Independent evaluation of the efficacy of the KS2 Reading Fluency Project on reading comprehension: a two-armed cluster randomised trial

Statistical Analysis Plan

Evaluator (institution): Verian Group

Principal investigator(s): Prof. Natalie Gold

PROJECT TITLE	**Independent evaluation of the efficacy of the KS2 Reading Fluency Project on reading comprehension: a two-armed cluster randomised trial**
DEVELOPER (INSTITUTION)	HFL Education
EVALUATOR (INSTITUTION)	Verian Group
PRINCIPAL INVESTIGATOR(S)	Prof. Natalie Gold
PROTOCOL AUTHOR(S)	Prof Natalie Gold, Dr Michael Ratajczak, Pieter Cornel, Rupert Riddle, Dr Janet Vousden, Prof. Kate Cain
TRIAL DESIGN	Two-arm cluster randomised controlled trial with random allocation at school level
TRIAL TYPE	Efficacy
PUPIL AGE RANGE AND KEY STAGE	10-11, KS2
NUMBER OF SCHOOLS	178
NUMBER OF PUPILS	1239 (Based on the data submission of 169 schools)
PRIMARY OUTCOME MEASURE AND SOURCE	Reading comprehension (subtest of the York Assessment of Reading for Comprehension, or YARC)
SECONDARY OUTCOME MEASURE AND SOURCE	Reading accuracy (subtest of the YARC), reading rate (subtest of the YARC), reading prosody (Multidimensional Fluency Scale, or MDFS), and KS2 National Curriculum Reading Test score (NPD, KS2_READSCORE variable, KS2_READMRK)

SAP version history

VERSION	DATE	REASON FOR REVISION
1.0 [*original*]	26/02/2025	*N/A*

Table of contents

Introduction

This statistical analysis plan (SAP) sets out the intended impact evaluation for the independent evaluation of the KS2 Reading Fluency Project, which will be conducted in 168 schools in England. The KS2 Reading Fluency Project is an educational programme designed to improve pupils' reading comprehension by improving their reading fluency, making their reading more automatic and improving their prosody – the patterns of stress and intonation in a language. It targets Year 6 (Y6) pupils who are assessed as 'not on track' to meet the expected reading standards at the end of KS2. The programme supports effective fluency instruction through six strategies:

- Modelled expert prosody – the teacher reads the text with expert prosody;
- Echo reading – the pupils read the text mimicking the teachers' expert prosody;
- Repeated reading – the teacher facilitates re-reading of the text towards increasingly fluent reading;
- Text marking – the pupils track the text and mark it to support increased prosody;
- Performance reading – the pupils work in small groups to practise and performance read sections of text; and
- Modelling comprehension strategies – the teacher models comprehension responses and supports pupils' skills in navigating a text and retrieving information.

Year 6 teachers will deliver the programme directly to pupils, following training from HFL. First, teachers will conduct a YARC assessment with pupils. This pre-test allows teachers to gain understanding of the challenges the children face with their reading in order to adjust their support. Pupils will then receive a 20-minute fluency session from their Y6 teacher, twice a week for 8 weeks, with re-reads in between these fluency sessions. For more details on the content of the programme, please refer to the Trial Protocol..

Verian is conducting an impact evaluation (IE) using a two-armed clustered randomised control trial (RCT) across 178 schools to assess the impact of the KS2 Reading Fluency Project on pupils' reading comprehension of age-appropriate texts. The evaluation will also examine the programme's impact on reading accuracy, reading rate, prosody (and their mediating effect on reading comprehension) and KS2 National Curriculum Reading Test results, as well as whether impact on reading comprehension differs by pupils' Free School Meals (FSM) status, English as Additional Language (EAL) status and number of fluency sessions received.

Alongside the IE, Verian is conducting a holistic implementation and process evaluation (IPE) to provide evidence about the intervention's implementation and to help us to understand fidelity, complementing our quantitative compliance and dosage analysis. The IPE will investigate the causal assumptions of the Logic Model that underpin the IE and provide information about what schools in the control arm do with their pupils, which will help us interpret the results of the IE. It will also provide evidence that could be used to improve the delivery of the intervention if it were scaled up, such as the appropriateness of the pupil selection criteria and whether the use of the YARC at baseline is helpful in tailoring the intervention. The use of the YARC at baseline is specified as a part of the intervention, hence it could only be administered in treatment schools and could not be used as a trial baseline.

Design overview

Trial design, including number of arms		**Two-arm cluster randomised controlled trial, with a 'business-as-usual' control**
Unit of randomisation		School
Stratification variables (if applicable)		(1) School level quartiles of EAL percentage (2) School level quartiles of FSM percentage
Primary outcome	variable	Reading comprehension
	measure (instrument, scale, source)	Reading comprehension standardised score (subtest of the YARC)
Secondary outcome(s)	variable(s)	(1) Reading accuracy (2) Reading rate (3) Prosody (4) KS2 Reading
	measure(s) (instrument, scale, source)	(1) Reading accuracy standardised score (subtest of the YARC) (2) Reading accuracy standardised score (subtest of the YARC) (3) Reading prosody score (MDFS) (4) KS2 National Curriculum Reading Test score (NPD, KS2_READSCORE variable, scaled score from 80-120)
Baseline for primary outcome	**variable**	Reading ability
	measure (instrument, scale, source)	Early Years Foundation Stage Profile Literacy – Reading result (NPD, FSP_LIT_G09 variable, categorical from 1-3)
Baseline for secondary outcome	**variable**	Reading ability
	measure (instrument, scale, source)	Early Years Foundation Stage Profile Literacy – Reading result (NPD, FSP_LIT_G09 variable, categorical from 1-3)

Research questions

- **RQ1:** What is the impact of the KS2 Reading Fluency Project intervention compared to a business-as-usual control on **reading comprehension** (as measured by the YARC) in eligible pupils?
- **RQ2:** What is the impact of the KS2 Reading Fluency Project intervention compared to a business-as-usual control on **reading accuracy** (as measured by the YARC) in eligible pupils?
- **RQ3:** What is the impact of the KS2 Reading Fluency Project intervention compared to a business-as-usual control on **reading rate** (as measured by the YARC) in eligible pupils?
- **RQ4:** What is the impact of the KS2 Reading Fluency Project intervention compared to a business-as-usual control on **prosody** (as measured by the MDFS) in eligible pupils?
- **RQ5(a):** What is the impact of the KS2 Reading Fluency Project intervention compared to a business-as-usual control on **performance on the KS2 National Curriculum Reading Test (as measured by the scaled score)** in eligible pupils?
- **RQ5(b):** What is the impact of the KS2 Reading Fluency Project intervention compared to a business-as-usual control on **the proportion of pupils reaching the expected reading standard on the KS2 National Curriculum Reading Test** in eligible pupils?
- **RQ6:** What is the impact of the KS2 Reading Fluency Project on reading comprehension **for FSM-eligible (in the last 6 years)** pupils who were eligible to receive the intervention in the treatment arm, **compared to FSM-eligible (in the last 6 years)** pupils who were eligible to receive the intervention in the control arm?
- **RQ7:** What is the impact of the KS2 Reading Fluency Project on reading comprehension **for EAL** pupils who were eligible to receive the intervention in the treatment arm, **compared to EAL** pupils who were eligible to receive the intervention in the control arm?

Randomisation overview

There were 181 schools initially recruited into the trial. The sample was recruited from across England.

Randomisation was carried out on the 25th of June using the R package 'randomizr'[2] Following EEFs guidance, randomisation was stratified by school level FSM and EAL percentage quartiles. These quartiles were calculated based on school-level characteristics data published by Explore Education Statistics. Per the trial protocol, we generated unique school IDs to conduct the randomisation so that researchers were blinded to the identity of the schools. See **Annexe 1 – Randomisation code** for details of how the randomisation was conducted)

At the time of randomisation, 178 schools were enrolled in the trial. 3 schools withdrew from the trial due to expected issues with project capacity. 89 schools were assigned to the treatment arm, and 89 schools assigned to the control arm. The baseline characteristics for each trial arm are reported below:

Table 1: Baseline characteristics of the sample

	Treatment N = 89		Control N = 89	
	Mean	Median	Mean	Median
Proportion of students eligible for FSM	28.6	26.1	28.4	25.4
Proportion of students with EAL	25.6	16.3	23.9	15.6
School size	321.9	285	342.6	346

Of the schools in control arm, 46 (51.7%) are located in an Education Investment Area (EIA), compared to 49 (55%) of the schools in the treatment arm.

Per the trial protocol, arm allocation was only communicated to schools in the control and intervention arm after the school had selected pupils to participate in the intervention. This approach was taken to reduce the risk of bias in the pupil selection process. Verian received pupil data from 169 schools. During the pupil selection process, nine schools withdrew from the trial due to concerns about teacher capacity or other existing commitments. During the intervention period, one additional school withdrew from the trial due to capacity issues. At the time of writing, 168 schools are expected to complete the endline assessment.

[2] Alexander Coppock et al., "Package 'Randomizr,'" CRAN, 2023, Available at: https://cran.r-project.org/web/packages/randomizr/randomizr.pdf.

Sample size calculations overview

Note on planned sample size calculations

Our power calculations were based on the assumption that the pre-test / post-test level 1 correlation would be 0.62. As indicated in the section below, this estimate was informed by previous research conducted by the EEF (see Planned sample size). However, this value (0.15) was not squared during the power calculation process. As a result, the MDES originally reported in the trial protocol for the full sample (0.19) and sub-group analysis (0.254) is incorrect.[3] We have updated the power calculations in Table 4 to reflect the true MDES at the point the trial protocol was published, and at the randomisation stage of the trial evaluation.

