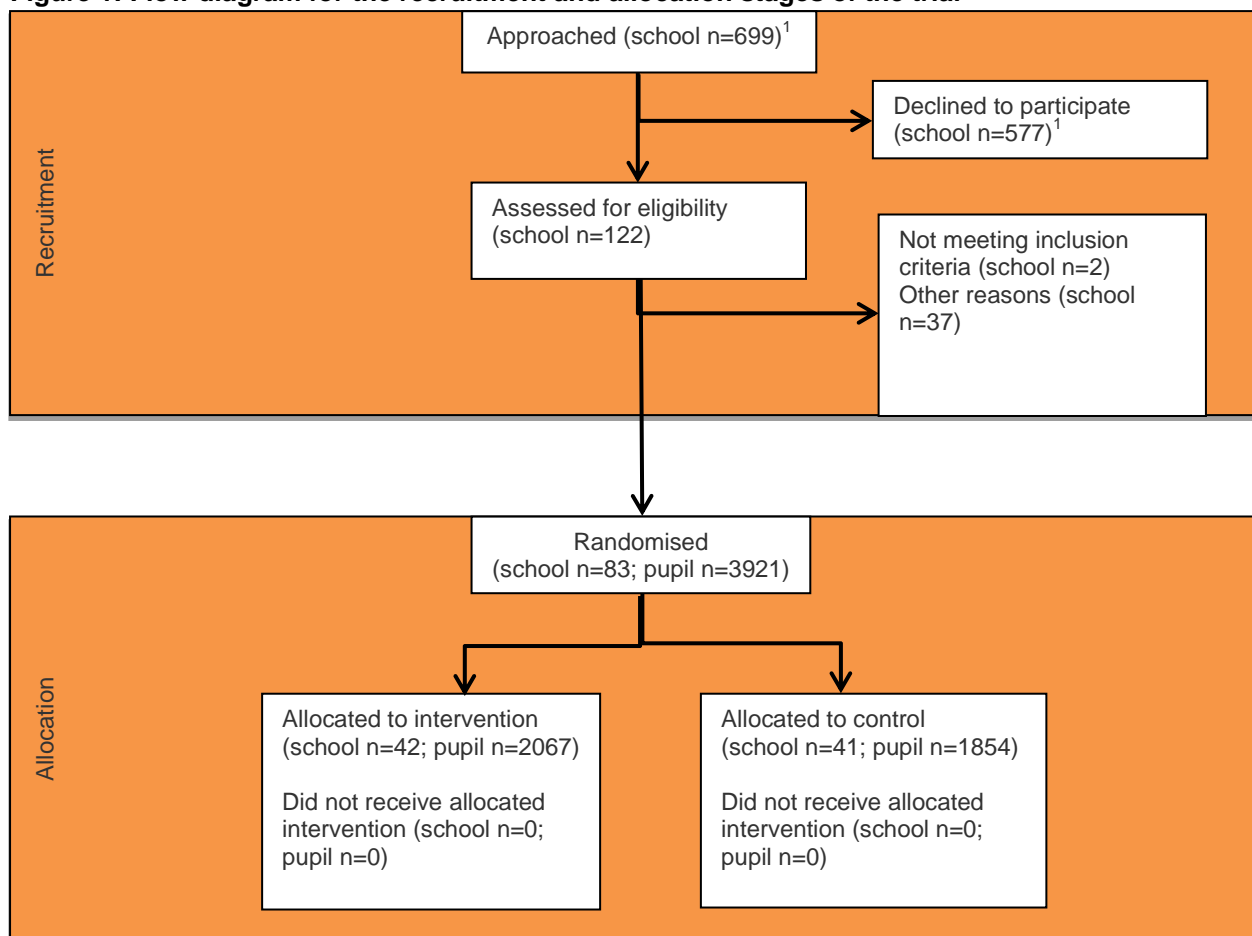
---

<sup>3</sup> Source: DfE (2015) Schools, Pupils and their Characteristics: January 2015 – National Tables (table 3b), available at <https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2015> (last accessed 12 May 2017)

<sup>4</sup> [educationendowmentfoundation.org.uk/public/files/Evaluation/Writing\\_a\\_Protocol/ICC\\_2015.pdf](http://educationendowmentfoundation.org.uk/public/files/Evaluation/Writing_a_Protocol/ICC_2015.pdf)

## Follow-up

Figure 1: Flow diagram for the recruitment and allocation stages of the trial



<sup>1</sup>At time of writing, figures for schools approached are outstanding for two of the five local authorities

## Outcome measures

### Primary outcome

There are two primary outcomes: reading attainment and emotional self-regulation.

