Words for All

Intervention Developer	Whole Education
Delivery Organisations	Whole Education
Evaluator	Evidence Development and Incubation Team Health and Social Care Workforce Research Unit
Principal Investigator	Dr Mary Baginsky, Susannah Hume
Protocol Author(s)	Miriam Styrnol, Dr Mary Baginsky, Susannah Hume, Omar Shakir
Type of Trial	Randomised controlled trial
Age or Status of Participants	School years 7-11
Number of Participating Local Authorities	16
Number of Children and Families	Approx. 2,400 pupils across treatment and control
Primary Outcome(s)	Educational attainment in verbal and reading skills with focus on improved vocabulary measured through Student Age Score (New Group Reading Test)
Secondary Outcome(s)	a) GSCE Attainment-8 b) English Language GCSE c) Total Difficulties Score (Strengths and Difficulties Questionnaire) d) Attendance
Contextual Factors	School size, school organisational structure, number of students in school with CSC experience, overall demand

	for CSC policy and practice in the LA, interface between CSC in LA and schools, COVID-19 impacts

Summary

This protocol describes a planned evaluation of the impact of Words for All, an intervention that trains teachers to deliver vocabulary-focused educational enrichment, on the educational attainment of CSC-experienced students. The Words for All programme builds on the key learning from the trials run by the Education Endowment Foundation in 2013/14 as well as wider emerging learning and evidence around effective CPD and implementation of evidence based practices.

Words for All will consist of bespoke enrichment programmes developed by delivery partner Whole Education and groups of three teachers (known as a 'triad') in schools delivered to CSC-experienced pupils, either by withdrawing them from class in groups, or via one-to-one support. The intervention under evaluation is the establishment of a triad of teachers and staff within each school and the resources and training are provided by Whole Education. This trial will be tested during the 2020-2021 academic year with the final report to be published in September 2022. Words for All will be delivered across Key Stage 3 and Key Stage 4 (Years 7 to 11).

The research design of the trial consists of a two-armed randomised controlled trial. Randomisation will occur at the level of family-group, stratified by school. This is to avoid spill-overs between siblings, who are very likely to influence each other's outcomes. Family group will be identified by the school and noted using a family identifier in the randomisation spreadsheet provided to WW-CSC.

All pupils who are identified by the Virtual School, Local Authority or school as having had a social worker in the last six years in these year groups are in scope. Pupils will be randomly allocated to treatment and control groups, with the clustering at the family level. Following this, pupils and parents/guardians will be provided with Information Sheets and given the opportunity to opt out of the evaluation. Those who opt out will be included in the randomisation, and, if allocated to treatment, will be enrolled onto the Words for All programme but will be excluded from evaluation data collection and analysis.

This trial will include both an impact and implementation/process evaluation. The primary outcome variable for impact will be the Standard Age Score (SAS), as measured by the New Group Reading Test (NGRT). The NGRT is a standardised reading and comprehension assessment split into three sections assessing phonics, sentence completion, and passage comprehension and was used in the original EEF evaluation as the primary outcome. The NGRT will be administered at three time points at the start, middle, and end of the trial to both treatment and control groups. Secondary outcome variables include GSCE attainment, attendance, and scores on the Strengths and Difficulties Questionnaire, a commonly-used behavioural screening questionnaire for children and young people.

Implementation and process evaluation will consist of surveys and interviews with professionals involved in delivering the enrichment or working with the youth involved

(heads, virtual school heads, teacher triads, social workers), observation of the launch event and teacher training events, and group interviews of a sample of students participating in the programme. Lastly, a cost evaluation will also be conducted.

Table of Contents

Background and Problem Statement

Children and young people with a social worker have worse educational outcomes than the student population at large[1]. The lack of good evidence on educational interventions to benefit this group represents a major challenge, as educational attainment is a strong indicator of success later in life. In order to remedy this evidence deficit, in 2019 WWCSC conducted a re-analysis of data from 63 education-related RCTs funded by the Education Endowment Fund (EEF).[2] The re-analysis was conducted in order to assess whether the educational interventions, which were not targeted at children with a social worker, had any impact on the on the attainment of young people who are classified as 'CSC-experienced' (those who are, or have been classified in the past six years as, Children in Need, subject of a Child Protection Plan, and/or Looked After).

The re-analysis identified ten interventions that showed *signs of potential*[3] for CSC-experienced children. Given the original studies were not powered for CSC-experienced children, the findings were exploratory, and therefore warrant further research. WWCSC has taken forward three of these interventions to test them out in the field, one of which was the Vocabulary Enrichment Full Programme (VEIP), which focused on improving language and literacy. The EEF project involved an RCT with 649 Year 7 pupils in 12 schools in Bolton over 2013-14, and did not find an impact on reading attainment. The WWCSC re-analysis indicated that it has potential for CSC-experienced participants specifically.

Following the VEIP project, Bolton Local Authority continued its commitment to improving language and literacy outcomes for all students, most recently culminating in the piloting and scaling of the Words for All (WFA) programme in 2018/19. The Words for All programme built on the key learning from the VEIP EEF trials in 2013/14, while incorporating additional synthesised evidence emerging since the trial, which is highly aligned to original VEIP. The Words for All programme also builds on wider emerging learning and evidence around effective CPD and implementation of evidence based practices, including incorporating findings from the DfE's own research on DfE Standards for Professional Learning that impact on young people.

Specifically, Words for All builds on certain aspects of the original VEIP intervention including a) allowing teachers flexibility in choosing aspects of interventions; and b) pupils being allowed to learn about vocabulary for subject areas other than English, which was an explicit part of the VEIP programme. It also addresses some of the key formative learning points from the original EEF trial in 2013/14, including addressing issues with planning and preparation time and being applicable for teachers with a lack of English specialism. This trial tests the impact of the WFA approach on the learning outcomes of CSC-experienced children.

[1] Department for Education (December 2018). Help, protection, education: concluding the Children in Need review. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/809236/190614_CHILDREN_IN_NEED_PUBLICATION_FINAL.pdf

[2] Sanders et al (2019) What works in education for children that have social workers?: summary report. Avaliable at: https://whatworks-csc.org.uk/wp-content/uploads/WWCSC_what_works_education_children_SWs_Feb20.pdf

[3] 'Studies with positive effects for young people who have had a social worker which are larger than the effect on young people without a social worker, and where there are no substantial complications to the sub- group analysis' (Sanders et al., 2019).

Intervention and Theory of Change

Words for All

The Words for All (WFA) programme is underpinned by four key components:

- Evidence based language acquisition practice – Program delivery informed by research about academic language acquisition and interventions that promote this.
- Focus on the social and emotional barriers – Program delivery informed by the social and emotional barriers to improving language acquisition.
- Evidence based CPD - Taking an evidence-based approach to teacher professional learning and practice development.
- Supporting sustainable change – Focus on how to successfully plan for and implement sustainable change.

