

**Reciprocal Reading effectiveness trial
Evaluation Protocol**
Evaluator (institution): Behavioural Insights
Team
Principal investigator(s): Patrick Taylor



Template last updated: December 2022

Evaluation summary

Project title	Reciprocal Reading effectiveness trial
Developer (Institution)	FFT (Fischer Family Trust)
Evaluator (Institution)	Behavioural Insights Team
Principal investigator(s)	Patrick Taylor
Protocol author(s)	Chiara Cappellini, Neus Torres Blas, Patrick Taylor
Trial design	Two-arm cluster randomised controlled trial with random allocation at school level
Trial type	Effectiveness
Pupil age range and Key stage	Key Stage (KS) 2 Years 5 and 6 (Ages 9 to 11)
Number of schools (at design stage)	300
Number of pupils (at design stage)	3600
Primary outcome measure and source	New Group Reading Test (NGRT) overall test score
Secondary outcome measure and source	NGRT passage comprehension score NGRT sentence completion score KS2 SATs reading score

Contents

Evaluation summary.....	1
Study rationale and background.....	4
Intervention.....	6
Impact evaluation design	15
Implementation and process evaluation (IPE) design.....	31
Cost evaluation design.....	43
Ethics and registration.....	45
Data protection.....	46
Personnel.....	50
Risks.....	51
Timeline	54
References	57
Appendix A: Changes since the previous EEF evaluation	59

Tables

Table 1: Role Descriptors for the project	8
Table 2: Time, location and duration of intervention activities	9
Table 3: Core components and acceptable adaptations	10
Table 4: Trial design	15
Table 5: Sample size calculations	22
Table 6: Key considerations for IPE research questions	32
Table 7: Sampling frame for the case studies	38
Table 8: IPE methods overview	40
Table 9: Project personnel	51
Table 10: Risk assessment & mitigation	52
Table 11: Timeline	55

Figures

Figure 1: Logic model: Training for teacher/TAs and school coordinators	13
Figure 2: Logic model: Reciprocal Reading pupil outcomes	14
Figure 3: Diagram of the NGRT online assessment structure	19
Figure 4: Triangulation of retrospective survey with treatment schools and case studies	36

Study rationale and background

Policy context

One of the Department for Education (DfE)'s current priorities is to 'level up education standards so that children and young people in every part of the country are prepared with the knowledge, skills and qualifications they need' (Department for Education, 2022).

This is an area where considerable work is needed. Key Stage 1 (KS1) reading attainment has reduced substantially in England between 2019 and 2022. In the 2021/22 academic year, 67% of pupils met the expected standard in reading, down from 75% in 2019 (Department for Education, 2022). During this period, pupils from disadvantaged backgrounds¹ have fallen even further behind in reading attainment than pupils not known to be disadvantaged (an 11 percentage point drop compared to a 6 percentage point drop) (ibid).

To help solve this problem, the Department for Education has re-endowed the Education Endowment Foundation (EEF) to produce high quality research on what works to improve attainment in schools, including literacy. As part of this vision, the DfE's Accelerator Fund has funded EEF research, including this effectiveness trial of the Reciprocal Reading targeted programme. One aim of the fund is to scale evaluations of educational interventions that have shown promise in previous trials. To fully evaluate whether these interventions are effective gap closers, the EEF will focus on scaling projects in the Education Investment Areas.²

Strength of evidence in this area

The following section outlines the solid foundation of evidence demonstrating the potential of Reciprocal Reading. However, the evidence-base lacks large-scale trials to consolidate its status as an effective intervention, achieving positive impact which can be consistently reproduced. This gap in the evidence-base explains why the EEF scores the intervention 3 out of 5 on its 'evidence strength' measure (EEF, 2019; EEF, 2020).

Until recently, there was a lack of experimental evidence rigorously evaluating the impact of Reciprocal Reading as a clearly defined intervention in the UK context. After being developed in New Zealand in the 1980s, the approach was evaluated primarily in the US and Australia. Specifically, rigorous quantitative evidence was collated to show the impact of reciprocal teaching, which has many parallels to Reciprocal Reading as an approach to improve educational attainment. In 1994, the available evidence from the international literature was integrated in a meta-analysis of 16 quantitative studies that computed an average effect size of +0.32 on reading comprehension (using ad hoc tests rather than standardised tests) (Rosenshine and Meister, 1994). However, while this meta-analysis shows that reciprocal approaches showed promise for improving educational attainment,

¹ The Department for Education defines a disadvantaged pupil as pupils who have been eligible for Free School Meals over the last 6 years or have spent at least one day in care with the Local Authority.

² The full list of Education Investment Areas can be found in the UK Government website: <https://www.gov.uk/government/publications/education-investment-areas/education-investment-areas>

they cannot be used as surrogates for studies evaluating Reciprocal Reading as a neatly defined intervention.

In 2014, the EEF conducted the first UK trial of Reciprocal Reading as a discrete intervention. It was a small-scale efficacy trial with 41 secondary schools, which yielded inconclusive results (Crawford and Skipp, 2014). Despite finding a small positive effect on Year 7 pupils, causality could not be inferred due to issues with outcome testing, high attrition rates and unbalanced samples (ibid). In 2019, the EEF and Queen's University Belfast conducted an efficacy trial to test Reciprocal Reading for Key Stage 2 (KS2) pupils in 98 English primary schools. This study investigated the effects of an intervention on the reading skills and comprehension of Year 4 pupils as a whole class, and as a 'targeted intervention' delivered to a smaller group of 12 Year 5 and 6 pupils (aged 9-12) who were identified as good at decoding but poor at comprehension. Although the results did not show evidence of impact on improvement, in months of progress, on reading skills and comprehension for the whole class, the intervention showed promise when implemented as a targeted intervention for the smaller targeted group. The targeted intervention improved overall reading scores by +0.24 standard deviations, as measured by the NGRT standardised tests (O'Hare et al., 2019). The targeted intervention had an even larger effect on pupils eligible for free school meals who experienced an improvement of +0.28 standard deviations in reading comprehension, equivalent to 2 additional months of progress (ibid).

Between 2021 and 2022, two more UK-based randomised controlled trials evaluated the impact of Reciprocal Reading on older pupils aged 11-12 years. The first included 315 secondary school pupils in 14 schools from areas of high social-economic deprivation in England (Thurston et al., 2020). Results were consistent with the 2019 efficacy trial, estimating a significant effect of +0.19 standard deviations on overall reading.³ Despite showing that these effect sizes and directions can be sustained across different settings and adaptations, the authors highlighted the need for a larger scale trial to see if effects generalise to a broader population (ibid). The second studied a small sample of 800 pupils over 20 English schools, and results are yet to be published (Cockerill et al., 2022).

Building on this evidence, a 2021 mixed-methods process evaluation conducted across 35 schools in an area of high deprivation suggested the intervention could be rolled out with high fidelity across KS1 and KS3 (Cockerill et al., 2021). The study claims to provide evidence that the intervention can be implemented in a variety of settings and in regions of high deprivation. However, due to the relatively small sample of schools in the study, these claims need to be assessed with caution.

Why is this evaluation needed?

A large-scale trial is needed to determine with confidence whether the targeted version of Reciprocal Reading can be impactful at scale across England. The EEF has commissioned the Behavioural Insights Team (BIT) to answer this outstanding question by independently evaluating Reciprocal Reading as a targeted intervention in 300 English primary schools. This trial protocol explains the proposed research design. We will run a two-arm clustered randomised controlled trial (RCT) in 300 primary schools across England, alongside a mixed-methods implementation and process evaluation (IPE). In this effectiveness trial, the aim is to test the targeted intervention at scale with minimal interaction from the developers

³ Measured with the NGRT tests, adapted for secondary school pupils.

to observe if it works in everyday conditions in a large number of diverse settings. To allow for a clear comparison of results between this study and the efficacy trial that preceded it, we will replicate aspects of the efficacy trial design, in particular the outcome measures.

Intervention

This intervention description uses an adapted version of the TiDieR framework (Hoffman et al., 2014) for its structure. An intervention description was developed by the Reciprocal Reading team at Fischer Family Trust (FFT) and researchers at Queen's University Belfast for the efficacy trial of Reciprocal Reading (O'Hare et al., 2019:11). This new version has been created to add detail and to reflect key changes to the programme since the efficacy trial.

Why does the programme exist?

The goal

The programme aims to improve pupils' reading comprehension and, in the longer term, their overall literacy.

The theory

It addresses this goal by teaching children to use four key reading comprehension strategies: predicting, clarifying, questioning and summarising. These strategies help children to understand the texts that they are reading in the intervention, and also provide a framework for them to apply to texts (both fiction and non-fiction) that they read in the future.

Who are the participants?

Participants need to meet the following criteria to be eligible for the programme.

Baseline demographics

Aged 9 to 11 (primary school Years 5 and 6).

Baseline reading proficiency

The intervention is designed for pupils who have good decoding skills but poor comprehension skills. These are pupils who struggle to make sense of the texts they read, particularly struggling to understand anything which is not explicitly stated in the text. The teacher or teaching assistant (TA) will identify these pupils based on their perceptions of the pupil's reading proficiency and on a checklist of characteristics provided by FFT as part of the introductory webinar, which includes a guidance document on screening pupils.

Teachers and TAs are asked to supplement their perceptions with existing school assessments, looking for results showing high scores on reading accuracy but significantly lower on reading comprehension, and using the Simple View of reading model to reflect upon the information gathered (Gough & Tunmer, 1986). FFT provides a guidance document on how to make an informed selection, but is not compulsory for schools to use.

The rationale for these eligibility criteria is that there are already explicit interventions for children whose decoding skills are weak, and children who are average or good readers can

access whole class teaching easily and read independently. Children who can decode but struggle to understand are a hidden group who are speculated to be functioning just below age-related expectations and not reading for pleasure. Intervention developers hypothesise that pupils who can decode but struggle to comprehend passages may find it harder to draw pleasure from reading, and thus may read less - which in turn may affect their reading attainment (Clark and De Zoya, 2011). This creates a cyclical issue, which links poor comprehension to less engagement with reading, leading to lower reading frequency, resulting in lower reading attainment. By identifying these readers, teaching staff are supported to think about their needs and begin to address their weaknesses, improving their ability to read with understanding and to read for pleasure.

What materials are needed?

FFT's Reciprocal Reading trainers will provide the teachers and TAs delivering the sessions and school coordinators with a resource pack that will help them deliver the intervention in school. The pack will contain the following:

1. A handbook that contains:
 - a. A summary of the evidence behind the intervention
 - b. Detailed descriptions of the 4 strategies
 - c. Formats and templates for planning sessions
 - d. Examples of session plans
 - e. Advice on how to choose appropriate texts for pupils to read
 - f. Support materials for training days 1 and 2
 - g. Notes on the core components of the training, cross-referenced with the materials given to schools
 - h. Advice on how to conduct follow-up support meetings with schools, including frequently asked questions
2. Advice on how to choose the pupils for the programme
3. A manual on programme implementation, outlining the main expectations of the programme for the school when delivering the programme, the delivery timeline and the roles and responsibilities for the school staff involved in the programme
4. An anthology of short stories to use in the sessions
5. Dictionaries

In addition to the resource pack, schools are provided with access to a website with video material of lessons and additional planning examples.

What does the programme entail?

The programme has two key components:

1. A reading programme for pupils
2. Training and support for teachers and TAs who deliver the reading programme

These are described in turn below.

Reading programme for pupils

The programme for pupils involves a series of reading sessions for small groups (approx. 6 pupils per group, with a minimum of 4 pupils and maximum of 8). In each session, pupils read part of a text, individually, in pairs or as a whole group. Pupils are supported to read the text using four reading strategies:

1. **Predict:** Pupils make a prediction about the text; for example, what will happen to a character in a story.
2. **Clarify:** Pupils identify words or phrases that they do not fully understand and work out their meaning through discussing the unfamiliar words in context.
3. **Question:** Pupils ask questions of the text to deepen their understanding; for example, why a character is behaving in a certain way, and are encouraged to answer the questions through group discussion.
4. **Summarise:** Either the teacher/TA leading the session or a pupil will summarise what has been read and other pupils will be invited to add information.

