

Statistical Analysis Plan: Lexia

Evaluator (institution): York Trials Unit, University of York



Principal investigators: Louise Tracey and Louise Elliott

PROJECT TITLE	Lexia Reading Core5®
DEVELOPER (INSTITUTION)	Queen's University Belfast, Lexia UK
EVALUATOR (INSTITUTION)	University of York
PRINCIPAL INVESTIGATOR(S)	Louise Tracey & Louise Elliott
TRIAL (CHIEF) STATISTICIAN	Caroline Fairhurst
SAP AUTHOR(S)	Caroline Fairhurst & Saleema Rex
TRIAL REGISTRATION NUMBER	TBC
EVALUATION PROTOCOL URL OR HYPERLINK	Please see the latest protocol here .

Changes from current version of the protocol included in this SAP

- In the protocol, pupils who received a standard age score (SAS) of 85 or less in any of the three subtests of the WRMT-R/NU at pre-test were planned to be recruited but this was amended to include all pupils with pre-test WRMT-R/NU score less than or equal to the 12th lowest ranking pupil's score for that school (up to a maximum of 14 per school). This change was carried out because fewer children than anticipated were fulfilling the original selection criteria.
- In the protocol, we estimated to recruit an average of 9 pupils per school based on the original sample size calculation to detect a MDES of 0.2 with 80% power but this was increased to recruit up to 14 pupils per school. This was done because schools suggested that they could manage up to 7 Lexia pupils.
- The original protocol indicated that variable block sizes would be used in the stratified block randomisation, but a fixed block size of 2 was implemented in practice to ensure no more than 7 pupils were allocated to receive the Lexia intervention in any one school.
- In the protocol, it is stated that the treatment effect size would be calculated based on the adjusted mean difference between the intervention and control group and the total variance obtained from the same linear mixed model. However, this does not follow the most recent EEF analysis guidance. Therefore, effect sizes will be calculated by dividing the adjusted mean difference between the intervention and control group by the pooled variance obtained from an unconditional model. A 95% CI for the effect size will be calculated by dividing the 95% confidence limits for the adjusted mean difference by this same variance. All parameters used in these calculations will be provided in the final report.

Table of contents

Table of contents	2
SAP version history.....	2
Introduction.....	2
Design overview.....	3
Follow-up and outcome measures	5
Sample size calculations overview	8
Randomisation.....	9
Analysis	9
Primary outcome analysis.....	10
Secondary outcome analysis	11
Interim analyses	11
Subgroup analyses.....	11
Additional analyses.....	11
Imbalance at baseline.....	11
Missing data	12
Compliance	12
Intra-cluster correlations (ICCs)	13
Effect size calculation	13
References	13

SAP version history

VERSION	DATE	REASON FOR REVISION
1.0	16/11/2018	N/A

This analysis plan was written post-randomisation but prior to receipt of any outcome data, and deals only with the statistical analysis of the impact evaluation.

Introduction

Lexia, a computer-based independent learning system (ILS), was originally developed in the US to help pupils with dyslexia. Lexia Reading Core5® is designed for wider use and provides personalised, adaptive learning for a wide range of ability levels at primary school age.

Pupils begin with a diagnostic test and are placed at an appropriate level and work independently, typically having three to four 20-minute sessions per week (not including

set-up time). The system is able to keep track of users' progress in real-time. The tracking data will be shared by the developer and will be used in the compliance analysis

Facilitators (Teachers and/or teaching assistants) are provided with reports to monitor pupils' performance and, where appropriate, paper-based activities are suggested from within the system. Facilitators need to give pupils initial guidance on using the programme, teach and reinforce some units, and oversee and monitor pupil progression. Online training and support is offered to facilitators.

The programme is most commonly used in UK schools as an in-school supplement to target struggling readers although it can also be used as a whole class intervention or as a school-provided, home-use supplement to teaching. Lexia Reading Core5® has been adapted to the UK context (eg. using UK-English audio and spellings) and is currently being used in over 3,000 schools.

Although some studies of Lexia have been conducted in the UK these have generally been small-scale (Brookes 2016). There has been one randomised controlled trial (RCT) of Lexia Reading Core5® in Northern Ireland which found effects for blending and non-word reading in 4-6 year old pupils (O'Callaghan et al. 2016). The proposed efficacy trial provides an opportunity to evaluate the programme using a large-scale RCT in England (e.g. intervention adapted to include UK-English audio and spellings).