Whole Education will work with groups of three teachers (known as a 'triad') in schools to design a version of the Words for All programme best suited for their school, meaning that the intervention under evaluation is the establishment of a triad of teachers and staff within each school and the resources and training are provided by Whole Education. Decisions about the nature of the programme delivery will be made by teachers. The interventions developed by teachers will be in the range of the following:

- Reading out loud – treatment group students will be withdrawn from class to participate in a session where the school librarian reads out loud extended texts to students (e.g. a novel), rich with complex vocabulary, to expose students to new words and meanings.
- Academic Vocabulary Building – treatment group students will be withdrawn from class at the beginning of the week to be pre-taught pre and parallel instruction of vocabulary to support students to access new curriculum content, including exploring etymological roots to aid comprehension and application.
- Academic non-fiction - Pre and parallel teaching of high-quality non-fiction texts to support curriculum units, to develop wider context and cultural understanding and to develop strategies for learning new vocabulary students encounter.
- Talking about learning - Support for students to articulate ideas and spoken expression through e.g. structured questioning to develop reading comprehension and to extend spoken vocabulary. Intentional and planned for opportunities to talk.
- Reading for pleasure - Explicit support and mentoring for students to find more enjoyment in reading by making informed recommendations (based on knowledge of age appropriate literature) engaging in 'book talk' and creating space for independent reading time within a highly social reading environment.

As can be seen from the above, interventions will involve targeted work with CSC-experienced pupils, either by withdrawing them from class in groups, or via one-to-one support. The risk of spillover effects from the treated to control pupils is considered minimal because of the highly targeted nature of the intervention to a small subset of the large school populations. Components of the intervention are at the teacher level, and an explicit longer-term goal of WFA's is to embed cultural change around using evidence to support students with lower levels of vocabulary development. However, the specificity of the interventions, size of the schools involved and the time horizon of the evaluation mean that we do not expect there to be substantial spillovers between treated and control pupils during the intervention period. The logic model for Words for All is provided in Figure 1 overleaf.

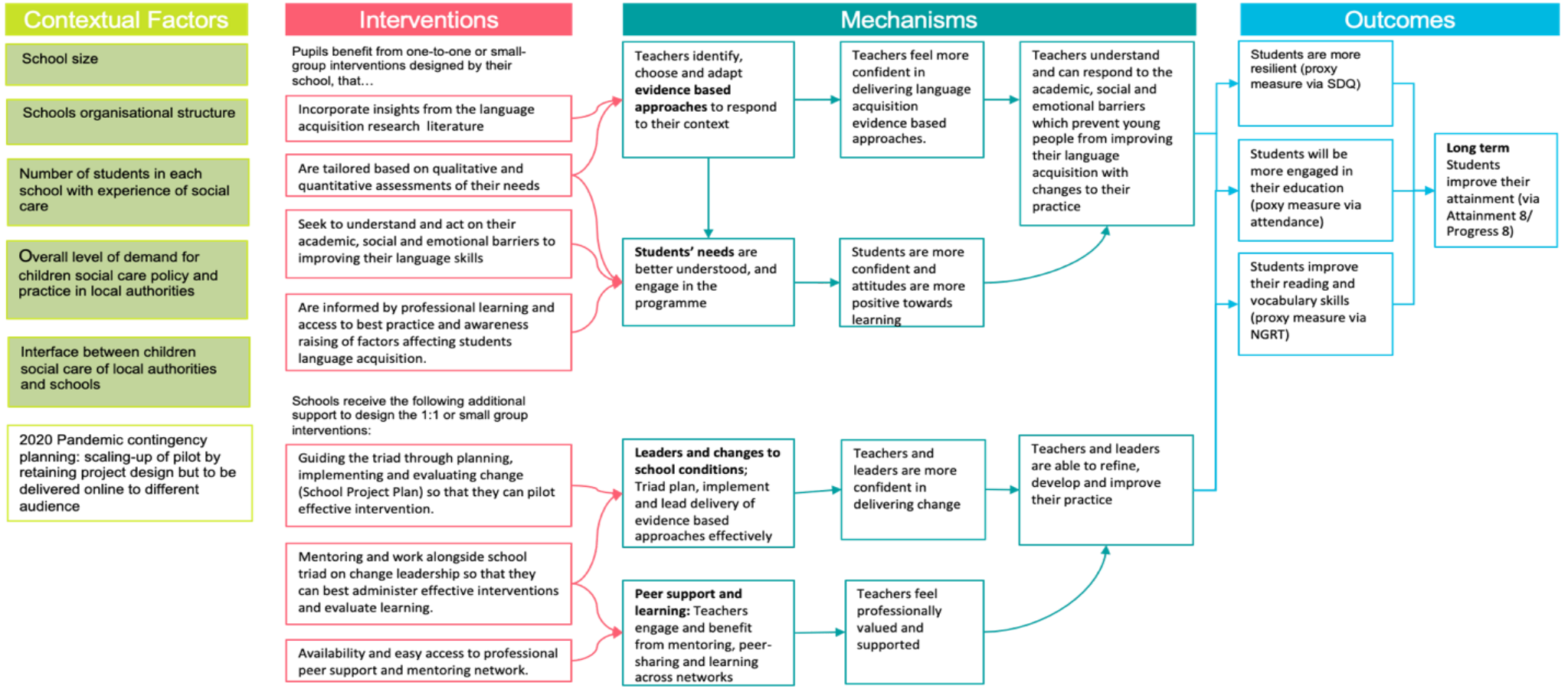
What Works for Children's Social Care

Trial Evaluation Protocol
Words for All

Evaluator (institution): King's College London

Principal investigator(s): Dr Mary Baginsky

Figure 1: Logic Model for Words for All



Situation & problem statement: Children and young people with a social worker have worse educational outcomes than the student population at large. Enhancing literacy is important for improving educational attainment for children; particularly the project's target group of children with a social worker. The Words for All (WFA) project provided by Whole Education (WE) works with secondary schools to deliver vocabulary and reading enrichment interventions to improve student literacy and academic outcomes, with a particular focus on enhancing vocabulary. The project will work with groups of three teachers (triads) in schools to design a version of the programme best suited for their school, meaning that the intervention held constant is the triad of teachers and staff within each school and the resources and training provided by WE.

NB: This logic model focuses on how Words for All may influence the outcomes; however there are many other factors that also influence the outcomes that are not included in the model.

Impact Evaluation

Research questions

The primary research question focuses on the impact of the Words for All programme on the target group:

- What is the impact of Words for All on the average reading skills as measured by the New Group Reading Test for Year 7 to 11 pupils who have been considered to be Children in Need, subject to a Child Protection Plan, and/or Looked After in the past six years?

The secondary research questions are as follows:

- What is the impact of Words for All on GCSE attainment (as measured by 1. Attainment-8, and 2. English Language GCSE grade) for Year 11 pupils who have been considered to be Children in Need, subject to a Child Protection Plan, and/or Looked After in the past six years?
- What is the impact of Words for All on social and emotional skills (as measured by the Strengths and Difficulties Questionnaire) for Year 7 to 11 pupils who have been considered to be Children in Need, subject to a Child Protection Plan, and/or Looked After in the past six years?
- What is the impact of Words for All on number of sessions absent for Year 7 to 11 pupils who have been considered to be Children in Need, subject to a Child Protection Plan, and/or Looked After in the past six years?