The aim is for pupils to be taken through this cycle of strategies 3 or 4 times per session. The texts that pupils read in the sessions can come from the anthology of short stories provided in the programme materials or be chosen by the teacher/TA following the guidance in the resource materials.

Training and support for school coordinators and teacher/TAs:

The teachers and TAs who coordinate and deliver the reading programme are given the following training and support:

1. **Programme briefing:** An initial briefing for the teacher responsible for coordinating the intervention in their school that includes guidance on how to choose the pupils for the intervention.
2. **Training:** Two days of training for teachers and TAs responsible for delivering the programme with pupils.
 - a. Day 1 covers the principles behind the intervention, the programme process, resources for delivery, planning for implementation, and guidance and tools for quality assurance (QA).
 - b. Day 2 focuses on extending practice so that it continues to meet the needs of more experienced and improving readers by challenging them further.
3. **Online support meetings:** To provide guidance on implementing and sustaining the intervention.
4. **Ad-hoc support:** An email and telephone helpline which School Coordinators can contact for advice on implementation.

Who delivers the intervention?

Table 1 gives an overview of the roles involved in delivery of the intervention.

Table 1: Role Descriptors for the project

Role title	Role description	Organisation
------------	------------------	--------------

School Coordinator	A member of teaching staff (ideally a senior manager): • Manages the intervention in a school. • Supports the selection of teacher/TA. • Support the selections of target pupils, which will be led by their current teacher or another member of the SLT. • Monitors attendance to ensure minimum dosage is delivered. • Offers ad hoc support to teacher/TAs on session delivery. • Advocates for the intervention with senior management and board of governors. • Encourages implementation of Reciprocal Reading practices after the intervention has ended.	School
Teacher/TA	A teacher or TA: • Delivers the programme reading sessions with pupils. Keeps attendance register.	School
FFT Trainer	• Delivers the training and support programme for School Coordinators and Teacher/TAs.	FFT
FFT School Support Team	• Provides ad hoc email and telephone advice to School Coordinators and Teacher/TAs. This would then be triaged to the appropriate support resource.	FFT
FFT Senior Trainer	• Trains the FFT Trainers. • Facilitates the community of practice for FFT Trainers. • Monitors implementation data and the quality of delivery in schools.	FFT

How is the programme delivered?

The **programme for pupils** is delivered in curricular time. Each school will decide how to integrate it into the timetable, but FFT recommends avoiding teaching them in place of curricular English lessons or individual reading lessons. In fact, the intervention should not replace other curricular lessons or activities generally even from other subjects.

The **training for teachers and TAs** is delivered in addition to schools' standard programme of in-service training (INSET) days.

When, where and how much?

Table 2 summarises the location, timing and duration of all programme activities.

Table 2: Time, location and duration of intervention activities

Activity	When	Where	Duration
Reading programme for pupils			
Reading sessions	2 sessions per week, for 12 weeks, during the Autumn term of 2023 and Spring term of	In-school, in a reasonably quiet and uninterrupted working environment	20-30 mins per session

	2024		
Training and support for teachers and TAs			
Programme briefing	April to September 2023	Online	90 mins
Training day 1	October and November 2023	Out of school training facility	1 day
Training day 2	February 2024	Out of school training facility	1 day
Support meetings	One between training day 1 and 2. One after training day 2 (end of Spring term / beginning of summer term).	Online	2 x 60 mins
Ad-hoc support	When needed during delivery period	Email or telephone	As needed

Tailoring

Table 3 indicates which of the programme activities are core (should be delivered always) and what kind of adaptations are expected and acceptable.

Table 3: Core components and acceptable adaptations

Activity	What's core?	Acceptable adaptations
Reading programme for pupils		
Reading sessions	Core delivery: 2 weekly 20-minute sessions over 12 weeks - At least 4 pupils per group (with a maximum of 8) - Sessions follow the 4 strategies structure, repeated 3 or 4 times per session - Must read an appropriate text as defined in the intervention guide - All pupils use a dictionary - Working environment must be reasonably quiet and uninterrupted Core management: - Selecting pupils according to criteria - Monitoring of session delivery and attendance - Ad hoc support from School Coordinators to Teacher/TAs on session delivery - School Coordinator must support timetabling - Minimum of half-termly meetings between School Coordinators and	Delivery adaptations: - Delivering more than the minimum dosage (in terms of session length and number) - Pupil groups can either be separated by year group or be mixed year group (within pupil eligibility criteria) - Minimal changes to the small group composition, always following pupil selection guidance for newcomers to avoid altering the small group dynamic. - Reading texts not in the anthology, using text selection guidelines - The extent to which ad-hoc programme support is drawn upon - The extent to which teachers and TAs use the 4 reading strategies in other subjects (embeddedness) Management adaptations: - Level and type of QA and support from School Coordinators - Whether or not School Coordinator does intervention profile raising work with SLT and board of governors

	Teacher/TAs	
Training and support for teachers and TAs		
Programme briefing	• All of session • School coordinator and sessions leads must attend •	• Replacement teachers/TAs receive online training from FFT
Training day 1	• All of day	• Replacement teachers/TAs receive online training from FFT
Training day 2	• All of day	• Replacement teachers/TAs receive online training from FFT
Support meetings	• Both meetings	• Can last between 60-90 mins depending on need • Meetings can happen in-person at the school if the trainer lives nearby
Ad-hoc support	• Must be available to all schools	• Optional for schools to use

Tailoring delivery to scale: Training the trainers

To scale the intervention delivery, FFT will train a new cohort of trainers. While this is not a key component of the intervention, it is a key activity to tailor the intervention to this effectiveness trial. Before delivering training in schools, trainers will complete the following development activities:

1. **Initial training:** 3.5 days training on the principles of Reciprocal Reading, and how to deliver the training and support programme for teachers and TAs.
2. **Observations:** At least one observation of an experienced trainer delivering the two days of training for teachers and TAs.
3. **Co-training:** Delivering at least one full training programme for teachers and TAs alongside an experienced trainer.
4. **Independent delivery of training pre-trial:** Delivering the full training programme for teachers and TAs at least once before starting training trial schools.
5. **Community of practice:** Facilitated discussions to share practice between trainers during the delivery period.

The trainers will be trained by experienced staff from FFT, “FFT Senior Trainer”. Two cohorts of trainers will be trained for the effectiveness trial. Cohort 1 will receive all 5 elements of the training and support described above before the trial begins. Cohort 2 will receive an equivalent training that will start before the trial begins and will continue in parallel to the trial. Trainers will not be able to start training teachers and TAs on Reciprocal Reading until they have completed their own training. Important to note, in the efficacy trial, the training was delivered by a small team of highly experienced trainers (who now act as FFT Senior Trainers), so this represents an important change in the delivery approach.

Time period: Cohort 1 trainer training will take place before the start of the trial, and will be concluded by the start of school recruitment in February 2023. For this reason, the evaluation will not focus on this element of the intervention.

Managing quality

FFT has the following strategies in place to support effective implementation:

- A clear set of written programme materials, including a manual, a set of suggested session plans and anthology of short stories.
- A comprehensive and standardised training and support programme for school coordinators and teacher/TAs.
- Monitoring and support from the FFT Senior Trainer and the wider FFT education team, including the following quality assurance observations by the FFT Senior Trainer:
 - At least 1 observation of each FFT Trainer covering Day 1 and Day 2 and support meetings
 - Observations of a sample of schools, inspecting the programme set up
- Guidance on pupil selection in the written materials and webinars, and additional advice and support on the choice of pupils available to schools upon request.
- Schools are given an online tool to register attendance and dosage of the Reciprocal Reading sessions. Teachers/TAs will keep registers for all of their sessions. FFT and the School Coordinator will have access to these for monitoring.
- A programme of online training to train replacement Teacher/TAs in the event of staff turnover during intervention delivery.

Logic Models

Figure 1: Logic model: Training for teacher/TAs and school coordinators

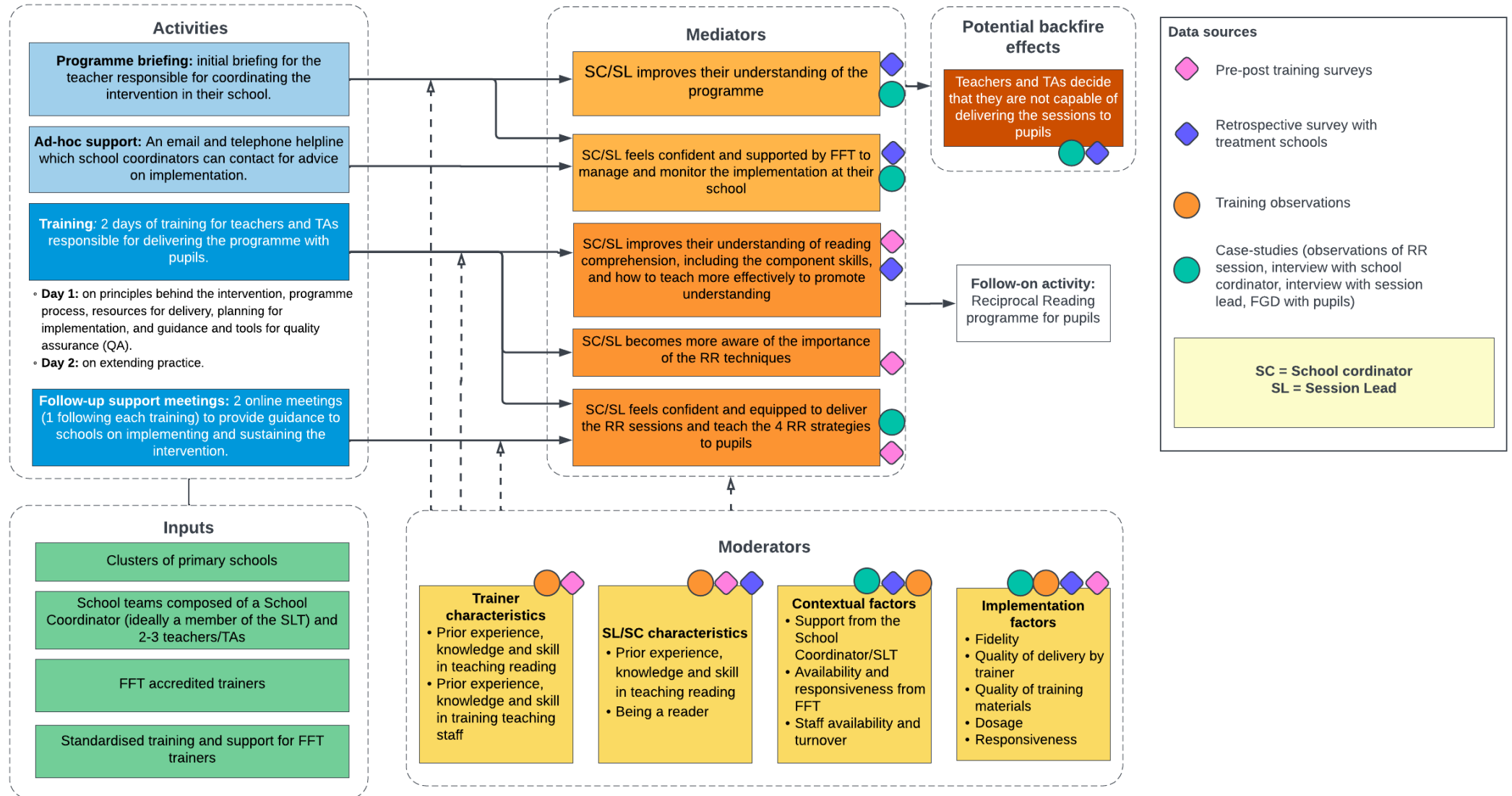
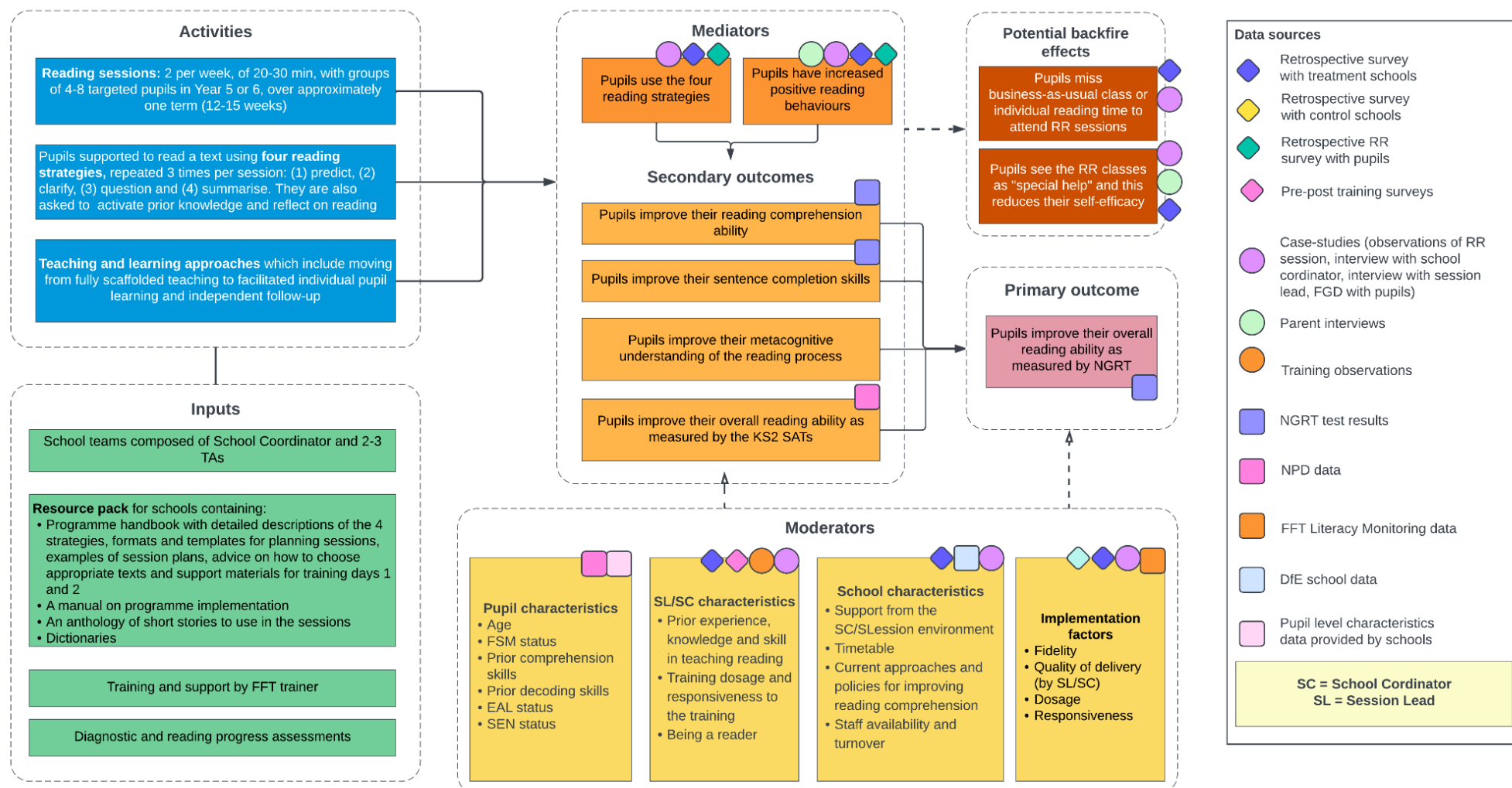