Reading attainment was measured at pre-test using the Hodder Oral Reading Test Form A (Hodder Education), which was administered by Year 1 class teachers (or teaching assistants) during the pre-test period and prior to randomisation.

Reading attainment will be measured at post-test (June 2017) using the paper version of the Hodder Group Reading Test (Hodder Education). This will be administered on a whole-class basis under test conditions, independently by fieldworkers (blind to group allocation) who will be trained, overseen and monitored by the evaluation team. Completed test papers will be returned by the fieldworker on the day of testing directly to Hodder for marking, which again will be blind to group allocation. The raw, unstandardized score will be used in analysis.

Emotional self-regulation was measured at pre-test using teacher-report on the 'self-regulation' subscale of the Emotional Literacy Checklist (Faupel, 2003; GL Assessment). This contains 4 items (e.g. loses temper when loses at a game or in a competition), each rated on a 4 point scale (0=very true; 1=somewhat true; 2=not really true; 3=not at all true). Total scores on this scale can range from 0 to 12, with higher scores reflecting greater emotional self-regulation.

Pupil self-report of emotional self-regulation at pre-test was not collected due to the pupil age at pre-test (five to six years) and the lack of reliable self-report measures for children of this age. However self-report of emotional self-regulation will be collected at post-test (when pupils are aged six to seven years), using the Child Anger Management Scale (CAMS; Zeman et al., 2001). This measure was selected over the Emotional Literacy Checklist as the self-regulation subscale in the pupil version demonstrates poor reliability when used as a single subscale (Cronbach's  $\alpha=0.52$ ; Faupel, 2003). The CAMS contains 11 items covering three areas: inhibition of anger expression (four items, e.g. 'I hold my anger in'); coping, or anger control (four items, e.g. 'I try to calmly deal with what is making me feel mad'), and dysregulation of anger expression (three items, e.g. 'I say mean things to others when I'm mad'). Each item is rated on a 3-point scale (1=hardly ever; 2=sometimes; 3=often), meaning total scale scores can range from 11 – 33 with higher scores reflecting greater emotional self-regulation.

The CAMS has been used with children aged six years and over and has demonstrated good internal consistency (Cronbach's  $\alpha=0.62-0.77$ ) as well as convergent and discriminate validity with measures of emotion, psychopathology, and social functioning (Zeman et al. 2001; Zeman et al., 2002; Suveg & Zeman, 2004; Sim & Zeman, 2006; Suveg et al., 2009). The scale was recently used by the evaluators in an evaluation of another social and emotional learning programme (Roots of Empathy) in a UK setting, in which it demonstrated good reliability (Cronbach's  $\alpha=0.69-0.77$ ) with pupils aged eight years and over. This measure will be included within a questionnaire that will be administered on a whole-class basis by the fieldworkers.

### **Secondary outcomes**

The two secondary outcomes are self-regulated learning and social skills.

Self-regulated learning was measured at pre-test using teacher report on the 'attention/persistence' and the 'emotional control' subscales of the Learning Behaviour Scale (McDermott et al., 2001). These subscales contain 12 items in total (e.g. doesn't stick to tasks), each rated on a 3-point scale (0=doesn't apply, 1=sometimes applies; 3=most often applies). The same measure will be used again at post-test and again using teacher report. Pupil report will not be measured due to the lack of existing self-report measures available for use with pupils aged five to seven years.

Social skills will be measured at post-test using the social skills subscale of the Social Skills Improvement System rating scale (SSIS; Gresham & Elliott, 2008). This scale measures aspects of social skills, including communication, cooperation, assertion, responsibility, empathy, engagement and self-control. It contains 46 items which are rated on a 4-point scale (0=never; 1=seldom; 2=often; 3=always). This measure was chosen as it is well established in the academic literature, is standardised for use with children and young people aged three to 18 years, includes both teacher- and pupil-report options, and can be completed by pupils aged eight years and over. At pre-test, due to pupil age, social skills will be measured using teacher-report only. At post-test, social skills will be measured using pupil self-report only.