Our power calculations assumed an intra-cluster correlation (ICC) of 0.15 at school level. Our estimate for the ICC is based on a previous trial conducted with the EEF titled 'Abracadabra (ABRA) trial: An online reading programme aimed at improving early literacy' [4]. The estimate for the pre-/post-test correlation is drawn from the median of 32 KS1 and KS2 literacy trials conducted by the EEF (based on helpful information that the EEF has since published to inform power calculations for future trials). Our recommended sample size remains reasonably robust even after conducting sensitivity analyses that vary key parameters (see Table 4 in the Trial Protocol for details).

Updated sample size calculations

Table 2 presents the sample size calculations from the trial protocol (Columns 1 and 2), along with updated power calculations that reflect the sample at randomisation (Columns 3 and 4). The sample size calculations in the randomisation columns are based on the 178 (89 control, 89 intervention) schools enrolled in the trial at the point of randomisation.

As stated above, 169 schools completed the pupil selection process and shared this data with Verian. The average number of pupils selected for inclusion in the trial per school is 7.3. Following discussions with EEF, we used this value to conduct indicative power calculations at the point of randomisation (with 178 schools), assuming that the remaining 9 schools who were randomised also selected, on average, 7.3 pupils each. Based on the updated power calculations at the point of randomisation, we are sufficiently powered to detect a MDES of 0.199 in our primary analysis.

The percentage of pupils eligible for FSM in schools is estimated at 23.8%.[5] If we were to assume this percentage of our small group then we would have approximately 1.737 pupils per group – which results in an effect size of 0.284 as shown in Column 4, Table 22. To be powered to detect a MDES of 0.284 in our sub-group analysis of pupils eligible for FSM (or pupils with EAL), approximately 59% of the pupils (cluster size of 4.3 pupils per school)

[3] When designing the trial, we determined that a minimum sample size of 162 schools (81 per arm) was required to achieve an MDES of 0.20. However, to account for potential attrition, we assumed that 10% of recruited schools would drop out throughout the trial and were conservative with our average cluster size estimate of 6 pupils per school cluster. As a result, we intended to recruit 180 schools into the trial, resulting in a Minimum Detectable Effect Size (MDES) of 0.19 standard deviations for the primary analysis of the intervention's effect on reading comprehension amongst Year 6 pupils, and 0.254 standard deviations for FSM-eligible pupils.

[4] K. Bell et al., "ABRACADABRA (ABRA) and Reading and Understanding in Key Stage 1 (RUKS)." (EEF, 2022), ABRA-Report-Final.pdf (d2tic4wvo1iusb.cloudfront.net).

[5] This estimate comes using National Schools Pupils and Characteristics 2022/2023: GOV.UK (explore-education-statistics.service.gov.uk).

selected for inclusion in the trial would have to reflect these characteristics (assuming 178 schools are included in the trial and each school selects an average of 7.3 pupils).

Table 2: Updated sample size calculations

		Protocol		Randomisation	
		OVERALL	FSM / EAL	OVERALL	FSM
Minimum Detectable Effect Size (MDES)		0.205	0.297	0.199	0.284
Pre-test/ post-test correlations	Level 1 (pupil)	0.3844	0.3844	0.3844	0.3844
	Level 2 (class)	0	0	0	0
	Level 3 (school)	-	-	-	-
Intra-cluster correlations (ICCs)	Level 2 (class)	-	-	-	-
	Level 3 (school)	0.15	0.15	0.15	0.15
Alpha		0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8
One-sided or two-sided?		2-sided	2-sided	2-sided	2-sided
Average cluster size		6	1.5	7.3	1.737
Number of schools	Control	90	90	89	89
	Intervention	90	90	89	89
	Total	180	180	178	178
Number of pupils	Control	540	135	649	162
	Intervention	540	135	650	162
	Total	1,080	270	1299	324

Achieved sample size

At the protocol stage, the trial was powered to detect an MDES of 0.205 standard deviations for the primary analysis of the intervention's effect on reading comprehension amongst Year 6 pupils, and 0.297 standard deviations for FSM-eligible pupils. At the point of randomisation, this trial was powered to detect a MDES of 0.199 standard deviations for the primary analysis, and 0.284 standard deviations for sub-group analysis for FSM-eligible pupils, which is in accordance with the EEF's guidance that MDES should be less than 0.2. Additional schools and pupils are expected to withdraw from the trial across the trial period, meaning that the true MDES at the point of analysis is expected to increase slightly.

This trial is not powered to detect an effect of 0.2 on the FSM or EAL subgroups of interest– unless the highly unlikely assumption is met that all pupils taking part are also eligible for FSM (or have been eligible in the last 6 years) or are EAL. FSM or EAL status is not a criterion which teachers will be using to select pupils, although it may be correlated with low reading ability.

Data Collection

Overview of data collection process

Data collection in schools will take place from 10th of February to end of March 2025. To mitigate fatigue we will divide the assessment into two short (~15-20 minute) assessments, one for the YARC and one for the MDFS, for the pupils with a few hours between them. This assessment will be conducted by 26 independent assessors who are blind to the school's assignment to either the treatment or control arm. Each assessor has a professional background working in schools and delivering assessments to pupils, and has an up-to-date DBS check.

Prior to data collection, Verian will provide comprehensive training to assessors on how to administer the YARC and MDFS to pupils, and how to score the assessments and share the data securely with Verian. As part of this training, assessors will complete example assessments for the YARC and MDFS, and we will ensure consistency in scoring by providing model answers and an opportunity for assessors to compare responses.

After the training period, we will ask assessors to complete a final YARC and MDFS assessment to ensure that they understand the process and reduce the risk of inconsistent scoring. Any assessor who deviates significantly from the model answers provided will be asked to review the training materials.

At the end of the data collection period, assessors will also re-grade an MDFS assessment to assess each assessor's deviation from the gold standard (i.e., the score they should have assigned during the endline assessment) in their scoring.

YARC procedure

At the start of the YARC assessment, assessors will ask pupils to confirm their name and date of birth. Next, assessors will complete a short 'icebreaker' task with pupils which involves asking them about their attitudes towards reading (See 'Reading Enjoyment Survey).

To obtain reliable estimates for Reading accuracy, rate and comprehension, the YARC requires pupils to read **two consecutive level passages** to completion and answer 8 comprehension questions per passage. Whilst the pupil reads a passage, the assessor will record the number of reading errors (irrespective of the type of error made) and the time taken to read the text. After the pupil has finished reading the text, the assessor will ask the pupil 8 comprehension questions based on the passage and score the number of correct responses.

All pupils will start the YARC assessment by reading passage 5B.

YARC Passage Selection Rules

After a pupil completes their **first passage reading** of the YARC (**Passage 5B**), their performance on that passage is used to determine the next passage they read as part of the assessment. The pupil will progress to a **higher-level passage** for their **second passage reading** only if the following criteria are met:

1. The pupil made **15 or fewer reading errors** during the passage reading, **and**
2. The pupil answered **5 or more comprehension questions correctly**.

The pupil will progress to a **lower-level passage** for their **second passage reading** if either:

1. The pupil made **16-20 reading errors** during the passage reading**, or**
2. The pupil answered **4 or fewer comprehension questions correctly**.

If a pupil makes **21 or more** reading errors during a passage reading, the assessor will allow the pupil to finish reading the passage (if they are able and willing to do so), but not record the time taken to complete the passage, nor ask the comprehension questions associated with the passage. The assessor will then 'restart' the assessment by asking the pupil to read the next lowest passage and treat it as the pupil's **first passage reading**. In this case, the pupil will progress to a **lower-level passage** for their **second passage reading** (I.e., If a pupil makes 21 or more reading errors on Level 5, they will read passage level 4 and passage level 3).

Some pupils may be unable to access passages from levels 3-6. If the pupil regresses to passage level 2 and below during the YARC assessment, there are unique discontinuation rules at each level which we will employ. Specifically:

1. When reading **passage level 2**:
   a. If the pupil makes **16 or more** reading errors during the passage read, the assessor will allow the pupil to finish reading the passage (if they are able and willing to do so), but not record the time taken to complete the passage, nor ask the comprehension questions associated with the passage. In this case, the pupil will read Passage Level 1 as their first passage.
2. When reading **passage level 1**:
   a. If the pupil makes **16 or more** reading errors during the passage read, the assessor will encourage the pupil to finish the text and ask comprehension questions.
   b. If the pupil is unable to complete the passage, the test will be discontinued, and the pupil will score zero for comprehension on this passage.
3. When reading the **beginner level** passage:
   a. If the pupil makes 16 or more reading errors during the passage read, the assessor will encourage the pupil to finish the text and ask comprehension questions.
   b. If the pupil is unable to complete the passage, the pupil will score zero for comprehension on this passage and the test will be discontinued entirely.

The assessment will begin with pupils reading passage Level 5 – B 'Walk in the Fog '. If the pupil makes fewer than 15 reading errors on their first passage, they will read either passage Level 6 – B 'Shoes', or Level 4 – B 'Bees', depending on their comprehension score.

After completing the YARC, the assessor will use the number of reading errors, the time taken to complete the passage(s) and the comprehension scores to generate standard scores using the YARC scoring tool. The assessors will then complete an excel template which includes the reading error, rate and comprehension scores for both passages, the standard scores generated by the scoring tool, and the passages levels used in the assessment.

MDFS procedure

The MDFS assessment requires pupils to read two short passages – one narrative, one expository – first silently and then aloud for approximately one minute. If the pupil has not finished reading the passage within 1 minute, the assessor will ask the pupil to pause at an appropriate break in the text. As the pupil reads each passage aloud, the assessor will score the pupil on 'Expression and volume', 'Phrasing', 'Smoothness' and 'Pace' using the MDFS scoring rubric (See Trial protocol for details). Each category is scored 1-4.