Figure 2: Logic model: Reciprocal Reading pupil outcomes



Impact evaluation design

Research questions

The **primary research question** of the impact evaluation is the following:

1. What is the difference in **reading proficiency** (as measured by the NGRT overall reading score) of pupils who receive the Reciprocal Reading programme, as compared to pupils in the control group?

The evaluation will also address the following **secondary research questions**:

2. What is the difference in **reading proficiency** (as measured by the NGRT overall reading score) of **FSM**-eligible pupils who receive the Reciprocal Reading programme, as compared to FSM-eligible pupils in the control group? Is the treatment effect for FSM-eligible pupils different from the effect for the full sample of pupils?
3. What is the difference in **reading comprehension** (as measured by the NGRT passage comprehension score) of pupils who receive the Reciprocal Reading programme, as compared to pupils in the control group?
4. What is the difference in **sentence completion skills** (as measured by the NGRT sentence completion score) of pupils who receive the Reciprocal Reading programme, as compared to pupils in the control group?
5. What is the difference in **reading attainment** (as measured by the KS2 English SAT reading score) of pupils in Year 6 who receive the Reciprocal Reading programme, as compared to pupils in the control group?

Design

Table 4: Trial design

Trial design, including number of arms		Two-arm, cluster randomised
Unit of randomisation		School
Stratification variables (if applicable)		Batched randomisation
Primary outcome(s)	Variables	Reading proficiency
	Measures (instrument, scale, source)	New Group Reading Test (NGRT) overall reading score, GL Assessment (1)
Secondary outcome(s)	Variable(s)	Reading comprehension (2) Sentence completion skill (3) Reading attainment (4)
	Measure(s) (instrument, scale, source)	NGRT passage comprehension score, GL Assessment (2) NGRT sentence completion score, GL assessment (3) KS2 SAT reading score (4)
		Variable Reading proficiency (1)

Baseline for primary outcome	Measure (instrument, scale, source)	New Group Reading Test (NGRT) overall reading score, GL Assessment (1)
Baseline for secondary outcome	Variable	Reading comprehension (2) Sentence completion skill (3)
	Measure (instrument, scale, source)	NGRT passage comprehension score, GL assessment (2) NGRT sentence completion score, GL assessment (3)

This effectiveness trial will be a two-arm cluster randomised trial with randomisation at the school level. The two arms will be:

- A. A treatment arm of schools in which nominated teachers receive Reciprocal Reading training and deliver Reciprocal Reading programme lessons to a group of selected pupils in Year 5 and 6. These lessons are additional to the school curriculum. These schools will receive a subsidy from EEF for the programme fee.
- B. A control arm of business-as-usual, in which schools continue as they otherwise would have, with no changes in their curricular activities. Control schools will receive a financial incentive of £1000 to participate in the trial, which will be paid after completion of the post-intervention test.

School-level randomisation was chosen to minimise the risk of spillovers between the control and treatment arms, as opposed to pupil or class-level randomisation. Control schools will not participate in the training for teachers and the risk of communication between treatment and control schools is low.

The impact evaluation will estimate the effect of the Reciprocal Reading programme on the overall reading proficiency of KS2 pupils. This will be measured by the overall test score of the NGRT, our primary outcome. Additionally, the evaluation will estimate the impact of the programme on **four secondary outcomes**: pupils' sentence completion skills, pupil's reading comprehension, reading attainment in KS2 SATs for Year 6 pupils, and pupil's metacognitive understanding of the reading process. The NGRT sentence completion score will be used to measure sentence completion skills, while KS2 reading attainment will be measured by the English reading test results of the KS2 SATs.

Participant selection

Pupil eligibility

The study participants will be pupils in Years 5 and 6 (KS2), aged between 9 and 11 years old, that fulfil the eligibility criteria in the recruited schools and are nominated to participate in the trial.

Selected pupils will have relatively good decoding skills but poor comprehension skills. They will be identified as candidates by their teacher, with the support of the School Coordinator, using guidance and materials provided by FFT when the school signs up for the trial.

The guidance consists of two sets of eligibility criteria: 1) the reading skills that must be held by a child to be considered as having good decoding skills, and 2) a list of difficulties that may

be experienced by a child struggling with reading comprehension. The teacher will compare each child against the two and identify those children that match the criteria. They will base their assessment on their own subjective judgement, but are encouraged to supplement this with existing school assessment data. These assessments should show good scores on reading accuracy and lower scores on comprehension tasks for selected pupils. **Each school will select two groups of ideally 6 children in Year 5 and Year 6** that fulfil both sets of criteria (with a minimum of 4 pupils and a maximum of 8 pupils per group).

Children in control group schools will be selected in the same way for the purpose of completing outcome testing, but they will not receive the Reciprocal Reading programme.

School eligibility

To participate in the trial, a school should have at least 4 pupils in Year 5 and at least 4 pupils in Year 6 that fulfil the eligibility criteria and can make 2 delivery groups; smaller schools that cannot reach these numbers will not take part in the trial.

Other school requirements to participate in the trial are:

- a) having a minimum of one full class of Year 5 pupils and Year 6 pupils (because it is unlikely that a mixed year group class would have sufficient pupils who will fit the criteria to be able to take part);
- b) not having been involved in the previous trials of Reciprocal Reading or used FFT Reciprocal Reading; and
- c) not being involved in other trials similar to Reciprocal Reading which focus on improving reading skills, reading attainment or reading comprehension.

Any schools belonging to a multi-academy trust (MAT) will have to agree to not share information about the programme or the teaching of reading comprehension with other schools in the same MAT during the period of the trial, to prevent spillovers between treatment and control schools.

School recruitment will be managed and carried out by FFT, and will aim to have at least 50% of the recruited schools in [Education Investment Areas](#), as defined by the DfE. Recruitment efforts will try to prioritise schools with a higher-than-average proportion of FSM-eligible pupils, although this requirement is not prescriptive and will be secondary to achieving the desired recruitment goal of 300 schools needed for statistical power. This represents planned recruitment of approximately 3,600 pupils (12 per school, on average), with approximately 1,800 receiving Reciprocal Read and 1,800 receiving business as usual. In treatment schools, recruitment also includes one school coordinator per school and two other teacher/TAs (likely to be Teaching Assistants) that will be trained and will deliver the intervention as teacher/TAs. For control schools we will have to recruit 1-2 members of school staff (senior leadership, teachers or Teaching Assistants) which will support activities related to the trial.

The section of this protocol titled “Sample size” describes the power calculations and the justification for the targeted sample size.

FFT will be recruiting schools from January to July 2023, so programme delivery can start in schools at the beginning of the next academic year (October - November 2023). Once recruited, schools will carry out the baseline testing on the primary outcome with nominated

pupils. They will then be randomised into treatment and control by BIT, ideally before September, so they can start the programme when the academic year begins.

Since FFT is planned to be recruiting until the very end of the academic year in July, which could mean that baseline testing and treatment allocation might not be completed before the end of the Summer term, randomisation will be in two batches to allow for a wide recruitment window: we aim for 80% of the recruited schools to be randomised in July 2023, and the other 20% to be randomised in September 2023. In the event that recruitment is completed earlier than July 2023⁴, randomisation will still be separated to allow for replacement of any schools that drop out of the trial during the Summer before the programme starts. This will help minimise school attrition early in the programme.

Schools wishing to participate in the trial will be asked by FFT to sign a Memorandum of Understanding (MOU) before enrolment, agreeing to the required activities for both the intervention and evaluation. This includes the requirement for each school to assess their pupils, select a group of candidates and allow administration of the baseline NGRT assessment before they are randomised into a trial arm. They will also have to repeat the NGRT test with those pupils at the end of the trial.

Recruited schools will also have to agree to share the relevant pupil data with BIT (via FFT) for evaluation purposes.

Outcome measures

Primary outcome

The trial's primary outcome is the overall score in the digital [New Group Reading Test \(NGRT\)](#) by GL Assessment⁵, which measures overall reading proficiency. The NGRT is a standardised assessment that measures skills in sentence completion and reading comprehension and, unlike other national reading assessments like the KS2 SAT reading test (which is done in May of Year 6), it can be taken by Year 5 and Year 6 pupils immediately after the intervention. The NGRT overall score was one of the two primary outcomes in the efficacy trial, so using it as the primary outcome measure for the effectiveness trial will allow for better comparison between the two trial results, removing the likelihood of obtaining a differing result due to the measure used.