## **Analysis**

### **Primary intention-to-treat (ITT) analysis**

The primary analysis will be conducted in Stata version 14 (StataCorp, 2015) on an intention to treat basis. Multilevel random-effects linear regression modelling will be used to determine the effects of the programme on the primary outcomes (reading attainment and emotional self-regulation). A separate model will be constructed for each outcome, with the outcome as the dependent variable and the following covariates:

- Dummy variable representing membership of the intervention group (1=intervention; 0=control); and

- Variable representing the score at pre-test for the outcome.

### *Imbalance at baseline*

The extent to which the control and intervention groups are balanced will be assessed by comparing:

- school characteristics: school type, locality (urban/rural), Ofsted rating, school size (based on total number of pupils on the register);
- school-level pupil characteristics: proportion attaining level 4 or above at Key Stage 2 in reading, writing and maths (in 2015), proportion with English as an additional language, proportion FSM eligible and proportion with statements of special educational needs or education and health care plans;
- pupil characteristics: gender, FSM eligibility and EAL;
- primary and secondary outcomes measured at pre-test.

Linear and logistic multilevel regression models will be used to model the imbalance between the intervention and control groups, accounting for the clustered nature of the data. Means and standard deviations of these characteristics will be presented in tabular format, disaggregated by control and intervention group. Imbalance at baseline attainment will be reported an effect size.

### *Missing data*

Missing post-test data for reading attainment may occur if pupils are absent from school on the day of testing. This will be minimised by a follow-up visit to any school with pupil absences. Missing data may also occur if a pupil leaves the school completely before the post-tests are administered, if the child does not assent to participate on the day of testing, or if parents withdraw consent.

For the three other outcomes measured using multi-item scales (i.e. emotional regulation, self-regulated learning and social skills), there will be two types of missing data: (1) complete missing data (e.g. due to pupil absence from school during data collection or refusal to participate) and (2) partial missing data where the pupil has completed some but not all items of the measure. Again, complete missing data will be minimised by a second visit to the school to obtain data for any pupils who were absent on the main day of testing. Partial missing data will be minimised during administration, as the fieldworker will ask pupils to check over their questionnaires for any items that may have accidentally missed out. The extent of missing data within each scale will be checked and cross-referenced with paper questionnaires to check for data entry errors.

For all variables, the proportion of and reason for missing data will be assessed and reported. The proportion of each outcome lost to follow-up in the control and intervention group will be examined through cross-tabulations. If missing data are less than five percent then a complete case analysis will be undertaken.

Multiple imputation will be conducted as a sensitivity analysis if there is a high level of missing data (i.e. over 5%). The pattern of missing data will be explored by comparing the proportion of missing data in each of the control and intervention groups in addition to exploring how missingness is related to the outcomes in question. If the data can be assumed to be MCAR then imputation is not required. If the data are assumed to be MAR, this assumption renders the missing mechanism ignorable, simplifying the imputation step whilst ensuring correct inference. The imputation model will impute data separately for the control and intervention groups and will include all relevant variables and auxiliary variables involved in the analysis and sampling design. The imputation will be performed using chained equations which fills in missing values in multiple variables iteratively by using a sequence of univariate imputation methods with fully conditional specification of prediction equations. This method accommodates arbitrary missing-value patterns. Twenty imputations will be conducted in order to lessen the simulation (Monte Carlo) error. The analysis using

the imputed datasets will then be compared to the complete case analysis. The sensitivity analysis would constitute a secondary analysis and the primary outcome analysis will be presented without the multiple imputation.

### Secondary outcome analyses

Analysis of secondary outcomes (self-regulated learning and social skills) will be conducted using the same procedure as for the primary outcome analyses (described above).

### On-treatment analysis

Teachers complete a module report and return this to the local coordinator at the end of every module (6 modules in total). Module reports are collated by the programme developer and shared with the evaluator. These reports collect information on the delivery of each session, including whether the session was delivered and whether all, some or none of the activities for that session were followed. This information will be used to create a dosage variable, based on the following coding strategy:

- 0 = Session not delivered
- 1 = Session delivered, no activities followed
- 2 = Session delivered, some activities followed
- 3 = Session delivered, all activities followed

Codes will be applied across all 24 sessions for each class within each school in the intervention group, and summed to create the dosage variable (possible range 0 - 72). The primary and secondary outcome analyses outlined above will be repeated using this dosage variable in place of the intervention group variable.