Design

Trial type and number of arms		Two-arm randomised controlled trial
Unit of randomisation		Family
Stratification variables (if applicable)		School
Primary outcome	variable	New Group Reading Test (NGRT) Standard Age Score (SAS)
	measure (instrument, scale)	NGRT scoring returns the SAS based on the student's raw score which has been adjusted for

		both age and the difficulty of the test taken and placed on a scale that makes a comparison with a nationally representative sample of students of the same age across the UK and a Stanine which places the student's SAS score on a scale of 1 (low) to 9 (high) and offers a broad overview of their performance.
Secondary outcome 1	variable(s)	GCSE Attainment
	measure(s) (instrument, scale)	Attainment-8 is a score between 1 and 9 representing a student's average attainment across a 'well-balanced' set of GSCE exams including mathematics (double weighted), English Language (double weighted if English Literature is also taken), three further qualifications that are part of the English Baccalaureate (EBacc) and three further qualifications that can be GCSE qualifications or other non-GCSE qualifications on the DfE approved list
Secondary outcome 2	variable(s)	GSCE English Language grade
	measure(s) (instrument, scale)	This will be a continuous variable taking a value of between 1 and 9 representing the student's grade in GCSE English Language only
Secondary outcome 3	variable(s)	SDQ Total Difficulties Score
	measure(s) (instrument, scale)	The Strengths and Difficulties Questionnaire (Goodman et al. 1998)
Secondary outcome 4	variable(s)	Attendance
	measure(s) (instrument, scale)	Number of sessions of unauthorised absence in AY2020-21, taken from the National Pupil Database (NPD). This will be calculated as the sum of the instances recorded across Autumn, Spring and Summer term

Randomisation

Randomisation will occur at the level of family-group, stratified by school. This is to avoid spill-overs between siblings, who are very likely to influence each other's outcomes. Family group will be identified by the school and noted using a family identifier in the randomisation spreadsheet provided to WW-CSC. The specifics of how schools identify sibling groups may vary, so these will be documented as part of the evaluation.

There is a risk of spillovers between treatment and control pupils, given treated teachers may use their W4All training to help those who have not been allocated treatment in other classes. We plan to mitigate this issue by getting teachers participating in W4All to agree to isolate their knowledge to only family-groups who have been allocated treatment.

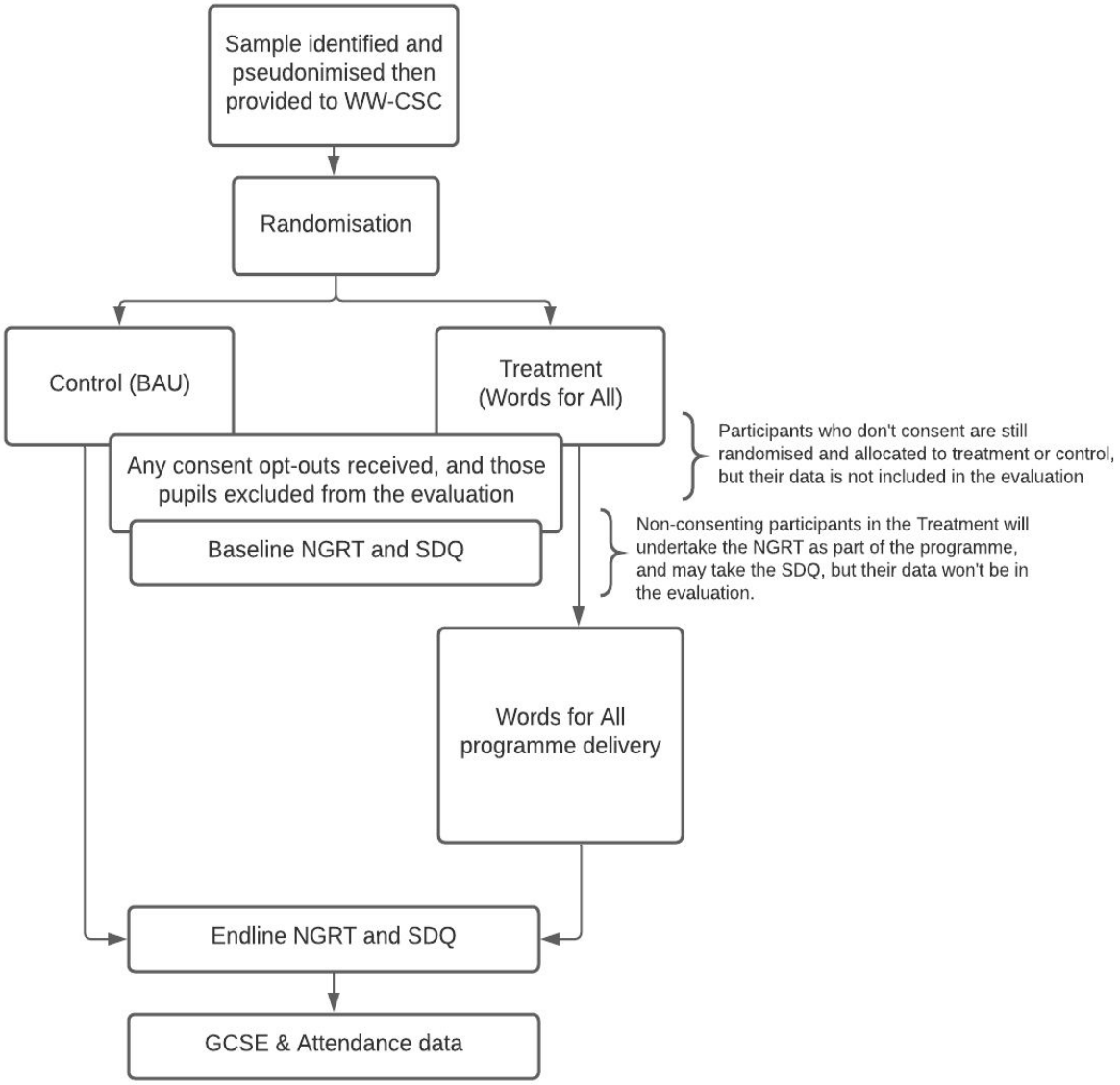
Balance checks on observables will also be conducted to confirm randomisation was successful by ensuring there are no concerning differences (i.e., of more than 5pp) on gender and year group composition, type of CSC-experience between the treatment and control groups. Randomisation will be conducted by WW-CSC.

Participants

Words for All will be delivered across Key Stage 3 and Key Stage 4 (Years 7 to 11). All pupils who are identified by the Virtual School, Local Authority or school as having had a social worker in the last six years in these year groups are in scope. Pupils and parents/guardians will be provided with Information Sheets and given the opportunity to opt out of the evaluation.

Those who opt out will be included in the randomisation, and, if allocated to treatment, will be enrolled onto the Words for All programme. However, they will be excluded from the following evaluation data collection and analysis. Figure 2 below gives the trial flow.

Figure 2: Trial flow diagram for Words for All



There may be instances where randomisation results in a single pupil from a year group being allocated to the treatment. If schools consider this arrangement unworkable, then they will be able to select a 'buddy' from out of sample to include in the intervention. These out of sample buddies will not be part of the trial, even though they are part of the intervention.