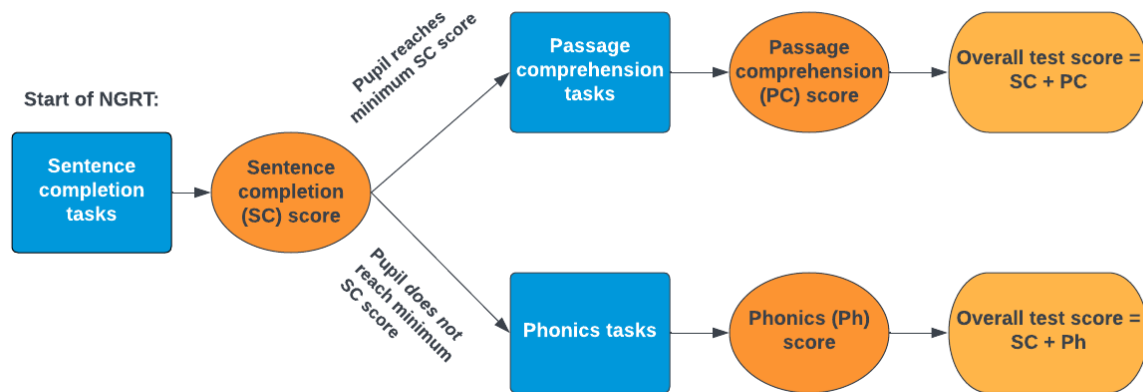
The NGRT test was selected for use in the efficacy trial because it has a subscale specifically for passage comprehension. This makes it more accurate in measuring the desired outcome than other national English reading assessments and can give more detailed information to teachers about the decoding and passage comprehension skills of a child. The digital version of the test is adaptive, so it responds to the level of the child as they take the test. If a child does not get a minimum score in the first sentence completion tasks, they will not progress to passage comprehension and will do phonics tasks instead. This means that pupils taking the test receive two subscales that make up the total score: a sentence completion score, and a passage comprehension or phonics score, depending on their performance (see Figure 3 below). Overall reading score is calculated by GL and available for all pupils, but reading

⁴ In practice, the recruitment goal was achieved in May 2023.

⁵ The NGRT is commercial and owned by GL Assessment so we are unable to include a copy of it in the Appendix.

comprehension scores are not available for those who scored very low on sentence completion tasks, which identify vocabulary gaps. Not being able to achieve the minimum score in sentence completion is likely to mean that a pupil does not have the required decoding skills to access the programme, according to FFT's guidance on pupil selection.

Figure 3: Diagram of the NGRT online assessment structure



Unlike the efficacy trial, we will not use the NGRT reading comprehension score as a dual primary outcome because it will not be available for all pupils taking the test. Firstly, comprehension score missingness is not at random, but correlated with a pupil's reading accuracy skills. This could bias our estimates for the change in comprehension score because, assuming the programme is effective, more pupils in the treatment group will be expected to have comprehension scores at the endline than pupils in the control group. This bias is likely to be in the downwards direction; the variable is censored and absent for the lowest achieving pupils, so the estimated effect will not reflect the full increase for pupils at the lower-end of the distribution. This problem affected only 1% of the final sample in the targeted intervention in the efficacy trial, but since the sample of schools for the effectiveness trial will be three times larger, we can expect more variability in pupils' outcomes.

Lastly, having only one primary outcome allows us to retain more statistical power than having two, as we do not have to adjust our results for family-wise error rate. This is more important in an effectiveness trial, as we are testing the intervention under everyday conditions and at a larger scale. The intervention effect is therefore expected to be smaller than in the efficacy trial, where the intervention delivery was led by its developers in a smaller number of schools (that were likely more responsive to the intervention) and the Teacher/TAs were trained by the developers. In this trial, Teacher/TAs will be trained by FFT staff that will have been trained by the developers in turn, which can also dilute the intervention effect. We have nevertheless kept the passage comprehension score as a secondary outcome in the Impact evaluation.

The NGRT overall test score for the digital version ranges between 0 and 500 points. The test has 47 items, although the total number depends on pupil performance, and a high reliability with a Cronbach's alpha above 0.9.

The baseline and endline tests will be delivered digitally in schools under exam conditions by Qa Research, a BIT subcontractor. The baseline assessment will be done in two batches,

along with randomisation (we provide more details on the batched randomisation in the corresponding section), and the endline will be done at the end of the intervention. BIT will set up the initial assessment test for schools and the accounts with unique IDs for each pupil taking the test. Qa Research will be the exam invigilators, and will help pupils to use the correct ID to take the test.

BIT and Qa Research will work with schools to minimise pupil attrition through repeat mop-up assessments, which may happen with repeat visits from Qa Research or coordinated remotely and administered by schools. For the pre-assessments, Qa Research assessors will identify missing pupils on testing days, and demonstrate to school staff how to administer the tests for these missing pupils. The school staff will then have 1 month to complete these mop-ups. During this time, Qa Research will be in contact with schools to nudge them to complete the tests, and support them with any issues. In schools with a large percentage of missing pupils, Qa Research will re-visit the school in person to administer mop-ups. For the post-assessment stages, Qa Research will re-visit schools if there were absent pupils on the testing days. In specific circumstances, where schools require long travel times to reach for assessors, and with a small proportion of absent pupils (1-3 pupils) these schools will be instructed to self-administer the mop-up tests using the same procedure as the pre-assessments.

Once the tests are done, GL Assessment will compute the scores for each pupil, which BIT will access through an online platform. Baseline and endline pupil data will be shared with schools after the end of the trial, to minimise the risk of control schools diverging from business-as-usual on the basis of the data.

Secondary outcomes

We will also study the impact of the programme on three secondary outcomes: reading comprehension skills (measured by the NGRT passage comprehension scores), sentence completion skills (measured by the NGRT sentence completion score) and reading attainment (measured by the KS2 SAT reading score).

NGRT passage comprehension score

The NGRT passage comprehension subscale measures performance in reading comprehension tasks. The improvement of reading comprehension is the main target of the 4 reading strategies of the Reciprocal Reading programme and therefore the hypothesised main pathway to increase reading proficiency in the intervention participants (see Logic Model in Figure 2). This measure was one of the dual primary outcomes in the efficacy trial so keeping it as a secondary outcome will allow us to compare results between the efficacy and effectiveness trials.

The NGRT passage comprehension subscale will only be available for pupils that score a minimum value in the sentence completion tasks of the test and can progress to do passage comprehension exercises. For this reason, we are not using this measure as a primary outcome, even though it is a direct assessment of the target of the intervention.

The NGRT passage comprehension subscale will be obtained together with the NGRT overall score, at baseline and endline. The passage comprehension part of the test has 27 items, although the total number of items depends on pupil performance. This is because of the adaptive nature of the test: it presents the next question to pupils automatically based on the pupils' performance as they complete them. This way, higher ability pupils can be

challenged while weaker readers also remain engaged with accessible questions (GL Assessment, 2023).

NGRT sentence completion score

The NGRT sentence completion subscale measures reading accuracy and basic comprehension through performance on the sentence completion tasks of the test. Although the main focus of the intervention is reading comprehension, the Reciprocal Reading programme also improves overall reading proficiency by consolidating sentence completion skills (see Logic model in Figure 2). We are including it as a secondary outcome to test this causal pathway.

The NGRT sentence completion subscale will be obtained together with the NGRT overall score and the passage comprehension subscale, at baseline and endline. The sentence completion part of the test has 20 items, although the total number of items depends on pupil performance.

KS2 SAT reading score

We will estimate the impact of the programme on reading attainment in the KS2 SAT English reading test. The KS2 SATs are standardised national assessments on the English national curriculum for KS2. They are sat in May by all Year 6 pupils in English primary schools under exam conditions.

The English reading SAT score measures overall reading proficiency, but the data will only be available for pupils in Year 6 at the end of the trial. This makes it inadequate as a primary outcome. However, given its importance and widespread use by teachers and practitioners to measure reading attainment of KS2 pupils, we will evaluate the impact of the programme on KS2 SAT scores.

The KS2 English reading test comprises: a reading booklet containing three texts that increase in demand throughout the booklet, and a reading answer booklet containing questions (50 marks). Pupils' test papers are scanned so that marking can be conducted on screen by trained markers.

BIT will obtain the KS2 English reading SAT scores from the National Pupil Database (NPD) after the endline NGRT tests have been administered, using unique pupil numbers (UPNs) to match the pupils in the trial with the dataset. We will use the unamended version of the KS2 attainment data, released in September 2024. This dataset will not include the outcome of any marking reviews, but these reviews usually affect approximately 0.1% of the data so they would have a negligible effect on our analysis.⁶

The NPD variable that we will use for this outcome is KS2_READSCORE. This is a conversion of the raw score that allows for comparison between years of the KS2 tests. The scaled score can go from 80 to 120. Pupils achieving a scaled score of over 100 are considered as having met the expected standard in the test.

Some pupils in the trial may not have a scaled score. This is because pupils need a minimum raw score before they can be awarded the lowest scaled score. Pupils who do not

⁶ See 'Annex A: KS2 Review Outcomes' at: <https://explore-education-statistics.service.gov.uk/methodology/key-stage-2-attainment>. Last accessed 17 April 2023.

achieve the lowest scaled score on the test have not demonstrated sufficient understanding of the KS2 curriculum in the subject and they are considered as “working below the standard of the reading test” (Department for Education, 2019). In 2021/2022, this amounted to 4% of pupils in the KS2 assessments (Department for Education, 2022). Raw scores would therefore be better for trial analysis purposes, but they are not available in the NPD.

Baseline measures

The NGRT will be taken by all pupils in the trial at baseline before treatment status is allocated to schools. We will use the baseline measures of the overall test score, sentence completion and passage comprehension subscales to control for baseline attainment in their respective analyses.

Schools will have access to the baseline and endline scores only after the trial has finished. This has two purposes: first, to minimise the risk of altering business as usual in control schools as much as possible; and second, to prevent NGRT results to influence delivery of Reciprocal Reading in intervention schools and make the treatment conditions as close as possible to non-trial conditions, as the NGRT scores may not be available outside of the evaluation.

Sample size

Table 5: Sample size calculations

		Overall	FSM
Minimum Detectable Effect Size (MDES)		0.13 (Hedges' G)	0.16 (Hedges' G)
Baseline/ endline correlations	level 1 (pupil)	0.56	0.58
	level 2 (class)	-	-
	level 3 (school)	-	-
Intraclass correlations (ICCs)	level 2 (class)	0.15	0.18
	level 3 (school)	-	-
Alpha		0.05	0.05
Power		0.8	0.8
One-sided or two-sided?		Two-sided	Two-sided
Average cluster size		12	6
Number of schools⁷	Intervention	150	150
	Control	150	150
	Total	300	300
Number of pupils	Intervention	1800	900
	Control	1800	900

⁷ The number of schools and pupils in this table reflects the target numbers at *the design stage*, before attrition. If an amended protocol is published later in the trial cycle and the number of schools and pupils *post-randomisation* are known, this will be explicitly stated next to both numbers.

	Total	3600 (before attrition)	1800 (before attrition)
--	--------------	----------------------------	----------------------------

Table 5 shows that with 300 schools participating in the study, we will be powered at 80% to detect an effect of 0.13 in Hedges' G for the NGRT overall test score (the primary outcome). This calculation is based on the following assumptions:

- A baseline correlation between baseline and endline primary outcome of 0.56 for the full sample and 0.58 for the FSM subsample, as reported in the efficacy trial (targeted intervention);
- Intraclass correlation of 0.15 for the full sample and 0.18 for the FSM subsample, as reported in the efficacy trial (targeted intervention);
- Pupil attrition of 15%, as reported in the efficacy trial (targeted intervention);
- School-level clustered randomisation with a cluster size equal to 12 pupils, with 2 delivery groups of 6 pupils per school, before attrition;
- A proportion of FSM pupils of 50%, for the subgroup analysis (a cluster size of 6 FSM-eligible pupils). The proportion of FSM pupils in the targeted intervention of the efficacy trial according to the EVERFSM_6_P variable in the NPD was 54%. We are assuming a slightly smaller percentage of FSM-eligible pupils because we expect more variability in our sample. The national average rate of FSM pupils in state-funded primary schools is 23.1% (ONS, 2022);
- Alpha of 0.05 and statistical power of 80%;
- Equal-sized treatment and control groups, where all pupils in the intervention and control groups will be tested.

The trial will also be powered to detect an effect for the FSM sub-group of 0.16 in Hedges' G for the NGRT overall test score under the same assumptions.

To detect the effect on the primary outcome that was found in the efficacy trial (0.14 in Hedges' G), we would need a sample of 257 schools⁸. For the subgroup analysis, we would need 179 schools to detect the effect found in the efficacy trial on FSM pupils (which was 0.20 in Hedges' G).