### Subgroup analyses

Subgroup analyses will be conducted to determine whether the programme works differently for different subgroups of children. One subgroup was specified in the protocol, namely, pupils eligible for FSM. The main analyses outlined above will be repeated on a subsample of children who are identified as ever eligible for FSM based on data obtained from the NPD.

### Effect size calculation

Effect size (Hedges' g) will be calculated as the standardised mean difference between the control and intervention groups, using the pooled standard deviation. The pooled standard deviation will be calculated using the formula:

$$s = \frac{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}}{n_1 + n_2 - 2}$$

Hedges' g will then be calculated as:

$$g = \frac{\text{coefficient}}{\text{pooled standard deviation}}$$

## Report tables

Table 1: Summary of impact on primary outcomes

Group	Effect size (95% confidence interval)	Estimated months' progress	EEF security rating	EEF cost rating
-------	--	-------------------------------	---------------------------	--------------------



Treatment vs. control (Reading)				
Treatment vs. control (Emotional self-regulation)				
Treatment FSM vs. control (Reading)				
Treatment FSM vs. control (Emotional self-regulation)				

**Table 2: Timeline**

Date	Activity

**Table 3: Minimum detectable effect size at different stages**

Stage	N schools (n=intervention ; n=control)	Correlation between pre- test (+other covariates) & post-test	ICC	Blocking/ stratification or pair matching	Power	Alpha	Minimum detectable effect size (MDES)
<b>Protocol</b>	70 (35; 35)	0.60	0.16	Minimisation (on local authority and KS2 attainment)	80%	0.05	0.20
<b>Randomisation</b>	83 (42; 41)	0.70	0.106	Minimisation (on local authority and KS2 attainment)	80%	0.05	0.17
<b>Analysis (i.e. available pre- and post-test)</b>							

**Table 4: Baseline comparison**

Variable	Intervention group		Control group		
	School-level (categorical)	n/N (missing)	Percentage	n/N (missing)	Percentage
<b>School type</b>					
Academy Converter		10/42 (0)	24	11/41 (0)	27
Academy Sponsor-led		4/42 (0)	10	3/41 (0)	7
Community School		16/42 (0)	38	17/41 (0)	41

Foundation School	1/42 (0)	2	3/41 (0)	7
Voluntary Aided School	8/42 (0)	19	4/41 (0)	10
Voluntary Controlled School	3/42 (0)	7	3/41 (0)	7
<b>Location</b>				
Urban	37/42 (0)	88	33/41 (0)	80
Rural	5/42 (0)	12	8/41 (0)	20
<b>Ofsted rating</b>				
Outstanding				
Good				
<b>School-level (continuous)</b>	<b>n (missing)</b>	<b>Mean</b>	<b>n (missing)</b>	<b>Mean</b>
<b>School size</b>				
Total number of pupils enrolled	42 (0)	327	41 (0)	319
<b>Attainment</b>				
% attaining level 4 in KS2 Reading, Writing and Maths	42 (0)	80	41 (0)	79
<b>Free School Meal eligibility</b>				
% eligible for FSM	42 (0)	18	41 (0)	15
<b>English as an Additional Language</b>				
% with English as an additional language	42 (0)	17	41 (0)	14
<b>Special Educational Needs</b>				
% with statements of SEN or EHC plans	29 (13)	2	29 (12)	2
<b>Pupil-level (categorical)</b>	<b>n/N (missing)</b>	<b>Percentage</b>	<b>n/N (missing)</b>	<b>Percentage</b>
<b>Male</b>				
<b>Eligible for FSM</b>				
<b>English as an Additional language</b>				
<b>Pupil-level (continuous)</b>	<b>n (missing)</b>	<b>Mean</b>	<b>n (missing)</b>	<b>Mean</b>
<b>Reading attainment (pre-test)</b>				
<b>Emotional self-regulation (pre-test)</b>				
<b>Self-regulated learning (pre-test)</b>				
<b>Social skills (pre-test)</b>				

**Table 5: Primary analysis**

Outcome	Raw means				Effect size		
	Intervention group		Control group		n in model (intervention; control)	Hedges g (95% CI)	p- value
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)			
<b>Reading attainment</b>							
<b>Emotional self-regulation</b>							
<b>Self-regulated learning</b>	...	...	...	...	...	...	...
<b>Social skills</b>							