All pupils will read passage 1 and passage 2 regardless of their performance on passage 1.

MDFS Passage Selection

To develop appropriate passages for this assessment, we used nine 150-word excerpts from Rising Star's texts – a resource which contains texts designed for pupils in years 4-6 – to establish a baseline difficulty across 4 key readability metrics: Flesch-Kincaid Grade level, Flesch Reading Ease, Coh-Metrix L2 Readability and CELEX Log frequency for all words for our passages (Average Word Frequency). The mean readability scores across each different metric for the Rising Star texts for each year group are reported in Table 3.

Table 3: Mean Readability Metrics of sample texts, split by suggested year group.

Year group	Fleisch-Kincaid Grade Level	Fleisch Reading Ease	Average Word Frequency	Coh-Metrix Readability
Year 4	3.955	83.989	2.603	15.358
Year 5	5.425	76.660	2.886	22.047
Year 6	7.071	74.111	3.105	16.588

We then developed two short passages (of approximately 150 words) to texts which had a comparable readability rating to the average of the Year 5 Rising Stars texts (Passage 1), and Year 6 Rising Stars texts (Passage 2). Passage 1 is a non-fiction passage about The Great Fire of London, which was written by a researcher. Passage 2 is an excerpt from Lewis Carol's "The Nursery 'Alice'", edited by a researcher.

Table 4: Readability metrics of MDFS passages

Passage	Fleisch-Kincaid Grade Level	Fleisch Reading Ease	Average Word Frequency	COH-Metrix Readability
Passage 1	5.235	77.568	2.994	17.400
Passage 2	7.435	81.166	3.158	19.965

After completing the MDFS, the assessor will record the ‘Expression and volume’, ‘Phrasing’, ‘Smoothness’ and ‘Pace’ scores (from 1-4) for each passage in the same excel template in which the YARC scores were recorded.

Analysis

Primary outcome analysis

Outcome

The primary outcome of interest is each pupil's standardised reading comprehension score from the YARC. The primary outcome will be collected during an endline assessment at the end of the intervention period, scheduled for February/March 2025. This assessment will be conducted by independent assessors who are blind to the school's assignment to either the treatment or control arm.

Note that the YARC uses combined norms for calculating age-standardised scores across different passage levels. This means that the standardisation process accounts for differences in passage difficulty, allowing for direct comparability of age-standardised scores regardless of the passage level a student reads.

Analysis

The primary analysis for RQ1 will follow an intent-to-treat approach, employing a linear mixed-effects model to account for the clustering of pupils within schools, with pupils' YARC reading comprehension standardised score as the response variable. The arm allocation indicator (comparing treated and control pupils) and pupils' Early Years Foundation Stage Profile (EYFSP) Literacy – Reading score (baseline) will be included as fixed effects in the model, with school treated as a random effect. Specifically, we will estimate a random intercept for each school in the trial to account for school-level differences in our sample.

In line with EEF guidance, we will include stratification variables as covariates in the model as fixed effects, specifically the quartiles of FSM and EAL percentages at the school level. These quartiles were calculated based on school-level characteristics data published by Explore Education Statistics.

We propose to use the following model:

$$Y_{ij} = \beta_0 + \beta_1 Intervention_j + \beta_2 Baseline_{ij} + \beta_3 FSM_j + \beta_4 EAL_j + \gamma_j + \varepsilon_{ij}$$

Where Y_{ij} is the YARC reading comprehension standardised score for the i^{th} pupil ($i = 1, \dots, n$, where n is the number of participating pupils) in the j^{th} school ($j = 1, \dots, m$, where m is the number of participating schools). $Intervention_j$ is an indicator for arm allocation for the j^{th} school (0 = Control, 1 = Intervention) and $Baseline_{ij}$ is the EYFSP Literacy – Reading result of the i^{th} pupil's in j^{th} school. FSM_j and EAL_j are variables which reflect the quartile of FSM and EAL percentage for the j^{th} school. These variables are included in the model as fixed effects. γ_j is the deviation of school j's mean z-score from the grand mean ($\gamma_j \sim N(0, \sigma_1^2)$), and ε_{ij} is the residual error term for the i^{th} pupil in j^{th} school ($\varepsilon_{ij} \sim N(0, \sigma_2^2)$).

Unless otherwise stated, this and other analyses will be conducted in R using the lme4[6] package.

[6] Douglas Bates et al., "Lme4: Linear Mixed-Effects Models Using 'Eigen' and S4," July 3, 2024, https://cran.r-project.org/web/packages/lme4/index.html.

Secondary outcome analysis

Outcomes

The secondary outcomes of interest are each pupil's YARC reading accuracy standardised score (RQ2), YARC reading rate standardised score (RQ3), MDFS prosody score (RQ4) and KS2 National Curriculum Reading Test Score (RQ5). The YARC reading accuracy and reading rate standardised scores and the MDFS prosody score will be collected alongside the primary outcome (YARC reading comprehension standardised score) by independent assessors who are blind to the condition of the school. The KS2 National Curriculum Reading Test Score will be collected via the National Pupil Database (NPD). Depending on data availability due to missingness (See Trial Protocol pp 17-18), we will use either KS2_READSCORE or KS2_READMRK variable

Analyses

The specification of the models for RQ2 to RQ5 is the same as for the model of the primary analysis

The model for RQ2 is:

$$Y_{ij} = \beta_0 + \beta_1 Intervention_j + \beta_2 Baseline_{ij} + \beta_3 FSM_j + \beta_4 EAL_j + \gamma_j + \varepsilon_{ij}$$

Where Y_{ij} is the YARC reading accuracy standardised score for the i^{th} in the j^{th} school, with the remaining terms defined as in the RQ1 model.

The model for RQ3 is:

$$Y_{ij} = \beta_0 + \beta_1 Intervention_j + \beta_2 Baseline_{ij} + \beta_3 FSM_j + \beta_4 EAL_j + \gamma_j + \varepsilon_{ij}$$

Where Y_{ij} is the YARC reading rate standardised score for the i^{th} pupil in the j^{th} school, with the remaining terms defined as in the RQ1 model.

The model for RQ4 is:

$$Y_{ij} = \beta_0 + \beta_1 Intervention_j + \beta_2 Baseline_{ij} + \beta_3 FSM_j + \beta_4 EAL_j + \gamma_j + \varepsilon_{ij}$$

Where Y_{ij} is the MDFS prosody score for the i^{th} pupil in the j^{th} school, with the remaining terms defined as in the RQ1 model.

The model for RQ5 is:

$$Y_{ij} = \beta_0 + \beta_1 Intervention_j + \beta_2 Baseline_{ij} + \beta_3 FSM_j + \beta_4 EAL_j + \gamma_j + \varepsilon_{ij}$$

Where Y_{ij} is the KS2 National Curriculum Reading Test scaled scores (KS2_READSCORE) for the i^{th} pupil in the j^{th} school, with the remaining terms defined as in the RQ1 model.

The scaled score has the added benefit of comparability with previous and future KS2 scores and can be used to examine whether there is a difference between control and treatment arm pupils in terms of meeting the reading expected standard (overall aim of the programme). Specifically, pupils reach the expected standard by achieving a scaled score of 100 or higher. Critically, the scaled scores are derived from raw scores (KS2_READMRK) and if a pupil scores less than 3 on the test (the raw score range if 0 to 50), they do receive a scaled score.[7] Thus, if we find the scaled score to contain more missing data than the raw score, we will use the raw score as the outcome to answer RQ5, and the scaled score to look at differences in expected standard between arms descriptively.

Subgroup analyses

We will examine the impact of the intervention on our primary outcome for two subgroups of interest – pupils with FSM (RQ6) and EAL (RQ7) status. To answer RQ6, we will re-run the primary model including an interaction term for pupil-level FSM status.

$$Y_{ij} = \beta_0 + \beta_1 Intervention_j + \beta_2 Baseline_{ij} + \beta_3 FSM_j + \beta_4 EAL_j + \beta_5 FSM\_Indicator_{ij} + \beta_6 (Intervention_j \cdot FSM\_Indicator_{ij}) + \gamma_j + \varepsilon_{ij}$$

Where the $FSM\_Indicator_{ij}$ is a pupil-level indicator (0 if not eligible for FSM; 1 if β_6 is the interaction between YARC reading comprehension standardised score for the i^{th} j^{th} school and that pupil's FSM status. This indicator will be taken from the National Pupil Database, using the variable '.EVERFSM_6_P'.

To answer RQ7, we will re-run the primary model including an interaction term for pupil level EAL status.

$$Y_{ij} = \beta_0 + \beta_1 Intervention_j + \beta_2 Baseline_{ij} + \beta_3 FSM_j + \beta_4 EAL_j + \beta_5 EAL\_Indicator_{ij} + \beta_6 (Intervention_j \cdot EAL\_Indicator_{ij}) + \gamma_j + \varepsilon_{ij}$$

Where the $EAL\_Indicator_{ij}$ is pupil-level binary EAL indicator (0 not EAL or EAL status is unknown; 1 if has EAL) and β_6 is the interaction between YARC reading comprehension standardised score for the i^{th} pupil in the j^{th} school and the pupil's EAL status. This indicator will be taken from the National Pupil Database, using the variable 'KS2_EALGRP'.

In line with EEF Analysis guidance, we will also produce split sample analysis results in the appendix of the evaluation report for each subgroup, containing only pupils from that subgroup as a sensitivity check [8]. If we find an inconsistent pattern of results, the evaluation will explore possible reasons for this divergence.

It should be noted that all subgroup analyses are exploratory, as we do not expect to be powered to detect differences within these subgroups. As previously mentioned, this limitation affects the study's ability to ascertain the intervention's impact on specific subgroups.