The trial therefore appears to be well-powered. Together with the delivery team, we decided to be conservative and aim for a higher sample of 300 schools with two goals: i. to allow for the possibility of not hitting the school recruitment target; and ii. to account for smaller effect sizes than those found in the efficacy trial (as we would expect from a scaled-up intervention).

The power calculations were done using R studio.

Randomisation

Randomisation will be clustered at the school level (as opposed to pupil or class level) to minimise the risk of spillovers between pupils in treatment and control groups and to facilitate delivery.

It will be carried out by BIT in two batches:

⁸ The effect size actually reported for the efficacy trial was 0.14 in Hedges' G. We converted it to Cohen's D using the standard deviations from the trial report.

- A first batch will randomise approximately 80% of schools that are recruited between January and July 2023;
- A second batch will randomise all remaining schools and those recruited between then and September 2023.

This two-batch process was decided with FFT in order to randomise and communicate treatment assignments to the biggest possible number of schools before the Summer break, so they can start delivery when the new academic year begins, while giving the delivery team room to extend recruitment efforts until the end of July. This approach will be kept even if recruitment is completed earlier than planned, as it will allow the delivery team to replace any schools dropping out of the trial during the Summer with candidate schools from the waiting list in September, in order to minimise school attrition before delivery starts.

Recruited schools will be asked to send FFT a list of the 8 to 16 nominated pupils and their characteristics: unique pupil reference number, first and last name, school, date of birth, gender, year group, FSM status, EAL status, and SEN status and category. BIT will use a list of schools' IDs to conduct the randomisation, employing statistical software (Stata or R) to generate random numbers and using them to randomly assign schools to each trial arm. We will set and record the seed of the random generator so the randomisation results are reproducible.

The batch randomisation with batches of different sizes increases the risk of imbalance in covariates like FSM eligibility, which can reduce the statistical power for the FSM subgroup analysis. We will address this issue with balance checks and additional covariate adjustments in our primary and secondary analyses. These approaches are described in more detail in the Statistical Analysis section below.

We will not use stratification and minimisation by FSM eligibility and/or baseline attainment for the following reasons. Firstly, the number of schools that will be recruited for each batch is unknown, and it is uncertain whether the strata for the second batch will have a sufficient number of schools. To ensure a timely delivery, it is important to randomise the highest number of schools as early as possible so that teacher training can begin as soon as the academic year starts. Thus, setting as few logistical restrictions as possible during recruitment allows for more flexibility in the number and size of randomisation batches.

Secondly, since randomisation will be at the school level but balance is required on pupil-level characteristics, stratification at the school or group level will not guarantee balance on pupil covariates. The accuracy of the data on free school meal (FSM) eligibility from schools cannot be guaranteed, (there was a considerable difference between the number of FSM-eligible pupils reported by schools and the number of FSM pupils shown in the NPD data in the previous efficacy trial of Reciprocal Reading).

Lastly, stratification by baseline attainment would require waiting for the baseline test results, which could result in the first batch having fewer schools due to school delays in testing, or delays in the start of delivery.

Therefore, the trial will not use stratification or minimisation on FSM eligibility or baseline attainment, allowing for a more flexible and timely recruitment process. Not stratifying will increase the probability of getting a better balance in school-level unobservables, and we will be able to control for pupil level covariates in our statistical analysis.

Stratification by pupil year group is not deemed necessary because FFT will ask schools to select 6 pupils from both year groups. If schools follow that guidance, the sample should be balanced across year groups. In practice, we expect a small amount of variability in the number of pupils from each year group (Y5 or Y6) across schools. We will assess the distribution of pupils from Y5 and Y6 across the sample of recruited schools before randomisation. If we observe a high degree of variability in the composition of selected pupils from each year group, we may decide to stratify by the amount of pupils in Year 6. In that case, we will classify the schools by whether the number of nominated Y6 pupils is above or below the sample median.

Statistical analysis

Analysis will be conducted on an intention-to-treat (ITT) basis, including all selected pupils in schools participating in the programme. Analyses will be conducted in Stata or Rstudio, using 2-sided significance tests, at the 5% significance level.

The analysis requires a comparable group of pupils from schools assigned to the control group. Schools will be asked to nominate pupils to participate in the programme during the recruitment process (prior to being informed of their trial arm allocation). This will allow us to identify the pupils who would have received in the intervention, had their school been assigned to the treatment group.

In addition to the effect sizes, we will summarise the means and standard deviations of baseline and endline scores for each trial arm, together with histograms of baseline and endline data distributions.

Primary analysis

Analysis will be conducted on an intention-to-treat (ITT) basis to compare treatment and control arms and will include all nominated pupils in participating schools for whom we have complete data on the relevant variables.

Analysis will be carried out using an ordinary least squares (OLS) regression with clustered standard errors at the school level, to reflect the clustered design of the trial:

(Equation 1)

$$Y_{is} = \beta_0 + \beta_1 T_{is} + \beta_2 PreNGRT_{is} + \beta_3 X_{is} + \epsilon_{is}$$

Where:

- Y_{is} is the endline NGRT overall test score for individual i , in school s ;
- T_{is} is a binary indicator for the treatment for individual i , in school s (1 if the pupil is in a treated school and 0 if not);
- $PreNGRT_{is}$ is the baseline attainment for individual i , in school s , measured through the baseline NGRT overall test score;
- X_{is} is a vector of pupil and school-level covariates including FSM status, year group, gender, EAL status, and randomisation batch for individual i , in school s ; and
- ϵ_{is} is the cluster-robust error term, for individual i in school s , clustered at the school level (assuming the errors are correlated within school and reflecting the design of the study).

We have included a vector of individual-level covariates that are expected to be correlated with reading outcomes in order to improve the precision of the estimate for β_1 . This is expected because the year group, gender and EAL status of pupils have been reported by GL assessment as being determinants of statistically significant differences in NGRT scores⁹, while the efficacy trial found a higher treatment effect for FSM-eligible pupils in the targeted intervention. We also include as control a factor variable indicating the school randomisation batch to account for the batch stratification and for differences in timing for baseline tests between school batches.

Secondary analysis

NGRT passage comprehension subscale

For the analysis of the endline NGRT passage comprehension score, we will replace $PreNGRT_{is}$ and Y_{is} in our specification of the primary analysis (Eq. 1) with the baseline and endline NGRT passage comprehension subscale respectively.

The estimation of β_1 in this model may be biased due to sample selection, as missing endline passage comprehension scores for pupils with an endline NGRT overall test score will be missing not at random. Pupils who do not achieve the minimum score in the sentence completion section of the test will not be assigned passage comprehension tasks. As the treatment is anticipated to enhance sentence completion abilities, the treatment group pupils are more likely to achieve this outcome at endline, potentially introducing bias into the estimate for β_1 .

In the efficacy trial, 1% of the sample of pupils for the targeted intervention had overall reading score but not reading comprehension score at endline (O'Hare et al., 2019, Appendix H)¹⁰. We will conduct a missing data analysis for the passage comprehension score to assess the severity of the selection and interpret the results of this secondary analysis accordingly. This will include: a) formally testing if the probability of having an endline score is affected by the treatment, b) report how many pupils with no baseline passage comprehension score did have an endline score in the treatment and control groups, and c) report the percentage of pupils with an NGRT overall reading score that were missing the passage comprehension score, at baseline and endline.

If the selection into the sample is high enough to be significant, we will compare the results of estimating β_1 for the full sample against a subsample of pupils who were above the threshold to have a passage comprehension score before and after the trial. The approach to the missing data analysis is described in more detail below. The distribution of the NGRT overall score, passage comprehension and sentence completion subscales at baseline will also be used to assess whether the baseline reading proficiency of pupils obeyed the requirements to participate in the trial given by FFT in their guidance on pupil selection.

⁹From the NGRT Technical guidance by GL Assessment. Last accessed on 20/02/2023.

¹⁰ This was 5% and 6% at baseline and endline for the pupils in the universal intervention of the efficacy trial.

NGRT sentence completion subscale

For the analysis of the endline NGRT sentence completion subscale, we will replace $PreNGRT_{is}$ and Y_{is} in our specification of the primary analysis (Eq. 1) with the baseline and endline NGRT sentence completion subscale respectively.

KS2 English reading SATs for Year 6 pupils

For the analysis of the KS2 English reading SAT scores for Year 6 pupils, we will replace Y_{is} with the KS2 English reading SAT scaled score obtained from the NPD database. $PreNGRT_{is}$ will be kept to control for baseline attainment. This analysis will compare the SAT scaled score of Year 6 pupils in the treatment and control arms.

We anticipate that around 3-4% of pupils in the sample will not have a KS2 reading score due to not meeting the minimum test standard, according to the national average. If the percentage is below 5%, this issue is unlikely to bias impact estimates, regardless of the pattern of missingness (Schaffer, 1999). We will conduct missing data analysis as described below to assess the extent of the issue and interpret the results accordingly.

Estimation of effect sizes

The primary and secondary analyses, including the FSM subgroup analysis, will convert effect sizes into Hedges' g using the following transformation:

(Equation 2)

$$\text{Hedges' } g = \frac{M_1 - M_2}{SD_{pooled}^*}$$

Where $M_1 - M_2$ is the difference in mean test scores between the treatment and control group, and SD_{pooled}^* is the pooled unconditional standard deviation.

The effect size for the primary analysis and FSM-subgroup analysis will also be converted into months of progress for the evaluation report.

We will report 95% confidence intervals for the effect sizes of all primary and secondary outcome analyses. The coefficients will also be reported together with one to three stars to show confidence of the estimate at different standard levels of significance (1%, 5%, 10%).

To conduct MCA (multiple comparison adjustment), we will apply the Benjamini-Hochberg method to correct the p-values reported for the two secondary analyses on NGRT scales. The passage comprehension and sentence completion subscales will be obtained from the same test and measure two important components of reading attainment, which increases the likelihood of family-wise error rate for the two comparisons.

By contrast, we will not do MCA for the other two secondary outcomes. The analysis on KS2 reading test results will be done with a sample subgroup so it is not an equivalent comparison, while the metacognition survey measures pupil behaviour, an outcome which is very different from reading attainment. The interpretation of the results for these two outcomes will be separate and distinct from the two NGRT subscales.

Subgroup analyses

We will do two subgroup analyses for FSM-eligible pupils to study any heterogeneous treatment effects in that group, following EEF statistical analysis guidelines:

1. Estimate the model specified in the primary analysis on the subsample of FSM-eligible pupils;
2. Estimate a similar model including an interaction term for FSM eligibility, using the entire sample (see Equation 3 below).

Both approaches will estimate the effect size for FSM pupils, but the latter uses information from the whole sample. Under ideal conditions, the total treatment effect in both should be analogous. The results for both will be compared and reported in the Appendix as a robustness check for this subgroup analysis. We will explore any difference and discuss its implications for the results.

(Equation 3)

$$Y_{is} = \beta_0 + \beta_1 T_{is} + \beta_2 EverFSM_{is} + \beta_3 (EverFSM_{is} * T_{is}) + \beta_5 X_{is} + \epsilon_{is}$$

Where:

- Y_{is} is the endline NGRT overall test score for individual i , in school s ;
- T_{is} is a binary indicator for the treatment for individual i , in school s (1 if the pupil is in a treated school and 0 if not);
- $EverFSM_{is}$ is a binary indicator equal to 1 if the pupil has ever been eligible for FSM and 0 if not);
- $PreNGRT_{is}$ is the baseline attainment for individual i , in school s , measured through the baseline NGRT overall test score;
- X_{is} is a vector of pupil and school-level covariates including year group, gender, EAL status, and randomisation batch for individual i , in school s ; and
- ϵ_{is} is the cluster-robust error term, for individual i in school s , clustered at the school level (assuming the errors are correlated within school and reflecting the design of the study).

We will use the variable “EVERFSM_6_P” in the NPD to determine FSM eligibility.

We will report and compare the effect sizes of both models, the restricted sample and the interaction term. The effect sizes will be reported in terms of Hedges’ G .

Analysis in the presence of non-compliance

Compliance for this analysis is defined at the pupil level.¹¹ A pupil should attend at least 20 sessions to be considered compliant in the treatment arm. Any amount below that will be considered as non-compliance¹².

¹¹ A separate analysis of compliance at the teacher level (e.g. attendance at training sessions) is specified in the IPE, but not used as part of the Complier Average Causal Effect (CACE) analysis specified here.