Sample size / MDES calculations

Calculations to determine minimum detectable effect size have been conducted in R using the pwr package (see Appendix). We find that 2,400 participants across intervention and control group should be sufficient to detect a small effect of 0.06 on our primary outcome, adjusting for baseline scores (0.12 unadjusted) with an alpha of 0.05 and 80% power. As the primary outcome is manually-collected, anticipate some attrition at both the school and pupil level. If attrition is above 5%, we would seek to impute outcomes for missing cases, as per the section on Missing Data, below.

		MDES (Proportion of a Standard Deviation)
MDES		0.06
Baseline/Endline correlations	Child	0.7[4]
	Family	NA
	Social Worker	NA
Intracluster correlations (ICCs)	Family	0.4[5]
	Social Worker	NA
	Team	NA
Alpha		0.05
Power		0.8
One-sided or two-sided?		Two-sided
Level of intervention clustering		Family
Average cluster size		1.3[6]
Sample Size (children)	Intervention	1,200 participants
	Control	1,200 participants
	Total	2,400 participants

[4] Based on https://educationendowmentfoundation.org.uk/public/files/Projects/Round_10_-_Reciprocal_Reading_SAP.pdf and https://www.gl-assessment.co.uk/sites/gl/files/images/Baseline-progress-TI.pdf

[5] Based on https://www.understanding-inequalities.ac.uk/types-of-inequality/disentangling-the-importance-of-individual-and-family-factors-on-educational-and

[6] The average number of siblings per family is 1.9, but cluster size is likely to be smaller than that given that only siblings in years 7-11 are in scope.

Outcome measures

Primary outcome

The Primary Outcome will be Standard Age Score (SAS), as measured by the New Group Reading Test (NGRT). The NGRT is a standardised reading and comprehension assessment split into three sections assessing phonics, sentence completion, and passage comprehension and was used in the original EEF evaluation as the primary outcome.

The SAS is based on the student's raw score which has been adjusted for both age and the difficulty of the test taken and placed on a scale that makes a comparison with a nationally representative sample of students of the same age across the UK, and a Stanine which places the student's SAS score on a scale of 1 (low) to 9 (high) to offer a broad overview of their performance.

Beyond the fact that using the NGRT will allow for some comparison with the previous EEF trial, the NGRT is a good tool for measuring reading skill because of its widespread use, the capacity to deliver it digitally, because of the relatively short time it takes to administer (about 20 minutes), minimising the administrative burden and testing fatigue.

The NGRT is planned to be administered by schools at three time-points to treatment-group pupils as part of the intervention. However, to ensure it is conducted blind to treatment assignment, invigilation of the actual tests at baseline and endline, for treatment and control pupils, will be conducted by RAs trained by EDIT. Only data from the first and third tests are included in the analysis, as baseline control and primary outcome variables respectively. The second test is taken as part of the WFA programme by some but not all of the cohort, and is not included as part of the analysis.

Secondary outcomes

GCSE attainment

The age-range in scope encompasses GCSEs, so we will take as a secondary outcome GSCE attainment specifically for the Year 11 subgroup as measured in a) Attainment-8 scores and b) grade in English Language GCSE only. These will form useful secondary measures of the impact of WFA and will require no additional administrative burden as it is already a requirement for Year 11 students.

The Attainment-8 is a score is intended to measure a student's performance across a 'well-balanced' set of GSCE exams including mathematics (double weighted), English Language (double weighted if English Literature is also taken), three further qualifications that are part of the English Baccalaureate (EBacc) and three further qualifications that can be GCSE qualifications or any other non-GCSE qualifications on the DfE approved list. GSCE scores will be collected from the National Pupil Database under the same permissions that allow us to collect attendance data. The English Language GCSE score will be a single score from 1 – 9 for that pupil.

Strengths and Difficulties questionnaire

We will also collect the self-directed Strengths and Difficulties Questionnaire scores (Goodman et al. 1998) in parallel to the first and last administrations of the NGRT. The SDQ is a commonly-used behavioural screening questionnaire for children and young people. It

has 20 questions divided into five scales including emotional symptoms, conduct problems, hyperactivity/inattention, and peer relationships problems.

Some students in this sample (those who are looked after by the Local Authority) receive the SDQ test from local authorities regularly; where the Local Authority advises that the SDQ has been run recently (in the past 7 days) with an in-scope pupil we will aim to utilise the existing test results in order to avoid over-burdening pupils with assessment.

Attendance data

We will use National Pupil Database data to measure total number of unauthorised absences in AY2020-21, operationalised as the sum of unauthorised sessions (half days) absent across the Autumn, Spring and Summer terms., using the following NPD variables:

- Absence.UnauthorisedAbsence_Autumn_ab20,
- Absence.UnauthorisedAbsence_Spring_ab20,
- Absence.UnauthorisedAbsence_Summer_ab20.

Analysis plan

Primary Analysis:

Analytical Specification

The primary outcome will be analysed using an Ordinary Least Squares (OLS) regression with the following specification: We propose to power the trial for a small effect size of 0.07 on our primary outcome, adjusting for baseline scores (0.12 unadjusted).

$$Y_{ic} = \alpha + \beta_1 D_c + \beta_2 A_i + \beta_{3:11} X'_i + \beta_{12} S_i + \varepsilon_{ic}$$

Where:

- Y_{ic} is individual i in family c's Standardised Age Score (SAS) at endline measurement;
- D_c is the treatment assignment, clustered at the level of the family/sibling group;
- A_i is individual i's SAS at baseline measurement;
- X_i is a vector of x characteristics of individual i, comprising gender (coded as male, female, or other/not specified, provided by schools), year group (coded as Year 7, Year 8, Year 9, Year 10 or Year 11 and provided by schools), and most recent type of CSC intervention (coded as Child in Need, Child Protection Plan, or Looked After, with Child in Needs as the reference category, accessed via the NPD);
- S_i is a school fixed effect; and
- ε_{ic} is a cluster-robust standard error term, clustered at the level of the family.

Effect sizes will be reported as a proportion of the standard deviation of the outcome in the control group (Glass's delta). No adjustment for multiple comparisons will be used, as there is one primary outcome.

Secondary Analysis

As there are four secondary outcomes, the Hochberg step up procedure will be used to adjust for multiple comparisons, as per the WW-CSC statistical guidance.

Secondary outcomes

For analysis of the GCSE Attainment-8 and GCSE English Language grade secondary outcomes, the analysis will use two OLS regressions with the following form:

$$Y_{ic} = \alpha + \beta_1 D_c + \beta_2 A_i + \beta_{3:11} X'_i + \beta_{12} S_i + \varepsilon_{ic}$$

Where:

- Y_{ic} is individual i in cluster c's attainment on the measure;
- A_i is individual i's attainment on the Key Stage 2 SATs, accessed via the NPD. For analysis of GCSE English Language grade, KS2 English only will be used.
- All other terms are as specified in primary analysis.