Additional analyses

Sensitivity analysis

Due to the fact the YARC and MDFS are administered by independent assessors, there may be variability in the data attributable to assessor-level differences. However, we do not have

[7] 2024 key stage 2 scaled score conversion tables - GOV.UK and Understanding scaled scores at key stage 2 - GOV.UK

[8] EEF, "Statistical Analysis Guidance for EEF Evaluations," 2022, https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1729585898.

strong theoretical reasons to include this variability in our primary and secondary models, and we are uncertain how it might influence the intervention estimates.

To consider this potential variability, we will run an additional sensitivity analysis for RQ1 and RQ4. In this sensitivity model, we will include an endline assessor-level covariate that approximates each assessor's deviation from the gold standard (i.e., the score they should have assigned during the endline assessment; see Overview of Data Collection). We will also add a random effect for assessors.

At this stage, we do not know whether the assessor random effect will be modelled as crossed or nested, since we are unsure if more than one assessor will visit the same school (due to possible mop-up visits). Consequently, the exact model specification will depend on the observed data.

Mediation analysis

The logic model assumes that the intervention will result in pupils reading age-appropriate texts more fluently which in turn will lead to improved text comprehension (see the evaluation protocol). To examine the underlying mechanism of how the intervention works, we will perform a mediation analysis to see whether and to what extent reading accuracy, reading rate and prosody contribute to the overall effect of the intervention on reading comprehension. To do so, we will use structural equation modelling (SEM). SEM allows for the simultaneous estimation of direct and indirect effects [9]. This analysis will be conducted in R, using the Lavaan package[10].

The model consists of:

1. Mediator equations:

We first model how the intervention affects each of the three reading skills (mediators):

- Reading accuracy ($rACC_{ij}$):

$$rACC_{ij} = \alpha_{0a} + \alpha_{1a} Intervention_j + \alpha_{2a} Baseline_{ij} + \alpha_{3a} FSM_j + \alpha_{4a} EAL_j + \gamma_{ja} + \varepsilon_{ija},$$

where $rACC_{ij}$ is the reading accuracy score for pupil i in school j, α_{0a} is the intercept term, and α_{1a} is the effect of the intervention on reading accuracy, controlling for the baseline score and covariates. It quantifies how much reading accuracy changes on average due to the intervention.

- Reading rate ($rRate_{ij}$):

$$rRate_{ij} = \alpha_{0b} + \alpha_{1b} Intervention_j + \alpha_{2b} Baseline_{ij} + \alpha_{3b} FSM_j + \alpha_{4b} EAL_j + \gamma_{jb} + \varepsilon_{ijb},$$

where $rRate_{ij}$ is the reading rate score for pupil i in school j, α_{0b} is the intercept term, and α_{1b} is the effect of the intervention on reading rate, controlling for the baseline score and covariates. It quantifies how much reading rate changes on average due to the intervention.

---

[9] Douglas Gunzler et al., "Introduction to Mediation Analysis with Structural Equation Modeling," *Shanghai Archives of Psychiatry* 25, no. 6 (December 2013): 390, https://doi.org/10.3969/j.issn.1002-0829.2013.06.009.
[10] Yves Rosseel et al., "Lavaan: Latent Variable Analysis," September 26, 2024, https://cran.r-project.org/web/packages/lavaan/index.html.

- Reading prosody ($rPros_{ij}$):

$$rPros_{ij} = \alpha_{0c} + \alpha_{1c}Intervention_j + \alpha_{2c}Baseline_{ij} + \alpha_{3c}FSM_j + \alpha_{4c}EAL_j + \gamma_{jc} + \varepsilon_{ijc},$$

where $rPros_{ij}$ is the reading prosody score for pupil i in school j, α_{0c} is the intercept term, and α_{1c} is the effect of the intervention on reading prosody, controlling for the baseline score and covariates. It quantifies how much reading prosody changes on average due to the intervention.

2. Outcome equation:

Next, we model reading comprehension as a function of the intervention and mediators:

$$Y_{ij} = \beta_0 + \beta_1 Intervention_j + \beta_2 rACC_{ij} + \beta_3 rRate_{ij} + \beta_4 rPros_{ij} + \beta_5 Baseline_{ij} + \beta_6 FSM_j + \beta_7 EAL_j + \gamma_j + \varepsilon_{ij}$$

Where Y_{ij} is the YARC reading comprehension standardised score for the i^{th} pupil ($i = 1, \dots, n$, where n is the number of participating pupils) in the j^{th} school ($j = 1, \dots, m$, where m is the number of participating schools). β_1 is the direct effect of the intervention on reading comprehension (the impact of intervention on comprehension that is not explained by changes in reading accuracy, rate, or prosody), while β_2, β_3, β_4 are the direct effects of reading accuracy, rate, and prosody on reading comprehension, respectively. The indirect effects of the mediators (the effects of the intervention on reading comprehension that occur through the mediators) are calculated by multiplying the effect of the intervention on each mediator (α_{1n}) by the effect of that mediator on the outcome (β_n) The remaining terms are defined as in the RQ1 model.

We will also derive the total direct effect, defined as the sum of direct and indirect effects, and calculate the proportion of the total effect mediated by each mediator in relation to the overall effect of the intervention on reading comprehension. We will use bootstrapping to derive standard errors, confidence intervals, and p-values for these estimates.

If the indirect effect of a mediator is significant but the direct effect is not, this can be interpreted as full mediation. If both the indirect effect and the direct effect are significant, this can be interpreted as partial mediation. If neither effect is significant, this can be interpreted as no evidence for mediation. However, while statistical significance is important, it's also valuable to consider the size and practical significance of the effects, as well as the confidence intervals, when interpreting mediation results. Additionally, non-significant results can sometimes be due to insufficient statistical power rather than the absence of an effect. Consequently, we will not interpret mediation output solely based on statistical significance.

Dosage

The logic model assumes that attending additional fluency sessions will lead to better outcomes, as pupils have more opportunities to practice and receive feedback (see the evaluation protocol). To test this assumption, we will run a dosage model where we examine whether there is a difference in the impact depending on the number of fluency sessions a pupil in the intervention arm receives during the trial. Data on pupils' attendance at fluency sessions will be collected by teachers and shared with the Evaluation Team.

The proposed dosage model is:

$$Y_{ij} = \beta_0 + \beta_1 Dose_{ij} + \beta_2 Baseline_{ij} + \beta_3 FSM_j + \beta_4 EAL_j + \gamma_j + \varepsilon_{ij}$$

Where $Dose_{ij}$ is a numerical variable which indicates the number of doses (i.e., sessions attended from 0 to 16) by pupils in the intervention condition. The remaining terms are defined as in the RQ1 model. The dosage coefficient will estimate the marginal impact of a pupil attending an additional training session. This analysis will be conducted in R using the 'lme4' package.

Longitudinal follow-up analyses

There is no longitudinal follow-up as part of this trial, but data will be archived to facilitate future longitudinal analyses.

Imbalance at baseline

A well-conducted randomisation ideally should create arms that are equivalent at baseline (at the point of randomisation), with any imbalance at baseline occurring by chance. However, to check for, and monitor, imbalance at baseline in the realised randomisation, we will conduct baseline equivalence testing using across the trial arms at the school and pupil level. Specifically, we will assess baseline equivalence by reporting descriptive statistics for each trial arm and calculating standardised mean difference (SMD) for the covariates. The SMD is a commonly used metric to assess baseline imbalance between arms in education trials[11]. In line with What Works Clearinghouse standards, we will include covariates with an SMD greater than 0.25 in a sensitivity analysis to adjust for substantial baseline differences.

At the school level, the unit of randomisation in this trial, we will report the following characteristics across trial arms:

- School size (number of pupils)
- Proportion FSM (whole school)
- Proportion EAL (whole school)
- Proportion of schools in an 'Education Investment Area'
- Average school performance at KS2
- Proportion of urban and rural schools
- Breakdown of Ofsted ratings
- Breakdown of school type

At the pupil level, we will report the following characteristics across trial arms:

- FSM status
- EAL status

[11] U.S. Department of Education, "What Works Clearinghouse Procedures and Standards Handbook, Version 5.0" (U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, ND).

- Gender
- EYFSP Literacy – Reading result.

We will also monitor potential imbalance in the final sample prior to analysis in order to identify any potential effect of attrition.

Missing data

Descriptive statistics

Attrition across both trial arms will be explored as a basic step to assess bias across arms. We will describe the extent of missingness of all variables to be used as part of the analysis, both overall and then by arm. We will provide cross-tabulations of the proportions of missing values on model covariates (at both pupil and school level), as well as on the primary outcome measure and secondary outcome measures.

Analyses

We will follow the EEF Statistical Analysis Guidance for assessing the type of missing data and conducting missing data analyses.[12]

Missing information generally is classified into the following situations:

- 'Missing completely at random' (MCAR): the data is missing in a way that is not correlated with either observed or unobserved variables.
- 'Missing at random' (MAR): the data is missing in a way that is correlated with observed variables.
- 'Missing not-at random' (MNAR); the data is missing in a way that is correlated with an unobservable variable, even after considering all information in the observed variables.

If the data is MCAR, complete case analysis will yield unbiased but less precise estimates. Consequently, for less than 5% missingness in our primary outcome we will proceed with a complete-case analysis.

In cases where there is over 5% of missingness in the primary outcome, we will explore the missingness mechanisms. Specifically, we will examine whether the missingness in the primary outcome is related to pupil-level FSM status, EAL status and gender (an auxiliary variable). Although the primary analysis uses school-level FSM and EAL covariates, we will use pupil-level indicators for the missingness analysis, as it is more appropriate for understanding the patterns of missing data at the individual level.