¹² As a reference for comparison, the average number of weeks of delivery in the efficacy trial for the targeted intervention was 26 and average time of delivery per week was 65 minutes, which represents more than twice the minimum recommended dosage by FFT (O’Hare et al., 2019).

We expect some degree of non-compliance from pupils assigned to treatment. This could also be due to teachers or schools not delivering the mandated amount of sessions, or due to pupil absence. As randomisation is at the school level, we do not expect non-compliance in the opposite direction (that is, pupils in the control group taking up treatment) or defiers (pupils that would attend sessions if they were placed in the control schools and would not attend sessions if they were allocated to the treatment group).

Pupil attendance data will be collected routinely by the teachers leading the sessions.

We will estimate the Complier Average Causal Effect (CACE)¹³ on the primary outcome using a two stage least squares (2SLS) approach. The estimation model will be specified in detail in the SAP. One of the assumptions of this instrumental variables (IV) model is the exclusion restriction: treatment assignment impacts reading attainment (the outcome of interest) only through treatment itself. This assumption would not hold, for example, if schools start giving pupils additional reading support to make up for not receiving the intervention in control schools, or if they substitute usual reading curricular activities for treated pupils. However, any extra provision they would have provided for struggling pupils regardless of the programme or testing will be considered as business as usual and should not affect this assumption. The IPE will examine the plausibility of this assumption.

The CACE analysis will be complemented with a descriptive analysis mapping the average increase in primary outcome by pupil dosage, defined as number of treatment weeks and also as number of lessons taken. Although we will not be able to make a causal inference about the relationship between dose and treatment effect, this analysis will be used to contextualise the findings from the CACE analysis and provide useful guidance for schools and others involved in delivery of the intervention.

Additional analyses and robustness checks

Imbalance at baseline

We will check for imbalance at baseline by comparing normalised differences in means on baseline test scores and demographic covariates correlated with the outcomes (year group, gender, FSM status, and EAL status) in the pre-attrition and post-attrition samples.

The normalised difference is defined as the difference in means between the two groups, divided by the pooled standard deviation. Normalised differences with a magnitude of 0.1 or less indicate a negligible correlation between the covariate and assignment to treatment group, which can usually be addressed through covariate adjustment in the regression (Austin, 2009). We will report on any differences and discuss the implications for the primary and secondary analyses.

Missing data analysis

We will use data from the NPD (as opposed to data from schools) to establish gender, FSM eligibility, EAL status and KS2 SAT scores of pupils where possible to minimise missingness. The primary outcome and other secondary outcomes (at pre-treatment and post-treatment) are survey or assessment-based and more susceptible to be missing.

¹³ Also known as LATE (Local Average Treatment Effect).

We will first report the number of missing observations for the following cases:

- A. Missing pre-treatment covariates
- B. Missing outcome data

We will also try to establish the mechanism behind missingness for each variable. Data can be missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR), and the approach to analysis will vary depending on the type. To establish the mechanism, we will use logistic regression models to model the presence of missing data with treatment assignment and other covariates. This exercise will be described in more detail in the SAP.

A. Missing pre-treatment covariates

Where pre-treatment covariates are missing, we will run two basic robustness checks on all analyses. First, we will create a new variable to indicate missingness and use this to re-estimate the effects. Second, we will re-run all analyses without covariates and report the unadjusted estimates.

We may also carry out imputation in line with EEF guidelines. Any imputation of covariates will be restricted to the primary analysis and will only be carried out when more than 5% of the data is missing. Schultz and Grimes (2002) suggest that, when less than 5% of data is missing, there is likely to be little bias introduced to estimated treatment effects, so we have adopted this threshold here.¹⁴ The approach to imputation will be specified in more detail in the SAP.

B. Missing outcome data

If outcome data is MCAR, we will run a complete case analysis as this will yield unbiased results. If it is MAR conditional on identified and available covariates, we will estimate a model including those covariates and interpret the results. If the data is MNAR, we will employ sensitivity analysis. These approaches will be described in more detail in the SAP.

CONSORT diagram

A CONSORT diagram will be used to present the summary of the flow of eligible pupils and their schools from recruitment through randomisation, post intervention assessment and analysis. The number of children and schools included or excluded at each stage will be clearly stated and the reasons for exclusion will also be stated.

¹⁴ This is also in line with the convention provided in EEF's guidance on statistical analysis.

Implementation and process evaluation (IPE) design

We propose a mixed method IPE to accompany the impact evaluation in this trial. The IPE methodology has been designed in collaboration with the developers and following the guidance provided by the EEF (2022). The nine different data-collection methods are carefully designed to triangulate with and integrate with each other to explore different IPE considerations and ensure rigour.

Research questions

The IPE accompanying the trial will aim to answer seven key research questions (RQs):

1. To what extent do trainers deliver the training as intended, and how well do they deliver it? (Fidelity/Quality)
2. To what extent do school coordinators and teacher/TAs implement Reciprocal Reading sessions as intended, and how well do they deliver it? (Fidelity/Quality)
3. What are the characteristics of the schools and pupils reached through scaled up delivery, and how do these characteristics moderate the effects of the intervention? (Reach/Moderators)
4. How much of the intervention do treatment group pupils receive? (Dosage)
5. How does the intervention work for pupils from disadvantaged socio-economic backgrounds? (Moderators/Causal mechanisms)
6. Does the intervention complement or replace existing efforts? What do teachers and teaching assistants perceive to be the distinguishing features of the intervention? (Programme differentiation/ Unintended consequences)
7. What reading interventions do selected pupils in the control group receive? And does the pupil selection process seem to influence this? (Adherence, Contamination)

To explain how these seven research questions tie to the overall aim of the study Table 6 presents the three key considerations which guided the IPE research question design.

Intervention at larger scale: RQs 1-4 aim to explore whether the results of the efficacy trial can be replicated on a larger scale. Challenges may arise from more schools, working in areas of greater economic deprivation, and a wider pool of trainers. Logistical complications on the delivery side may also affect delivery quality.

Intervention as gap closer: RQ5 examines potential mechanisms for treatment effects, or lack of treatment effects for less affluent pupils. The IPE explores the hypothesis that Reciprocal Reading can have a levelling effect on inequalities in the home reading environment by providing disadvantaged pupils with a stimulating group reading environment, and techniques to comprehend texts more effectively. RQ5 also explores whether Reciprocal Reading triggers specific mechanisms for change, to support underperforming pupils' literacy attainment. It will look at whether there are specific components of Reciprocal Reading which make it more effective as a gap closer than

standard extra-curricular literacy activities adopted by schools, such as unstructured reading time.

Changes in usual practice: RQ6 and RQ7 examine usual practice and unintended consequences in schools. RQ6 investigates if Reciprocal Reading is additional to existing interventions and curricular activities, and RQ7 investigates adherence for the control group. The adherence of the control group will be explored by monitoring spillovers resulting from the trial. Specifically, before randomisation, a selection process will identify suitable pupils for the intervention. This could lead control group teachers to offer targeted support for control group pupils during the trial period as a direct consequence of the trial which may contaminate findings¹⁵.

Table 6: Key considerations for IPE research questions

Links to IPE research questions	Consideration	Implications for IPE	Links to IE
RQ1 RQ2 RQ3 RQ4	Intervention being delivered at a larger scale	Important to capture the changes between the efficacy and effectiveness trial, and implications of these changes (e.g. wider pool of trainers). A higher level of variability and adaptations is expected for both pupil characteristics and school contexts (e.g., higher levels of pupils eligible for FSM or higher levels of school staff turnover). Implementation logistics can become harder to manage at scale, affecting quality and fidelity of delivery.	Provides potential explanation for changes in estimated effects from efficacy to effectiveness trial.
RQ5	Evidence of promise for intervention as gap closer (reducing the gap in reading attainment between FSM pupils and non-FSM pupils)	If intervention is more effective for pupils from disadvantaged socio-economic backgrounds, we want to understand why and what can be done to maximise these effects in future iterations of the intervention. Our proposed approach is outlined in the methodology of this IPE. We also want to investigate whether Reciprocal Reading triggers specific mechanisms to improve literacy attainment for disadvantaged pupils, which are more effective than standard extra-curricular literacy initiatives in schools, such as unstructured reading time.	Provides potential explanation for moderators that may lead to greater treatment effects for a sub-group of pupils.
RQ6 RQ7	The intervention leads to changes in usual practice that are not intended in the trial	In treatment schools, Reciprocal Reading sessions may replace other interventions or curricular activities, which may have unintended consequences for selected pupils. Both treatment and control group schools carry out a pupil selection process pre-randomisation. This introduces the potential for control group schools to deliver additional reading interventions to pupils who are identified as in	Provides potential explanation for treatment effects and may provide evidence of backfire effects.

¹⁵ FFT included guidance for treatment and control schools about the support that control schools should be offering to the children they selected to participate.

		need of extra support. This contamination could lead to an underestimation of the treatment effects.	
--	--	--	--

Research methods

The IPE will consist of nine different data-collection methods as follows.

Monitoring data

FFT's data team will share data on the **number of sessions delivered in schools and pupil attendance**. This will be collected through FFT's digital platform, where teacher/TAs and session coordinators complete session registers on a regular basis (they will be encouraged to log session attendance weekly). The data will be shared with BIT at the end of programme delivery. However, the delivery team will also periodically monitor data collection through check-ins to ensure quality and completeness of the data. We expect to have data for the majority of the 3,600 pupils across 300 schools, for 12 sessions or more.

We will report descriptive statistics on these data and use them as part of the compliance analysis. Specifically, we will do a descriptive analysis cross-tabulating average outcomes with different thresholds for intervention dosages. This will help understand how outcomes may vary across different levels of intervention dosage, important to note this will not be a causal analysis.

As specified in the impact evaluation, pupils will be considered compliant if they have attended a minimum of 20 sessions.

Socio-demographic analysis

BIT will collect pupil level characteristics through the DfE's National Pupil Database (NPD) including:

- FSM status
- EAL status
- SEN status
- SEN category
- Ethnicity
- Gender

BIT will access the NPD after the post-NGRT tests have been administered, using unique pupil numbers (UPNs) to match the pupils in the trial with the database. These same data fields will also be shared with us by FFT, who will collect these data directly from schools. This will help us mitigate any risks of missing data, incomplete fields on the NPD or other unexpected circumstances which may impact the analysis. These data should cover the entire study population of 3,600 pupils.

FSM status will be used for the sub-group analysis in the impact evaluation. The other data will be used to compute descriptive statistics and report on the sample's composition and compare it to the general population of English school pupils and the sample from the efficacy trial. This analysis will be key in understanding the generalisability of the impact findings.

School data analysis

Using the DfE's Get Information about Schools portal we will collect quantitative data on schools in the treatment and control group, including:

- % of FSM
- Location
- Ofsted Rating

BIT's evaluation team will obtain these data for all 300 schools. These data will be used to compute descriptive statistics on the trial sample, and compare it to the general population of English schools and the sample from the efficacy trial. This analysis will help us to understand generalisability, and could provide potential explanations of why outcomes from the efficacy trial were not sustained, or evidence that outcomes remain consistent even when implementation settings differ.

Observation of training sessions

We will observe training delivery and collect data using an observational framework to assess quality and fidelity. Achieving effective training outcomes for teacher/TAs and school coordinators is an essential step in triggering the theory of change for the intervention and achieving impact on pupils' reading outcomes. Without high quality training outcomes for school coordinators and teacher/TAs, it will not be possible to implement Reciprocal Reading effectively in schools. This is an area of particular interest as the intervention scales, as in previous trials the training was delivered by a small group of highly experienced trainers who were also close to the design of the intervention. Given that this trial will have to deliver the programme on a larger scale compared to the previous Reciprocal Reading trials funded by the EEF, new trainers will need to be trained. This may cause higher variability in the quality of training which may moderate pupil outcomes.