For the SDQ Total Difficulties Score (TDS), the analysis will again use an OLS regression, with the following form:

$$Y_{ic} = \alpha + \beta_1 D_c + \beta_2 A_i + \beta_{3:11} X'_i + \beta_{12} S_i + \varepsilon_{ic}$$

Where:

- Y_{ic} is individual i in cluster c's TDS at endline;
- A_i is individual i's TDS at baseline.
- All other terms are as specified in primary analysis.

Finally, for attendance, the analysis will use a negative binomial regression model (NBMR). This takes the form of a quasi-poisson, log-linear regression but accounts for overdispersion in the dependent variable (i.e. unauthorised absences) which crucially allows variance to exceed the models predicted value of expected absences, λ_{ict}. This improves on a standard OLS as we expect unauthorised absences to be non-linear and have a one-tailed cluster around zero, closely following a NBRM distribution which better reflects a count variable and the repeated zeros in our outcome we expect to see. The NBRM takes the following form:

$$Ln(\lambda_{ict}) = \alpha + \beta_1 D_c + \beta_{3:11} X'_i + \beta_{12} S_i$$

Where:

- $Var(Y_{ict}) = \phi\lambda_{ict}$
- $\lambda_{ict} = E[Y_{ict}|X'_i]$ refers to unauthorised absences predicted by the model conditionally based on the set of X'_i covariates for a given individual-i clustered within a family c's number of sessions of unauthorised absence in a given unit of time exposure t. This model prediction is defined below.
- D_c is the treatment assignment, clustered at the level of the family/sibling group;
- Other terms are as specified in primary analysis

Here, λ_{ict} is defined as per Allison & Waterman (2002)[7]. Within this definition of our predictor term is a parametrisation of conditional overdispersion, ϕ. A standard Poisson model is one that constrains this value of ϕ to zero. NBRM allows $\phi > 0$ and allows for overdispersion (i.e. variance greater than the mean). We will first compute a one-tailed likelihood ratio test that ϕ equals to zero and compare the chi-squared value to test if the NBRM is indeed a

[7] Available at: https://www.sas.upenn.edu/~allison/FENB.pdf

better fit of the data than the Poisson model. If, unexpectedly, ϕ equals zero we will instead use a standard Poisson model.

No baseline measure of attendance will be used, owing to the volatile nature of attendance in 2019-20 and likely 2020-21 given the ongoing COVID-19 pandemic. Also, the NBRM naturally generates the expected attendance for an individual based upon their set of covariates in a given time period, and then compares their actual observed attendance to this expected value, so our model is comparatively more robust to limited baseline data than an OLS. Beneficially, our t subscript means our time exposure under consideration can vary for each clustered individual and our model will predict their expected mean attendance within that given timeframe. This is useful to compare outcomes between individuals who had varying degrees of COVID-19 disruption and potentially shorter or reduced academic years.

Robustness checks

Schools have self-identified pupils that are in scope so there is a slight chance that pupils who have been included in the study will be classed as out-of-scope when their NPD record is accessed. These pupils will be excluded from the analysis, but we will rerun primary and secondary analysis with all pupils for whom we have data (including pupils whose NPD records suggest they should not have been nominated by schools as being in scope).

In addition, there may be a small number of pupils who received the intervention 'buddied' with an out-of-sample pupil (see 'Participants' for reasons). 'Buddy' pupils will not be in our sample, but there is a minor risk that the treatment effect for pupils who were 'buddied' in this way may be different as a result of their receiving the intervention in conjunction with a pupil who was not within the target group for the intervention. We will therefore re-run all analyses excluding pupils who were buddied with a pupil from out of sample.

Missing Data

It is very likely that there will be missing data in the baseline and endline NGRT scores where pupils were absent on the day the test was administered, or in some cases where they have transferred schools or left the Local Authority area. Where baseline data is missing, this will be imputed using multiple imputation, in order to preserve power. For other covariates, which are all factor variables, missingness will be coded as a separate factor level.

Where endline data is missing, particularly for the NGRT SAS, if missingness is uncorrelated with treatment assignment, and attrition is <5%, complete case analysis will be used. If attrition is greater than or equal to 5%, Multiple Imputation by Chained Equations within the experimental group will be used to impute the primary outcome, using available outcomes (e.g. attendance) and covariates. If missingness is correlated with treatment assignment, then it is Missing Not At Random, then no form of imputation is able to adjust for this. Complete case analysis will be used, with discussion of the implications of the missingness for interpreting treatment estimates.

Exploratory Analysis

Variations in treatment effect by type and relative duration of CSC experience

We are interested in the extent to which the intervention impacts participants with different types of CSC experience differently. In this, we are interested both in whether the time since last CSC experience (i.e. those currently subject to a particular CSC intervention versus experiences in the more recent or more distant past) affects the impacts of Words for All; and in whether the specific CSC intervention affects the impacts.

For the first of these purposes, the sample will be split into three sub-samples, with pupils assigned to the sub-sample representing their current or most recent level of CSC experience (Child in Need, Child Protection Plan, or looked after), I'_i. This sub-sample will also be categorised by time to see whether those with more recent CSC experiences responded differently to the programme. This is reflected in the ordinal time variable T_i. Measuring time ordinally across the three categories (I.e., that the three categories can be ranked in a sequential order) will allow us to more confidently test the overall trend and whether time can predict our outcomes. We can then start to assess whether T_i is a factor that varies the general effectiveness of W4All compared to using individual dummy variables which, if significant, only provide explanation for that dummy's given time period. Our approach instead means we can estimate marginal effects and suggest whether, over time more generally, programme effects vary for those who are CSC experienced from an increasing amount of time ago. Firstly, including both type and time since CSC intervention explicitly will allow us to test if treatment effects differ across the time since CSC involvement, holding constant differences across the type of CSC.

We will use an interaction term between treatment allocation and time to do this and thus the specification is as follows:

$$Y_{ic} = \alpha + \beta_1 D_c + \beta_2 A_i + \beta_3 T_i + \beta_4 (D_c * T_i) + \beta_{5,6} I'_i + \beta_{7:13} X'_i + B_{14} S_i + \varepsilon_{ic}$$

Where:

- I'_i is a vector of two dummy variables indicating whether individual i was recorded as experiencing either of the other two forms of CSC intervention.
- T_i is the time since individual i experienced that given form of CSC, coded numerically as current experience = 3, experience in the past year = 2, or experience more than a year ago = 1.
- Other terms are as specified in primary analysis, with X'_i omitting the CSC indicators, which are now incorporated into I'_i.

Secondly, we are then able to adjust our regression to test whether treatment effects varied across the type of CSC an individual has experienced, holding constant the time since they last experienced that given form of CSC. This is important as it allows us to disentangle underlying causes and identify who is benefitting the most from W4All by isolating which variation in our treatment effect is coming from either recentness or type of CSC.

Therefore, we will conduct further exploratory analysis partitioning the sample into those who have current or very recent (within the last year) experience of any form of CSC intervention and interact this term with the cluster treatment assignment, as per the below:

$$Y_{ic} = \alpha + \beta_1 D_c + \beta_2 A_i + \beta_{3,4} I'_i + \beta_{5,6} (D_c * I'_i) + \beta_{7:13} X'_i + \beta_{14} T_i + B_{15} S_i + \varepsilon_{ic}$$

Where terms are as specified previously.