The trial aims to answer the following research questions:

1. How effective is Lexia Reading Core5® in improving struggling readers' reading skills during Year 2? [Primary Outcome]
2. How effective is Lexia Reading Core5® in improving struggling readers' word recognition skills during Year 2? [Secondary Outcome 1]
3. How effective is Lexia Reading Core5® in improving struggling readers' decoding skills during Year 2? [Secondary Outcome 2]
4. How effective is Lexia Reading Core5® in improving struggling readers' comprehension skills during Year 2? [Secondary Outcome 3]
5. How effective is Lexia Reading Core5® in improving struggling readers' fluency skills during Year 2? [Secondary Outcome 4]
6. How effective is Lexia Reading Core5® in improving struggling readers' outcomes in KS1 national reading assessments? [Secondary Outcome 5]
7. How effective is Lexia Reading Core5® in improving struggling readers' reading skills during Year 2 for FSM pupils? [Secondary Outcome 6]

Design overview

This is a two-armed within-school individual randomised controlled efficacy trial. Within-school randomisation was used as the intervention is delivered one-to-one via a computer; therefore, the possibility of diffusion is reduced and, where it is feasible, individual randomisation is more efficient (i.e. requires a smaller sample size to detect the same treatment effect) than cluster randomisation (by class or school for example). Our sample comprises of struggling readers within Year 2 (Key Stage 1) in evaluation schools in the academic year 2018-19. Eligible pupils from eligible schools were randomly allocated 1:1 to receive either the intervention or teaching as usual.

Schools were eligible to participate if they:

- Had approximately 50 pupils per year group;

- Were not involved in another EEF trial focusing on KS1 literacy or aiming to achieve change at a whole school level;
- Were not currently using Lexia Reading Core5®, or had used Lexia Reading Core5® in the previous 12 months;
- Met the technological requirements to support an IT-based intervention (the intervention can be run on iPads); and
- Were willing to implement the intervention with respect to the random allocation (i.e. only with those pupils assigned to the intervention group).

Schools were recruited by the delivery team, with support from the evaluation team. Fifty-seven eligible schools were recruited to take part in the trial.

Class teachers were asked to provide the names of the half of the Year 1 pupils with the lowest attainment at the end of the school year 2016-17 (based on their professional judgement). In the following year, 2017-18, these children were independently assessed by the administration of the three subtests (Word Identification, Word Attack and Passage Comprehension) of the Woodcock Reading Mastery Tests – Revised Normative Update¹ (WRMT-R/NU (Sutton 1999)) by assessors recruited and trained by the University of York. Children for whom a withdrawal of data form was received from their parents were not tested.

In the original protocol, it was planned that pupils who received a standard age score (SAS) of 85 or less (expected range is 74-166) in any of the three subtests of the WRMT-R/NU (word recognition, decoding and comprehension - corresponding to the subsets named above) at pre-test would be eligible to participate in the trial. However, when initial scores came in it became apparent that far fewer children than anticipated were fulfilling this criteria. Therefore, it was decided to include the group of lowest performing pupils (i.e. the most struggling readers) in each school. Each of the three subtests were scored and converted to a SAS as described in the WRMT-R/NU Examiner's Manual. The average of the SAS subtest scores was calculated and used to identify pupils to take part in the evaluation. An average SAS was calculated if pupils had completed at least two out of the three subtests.

Schools indicated that the maximum number of Lexia pupils they could manage was 6-7. Therefore, it was decided to include the 12-14 pupils in each class with the lowest pre-test score. If there was more than one pupil with the same score as the 12th pupil then up to 14 pupils were included. If more than 14 pupils would have been included by having the same score as the 12th ranking pupil then we planned to randomly select those with this score for inclusion such that a maximum of 14 were included per school so that there was no potential for more than 7 pupils to be randomised (1:1) to the Lexia intervention. However, in the event, this scenario did not arise.

The programme will be delivered to the pupils randomised to the intervention group from September 2018 for a period of 24 weeks/two full terms.