The observational framework will be co-developed with FFT to ensure that all relevant dimensions of the training are captured. This will be a semi-structured observation framework, including both open-ended and close-ended questions. As the programme developers and experts in the content, FFT's assessments about the quality of training sessions will be more nuanced than those of BIT's independent researchers, so FFT's input will be considered to define the key indicators of effective delivery. BIT researchers will use this framework to check that the sessions are meeting fundamental expectations of quality (as opposed to being a comprehensive assessment of quality). Two researchers will collaborate on these observations. The first observation will act as a pilot and will be conducted by both researchers together. After the pilot, the researchers will debrief to see if there were major discrepancies in their observations. If needed, the framework will be adapted after this pilot to minimise data collection variability. When logistically feasible, BIT's observations of the training sessions will also be compared with FFT's internal QA observations for the same sessions.

A purposive sample of 5 training sessions will be observed, each delivered by a different trainer. 4 sessions will be observed between October and November 2023, and another two sessions from January to February. Depending on training delivery logistics the observations will be in-person or remote.

Training survey

A pre- and post-survey completed by training participants (school coordinators and teacher/TAs) will be used to capture short-term training outcomes by measuring:

- participants' attitudes to the training
- changes in self-reported confidence in using Reciprocal Reading techniques
- changes in self-reported knowledge of Reciprocal Reading techniques
- changes in knowledge of Reciprocal Reading (using a knowledge test as a direct measure)
- characteristics about the trainee, such as years of experience in teaching, highest level of teaching qualification, and attitudes to towards reading
- attendance at the training session (partial which is 1 day of training, or full which is 2 days of training)
- attendance at the online support session

The survey questionnaire will be developed by BIT and shared with the EEF and FFT for editing before implementation. We will prototype it with two ex-primary teachers in BIT's education team to ensure it is clear and to estimate completion time. The survey will be administered to all school coordinators and teacher/TAs participating in the training, estimated to be 2 respondents each for 150 treatment schools (n~300). It will be a short online questionnaire hosted on SmartSurvey, completed at the start and end of the training sessions to ensure high completion rates and reduce the impact of confounding variables. To capture medium-term learning outcomes, some of the survey measures will be repeated in a retrospective survey (see below).

Retrospective survey of school coordinators and teacher/TAs

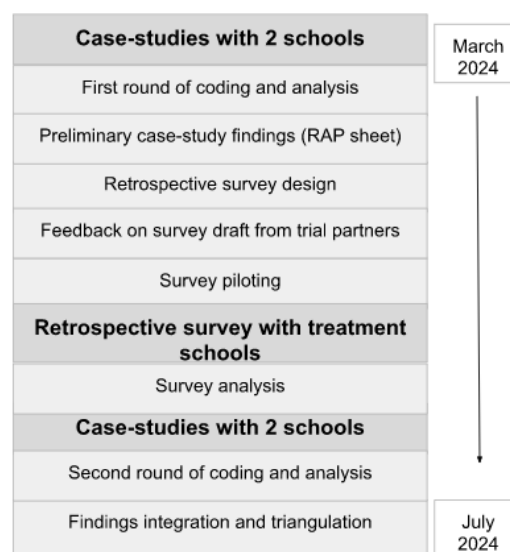
A survey of School Coordinators and Teacher/TAs involved in delivering the intervention will give us insight into responsiveness, adaptations, moderators, and unexpected impacts on usual practice. Retrospective online surveys will be shared with all school coordinators/teacher/TAs in the treatment group schools. The survey will be administered at the end of the delivery period, which is expected in March 2024. The questionnaire will include both close and open-ended questions, which will measure the following:

- repeat measures from training survey to monitor medium-term learning outcomes,
- self-reported attitudes to Reciprocal Reading,
- perceived impact on pupils of Reciprocal Reading (with specific questions on the impact of the intervention on FSM pupils),
- barriers and enablers to quality implementation,
- how the intervention was received by pupils eligible for FSM,
- adaptations to the delivery of the intervention,
- perceived impacts of the intervention on usual practice, i.e. has Reciprocal Reading replaced any curricular activities or alternative interventions,
- any other reading interventions or activities delivered to the treatment group during the trial period (using the same questions as those in the control group survey for comparability - see below),
- how Reciprocal Reading compares to other extra-curricular initiatives to promote literacy, such as unstructured reading time.

We expect 2 staff members in each of the 150 treatment schools to complete the survey, n~300. To ensure completion, we will use behaviourally informed communications¹⁶ and offer a monetary incentive (raffle with a £200 prize). This questionnaire will be administered via email using SmartSurvey, and designed to be as short and user-friendly as possible (no longer than 15 minutes of completion time). We will agree a time to launch the survey with FFT to establish a time when the majority of the schools will have completed >12 weeks of intervention (expected to be March 2024).

The questionnaire will be designed in collaboration with trial partners and piloted before launch. We will pilot it with two ex-primary teachers in BIT's education team to ensure it is clear and to estimate completion time. BIT will conduct two qualitative case-studies before launching the survey, which will inform its design. As will be explained in the subsequent section on school case-studies, the two case-studies conducted after the survey will enhance its findings (see Figure 4).

Figure 4: Triangulation of retrospective survey with treatment schools and case studies



Retrospective survey of school staff in control schools

The retrospective survey with school staff in control schools will monitor compliance by exploring any deviations from business-as-usual resulting from the trial. Teachers will have to select pupils for the trial and will identify pupils who are good decoders but poor comprehenders. There is a risk that, due to the selection of pupils and exposure to the concept of Reciprocal Reading through communications about the trial, teachers in the control group may provide their pupils with extra support or even self-implement aspects of the intervention. The survey will also capture any alternative literacy interventions that control schools may have received, which could result in improvements in reading attainment among control pupils. Therefore, the survey will include:

- Teacher/TA reports on any additional support that was provided to pupils selected to take part in the trial,

¹⁶ We will draw on behavioural science concepts to write effective emails to promote survey completion, e.g. using social norms by highlighting how a large portion of respondents has already answered.

- Teacher/TA reports (availability and perceived effect on pupil literacy) on current reading interventions delivered in schools,
- Teacher/TA reports (availability and perceived effect on pupil literacy) on current reading activities as part of the school activities, e.g. unstructured reading times or library visits.

The survey will be a short online questionnaire focused on measuring deviations from business-as-usual as a result of the trial, and capturing any concurrent interventions in the schools that will help us to interpret the impact evaluation results. One teacher per control school will complete the survey, which will result in approximately 150 respondents. To encourage high response rates, we will use behaviourally-informed¹⁷ communication and offer a raffle incentive worth £200.

This survey data will be used to contextualise the findings from the impact analysis. To add this context, we will conduct the following analysis on the data:

1. report descriptive statistics to show the number and percentage of participants that have participated in additional interventions during the intervention period (broken down by activity type and assignment); and
2. conduct balance checks (using a normalised differences approach) to see whether there are substantial differences on this point between the treatment and control groups.

We will not re-estimate treatment effects based on the outcome of this analysis, because conditioning on post-treatment variables can introduce bias (Montgomery et al. 2018).

Case-studies

This IPE will gather in-depth qualitative data from 4 schools, combining a range of data sources including school staff interviews, pupil interviews, and structured observations of Reciprocal Reading sessions. This case-study approach will help us understand how the intervention has been embedded in the school environment to generate evidence (Bradley et al. 2005). This is a case-study approach because of the sampling method, collecting data within the same settings, and because of the analytical strategy, which will look at cross-case comparisons between schools and pupils. The results will not be presented as narrative case-studies, but will be leveraged to explain the impact evaluation results as appropriate.

To triangulate and elaborate on the findings from the retrospective survey in treatment schools we will visit 4 schools as case-studies. We will purposely sample the schools to include 3 schools with greater-than-average % of FSM pupils, and one school with average % of FSM pupils. This will help us focus on how the intervention is delivered in areas of high deprivation, but also compare findings to the average school (Palinkas et al., 2015). These case-studies will help us investigate quality, fidelity, responsiveness, causal mechanisms and moderators. Building on the retrospective surveys for both control schools and treatment schools, the case-studies will explore how Reciprocal Reading is perceived to differ from other reading support, and whether it triggers causal mechanisms to improve literacy which

¹⁷ We will draw on behavioural science concepts to write effective emails to promote survey completion, e.g. using social norms by highlighting how a large portion of respondents has already answered.

are more effective than other extra-curricular literacy initiatives offered by schools. They will include the following qualitative data collection methods.

- **Observation of a Reciprocal Reading session** to observe fidelity, pupil responsiveness, and quality (including any differences in reception by pupils, identifying potential moderators).
- **Two in-person interviews with the school coordinator and a teacher/TA (one with each)** after delivery to explore their views of the programme, with a focus on how it impacted lower income pupils and how it compares to other extra-curricular initiatives to boost literacy attainment. Looking at whether the intervention particularly impacted pupils eligible for FSM, for example by exploring whether it was particularly effective at boosting reading enjoyment and exploring the mechanisms for this change, for example by exploring whether improved ability to comprehend texts led to heightened enjoyment or a higher amount of reading. These interviews will also ask school staff to compare Reciprocal Reading to alternative initiatives trialled in schools, such as unstructured reading time.
- **One qualitative focus group with pupils** who received the intervention. We will sample from a specific year group who took part. In two schools, we will sample Year 5 pupils and for the other two we will sample Year 6 pupils. This will be a structured interactive discussion led by a researcher experienced in working with children. We will recruit up to 6 pupils in each focus group, our analysis will focus particularly on the voice of pupils eligible for FSM to see how their experience compares to their peers. We will collect data on the pupils' home reading environment to understand how the intervention may benefit specific pupils.

Table 7. Sampling frame for the case-studies

Data source	Sampling approach
Schools case-studies	4 school visits: • 3 schools with higher than national average % of FSM pupils • 1 school with approximately average % of FSM pupils
Interviews with school staff	8 interviews with school staff • 4 interviews with school coordinators • 4 interviews with
Pupils focus groups	4 focus groups of 12 pupils per school, 1 per school visited: • 2 of these groups will be Year 5 pupils • 2 of these groups will be Year 6 pupils
Reciprocal Reading session observations	4 structured observations of Reciprocal Reading sessions, 1 per school visited.

The case-studies will inform the design of other IPE research instruments and the findings will be triangulated with those that come from the other methods used (refer to Figure 4). To achieve this integration and triangulation, two case-studies will be conducted in March 2024, and two more from June to July 2024.

BIT will collaborate with the delivery team to identify and recruit potential case-study schools. We anticipate sampling responsive and engaged schools who can effectively accommodate school visits, being mindful not to overburden schools with data collection, taking into

account the SAT testing period, and avoiding selecting the same schools for multiple IPE activities. We acknowledge this is a sampling limitation and will mitigate biases by corroborating findings using a rapid assessment procedure sheet (RAP sheet) with other data sources (e.g. the retrospective surveys with treatment schools, and parent interviews) (RReal, 2022)¹⁸. The RAP sheet is a tool for qualitative analysis, which asks researchers to concisely summarise findings in pre-defined themes, while also allowing researchers to revise the tool to add emerging fields. These summaries facilitate cross-source comparison and facilitate collaborative qualitative work across researchers.

The research instruments will be created by a qualitative researcher, who has extensive experience in designing and conducting qualitative evaluations, and will be peer-reviewed by BIT staff with first-hand experience as primary school teachers.. They will be guided by the logic model and the TiDier framework, with input from FFT and the EEF. To ensure high-quality data collection, the following measures will be taken:

- Engaging experienced qualitative researchers with a background in collecting data in schools and working with children.
- Developing detailed data-collection protocols to minimise variability.
- Holding debrief sessions after each data collection to review and revise research instruments as needed.
- Quality assuring all instruments, data-collection processes, and findings by the Head of Qualitative Research at BIT.

These focus groups and interviews will be recorded using encrypted dictaphones and professionally transcribed by our transcription partners McGowan. This will guarantee we obtain high quality transcripts that can be used during the analysis phase.

Parent interviews

BIT will conduct qualitative phone/video-call interviews with parents of pupils in the treatment group. Interviews will explore family-level attitudes to reading and pupils' home reading environment. Our respondents will be a purposive sample of 20 parents, using their children's FSM eligibility as a key criteria (50% will be families eligible for FSM and 50% will not be eligible). This sampling approach has been adopted to allow comparisons between the home reading environments of pupils eligible for FSM and non eligible pupils, and shed light on how Reciprocal Reading may act as a gap closer. We will also consider gender as a monitoring criteria, recruiting at least 25% of the sample from the least represented gender (e.g., if 75% of parents recruited identify as female, we will ensure the remaining 25% identify as male). This sampling frame will enable us to compare results across the two groups, exploring how socio-economic background may mediate the effects of the intervention.