We will fit a multilevel logistic regression model predicting missingness (1 = missing, 0 = complete) of the primary outcome from the pupil-level FSM, EAL, and gender variables, with pupils nested within schools. This model will be analogous to the primary analysis model (in terms of clustering) but will use a binary indicator of missingness as the outcome and pupil-level covariates rather than school-level covariates. If the effects of these pupil-level variables is significant at $p < 0.05$, we will compare the results of the complete case analysis against a model that includes these pupil-level covariates to adjust for the identified missingness mechanism. If the two sets of estimates are similar in direction and magnitude, we will proceed

[12] EEF, "Statistical Analysis Guidance for EEF Evaluations."

with the complete case analysis as the primary analysis is unlikely to be biased. However, if the results differ substantially, we will assume the data are MNAR. We will also assume MNAR if outcome missingness is not related to pupil-level covariates.

Second, we will examine missingness in the covariates (proportion FSM, proportion EAL, and baseline) themselves. Using a similar approach, we will fit separate multilevel logistic regression models for each covariate with an indicator variable (1 = missing, 0 not missing) as the outcome. For school-level covariates, we will use other school-level characteristics as predictors (e.g., school size, proportion FSM, proportion EAL, Education Investment Area status, average school performance at KS2, Ofsted rating where available, and school type). For the baseline measure, we will use pupil-level variables including FSM status, EAL status, and gender. If covariate missingness is significantly related to other measured variables at the corresponding level ($p < 0.05$), we will proceed to impute missing values (see below). We will then compare the model estimates using the imputed dataset against estimates from the complete case dataset. If these results are similar in direction and magnitude, we will rely on the complete case analysis. If they differ, we will assume MNAR. If no conditional relationship is found, we will also assume MNAR.

For MNAR scenarios and situations where covariates missingness is conditional on other variables, we will use multiple imputation by chained equations.[13] MICE works by iteratively fitting models through a multiple imputation strategy, assuming as the outcomes the variables containing missing data. Each variable containing missing data is modelled with the remaining variables as predictor variables. This is repeated for each variable with missing data and repeated, typically, until the estimates of the models stabilise. Recently, MICE has been found to have the lowest mean absolute error (MAE) on a variety of MNAR scenarios (e.g., MAE of 0.076 with 10% missingness based on unobserved values) with up to 60% of missingness according to a recent MNAR data generation and imputation method benchmarking study.[14] We do not expect to have more than 60% of missing data. Consequently, when covariates missing are conditional on other covariates and when we think data is MNAR we will report the pooled estimates generated by the model using imputed data alongside the complete case primary analysis. This will constitute a sensitivity analysis highlighting differences in coefficients, confidence intervals, and p-values, depending on the handling of missing data.

To perform MICE, if needed, we will use the ‘mice’ package in R. We will monitor convergence of the imputation process using Rhat.mice function from the ‘miceadds’[15] package. We will start with 100 imputations and, if the $\hat{R}$ values indicate that the mean and variance estimates have not stabilised, we will increase the number of imputations. Once the $\hat{R}$ statistics suggest that the estimates have stabilised, we will stop increasing the number of imputations.

---

[13] Stef van Buuren and Karin Groothuis-Oudshoorn, “Mice: Multivariate Imputation by Chained Equations in R,” *Journal of Statistical Software* 45 (December 12, 2011): 1–67, https://doi.org/10.18637/jss.v045.i03.
[14] Ricardo Cardoso Pereira et al., “Imputation of Data Missing Not at Random: Artificial Generation and Benchmark Analysis,” *Expert Systems with Applications* 249 (September 1, 2024): 123654, https://doi.org/10.1016/j.eswa.2024.123654.
[15] Alexander Robitzsch [aut,cre] (<https://orcid.org/0000-0002-8226-3132>), Simon Grund [aut] (<https://orcid.org/0000-0002-1290-8986>), Thorsten Henke [ctb], “Miceadds: Some Additional Multiple Imputation Functions, Especially for ‘Mice,’” February 19, 2014, https://doi.org/10.32614/CRAN.package.miceadds.

Compliance

We will define compliance at the pupil level, based on the following criteria:

- (1,0) - if the child has completed the baseline YARC, or not;
- (1,0) - if the child has attended at least 12 fluency sessions (out of a possible 16 fluency sessions), or not;
- (1,0) - if the child's teacher had attended at least 4 out of the 5 possible training sessions, or not.

These criteria were identified as core components of the intervention following consultation with the delivery team for this trial, HFL. For more detail on how these criteria were identified and the thresholds for compliance were set, please refer to the Trial Protocol.

We will estimate the Complier Average Causal Effect (CACE) using a Two Stage Least Squares (2SLS) approach, treating each of the compliance indicators as endogenous variables in the model[16]. Assuming that each binary indicator reflects the same underlying causal effect (i.e., the same underlying compliance construct) would be a stronger assumption than saying they capture distinct aspects of compliance. Furthermore, have no evidence that these three separate compliance measures are linearly related, and forcing them into a single measure could introduce bias or misrepresent their individual contributions. By modelling each compliance indicator separately, we can gain insight into the mechanisms by which the intervention affects outcomes. Combining all indicators into a single measure would make it impossible to discern whether improvements in the outcome are primary driven by pupil attendance at sessions, teacher training attendance, or completion of the baseline assessment.

Despite these variables being binary, we will use linear probability models in the first stage.[17]

First step models for each compliance measures

We will model each compliance measure using a linear regression with a random intercept for schools:

$$Y_{ij} = \beta_0 + \beta_1 Intervention_j + \beta_2 Baseline_{ij} +$$
$$\beta_3 FSM_j + \beta_4 EAL_j + \gamma_j + \varepsilon_{ij},$$

where:

- Y_{ij} represents each compliance measure (B_{ij}, A_{ij}, and T_{ij}):
  - B_{ij}: Indicates whether the i^{th} pupil in the j^{th} school completed the baseline YARC assessment (1) or not (0).
  - A_{ij}: Indicates whether the i^{th} pupil in the j^{th} school attended at least 12 fluency sessions (1) or not (0).

[16] Joshua D. Angrist and Jörn-Steffen Pischke, *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, 2009), https://doi.org/10.2307/j.ctvcm4j72.
[17] Joshua D. Angrist and Alan B. Krueger, "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives* 15, no. 4 (December 2001): 69–85, https://doi.org/10.1257/jep.15.4.69.

- T_{ij}: indicates whether the teacher of the i^{th} pupil in the j^{th} school attended at least 4 training sessions (1) or not (0).

In addition, for each compliance component, we will report the proportion of pupils and/or teachers in the intervention arm who complied and did not comply.[18]

Second stage model

We will the model the primary outcome, YARC reading comprehension standardised score (Y_{ij}), including the predicted compliance measures from the first stage:

$$Y_{ij} = \beta_0 + \hat{B}_{ij} + \hat{A}_{ij} + \hat{T}_{ij} + \beta_2 Baseline_{ij} + \beta_3 FSM_j + \beta_4 EAL_j + \gamma_j + \varepsilon_{ij}$$

where $\hat{B}_{ij}, \hat{A}_{ij}, \hat{T}_{ij}$ are the predicted values from first stage regressions. This model will tell us how much compliance with different components of the intervention affects reading comprehension.

Intra-cluster correlations (ICCs)

We will report the school-level ICCs based on the primary outcome measure. These will be calculated using the primary outcome analysis model, and the primary outcome analysis model with no predictors, accounting for the clustering of pupils within schools (the so-called empty model), and an empty model with EYSFP as the outcome. The formula for calculating the ICC is as below:

$$ICC = \frac{\widehat{Var_{school}}}{\widehat{Var_{school}} + \widehat{Var_{residuals}}}$$

Effect size calculation

We will estimate the effect size of the intervention using an adaptation of Hedges' g,[19] as done in previous EEF efficacy trials involving measures of reading attainment, comprehension, and fluency.[20] Specifically,

$$g = \frac{(\hat{Y}_T - \hat{Y}_C)_{adjusted}}{\sqrt{\sigma_s^2 + \sigma_{error}^2}}$$

where $(\hat{Y}_T - \hat{Y}_C)_{adjusted}$ is the mean difference between treatment and control arms adjusted for baseline EYFSP Literacy – Reading and $\sqrt{\sigma_s^2 + \sigma_{error}^2}$ is an estimate of the treatment and

[18] The EEF's guidance suggests reporting an F-test even for binary compliance indicators. However, we will report only the proportions of compliance, because it is highly likely that the p-values from the F-tests would be significant. This likelihood arises from the nature of our three binary compliance variables: each will be zero in the control arm and have some nonzero proportion in the intervention arm. In other words, the model is almost certain to reject the null hypothesis that the treatment and control arms are identical, making the F-test uninformative. Reporting proportions alone provides a clearer picture of compliance differences.

[19] L. Hedges, "Effect Sizes in Cluster-Randomized Designs," *Journal of Educational and Behavourial Sciences* 34, no. 4 (2007): 341–70, https://doi.org/10.3102/1076998606298043.