¹ <https://www.pearsonclinical.com/education/products/100000647/woodcock-reading-mastery-tests-revised-normative-update-wrmt-r-nu.quick.html>

Statistical Analysis Plan: Lexia

Evaluator (institution): York Trials Unit, University of York



Principal investigators: Louise Tracey and Louise Elliott

Table 1: Outcome measures

Trial type and number of arms		Two-armed within-school individual randomised controlled trial (efficacy)
Unit of randomisation		Pupils
Stratification variables (if applicable)		School
Primary outcome	variable	Reading ability
	measure (instrument, scale)	WRMT-III composite of the scores of four subtests (Word Identification, Word Attack, Passage Comprehension, and Oral Reading Fluency)
Secondary outcome	variable	Word recognition
	measure (instrument, scale)	WRMT-III word identification subtest score
Secondary outcome	variable	Decoding
	measure (instrument, scale)	WRMT-III word attack subtest score
Secondary outcome	variable	Comprehension
	measure (instrument, scale)	WRMT-III passage comprehension subtest score
Secondary outcome	variable	Fluency
	measure (instrument, scale)	WRMT-III oral reading fluency subtest score
Secondary outcome	variable	Reading attainment
	measure (instrument, scale)	KS1 reading raw score

Follow-up and outcome measures

Measure of prior attainment

Pre-test WRMT-R/NU was administered one-to-one to all the participating pupils, by trained assessment administrators before randomisation, at the end of the school year 2017-18.

Scores from this will be used as the baseline measure of prior attainment (Table 2).

The WRMT-R/NU is the previous version of the Woodcock Reading Mastery Test (its content was developed in 1987), which was superseded by the third version (WRMT-III) in 2011. The initial hope was to use the WRMT-III at pre-test; however, upon ordering the tests the evaluation team were informed that they could not be delivered in time for the testing period, and so an older version of the test had to be administered (the evaluation team, by

coincidence, already had copies of this test from a previous research project). The WRMT-III, however, will be available for the post-test.

All the subtests were fully revised/updated between the NU version and the III version. The III version used current research to ensure it covered all areas of reading ability, and it was re-normed and can be used from an earlier age (4y 6m rather than 5y). The newer version contains more engaging content, has two equivalent test forms, and the item content was updated to ensure no bias (gender, ethnicity, socioeconomic background, etc). For the subtests we are using, there are fewer items on each test (about half as many), about 40% of the items are new for Word Identification and Word Attack, while for Passage Comprehension all, but 1, items are new. Now the rule is to discontinue the test if 4 consecutive items are incorrect, rather than 6 consecutive items at the end of a page with the NU version. The WRMT-R/NU does not include a fluency subtest like the WRMT-III; therefore, the composite score will be used as the measure of prior attainment for the fluency subtest of the post-test (Table 2). The use of different WRMT test versions for the pre and post test will not affect the validity of the research.

Primary outcome measure

Post-test WRMT-III will be administered one-to-one to all the participating pupils by trained administrators who are blind to group allocation to avoid the potential for ascertainment bias. The assessment administrators will be recruited and trained by the evaluation team.

The primary outcome measure is the composite of the standard age scores of four subtests obtained from this measure (Woodcock 2011). The WRMT-III subtests measure word recognition (Word Identification), decoding (Word Attack), comprehension (Passage Comprehension) and fluency (Oral Reading Fluency). This is considered an appropriate measure as these subtests identify the key areas in which readers typically struggle and those that Lexia Reading Core5® targets. The composite score constructed from these subtests will reflect overall reading ability. This score will be obtained from summing the standard age scores for the four subtests.

We considered adding the 'Word Comprehension' subtest to enable us to use the 'Total Reading Cluster' as an overall measure. However, we decided not to do this as 'Word comprehension' would replicate some of the skills involved in 'Passage Comprehension' and additional time required to conduct would also result in additional burden on children and schools and increased costs.

Secondary outcome measures

The secondary outcome measures are the standard age scores of the individual subtests of the post-test WRMT-III and the KS1 reading raw scores (Table 1). The KS1 reading scores will be securely transferred from schools in an encrypted excel spreadsheet for the Year 2 pupils as the National Pupil Database only holds data on whether pupils are 'working towards', 'working at' or 'working above' the standard expected at the end of KS1.

In the protocol, we state that *raw* scores will be used for the post-test WRMT-III, by this we meant fine graded scores, and will actually use the standard age scores, following guidance from the user manual, as the composite score is obtained by first calculating standard age scores and second by summing these scores. 'Raw' scores have no meaning in this instrument.