In addition, we may conduct further exploratory analysis using causal forests (Wager and Athey 2018)[8] to explore which variables contribute most to treatment effects, and if there are any heterogeneous treatment effects by CSC type and duration, if sample size permits.

Variations in treatment effect by type of activities chosen by triad

A major component of the design of the Words for All intervention is that triads of teachers within each school will work to design a bespoke programme for their school's and individual pupils' needs. We therefore propose to run secondary analysis splitting the treatment variable into multiple dummies depending on the characteristics of support the school has chosen. It is not currently possible to specify these as they will need to be developed from information collected as part of the IPE. Appropriate categories and criteria for this variable will be determined by the process evaluation lead and levels assigned to schools before impact data is collated or analysed. It is not currently possible to define how this indicator will be categorised as this depends on implementation decisions not yet taken. We will then compare each treatment, both against the control, and against each other.

Impact on attainment in English in Years 7 - 10

As this project will randomise within-school, it may also be possible, in a subset of schools, to use school attainment records to explore improvements in attainment for pupils outside Year 11. The extent of this analysis will depend on the quality and comparability of attainment data available from schools.

If data is available, it will be analysed using the specification outlined for the primary analysis, with

- Y_{ic} being individual i's attainment in school-based English assessment(s) in 2020-21 and
- A_i being a measure of prior attainment.

COVID-19 introduces some complexity into the selection of a prior attainment measure as attainment in 2019-20 is likely to have been shocked in different ways for different pupils and to therefore not represent a good baseline. Further, for 2020-21 Year 7s, their prior attainment would have been in Primary School (or in the SATs, which were cancelled in 2019-20) and therefore held by a different school, and not disaggregated by subject. Accordingly, the exploratory analysis will focus on 2020-21 Year 9s and 10s, using their attainment in 2018-19 (when they were in Year 7 or Year 8) as the prior attainment measure.

Analysis of Harms

There are a number of potential harms that could arise from this intervention. These include making CSC-experienced pupils more visible to themselves, each other and the school community, which could result in pupils feeling singled out or being treated differently. There is also the potential that being withdrawn from class (if this is part of the programme in that school) could result in students missing important content. If these impacts feed through to the attainment, attendance or the SDQ, then they will be identified via the analysis specified above. In addition, the process evaluation will explore possible harms arising through qualitative work with pupils, teachers and social workers.

[8] Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. Journal of the American Statistical Association, 113(523), 1228-1242.

Contextual Factors Analysis

A number of contextual factors may be relevant to the effectiveness of Words for All. We control for school-level factors by including a school-level fixed effect. To the extent that different schools implement differently, we consider this in the exploratory analysis outlined above. There may be elements of the school environment in particular that affect effectiveness of the intervention; for example, school size, school organisational structure, number of pupils in the school with CSC-experience. There may also be contextual factors at the local authority level, such as overall levels of demand for CSC, CSC policy and practice, and how well the CSC-school interface functions.
We will explore the possibility of analysing the impact of these contextual factors on the effectiveness of the intervention. However, we expect that this will be mainly considered as part of the process evaluation, rather than the analytical strategy.

Implementation and process evaluation

Aims

The implementation and process evaluation (IPE) has the following aims:

- To evaluate the factors that facilitate and hinder the implementation of Words for All, including the extent to which fidelity is maintained and, indeed, what constitutes fidelity.
- To assess the confidence of teachers and other school-based professionals in the programme.
- To understand the response of students to the programme.

Research Questions

The research questions are provided below.

Fidelity

1. What are the expectations placed on schools involved in the Words for All programme?
2. What are schools' understanding of the essential components of the programme?
3. How have schools:
    a. Adhered to the essential components of the programme?
    b. Supplemented them?
4. What factors have hindered and/or facilitated the implementation of the programme?

Staff Experience

5. To what extent do staff feel that their participation on the programme has improved their skills and confidence in identifying and responding appropriately to the needs of students?
6. To what extent do staff perceive the programme to have been effective in addressing the needs of students?

Pupil experience

7. What are the experiences of students who have participated in the programme and what, if any, benefits did they perceive from it?

Research question(s)	Indicator(s)	Method
What are the expectations placed on schools involved in the Words for All programme?	• Expectations on schools outlined in the initial onboarding • Availability and access to expert and peer support for school triads • Provision of easy-to-understand data and evidence to schools on effective interventions	• Observation of launch event • Pre- and post-intervention surveys to all participating schools • Observation of hub meetings (subject to practicalities) • Attendance at a sample of 'expert seminars'
What are schools understanding of the essential components of the programme?	• Development of triad arrangements in schools[9] • Adherence to the key elements of the programme (to be developed)	• Pre- and post-intervention surveys to all participating schools • Observation of hub meetings (subject to practicalities) • Attendance at a sample of 'expert seminars' • Examination of project plans in 10 case study schools • Interviews in 10 case study schools with key members of staff (Triads and others as appropriate)
How have schools a) adhered to the essential components of the programme? b) supplemented them?	• Use of evidence-based interventions to improve teaching	• Pre- and post-intervention surveys to all participating schools • Interviews with triads in 10 case study schools

[9] Triads are made up of three school-based staff supported by Whole Education to lead the development and implementation of the programme within schools

	• Triad engagement in networks and peer information sharing • Training and support in place across school • Tracking and monitoring of teaching practice • Additional curricular materials • Whole school initiatives such as use of library, whole form / year / school activities	• Examination of teacher logs in 10 case study schools • Observation of use of intervention in 10 case study schools • Interviews with trained teachers and pastoral staff in 10 case study schools (not triad member) • Interviews with LA lead virtual heads linked to case study schools
What factors have hindered and/or facilitated the implementation of the programme?	• Time allocated to triad members for implementation of the programme • Time allocated for staff training within school • Tracking and monitoring systems for student achievement and teaching quality within schools p • Support and engagement of LA leads and virtual heads	• Pre- and post-intervention surveys to all participating schools • Interviews with triads in 10 case study schools • Interviews with non-triad trained teachers and pastoral staff in 10 case study schools (not triad member) • Examination of teacher logs in 10 case study schools • Interviews with LA lead virtual heads linked to case study schools

Design and Methods

It is important to note that Whole Education will work with individual schools to develop the most appropriate approach to support staff and students in each school. In this sense Words for All is a fluid rather than a one size fits all programme. Thus, only when the evaluation team gains an understanding of the profile of students enrolled on the programme **and** details of the approaches adopted in schools to support them, will we be able to scope the stages in more detail.

Observations

- Observation of the launch will allow the evaluators to assess the explanation of purpose and expectations provided to participants in the programme.
- Observation of the regional hub training session will allow an initial assessment to be made of sufficiency of training and appropriateness of methods within constraints of virtual delivery.
- Observation of expert webinars will provide details of how implementation is progressing alongside the expectations of the delivery partner in relation to specific aspects of the programme.
- Observations of elements of the intervention being applied by teachers in the case study schools.

Detailed schedules for all the observations will be developed to ensure that data are collected consistently. They will be used alongside field notes on the observations to provide more detailed descriptions of practice and to help contextualise findings.