[20] S. Dimova and S. Illie, "Peer Assisted Learning Strategies UK: Statistical Analysis Plan.," 2021, PALS_SAP_update_2021.pdf (d2tic4wvo1iusb.cloudfront.net).

control arms unconditional standard deviation (school level σ_s^2 and individual level σ_{error}^2 variance).[21]

We will calculate confidence intervals for the effect size using:

$$\hat{g} - c_{\alpha/2}\gamma_{\hat{g}} \leq \hat{g} \leq \hat{g} + c_{\alpha/2}\gamma_{\hat{g}}$$

where $\hat{g}$ is the estimated effect size, $\gamma_{\hat{g}}$ is the estimated standard error, and $c_{\alpha/2}$ is the critical value from the standard normal distribution corresponding to the desired confidence level (e.g., 1.96 for a 95% confidence interval). We will calculate $\gamma_{\hat{g}}$ using the following formula:

$$\gamma_{\hat{g}} = \omega\sqrt{\left(\frac{SE}{\sqrt{\sigma_s^2 + \sigma_{error}^2}}\right)^2 \gamma + \frac{g^2}{2df}}$$

where SE is the cluster-corrected standard error for the adjusted mean difference $(\hat{Y}_T - \hat{Y}_C)_{adjusted}$, ω is the small-sample correction factor, γ is the small number of clusters adjustment, and df is the degrees of freedom. We will calculate ω, γ, and df following Equation E.3, E.20, and E.21 in What Works Clearinghouse Procedures and Standards Handbook, Version 5.0.[22]

Per EEF's statistical guidance, we will also report Effect Sizes in terms of months of additional progress using EEF's conversion tool.[23]

---

[21] For subgroup analyses (e.g., FSM vs. non-FSM), we will standardise the subgroup-adjusted mean differences by the *subgroup's own unconditional standard deviation* to derive Hedges' g.
[22] U.S. Department of Education, "What Works Clearinghouse Procedures and Standards Handbook, Version 5.0."
[23] EEF, "Statistical Analysis Guidance for EEF Evaluations."

References

Alexander Robitzsch [aut,cre] (<https://orcid.org/0000-0002-8226-3132>), Simon Grund [aut] (<https://orcid.org/0000-0002-1290-8986>), Thorsten Henke [ctb]. “Miceadds: Some Additional Multiple Imputation Functions, Especially for ‘Mice,’” February 19, 2014. https://doi.org/10.32614/CRAN.package.miceadds.

Angrist, Joshua D., and Alan B. Krueger. “Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments.” *Journal of Economic Perspectives* 15, no. 4 (December 2001): 69–85. https://doi.org/10.1257/jep.15.4.69.

Angrist, Joshua D., and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press, 2009. https://doi.org/10.2307/j.ctvcm4j72.

Bates, Douglas, Martin Maechler, Ben Bolker [aut, cre, Steven Walker, Rune Haubo Bojesen Christensen, Henrik Singmann, et al. “Lme4: Linear Mixed-Effects Models Using ‘Eigen’ and S4,” July 3, 2024. https://cran.r-project.org/web/packages/lme4/index.html.

Bell, K., H. Ainsworth, L. Elliott, C. Fairhurst, Louise Gascoine, P. Hanley, C. Hewitt, et al. “ABRACADABRA (ABRA) and Reading and Understanding in Key Stage 1 (RUKS).” EEF, 2022. ABRA-Report-Final.pdf (d2tic4wvo1iusb.cloudfront.net).

Buuren, Stef van, and Karin Groothuis-Oudshoorn. “Mice: Multivariate Imputation by Chained Equations in R.” *Journal of Statistical Software* 45 (December 12, 2011): 1–67. https://doi.org/10.18637/jss.v045.i03.

Coppock, Alexander, J. Cooper, N. Fultz, and G. Blair. “Package ‘Randomizr.’” CRAN, 2023. Available at: https://cran.r-project.org/web/packages/randomizr/randomizr.pdf.

Dimova, S., and S. Illie. “Peer Assisted Learning Strategies UK: Statistical Analysis Plan.,” 2021. PALS_SAP_update_2021.pdf (d2tic4wvo1iusb.cloudfront.net).

EEF. “Statistical Analysis Guidance for EEF Evaluations,” 2022. https://d2tic4wvo1iusb.cloudfront.net/production/documents/evaluation/evaluation-design/EEF-Analysis-Guidance-Website-Version-2022.14.11.pdf?v=1729585898.

Gunzler, Douglas, Tian Chen, Pan Wu, and Hui Zhang. “Introduction to Mediation Analysis with Structural Equation Modeling.” *Shanghai Archives of Psychiatry* 25, no. 6 (December 2013): 390. https://doi.org/10.3969/j.issn.1002-0829.2013.06.009.

Hedges, L. “Effect Sizes in Cluster-Randomized Designs.” *Journal of Educational and Behavourial Sciences* 34, no. 4 (2007): 341–70. https://doi.org/10.3102/1076998606298043.

Pereira, Ricardo Cardoso, Pedro Henriques Abreu, Pedro Pereira Rodrigues, and Mário A. T. Figueiredo. “Imputation of Data Missing Not at Random: Artificial Generation and Benchmark Analysis.” *Expert Systems with Applications* 249 (September 1, 2024): 123654. https://doi.org/10.1016/j.eswa.2024.123654.

Rosseel, Yves, Terrence D. Jorgensen, Luc De Wilde, Daniel Oberski, Jarrett Byrnes, Leonard Vanbrabant, Victoria Savalei, et al. “Lavaan: Latent Variable Analysis,” September 26, 2024. https://cran.r-project.org/web/packages/lavaan/index.html.

U.S. Department of Education. “What Works Clearinghouse Procedures and Standards Handbook, Version 5.0.” U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, ND.

Annexe 1 – Randomisation code

```
---

title: "262400785 Reading Fluency School Randomization"

output: html_document

date: "2024-06-24"

---


###### Reading Fluency School Randomisation #######


## Clear Workspace

rm(list=ls())


## Import required libraries

library(readxl)

library(dplyr)

library(psych)


## Set WD

setwd("….")


## Load school level data


## load school info data <- use an updated school list

# Source: https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics - under 'Additional supporting files'. More information about this dataset is available here: https://explore-education-statistics.service.gov.uk/find-statistics/school-pupils-and-their-characteristics/data-guidance


school_data = read.csv("spc_school_level_underlying_data_19062024.csv")


## Load final school list provided by HFL - inclusive of all schools in the trial # Note - HFL provided an updated data sheet with no schools not included in the trial.

final_trial_data = read_xlsx('Verian HFL EEF RFP trial schools 190624.xlsx')
```

length(unique(final_trial_data$URN))

#[1] 179

Update details of schools based on feedback from selection guidance (05/06/2024)

#1 Update contact details for URN: 146532

final_trial_data$`Contact email` <- ifelse(final_trial_data$URN == 146532,

"", final_trial_data$`Contact email`)

#2 Update contact details for URN:147620

final_trial_data$`Contact email` <- ifelse(final_trial_data$URN == 147620,

"", final_trial_data$`Contact email`)

#3 Update contact details for URN:145560

final_trial_data$`Contact email` <- ifelse(final_trial_data$URN == 145560,

"", final_trial_data$`Contact email`)

#4 Update contact details for URN:110615

final_trial_data$`Contact email` <- ifelse(final_trial_data$URN == 110615, "", final_trial_data$`Contact email`)

##

As of 13/06/2024, 2 schools have dropped out (**HFL have already dropped these schools in the latest dataset used in the randomisation code**):

School 1: Bridge Learning Campus | URN: 139049

#final_trial_data = subset(final_trial_data,!final_trial_data$URN == 139049)

School 2: Queensway Primary School | URN: 107858

#final_trial_data = subset(final_trial_data,!final_trial_data$URN == 107858)

On the 24/06/2024 one additional school dropped out

School 3: Sidney Stringer Primary school | URN: 141938

final_trial_data = subset(final_trial_data,!final_trial_data$URN == 141938)

length(unique(final_trial_data$URN))

[1] 178

Note, in previous versions we used these files (as not all info was available):

#w1: #dsa_data = read_xlsx("EEF Recruitment Tracker - COPY FOR VERIAN.xlsx")

#w2: #trial_schools = read_xlsx("Verian HFL EEF RFP trial schools 010524.xlsx")

```

```{r Trim school data to include only relevant info}

### Select relevant columns from the school info data file - from an initial inspection, these variables appear to match the info shown on GIAS (Note - GIAS appears not to have updated to 2024 school level characteristic data at this point)

# Note - X.. of represents % of

## EAL status - percentage known

## FSM status - known to be eligible - rather than actually taking free school meals (applications could be ongoing)

## school size - headcount at time of data collection - rather than stated size of school

school_data_t = select(school_data,c("urn","school_name","headcount.of.pupils","number.of.pupils.whose.first.language.is.known.or.believed.to.be.other.than.English","X..of.pupils.whose.first.language.is.known.or.believed.to.be.other.than.English","X..of.FSM.eligible.pupils.taking.free.school.meals","X..of.pupils.known.to.be.eligible.for.free.school.meals","X..of.pupils.known.to.be.eligible.for.free.school.meals..Performance.Tables."))