Table 2: Outcome measures and associated baseline measures of prior attainment (all scores are age-normed)

Time point (outcome measure)	Measure	Variable	Scoring range	Time point (baseline measure)	Measure	Variable	Scoring range
OUTCOME				BASELINE			
June 2019	WRMT-III composite reading score*	Reading ability	231-580	June/July 2018	WRMT-R/NU composite reading score**	Reading ability	74-166
	WRMT-III word identification score	Word recognition	55-145^		WRMT-R/NU word identification score	Word recognition	70-159
	WRMT-III word attack score	Decoding	64-145^		WRMT-R/NU word attack score	Decoding	81-163
	WRMT-III passage comprehension score	Comprehension	57-145^		WRMT-R/NU passage comprehension score	Comprehension	70-175
	WRMT-III oral reading fluency score	Fluency	55-145^		WRMT-R/NU composite reading score***	Reading ability	74-166
	KS1 raw reading score	Reading attainment	0-40		WRMT-R/NU composite reading score	Reading ability	74-166

*primary outcome; **the average SAS score calculated at baseline and used to identify pupils for the evaluation; ***no fluency subtest in pre-test; ^lower bound differs depending on age

Sample size calculations overview

Protocol

We proposed to recruit 57 schools. This number of schools was considered realistic given the capacity of the delivery team and the additional processes established to ensure compliance. Assuming an average of 56 pupils in the school (28 per class, 2-form entry), we estimated that an average of 9 pupils per school would be identified as struggling readers. This was based on administration of a similar assessment, the York Assessment for Reading Comprehension (Snowling 2011), with similar pupils in a previous study (Tracey et al. 2014), which found that 17% of pupils were struggling readers, as identified by the YARC. Assuming a pre and post-test correlation of 0.6 between the baseline and post-test of the WRMT, with 80% power, the minimum detectable effect size (MDES) with this sample size would be 0.20 allowing for 10% pupil level attrition (StataCorp. 2017). The model that will be used to analyse the data will adjust for any potential within school correlation but this was not incorporated in the sample size calculation.

The first 50 schools recruited to the trial had an average FSM of 29.3% overall. Based on this estimate, we might conservatively have expected 3 of the struggling readers per school to be eligible for FSM; therefore, in this subgroup, *ceteris paribus*, the MDES would be 0.36. However, it is likely that FSM status and being a struggling reader are correlated, so a higher proportion of the 9 identified struggling readers might have been eligible for FSM. For example, with an average of 6 FSM pupils per school, the MDES would be approximately 0.26 (*ceteris paribus*).

Randomisation

A total of 57 schools were recruited. All eligible pupils with pre-test WRMT-R/NU score less than or equal to the 12th lowest ranking pupil in the class were randomised (1:1) to either Lexia intervention or teaching as usual. Based on this selection criteria, 697 pupils in total (mean 12.2 per school, range 12-13) were randomised per recruited school. Assuming a pre- and post-test correlation of 0.6 and 10% pupil-level attrition, the MDES with this sample size would be 0.18 with 80% power for the overall study.

For the FSM subgroup, since the recruited schools had an average of 29.2% FSM pupils, we could conservatively estimate an average of four randomised pupils to be eligible for FSM in each recruited school. Using otherwise identical parameters to those described above, the MDES in this subgroup would be 0.31 with 80% power. However, there could be more than four randomised pupils in each school eligible for FSM as FSM status is correlated to pre-test scores. Depending on the percentage of FSM pupils among the randomised pupils, MDES with 80% power could be as low as 0.18.

Table 3: Sample size estimations

		Protocol		Randomisation	
		OVERALL	FSM	OVERALL	FSM
MDES		0.20	0.36	0.18	0.31
Pre-test/ post-test correlations	level 1 (pupil)	0.6	0.6	0.6	0.6
	level 2 (class)	-	-	-	-

	level 3 (school)	-	-	-	-
Intracluster correlations (ICCs)	level 2 (class)	N/A	N/A	N/A	N/A
	level 3 (school)	N/A	N/A	N/A	N/A
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		9	3	12.2	4
Number of schools	intervention	57	57	57	57
	control	57	57	57	57
	total	57	57	57	57
Number of pupils	intervention	257	86	348	114
	control	257	86	349	114
	total	514	172	697	228

Randomisation

Block randomisation, stratified by site, was used to randomly allocate pupils within schools to either the intervention or control group. The original protocol indicated that variable block sizes would be used, but a fixed block size of 2 was implemented in the end to ensure no more than 7 pupils were allocated to receive the Lexia intervention in any one school. The allocation schedule was generated by the trial statistician, Caroline Fairhurst. Pupils were ordered by pre-test score and copied across to the allocation schedule, thereby ensuring allocation concealment from the schools and no chance of subversion.