Surveys

The pre- and post-intervention survey will be directed to the lead member of each triad in participating schools (anticipated to be 79) with instructions that it should be completed as a group exercise, with the option to provide individuals' responses as appropriate. The surveys will be conducted using the Online Surveys platform for which King's College London has the required license and they will be designed to take between 15 and 20 minutes to complete, although this could be longer if respondents choose to provide additional information.

The development of both surveys will be informed by the observations detailed below and through discussions with Whole Education. In addition to questions arising from those activities the pre-intervention survey will collect details on:

- The number of students receiving the intervention
- Prior use of programmes to support literacy and vocabulary acquisition
- Reasons for participating in and expectations of the programme
- Experience of and assessment of sufficiency of the training provided
- Planned activities linked to the programme, that will be introduced
- Prior use of similar schemes
- Confidence level (on a scale) in relation to teaching literacy skills
- Confidence level (on a scale) in relation to responding to students' literacy needs
- Confidence level (on scales) in relation to implementing and embedding the Words for All programme.

In the pre-intervention survey schools will also be asked to state if they would be willing to be considered as a case study school. And a brief description of what this would involve will be provided.

The post-survey participation survey for triads will also include generic and bespoke questions covering:

- Description of the overall programme in place in the schools
- Experience and sufficiency of training

- An assessment of support available on the elements of the programme and on wider aspects of managing change
- Confidence level (on a scale) in relation to teaching literacy skills
- Confidence level (on a scale) in relation to responding to students' literacy needs
- Confidence level (on scales) in relation to implementing and embedding the Words for All programme
- Assessment of fidelity achieved (benchmarked against descriptions provided at pre-intervention stage)
- Assessment of the usefulness of the intervention, outcomes and, where appropriate, sustainability
- The extent to which the intervention group and the control group have been separated and what additional support (if any) students in the control group have received

All relevant staff (teaching and non-teaching) will be asked to complete a pre- and post-intervention rating of their awareness of the programme and their confidence in responding to literacy needs of students.

The draft survey instruments will be shared with Whole Education to confirm the areas covered and to address any omissions, but the final decision on the contents of the instrument will remain with the evaluators. The evaluation team will discuss with Whole Education the possibility of trialling the survey instruments with schools that participated in the pilot programme. The instruments will be redrafted as necessary before being sent out to all schools. This will allow the evaluators to test that instructions and questions are clear and the intended meanings are understood. The data from these schools will be absorbed into the main data set with any additional or amended data collected as necessary.

Case Studies

In the pre-intervention surveys schools will be asked if they might be willing to participate in the case study phase of this research. We anticipate that two schools will be randomly selected from 5 different local authority areas participating in the programme. The 5 local authority areas will be selected (from a total of 15 participating) to ensure a good representation of regional and socio-economic differences.

Information sheets will be shared with staff before interviews take place and staff will also be required to sign a consent form. Before children are invited to participate in the group discussions, consent will be sought from parents. Information sheets and consent forms will be sent to parents in line with school systems (we anticipate that in some schools this will be through email or an online portal). Only after consent has been granted will information sheets be shared with students to invite them to participate. Students who do choose to participate will also be required to complete a consent form[10].

Visits to the school will take place during the summer term of 2021, COVID-19 restrictions allowing. If this proved impossible the fieldwork would be adapted for use virtually. Two days will be spent in each of the 10 schools / authority during which time the following will take place:

- Group interviews with the triads to a) explore in more detail their experience of implementing and developing the programme and b) understand how they came to

[10] More information regarding information sheets and consent forms can be found in the ethics application.

understand the issues aligned with their practice and what they might need to do differently in future. This will include discussion relating to potential 'spill-overs' between the intervention and control groups.

- Interviews with any teachers involved in the programme to capture their feedback on any impact of the programme on:
    - Their understanding of how literacy is supported
    - Their teaching style and content
    - The students
- Interviews with other key members of staff such as Head of English, Head of Special Needs (or equivalent) and Head of Pastoral (or equivalent) if not involved in the triad to collect their views on its implementation, efficacy and sustainability.
- Group discussions with students who have participated in the programme[11] – the number of groups that are held will, in part, be determined by ages of those involved, but the discussion will cover students' views on participation as well as exploring their assessment of any impact on their reading habits in general and their ability to access curriculum materials and activities.
- Interviews will be held with the lead person for the programme to assess the sustainability and scalability of the programme.
- Interviews with the local authority virtual headteacher, a social worker[12] and an educational psychologist to explore their awareness of and role in the programme, as well as any emerging views both on the way the programme has developed and its future use.

The details of each interview schedule will be determined at a later stage.

The school project plan (or equivalent) will be examined to identify how the project has been implemented and managed in line with expectations and how any adaptations have been applied, including to meet the range of needs and abilities of students participating in the intervention. The teacher logs kept by teachers involved in delivering the programme will be examined to calculate the intervention dosage, as well as the methods used for delivery. The logs will also inform the assessment of fidelity.

Sample Size

Data collection method	Sample size	Collection time
Observations	• 1 national launch event and sample of 6 of the 12 regional hub initial training sessions (Sample to be determined when participating LAs known) • 6 regional hub meetings	October – November 2020

[11] The lead researcher Dr Mary Baginsky has extensive experience of conducting group discussions involving children.

[12] Whole Education have not yet confirmed if social workers will be included as part of the programme or other professionals with a caring responsibility.

	• A sample of expert webinars (number to be determined when details are available) • Observations of elements of the intervention being applied in case study schools	Spring term 2021 Expert webinars dates tbc
Surveys	• Pre-intervention survey to all schools (71 recruited so far) • Post-intervention survey to all schools	December 2020- January 2021 May – June 2021
Interviews/ group discussions	Interviews and group discussions at 10 case study schools. To include: • Interviews with triads • Interviews with a sample of staff trained within school (number to be determined after discussions with schools) • Interviews with key members of staff such as Head of English, Head of Special Education and Pastoral Head • Group discussions with participating students • Interviews with LA leads, social workers and Educational Psychologists	May-June 2021

Analysis

Qualitative Data

Detailed notes will be taken as part of the review of the on-boarding materials and the observations of: (1) the launch event; (2) regional hub training events; (3) expert webinars; and (4) elements of the intervention in case study schools. All interviews and group discussions with school-based staff, students, LA leads, social workers and educational psychologists as part of the case studies will be recorded (with permission) and transcribed in full.

This body of qualitative data will be analysed using a reflexive thematic approach, aided by NVivo software designed for this purpose. This approach is particularly suited to exploring people's experiences, views and perceptions (Braun and Clarke, 2006). Patterns are

identified through data familiarisation, data coding, theme development and revision. This aids the identification of patterns of meaning in the data which can address the research questions (Patton, 2015). Quality assurance of this analysis will be achieved by two researchers analysing the data separately and then discussing the emerging themes.

Survey Data

Data from the pre- and post-intervention surveys will be transferred from the Online Survey platform to SPSS to produce descriptive statistics to summarise the data.