## Rename relevant columns for easier interpretation

school_data_t$FSM = as.numeric(school_data_t$X..of.pupils.known.to.be.eligible.for.free.school.meals)

```
school_data_t$FSM_PT =
as.numeric(school_data_t$X..of.pupils.known.to.be.eligible.for.free.school.meals..Performance.Table
s.)

school_data_t$EAL =
as.numeric(school_data_t$X..of.pupils.whose.first.language.is.known.or.believed.to.be.other.than.Eng
lish)

school_data_t$school_size = as.numeric(school_data_t$headcount.of.pupils)

Join school data with the updated list of trial schools in our trial

paired_school_data = left_join(final_trial_data,school_data_t, by = c("URN" = "urn"))

check to see if there are NAs in any important categories

sum(is.na(paired_school_data$FSM))

[1] 1

sum(is.na(paired_school_data$FSM_PT))

[1] 1

sum(is.na(paired_school_data$EAL))

[1] 1

sum(is.na(paired_school_data$school_size))

[1] 1

sum(is.na(paired_school_data$`EIA school`))

[1] 0

we see that 1 school has NAs for FSM, EAL and school size - id that school

missing_school = subset(paired_school_data,is.na(paired_school_data$FSM))

missing_school$`School name`

#[1] "Dale Hall Community Primary School"

Update Dale hale data with new URN#
```

```
## School 1: Dale Hall Community Primary School (URN: 150424)

## according to the GIAS website (linked below) this school previously had the URN: 124668

## https://get-information-schools.service.gov.uk/Establishments/Establishment/Details/150424#school-links:  124668

Dale_Hale_s = subset(school_data_t,school_data_t$urn == 124668)

Dale_Hale_s$school_name

## Create a copy of the original URN for communications with school

final_trial_data$legacy_urn = final_trial_data$URN

## Update a 'replacement' URN in our records for pairing data between datasets

final_trial_data$legacy_urn = if_else(final_trial_data$legacy_urn ==150424,124668,final_trial_data$legacy_urn )

### Join school data with the updated list of trial schools in our trial

paired_school_data = left_join(final_trial_data,school_data_t, by = c("legacy_urn" = "urn"))

## check to see if there are NAs in any important categories

sum(is.na(paired_school_data$FSM))

# [1] 0

sum(is.na(paired_school_data$FSM_PT))

# [1] 0

sum(is.na(paired_school_data$EAL))

# [1] 0

sum(is.na(paired_school_data$school_size))

# [1] 0
```

```
sum(is.na(paired_school_data$`EIA school`))

# [1] 0


# Check correlation between the two FSM measures

cor(paired_school_data$FSM,paired_school_data$FSM_PT)

# [1] 0.9955563

# The two measures of FSM are practically the same, so we can use either one


## Re-format data for readability


clean_dataset = select(paired_school_data, c("URN","legacy_urn","DfE number","School name","Project contact name","Contact email", "Contact phone number","DSA returned","EIA school","FSM","EAL","school_size"))


## Mark any schools that have had an updated URN - we will need to ensure we use this URN in communications with schools

clean_dataset$re_assigned_school = if_else(clean_dataset$URN != clean_dataset$legacy_urn,1,0)


table(clean_dataset$re_assigned_school)

#   0   1

# 177   1


subset(clean_dataset, clean_dataset$re_assigned_school==1)$`School name`

# [1] "Dale Hall Community Primary School"


##### Verify that all schools have values for FSM, EAL, School size and EIA area #####


sum(is.na(clean_dataset$FSM))

#[1] 0


sum(is.na(clean_dataset$EAL))
```

````
#[1] 0


sum(is.na(clean_dataset$school_size))

#[1] 0


sum(is.na(clean_dataset$`EIA school`))

#[1] 0

```


```{r}

##### Randomization #####


## Using the 'clean dataset', conduct the randomization


# Import package specified in TP

library(randomizr)


### Ensure relevant vars are treated as numeric / factor where relevant

clean_dataset$FSM = as.numeric(clean_dataset$FSM)

clean_dataset$EAL = as.numeric(clean_dataset$EAL)

clean_dataset$school_size = as.numeric(clean_dataset$school_size)

clean_dataset$`EIA school` = as.factor(clean_dataset$`EIA school`)

### Determine quartiles for FSM % (for randomization) , EAL % (for randomization) and School size
(to check distribution)


### FSM ###

# Derive quartiles for FSM

clean_dataset$fsm_quartiles <- cut(clean_dataset$FSM,

          breaks = quantile(clean_dataset$FSM, probs = seq(0, 1, by = 0.25), na.rm = TRUE),

           include.lowest = TRUE,

           labels = c(1, 2, 3, 4))
````

```
# Create a scatter plot
plot_fsm_quartiles <- function(data) {

  # Create the scatter plot
  plot(data$FSM, data$fsm_quartiles, main = "FSM% vs FSM Quartile Ranking",
       xlab = "FSM%", ylab = "FSM Quartiles", col = "red")

  # Calculate quartile boundaries
  quartile_boundaries <- quantile(data$FSM, probs = seq(0, 1, by = 0.25))

  # Add vertical lines at quartile boundaries
  abline(v = quartile_boundaries[2], col = "blue", lty = 2)  # 1st quartile
  abline(v = quartile_boundaries[3], col = "blue", lty = 2)  # 2nd quartile (median)
  abline(v = quartile_boundaries[4], col = "blue", lty = 2)  # 3rd quartile
}

plot_fsm_quartiles(clean_dataset)

## Check fsm quartiles are equally distributed
table(clean_dataset$fsm_quartiles,useNA = 'always')

  #  1    2    3    4 <NA>
  # 46   43   44   45    0

### EAL
clean_dataset$eal_quartiles <- cut(clean_dataset$EAL,
                    breaks = quantile(clean_dataset$EAL, probs = seq(0, 1, by = 0.25), na.rm = TRUE),
                      include.lowest = TRUE,
                      labels = c(1, 2, 3, 4))

## Plot the distribution of EAL% and EAL quartile ranking
```

```
plot_eal_quartiles <- function(data) {

  # Create the scatter plot
  plot(data$EAL, data$eal_quartiles, main = "EAL% vs EAL Quartile Ranking",
       xlab = "EAL%", ylab = "EAL Quartiles", col = "grey")

  # Calculate quartile boundaries
  quartile_boundaries <- quantile(data$EAL, probs = seq(0, 1, by = 0.25))

  # Add vertical lines at quartile boundaries
  abline(v = quartile_boundaries[2], col = "black", lty = 2)  # 1st quartile
  abline(v = quartile_boundaries[3], col = "black", lty = 2)  # 2nd quartile (median)
  abline(v = quartile_boundaries[4], col = "black", lty = 2)  # 3rd quartile
}

plot_eal_quartiles(clean_dataset)



## Check eal quartiles are equally distributed
table(clean_dataset$eal_quartiles,useNA = 'always')
  # 1    2    3    4 <NA>
  # 46   43   44   45    0


### School size - Create this variable to determine whether school size is approximately equally distributed post-randomization. # Note we are not randomizing by school size officially


clean_dataset$school_size_quartiles <- cut(clean_dataset$school_size,
                    breaks = quantile(clean_dataset$school_size, probs = seq(0, 1, by = 0.25), na.rm =
TRUE),
                     include.lowest = TRUE,
                     labels = c(1, 2, 3, 4))
```

```
## Plot the distribution of school size and quartile ranking

plot(clean_dataset$school_size,clean_dataset$school_size_quartiles)


## check school size quartiles is equally distributed

table(clean_dataset$school_size_quartiles,useNA = 'always')

 #  1    2    3    4 <NA>

 # 45   44   44   45    0


##### Randomize #####


### set the seed


set.seed(262400785)#  project number as a seed: 262400785


# according to randomizR package, create unique blocks for each combination of our two stratification variables : FSM% and EAL%


clean_dataset$stratum <- paste(

  clean_dataset$fsm_quartiles,

  clean_dataset$eal_quartiles,

  sep = "_")


clean_dataset$stratum

#   [1] "1_1" "2_2" "4_4" "3_2" "2_1" "2_2" "2_3" "1_2" "3_4" "1_2" "2_2" "2_2" "1_3" "2_4" "3_4" "4_3" "4_3" "4_3"

#  [19] "4_4" "3_4" "1_1" "1_3" "3_4" "3_1" "3_4" "2_4" "4_1" "4_1" "3_2" "3_1" "1_2" "4_3" "1_1" "2_3" "1_2" "1_2"

#  [37] "3_2" "2_4" "1_2" "2_1" "1_1" "4_1" "3_4" "1_3" "1_2" "1_1" "2_4" "2_4" "3_1" "4_3" "3_4" "3_3" "2_3" "3_4"

#  [55] "3_2" "4_1" "2_1" "1_2" "3_4" "3_4" "3_2" "3_1" "1_4" "1_2" "1_4" "2_1" "1_1" "2_3" "1_3" "1_4" "2_2" "1_1"

#  [73] "3_2" "4_2" "1_1" "1_3" "4_3" "1_1" "2_1" "2_1" "4_3" "3_3" "4_4" "4_2" "3_4" "1_1" "4_3" "2_1" "3_3" "1_2"
```

```
#  [91] "4_4" "3_4" "4_4" "2_1" "3_4" "1_2" "2_1" "4_4" "3_2" "4_3" "3_4" "3_2" "4_2" "3_4" "1_2"
"3_1" "3_4" "4_3"

# [109] "3_2" "1_1" "2_1" "4_4" "1_1" "1_1" "1_1" "3_3" "4_4" "4_3" "4_2" "2_2" "3_3" "2_4" "4_4"
"3_3" "1_1" "1_3"

# [127] "4_4" "2_4" "1_3" "1_1" "4_4" "4_3" "2_4" "2_3" "4_3" "1_2" "2_1" "1_2" "4_3" "4_3" "4_4"
"2_3" "2_3" "3_3"

# [145] "4_4" "2_1" "1_2" "3_3" "1_1" "1_3" "2_2" "1_1" "3_1" "2_1" "4_4" "4_1" "4_2" "2_4" "2_2"
"3_2" "4_2" "4_4"

# [163] "2_1" "4_2" "3_3" "1_4" "2_3" "4_3" "2_1" "3_3" "4_2" "1_1" "3_3" "2_1" "4_2" "2_4" "3_3"
"2_2"


table(clean_dataset$stratum)

# 1_1 1_2 1_3 1_4 2_1 2_2 2_3 2_4 3_1 3_2 3_3 3_4 4_1 4_2 4_3 4_4

#  19  15   8   4  16   9   8  10   6  10  12  16   5   9  16  15


length(unique(clean_dataset$stratum))

# [1] 16


### Leads to 16 unique sets for randomisation



### Per the Trial Protocol, we will conduct randomisation on an anonymised school data set. Create a random 8 digit identifier for the purpose of randomisation and temporarily drop school specific identifiers ###

# Function to generate unique 8-digit identifiers

generate_unique_ids <- function(n) {

  # Generate a sequence of unique integer IDs

  ids <- sample(1e7:(1e8 - 1), n)


  return(ids)

}


# Generate the identifiers

clean_dataset$random_id <- generate_unique_ids(nrow(clean_dataset))
```

```
length(unique(clean_dataset$random_id))

# 178


id_data <- select(clean_dataset, c('random_id','URN','School name','DfE number','legacy_urn','Project contact name','Contact email','Contact phone number'))

random_data = select(clean_dataset, -c('URN','School name','DfE number','legacy_urn','Project contact name','Contact email','Contact phone number'))


###### Conduct the actual randomisation #####


### use block_ra to randomly assign within blocks (groups of similar FSM / EAL)

random_data$treatment <- block_ra(

  blocks = random_data$stratum,

  prob_each = c(0.5,0.5))


##### Post-randomisation, check vars are relatively equally distributed across treatment and control #####


## Check arms are equally distributed

table(random_data$treatment)

#  0  1

# 89 89


## check treatment is relatively balanced across stratum

table(random_data$treatment,random_data$stratum)

  #   1_1 1_2 1_3 1_4 2_1 2_2 2_3 2_4 3_1 3_2 3_3 3_4 4_1 4_2 4_3 4_4

  # 0  10   7   4   2   8   5   4   5   3   5   6   8   3   4   8   7

  # 1   9   8   4   2   8   4   4   5   3   5   6   8   2   5   8   8


## Check FSM, EAL, School size and EIA are balanced

## FSM
```

```
table(random_data$treatment,random_data$fsm_quartiles)

 #    1  2  3  4

 # 0 23 22 22 22

 # 1 23 21 22 23


## EAL

table(random_data$treatment,random_data$eal_quartiles)

 #    1  2  3  4

 # 0 24 21 22 22

 # 1 22 22 22 23


## School size

table(random_data$treatment,random_data$school_size_quartiles)

 #    1  2  3  4

 # 0 22 19 24 24

 # 1 23 25 20 21


## EIA

table(random_data$treatment,random_data$`EIA school`)

 #   No Yes

 # 0 43  46

 # 1 41  48


describeBy(random_data$FSM,random_data$treatment)

describeBy(random_data$EAL,random_data$treatment)

describeBy(random_data$school_size,random_data$treatment)



##### post-randomisation, rejoin with ID vars #####


clean_dataset = merge(id_data,random_data)
```

```
# Final checks comparing whether tables produced by two datasets are identical
table_treatment_random_data <- table(random_data$treatment)
table_treatment_clean_dataset <- table(clean_dataset$treatment)

identical(table_treatment_random_data, table_treatment_clean_dataset)
# [1] TRUE


table_stratum_random_data <- table(random_data$treatment, random_data$stratum)
table_stratum_clean_dataset <- table(clean_dataset$treatment, clean_dataset$stratum)

identical(table_stratum_random_data, table_stratum_clean_dataset)
# [1] TRUE


table_fsm_random_data <- table(random_data$treatment,random_data$fsm_quartiles)
table_fsm_clean_dataset <- table(clean_dataset$treatment,clean_dataset$fsm_quartiles)

identical(table_fsm_random_data, table_fsm_clean_dataset)
# [1] TRUE


table_eal_random_data <- table(random_data$treatment,random_data$eal_quartiles)
table_eal_clean_dataset <- table(clean_dataset$treatment,clean_dataset$eal_quartiles)

identical(table_eal_random_data, table_eal_clean_dataset)
# [1] TRUE


table_school_size_random_data <-
table(random_data$treatment,random_data$school_size_quartiles)
```

```
table_school_size_clean_dataset <-
table(clean_dataset$treatment,clean_dataset$school_size_quartiles)


identical(table_school_size_random_data, table_school_size_clean_dataset)

# [1] TRUE



table_eia_random_data <- table(random_data$treatment,random_data$`EIA school`)

table_eia_clean_dataset <- table(clean_dataset$treatment,clean_dataset$`EIA school`)


identical(table_eia_random_data, table_eia_clean_dataset)

# [1] TRUE



descriptives_fsm_random_data <- describeBy(random_data$FSM,random_data$treatment)

descriptives_fsm_clean_Data <- describeBy(clean_dataset$FSM,clean_dataset$treatment)


identical(descriptives_fsm_random_data, descriptives_fsm_clean_Data)

# [1] TRUE



descriptives_eal_random_data <- describeBy(random_data$EAL,random_data$treatment)

descriptives_eal_clean_Data <- describeBy(clean_dataset$EAL,clean_dataset$treatment)


identical(descriptives_eal_random_data, descriptives_eal_clean_Data)

# [1] TRUE



descriptives_school_size_random_data <-
describeBy(random_data$school_size,random_data$treatment)

descriptives_school_size_clean_Data <-
describeBy(clean_dataset$school_size,clean_dataset$treatment)
```

```
identical(descriptives_school_size_random_data, descriptives_school_size_clean_Data)
# [1] TRUE

# School allocations, where 0 is control and 1 is treatment
table(clean_dataset$random_id,clean_dataset$treatment)
 #          0 1
 # 10062444 0 1
 # 10237674 0 1
 # 10846362 0 1
 # 11406992 1 0
 # 11417591 0 1
 # 11725297 1 0
 # 12322418 0 1
 # 12453291 1 0
 # 12795192 0 1
 # 13792461 1 0
 # 14397830 0 1
 # 14934385 0 1
 # 15147216 1 0
 # 15789464 1 0
 # 16495649 1 0
 # 16766069 0 1
 # 17099985 1 0
 # 17806110 1 0
 # 17844888 1 0
 # 18259827 1 0
 # 18451854 0 1
 # 18654885 1 0
 # 18699946 0 1
 # 20045798 1 0
```