Analysis

The statistical analysis proposed follows the most recent *revised EEF Statistical Analysis Guidance (2018)* available [here](#).

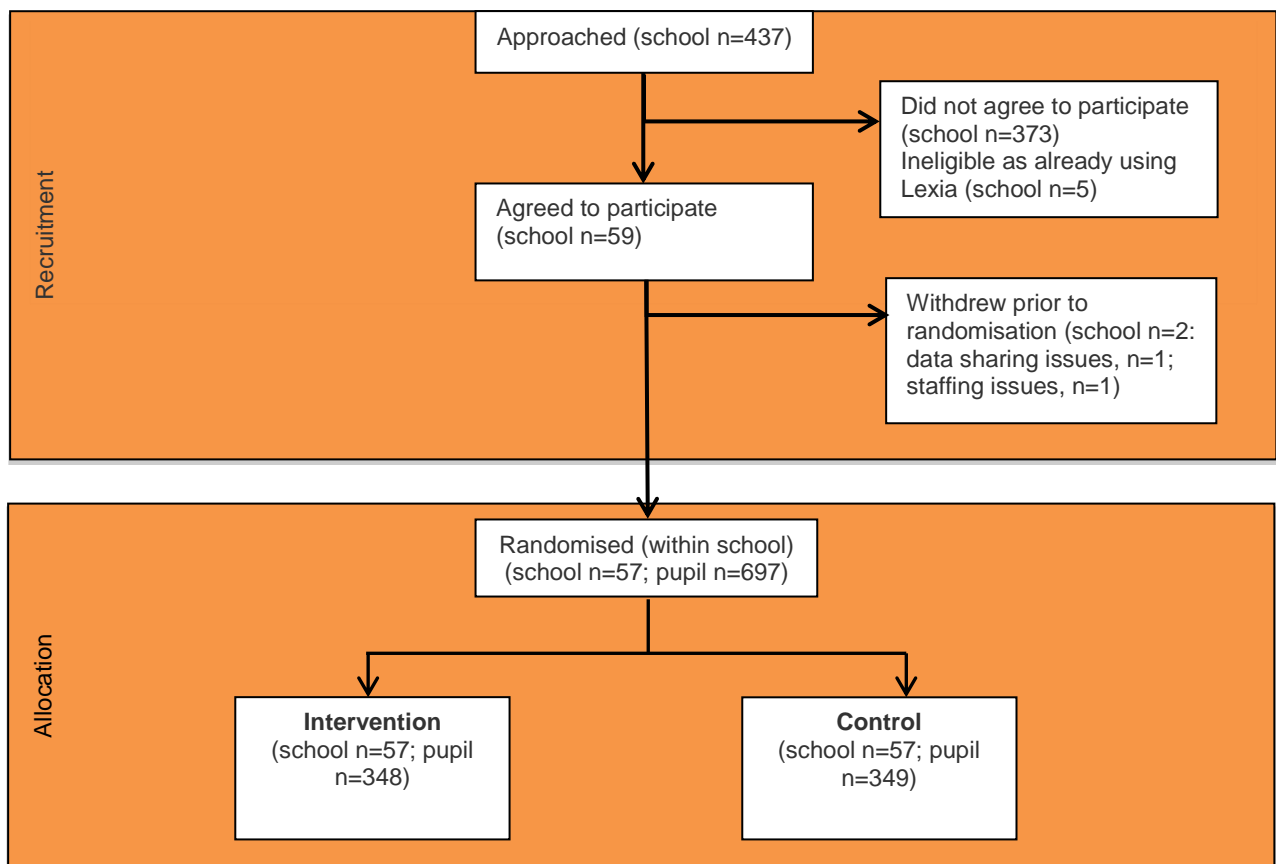
Analysis will be conducted in Stata v15 (or later, to be confirmed in the final report) using the principle of intention to treat (ITT), where data are available, including all pupils in the groups to which they were randomised irrespective of whether or not they actually received the intervention.

Statistical significance will be assessed using two-sided tests at the 5% level. Estimates of effect with 95% confidence intervals (CIs) and p-values will be provided. No formal comparison of baseline data will be undertaken, except to report the differences in WRMT-R/NU pre-test scores (composite reading score, word identification subtest score, word attack subtest score, and passage comprehension subtest score) as a Hedges' g effect size (Hedges 2007).