Cost evaluation

An analysis of the incremental resource implications of the project will be conducted focussing specifically on the cost of teaching time within schools allocated to programme, training for teachers, any other activities undertaken in implementing the programme and the delivery of the programme itself. Costs relating to programme delivery will include an assessment of the opportunity cost of teacher time allocated to preparatory and administrative activities related directly to the delivery of the programme and face-to face time spent with CSC children on core programme materials.

Data on time allocated to implementation and delivery by teaching and senior management staff will be collected via school surveys delivered post intervention combined with school-based records on the number of programme sessions actually delivered. A pre-intervention survey will seek to collect data on time allocated to existing activities that focus on the vocabulary for CSC children within participating schools, so that a baseline for existing activity can be estimated.

Resource use data will be utilised to model the likely incremental resource impacts of delivering the programme at scale, including assumptions made regarding the % of CSC children who are likely to participate in the programme at scale. Variation in survey responses between schools will be used to characterise uncertainty around average (mean) estimates.

Ethics & Participation

The research design has been reviewed by the WWCSC Research Ethics Committee. If any of the evaluation activities are considered to be in scope for the King's College London Research Ethics Committee (i.e. it is considered King's sponsored research) they will be reviewed and approved by the relevant King's College London research ethics board prior to commencement.

As the participant selection and randomisation documented in this protocol will be conducted by schools and WW-CSC respectively, the evaluation will seek consent to respondents' participation in the data collection and evaluation. Opt-out consent will be sought to access students' NPD data and where applicable school administrative data, while opt-in consent will be sought from participants in the qualitative research. Parents/guardians will receive a detailed Participant Information Sheet detailing the project and data handling procedures and given the opportunity to opt out by contacting us or returning a physical consent form. Opt-out consent will also be sought from pupils.

All staff or research assistants who interact with pupils will be DBS-checked and trained by the project team, and we will work with colleagues at What Works for Children's Social Care to ensure that safeguarding is best-practice and appropriate for vulnerable participants.
In terms of identifying risks to participants, the following table outlines identified risks and notes about how they may be mitigated:

Possible risk	Notes on mitigation
CSC-experienced pupils are made more visible to themselves, each other and the school community, which could result in pupils feeling singled out or being treated differently	Teachers will be briefed to be mindful of this risk. Pupils' subjective experience of the programme will also be probed via the Implementation and Process Evaluation.
Students might feel coerced to take part when they are approached.	It is important we consider the power imbalance between pupils, parents/guardians, schools and the evaluation and project teams. Students will be assigned to be involved in the Words for All intervention via random lot, but otherwise consistent with the school's role to provide education, including targeted teaching. It will be made clear to students and parents/guardians that opting out of the evaluation will not affect their participation in Words for All or their standing at the school in any way.
When students are approached to participate in the evaluation, they may feel that their questionnaire responses or decision to withdraw their data might have consequences for their relationship with the school.	The participant information sheet will offer pupils a contact email they can reach out to withdraw their data if they wish, and make it clear that withdrawing their data from the evaluation will not impact their standing with the school or their participation in the Words for All programme.

Registration

The trial protocol has been pre-registered the protocols on OSF (add link here)

Data protection

All data will be held according to the King's Data Protection Policy and Procedure. All data collection will adhere to ethical practice ensuring the confidentiality of information shared and the secure handling of data in accordance with the GDPR and King's College London's Data Protection Policy. The legal basis for processing personal data will be under Article 6(1)(e), i.e. 'a task in the public interest'.

Pupils will therefore be given the opportunity to opt out of the evaluation, and to opt out of having their data matched into the NPD, and will be provided with an Information Sheet that will explain to them this legal basis for processing their data, detail how long it will be stored for and that if/how it will be shared with other parties, and provide them with the mechanism to ask that their data be removed or to raise a complaint[13].

The restrictions of COVID-19 mean that researchers at King's will be working at home for key portions of the project, and data security will need to be handled accordingly. Owing to the sensitive nature of the data, which will include care-status of individual pupils, before transfer from WWCSC to King's, data will be de-linked from both pupil PRN and school URN, with matching keys to be held separately. Data will then be stored on a secure section of the King's server, with access limited to those who have a direct purpose for using it as part of the project.

Data drawn from the National Pupil Database will be accessed in accordance with protocols set by the Department for Education, using the Secure Research Service. Access to individual files and folders will be on a by-permission basis only with higher restrictions for files including sensitive or individual-level data sources put in place. Rights to edit access permissions to those files and folders will be limited to personnel with a research need to access the data. The Principal Investigator will control access to the folder and will regularly review who has access and if it is still required.

All participants in Implementation and Process Evaluation activities will be provided with an information sheet prior to their participation in any observations, interviews, or group discussions and/or completion of the surveys. Data collection will only proceed with the explicit consent of participants. All participants will be informed that they have the right to withdraw from the study at any time and this will extend to the right to withdraw their data at any time up to June 2022. All necessary measures will be taken to preserve the anonymity of all individuals, schools and local authorities involved in the research including in any research publications.

Personnel

Name	Title	Responsibilities
Dr Mary Baginsky	Senior Research Fellow, HSCWRU	PI, lead on implementation and process evaluation
Susannah Hume	Director of Evaluation, Evidence Development and Incubation Team (EDIT) in the Policy Institute	PI, lead on impact evaluation

[13] King's provides a statement on the use of personal data, here: https://www.kcl.ac.uk/research/support/rgei/research-ethics/kings-college-london-statement-on-use-of-personal-data-in-research

Miriam Styrnol	Evaluation Associate, EDIT	Impact evaluation
Professor Gillian Manthorpe	Professor of Social Work, Health and Social Care Workforce Research Unit (HSCWRU)	Implementation and process evaluation, general project quality assurance and oversight
Dr Carl Purcell	Research Associate, HSCWRU	Implementation and process evaluation
Dr Andy Healey	Senior Health Economist, Centre for Implementation Science and King's Health Partners	Cost analysis for implementation and process analysis
Professor David Waugh	Durham University School of Education	Consultant on literacy and training content

Timeline

Task	Timing
Trial Protocol published	Jan 2021
Randomisation	Jan/Feb 2020
Review of training materials	Nov 2020/Mar 2021
First round of observations	Jan/Feb 2021
Intervention training and planning phase implemented	Dec 2020/Feb2021
Baseline data collection of NGRT and SDQ	Dec 2020/Jan 2021
Pre-intervention surveys	Jan 2021
Intervention implemented to students	Feb– May2021
Second round of observations	Mar/Apr 2021
Post-intervention surveys	April/May 2021
Interviews/group discussions	May/June 2021

Endline data collection of NGRT and SDQ	June 2021
Impact analysis complete	March 2022
Final Report	September 2022

Appendix: Power Calculation Code

```
library(pwr)
N <- 1200 # sample size per arm
icc <- 0.4 # intra-cluster correlation, set high as cluster is family
m <- 1.3 # avg cluster size
c <- 0.7 # baseline-endline correlation
D <- 1+(icc*(m-1)) # design effect because of clustering
n <- N/D # analytical sample size - actual N deflated by design effect
pwr <- pwr.t.test(n = n, d = NULL, power=0.8, sig.level=0.05)
d.raw <-pwr$d
d.adj <- (1-(c^2))*d
```