21044459 1 0

21943342 0 1

22390828 1 0

22770040 0 1

22784240 0 1

22939874 1 0

23505685 1 0

23612441 1 0

24373930 1 0

24547332 1 0

25134016 0 1

26564652 1 0

27692880 1 0

27710684 0 1

28180720 1 0

28328929 1 0

28638208 0 1

29140848 0 1

29286639 0 1

30052612 1 0

30616250 0 1

30758778 1 0

31499562 1 0

31756783 1 0

32237587 0 1

33158112 1 0

33280437 1 0

34338329 0 1

34606763 1 0

34862200 1 0

34880941 1 0

35559738 1 0

35623923 1 0

37395001 0 1

37850464 0 1

38322245 1 0

38632880 0 1

39058892 0 1

39540087 0 1

39993460 0 1

40099971 0 1

41400065 1 0

42122211 0 1

42280260 1 0

42804824 1 0

42962173 0 1

44190100 1 0

44886068 0 1

45037727 0 1

45207859 1 0

45381881 1 0

45509797 0 1

45851292 1 0

46026790 1 0

46415329 1 0

46452979 0 1

46663257 1 0

47155853 0 1

47565004 1 0

49393281 0 1

49884725 1 0

50641241 0 1

50772501 0 1

50892473 1 0

52328604 1 0

52947780 1 0

53994524 0 1

54112012 0 1

54285308 0 1

54552951 0 1

54836278 1 0

56260316 0 1

56656349 1 0

57589762 1 0

57713059 0 1

57813811 0 1

58457962 1 0

58582884 1 0

58761042 0 1

59145156 1 0

59992154 0 1

60206213 0 1

60393841 1 0

61146596 1 0

61689411 1 0

61714812 1 0

61769006 1 0

62096732 0 1

63135859 1 0

63138436 0 1

63492046 0 1

63500304 1 0

64089925 1 0

65885486 0 1

66037936 1 0

66241912 1 0

66260262 0 1

66792778 0 1

67113808 1 0

67879307 1 0

68674831 1 0

69223095 1 0

69453725 0 1

69843545 0 1

70508769 0 1

71480329 0 1

71612041 0 1

72212343 0 1

73405435 0 1

73971073 1 0

74448615 0 1

74888000 1 0

75685980 1 0

75821332 1 0

76453979 0 1

77301986 1 0

77646128 0 1

77657222 0 1

78133691 0 1

78282876 0 1

79352222 0 1

79769434 1 0

79895719 0 1

80853143 0 1

81693071 0 1

82050438 0 1

82206652 0 1

82299023 1 0

82948563 0 1

83436892 0 1

83962162 0 1

84801733 1 0

84834070 0 1

85384069 1 0

85620202 0 1

86448116 1 0

86476824 1 0

86904327 1 0

87706241 0 1

89304989 0 1

90848754 1 0

90880830 0 1

91230104 0 1

91534619 0 1

94054880 0 1

94688285 1 0

95126885 0 1

95917353 1 0

96902397 0 1

97428193 1 0

98334571 0 1

98831580 0 1

99081059 1 0

99361833 1 0

````
### Create a dataset of emails for mail merge (25/06/2024)

# write.csv(clean_dataset,"262400785_Reading_Fluency_school_trial_list_25_06_24_v3.csv")



# Final sense check

Sense_check = read.csv("262400785_Reading_Fluency_school_trial_list_25_06_24_v3.csv")


Sense_check_clean <- subset(Sense_check, select = -X)


# Let's compare both datasets

print(names(as.data.frame(lapply(clean_dataset, as.character))))

print(names(as.data.frame(lapply(Sense_check_clean, as.character))))


# The two datasets are identical

identical(as.data.frame(lapply(clean_dataset, as.character)),
as.data.frame(lapply(Sense_check_clean, as.character)))

# [1] TRUE


```
````
````