A CONSORT diagram will be produced to show the flow of schools and pupils through the trial. Since this statistical analysis plan was prepared following randomisation, the initial sections of the CONSORT diagram can be completed and are presented in Figure 1.

Pre- and post-test scores will be summarised for each trial arm, and presented using histograms.

Figure 1: CONSORT diagram up to, and including, randomisation of pupils



Primary outcome analysis

The primary analysis will investigate any difference in the WRMT-III composite reading score between the two groups. The analysis will take the form of a linear mixed model at the pupil-level with outcome score as the response variable. Group allocation and measure of prior attainment, will be included as fixed effects in the model, and school as a random effect.

Model equation:

$$Y_{ij} = \beta_0 + u_{0j} + \beta_1 \text{Prior}_i + \beta_2 \text{Rand}_i + e_{ij}$$

Y_{ij} = response of the i^{th} pupil in j^{th} school, $i=1, \dots, n$, $j=1, \dots, m$

n = number of participating pupils

m = number of recruited schools

Prior_i = score for measure of prior attainment for the i^{th} pupil

Rand_i = indicator variable for group allocation for i^{th} pupil (0=Control, 1=Intervention)

u_{0j} = random effect represents the deviation of school j 's mean z-score from the grand mean

$$u_{0j} \sim N(0, \sigma^2_2)$$

e_{ij} = residual error term for the i^{th} pupil in j^{th} school

$$e_{ij} \sim N(0, \sigma^2_1)$$

Model assumptions will be checked as follows: the normality of the standardised residuals will be checked using a histogram and qq plot, and the homoscedasticity of the residuals assessed using a scatter plot of fitted values against the residuals. Visual inspection of the plots only will be used (no formal statistical tests). If the model assumptions are in doubt, a sensitivity analysis will be conducted in which transformations of the outcome and/or covariate data will be tried to improve the model fit.

Secondary outcome analysis

The secondary outcomes of the individual WRMT-III subtest standard age scores and KS1 Reading raw score will be analysed as described for the primary outcome, using an appropriate pre-test WRMT-R/NU score as the measure of prior attainment in the model (e.g. composite WRMT-R/NU pre-test score will be included in the model for the outcome of KS1 Reading).

Interim analyses

No interim analyses will be undertaken.

Subgroup analyses

Pupil UPNs, as obtained during the recruitment period will be used to access additional pupil characteristics from the National Pupil Database. The effect of the intervention on pupils who are eligible for FSM² will be assessed via the inclusion of FSM status (using the EverFSM indicator EVERFSM_6_P in the NPD) and an interaction term between FSM status and allocation in the primary analysis model. This will be followed by repeating the primary analyses in the subgroup of pupils who have ever been eligible for FSM.

Additional analyses

There will be no additional analyses.

Imbalance at baseline

Characteristics of the recruited schools will be presented overall. Pupil characteristics (gender, age and FSM status) and measures of prior attainment will be summarised descriptively by randomised group both as randomised and as analysed in the primary analysis.

Continuous measures will be reported as a mean and standard deviation (SD) while categorical data will be reported as a count and percentage. The unadjusted difference between groups on the pre-tests for will be reported as a Hedges' g effect size with 95% CI.

² Note that, throughout, reference to 'FSM status' relates to the variable EVERFSM_6_P in the NPD, which indicates if the pupil has ever been recorded as eligible for free school meals at any time in the last 6 years up to the pupil's current year (not including nursery).

Missing data

The amount of missing baseline and outcome data will be summarised, and reasons for missing data explored and provided in the report, where available. Significant predictors and possible mechanisms for the missing data will be discussed in the report.

A mixed effect logistic regression model will be run to predict the presence of missing primary outcome (composite WRMT-III reading score) data including group allocation and WRMT-R/NU pre-test score.

The impact of missing data on the primary analyses will additionally be assessed using multiple imputation by chained equations, including all available baseline variables (pre-test WRMT-R/NU scores, age, gender and FSM status) (Azur et al. 2011).

A 'burn –in' of 150 will be used and 30 imputed datasets will be created. The primary analysis will then be rerun within the imputed datasets and Rubin's rules will be used to combine the multiply imputed estimates.

The analysis above will not account for the multilevel nature of the data. Sensitivity analyses will be attempted to be carried out by imputing the missing data using multilevel imputation (e.g. via the REALCOM-IMPUTE macro which is compatible with Stata) and repeating the primary analysis with this imputed dataset. The result from this will be compared against the result from the data imputed using MI by chained equations for major differences.

Compliance

The programme will be delivered from September 2018 for a minimum period of 24 weeks/two full terms. Schools are expected to schedule 4 sessions of 30 minutes (including 10 minutes setup time) per week. At the pupil-level, to meet the minimum requirements for compliance, pupils allocated to receive the intervention must do a minimum of 60 minutes (excluding setup time) per week for at least 12 non-consecutive weeks. Each pupil will work independently during the intervention with one adult (either a teacher or teaching assistant) supervising the pupils. The adult's role will be to ensure the children are on task, to monitor progress, scaffold learning with paper-based resources where necessary and manage the software. The intervention will be delivered in-school only as, although some schools currently provide the programme for use at home, we want to ensure that pupils without wider access to IT are not disadvantaged. In-school provision will also ensure more consistency of implementation fidelity.

Throughout the course of the trial, the question was raised about whether schools could continue using the Lexia programme at the end of the intended 24 week period, but before post-testing. It was agreed that this could happen as long as only the pupils allocated to receive the intervention continued to have access to it. Therefore, some pupils may receive more than the initially intended number of sessions. The number and length of sessions received by pupils will be summarised.

A Complier Average Causal Effect (CACE) analysis for the primary outcomes will be considered to account for compliance with the intervention. Three analyses will be conducted at the pupil level, considering compliance as:

1. a dichotomous variable (complied or not) with compliance defined as completing at least 60 minutes a week for a minimum of 12 non-consecutive weeks;

2. a dichotomous variable (complied or not) with compliance defined as completing at least 720 minutes in total; and
3. a continuous measure of number of hours of the intervention received.

An instrumental variable approach will be used, with random group allocation as the instrumental variable (Dunn, Maracy, and Tomenson 2005).

Intra-cluster correlations (ICCs)

The intraclass correlation coefficient (ICC) associated with school for the pre- and post-test outcomes will be provided with 95% CIs.

Effect size calculation

In the protocol, it is stated that the treatment effect size would be calculated based on the adjusted mean difference between the intervention and control group, and the total variance (between plus within school variance), obtained from the linear mixed model. However, this does not follow the most recent EEF analysis guidance. Therefore, effect sizes will be calculated by dividing the adjusted mean difference between the intervention and control group by the pooled variance obtained from the unconditional model. A 95% CI for the effect size will be calculated by dividing the 95% confidence limits for the adjusted mean difference by this same variance. All parameters used in these calculations will be provided in the final report.

References

- Azur, Melissa J., Elizabeth A. Stuart, Constantine Frangakis, and Philip J. Leaf. 2011. 'Multiple imputation by chained equations: what is it and how does it work?', *International journal of methods in psychiatric research*, 20: 40-49.
- Brookes, Greg. 2016. 'What works for children and young people with literacy difficulties? The effectiveness of intervention schemes'.
- Dunn, G., M. Maracy, and B. Tomenson. 2005. 'Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods', *Stat Methods Med Res*, 14: 369-95.
- Hedges, Larry V. 2007. 'Effect Sizes in Cluster-Randomized Designs', *Journal of Educational and Behavioral Statistics*, 32: 341-70.
- O'Callaghan, P., A. McIvor, C. McVeigh, and T. Rushe. 2016. 'A randomized controlled trial of an early-intervention, computer-based literacy program to boost phonological skills in 4- to 6-year-old children', *Br J Educ Psychol*, 86: 546-58.
- Snowling, M.J., Stothard, S., Clarke, P. Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K. & Hulme, C. 2011. *YARC: York Assessment of Reading for Comprehension. Passage Reading. 2nd Edition* (London: GL Assessment.).
- StataCorp. 2017. "Stata Statistical Software: Release 15. ." In.: College Station, TX: StataCorp LLC.
- Sutton, Joe. 1999. *Woodcock Reading Mastery Tests Revised/Normative Update (WRMT-R/NU)*.
- Tracey, Louise, Bette Chambers, Robert E. Slavin, Pam Hanley, and Alan Cheung. 2014. 'Success for All in England: Results From the Third Year of a National Evaluation', *SAGE Open*, 4: 2158244014547031.
- Woodcock, Richard W. 2011. *Woodcock reading mastery tests. Manual : WRMT-III* (Pearson).