These interviews will be recorded (if participants consent) and transcribed using automated Google transcription, which has produced high quality instant transcripts in other projects.

To achieve our target sample, BIT will design a recruitment strategy, informed by the knowledge of the trial partners. This approach will capitalise on specific touch-points in the intervention and draw on behavioural science, for example using social norms in

¹⁸ A RAP sheet is an approach to qualitative data analysis developed by RREAL (Rapid Research Evaluation and Appraisal Lab).

communications to incentive completion (e.g, 20 other parents already signed up to speak to us!). Additionally, participants will be offered a £20 voucher as compensation for their time.

Table 8: IPE methods overview

Research Method	Data collection methods	Sampling ¹⁹	Data analysis	IPE dimension	RQ
Monitoring data	Attendance and session register - collated onto FFT's portal	All 150 treatment schools	Descriptive statistics, CACE analysis ²⁰	Dosage, compliance	RQ2 RQ3 RQ4
Socio-demographic sample analysis	National Pupil Database	All 3600 pupils	Descriptive statistics	Moderators	RQ3
School data analysis	DfE's school data portal	All 300 schools	Descriptive statistics	Moderators, contextual factors	RQ3
Training observations	Structured observations	Purposive sample of 4 sessions in Sept-Oct and another 2 in Jan-Feb (each by a different trainer)	Observational framework, RAP sheet ²¹	Moderators, causal mechanisms, adaptations	RQ1
Training surveys (pre- post)	Online questionnaires	n~300 school coordinators / teacher/TAs	Pre-post analysis, inductive coding of free text responses	Moderators, quality, causal mechanisms	RQ1
Retrospective survey of treatment schools	Online questionnaire	n~300 school coordinators / teacher/TAs	Descriptive statistics, thematic inductive coding of free text	Quality, fidelity, adaptations, moderators, usual practice	RQ2 RQ5
Retrospective survey of control schools	Online questionnaire	n~150 school staff in control group schools	Descriptive statistics, thematic inductive coding of free text	Usual practice	RQ7
Case-studies	Semi-structured interviews, structured focus group discussion, semi-structured observations	Purposive sample of 4 schools (3 with high % of FSM pupils and one with national average level)	RAP sheet, thematic coding of transcripts, observation framework	Quality, responsiveness, fidelity	RQ2 RQ3 RQ5 RQ6

¹⁹ Samples for the whole trial or entire arms do not take into account attrition, which is expected to be up to ~20%.

²⁰ Complier Average Causal Effect

²¹ A RAP sheet is an approach to qualitative data analysis developed by [BREAL](#) (Rapid Research Evaluation and Appraisal Lab).

Parent interviews	Semi-structured interviews	Purposive sample of 20 parents from the treatment group schools (10 whose children are eligible for FSM and 10 whose children are not)	RAP sheet, deductive coding	Moderators	RQ5
-------------------	----------------------------	--	-----------------------------	------------	-----

Analysis

The IPE data will be integrated and triangulated using the RAP sheet, which will outline the research questions for the IPE and the corresponding data sources. The RAP sheet is a pre-designed tool, which asks researchers to concisely summarise findings in pre-determined themes. However, researchers are also able to revise the RAP sheet to add additional themes which may become apparent from the data. This tool will facilitate comparisons of findings across the data sources. When each analysis is completed, we will update the document to assess how the data answers the questions and whether they are in agreement or disagreement. We will also have bi-weekly IPE debrief sessions (between March and July 2024) where we will review the findings and consider how they can inform future instrument design or analysis. During these sessions, we will use the logic model as a guide to examine whether the findings align with the theory.

Analysis of pupil level characteristics

An analysis of pupil and school characteristics will be conducted to understand the composition of the sample and how it compares to the wider population and the sample from the efficacy trial. We will report descriptive statistics for these data, including:

- Frequency, reported as a percentage
- Minimum and maximum values, reported as a count
- Averages, reported as mean and median
- Distribution, reported as standard deviation

We will also conduct a sub-group analysis of pupils eligible for Free School Meals (FSM) as part of the impact evaluation. All of the descriptive statistics for pupil and school characteristics will be reported in tables, and bar charts will be added to represent frequency.

Analysis of monitoring data

The analysis of monitoring data will include computing descriptive statistics to report on the intervention dosage delivered to pupils in the treatment group. The same descriptive statistics outlined for pupil and school characteristics will be used (frequency, minimum and maximum, averages, and distribution).

Additionally, we will examine differences in delivery at the regional level. We will use the delivery clusters specified by FFT as the analytical unit and compute the frequency and mean number of sessions delivered in each cluster to monitor regional variations. These

data will also be used as part of the impact evaluation to estimate the CACE, which will give an indication of how dosage moderates treatment effects.

Training survey analysis

To evaluate the training outcomes for school coordinators and teacher/TAs, we will conduct a pre-post analysis complemented by qualitative coding. The post survey will also include 2-3 open text fields, which will be coded using an inductive approach and semi-quantification. Using the SmartSurvey coding function, we will code each response with a theme, and see the % of responses that relate to a specific theme.

The following model will be used to estimate the effects of the training on teachers' learning, using ordinary least squares (OLS) regression. This pre-post analysis will be conducted on an intention-to-treat basis, including all complete cases in the sample.:

$$y_{it} = \alpha + \beta post_t + \gamma a_{it} + \gamma b_{it}' \varepsilon_{it}$$

Here:

- y_{it} is the outcome for respondent i in period t ²²
- $post_t$ is a binary indicator of the sample (1 for endline, 0 for baseline survey)
- a_{it} is a binary variable recording whether the training was delivered by an experienced trainer or a newly trained trainer²³
- b_{it} is a binary variable recording whether the respondent attended the training in full or partially

Retrospective surveys analysis

To analyse the retrospective surveys administered to school staff in the treatment and control groups, we will use a mixed-methods approach. Quantitative results will be presented using descriptive statistics, specifically the percentage of respondents choosing specific responses.

Qualitative responses will be coded using a mixture of inductive and deductive coding. The analysis will be performed using Nvivo and the RAP sheet. The process will be as follows:

- Inductively code a random sample of 20% of the responses
- Run a word frequency analysis to explore frequently used words (single word analysis stemmed, using stop words)
- Revise the RAP sheet (which will act as a codebook) to code the remaining sample
- Code the remaining 80% of responses using the RAP sheet

²² The survey will be designed after the trial protocol. It is likely to include multiple outcome variables including self-reported confidence and knowledge of Reciprocal Reading techniques, which may be weighted to create an index. If weighting applies, we will consult the trial partner on what these weights should be.

²³ An experienced trainer is defined as a trainer who has delivered Reciprocal Reading training as part of a previous delivery and was not trained for the purpose of this trial.

The statistical results will be presented in tables, the qualitative results will be presented by summarising the key codes in the codebook with example quotes.

Case-study analysis

The case-studies will be analysed using a mixture of observational framework analysis and transcript coding (both inductive and deductive). The RAP sheet will be a crucial tool in integrating findings from case-study settings, and exploring overarching themes. The RAP is a tool developed by RREAL (Rapid Research Appraisal and Evaluation Lab) (UCL and RReal, 2020; Vindola-Padros and Johnson, 2022), which summarises data across participants and data collection methods according to research questions (RReal, 2022). Transcripts will be analysed using the RAP sheet, with the following approach:

- Creating an initial RAP sheet on the key research questions, which will act as a codebook, coding the first set of transcripts from the first two-case studies
- Exploring additional emerging themes and revisiting the codebook
- Creating a set of hypotheses, which can be explored using the retrospective surveys to school coordinators/teacher/TAs, and the parent interviews
- Coding the transcripts from the final two case-study schools

Qualitative researchers will have recurring debrief sessions to see if their findings corroborate or contradict each other and find ways to produce nuanced outputs that take these into account.

Parent interview analysis

The parent interviews will be analysed using deductive coding. They will be analysed in Nvivo using the codebook developed through the case-studies and summarised using the RAP sheet.

Cost evaluation design

The cost evaluation will follow the "ingredient method" as specified by EEF guidelines (EEF, 2022; Levin, 2019). The ingredients of the program are as follows:

- Personnel for the implementation of the program
- Personnel during training for the implementation of the program
- Program costs
- Facilities, equipment and materials
- Other program inputs.

Data Collection Methods

Data collection methods will focus on disaggregating program-specific costs and triangulating the cost estimates using multiple data sources (EEF, 2022; Vassall et al., 2017; World Bank, 2019).

Structured group interviews with delivery team

We will conduct two group interviews with FFT staff: one at the start of delivery and one after delivery is concluded. The first session will inform data collection tools for the cost

evaluation. Drawing on the TiDier framework and Logic Model, we will explore all potential disaggregated program-specific costs with key staff members from FFT. This group interview will be recorded and transcribed (using Google Transcripts). This interview will help the Evaluation Team build the other data collection tools for this cost evaluation. The second interview will inform the scale-up analysis to understand how programme costs differed from the efficacy trial as a result of expanding to more schools and over a broader geographical area.

Researcher-administered questionnaire to school coordinators

At the end of the intervention, we will conduct remote structured interviews with a sample of 15 school coordinators from the program. These will be school coordinators from the same sample of schools who completed the proformas to ensure that we have complete cost data for each of the sampled schools. The survey will be administered remotely in a video call following a standardised template. School coordinators will be compensated with a £50 cash voucher for their time.

Analysis

The analysis will draw on specified methods to construct the Cost Evaluation Tables specified by the EEF (EEF, 2022).

Triangulating Data Sources

Our analysis will triangulate findings using multiple data sources (EEF, 2022; Vassall et al., 2017). The structured group interview with the delivery team will help us build the instruments for data collection by identifying all the program ingredients. This will be reported as a list of ingredients and prerequisites for the implementation as specified by the EEF. The ongoing proformas will be analysed using descriptive statistics, with a focus on variance (mean, median, standard deviation, range and interquartile range). The final researcher-administered questionnaire will capture missing data, unforeseen costs, and validate existing findings by reporting on the same descriptive statistics.

Constructing a Pupil-per-School-Year Estimate of Costs over Three Years

The cost evaluation will calculate the cost per pupil if the intervention is implemented in schools over three years. Reciprocal Reading is not a whole-class intervention, but a targeted intervention delivered to a small group of pupils in schools. In the trial, 12 pupils per school will receive the intervention, meaning 36 pupils per school across three years. However, in subsequent years of delivery, schools may expect to deliver the intervention to a larger group of pupils. The number of pupils per school will be estimated by collecting data on the expected number of pupils included in the intervention per school.

Sensitivity Analysis

We will carry out sensitivity analysis on the findings to account for heterogeneity and parameter uncertainty (Franklin et al., 2019; Vassall et al., 2017). Heterogeneity will be accounted for by reporting on descriptive statistics of estimated costs, including the mean, median, range, interquartile range, and standard deviation. We will present a lower cost and upper cost estimate to account showing potential variable scenarios. Parameter uncertainty will take into account that some resources may have uncertain market prices, requiring BIT

to make assumptions and forecasts about these. This will be particularly important given the expected high inflation rates.

Scale-up cost analysis

Using the findings from all data sources and especially the final interview with the delivery team we will report on whether programme costs differed from the efficacy trial as a result of scaling.

Ethics and registration

Ethical approval

Ethical review has been conducted by the BIT internal ethics committee, and the project was approved on 14th November 2022. The review was conducted by a non-project member of staff, who blind assessed the research plan and data collection materials to ensure the project has considered all the potential risks and has appropriate procedures in place to mitigate these. In this case, the Ethical Review was conducted by Dr Giulia Tagliaferri, Head of Quantitative Research at BIT.

Following BIT's ethical framework, the project was assigned a medium risk level. This is due to the following key considerations:

- Low risk in terms of research methods, subject matter, legal exposure and unknown unknowns
- Medium risk in terms of participants (some children in a regular setting)
- High risk in terms of nature of data (some personal, individual level special category data collected for pupils)

Procedures for obtaining agreement to participate in the trial

As part of the recruitment process, schools will be sent the following documents to lay out the requirements and terms of the trial.

- Memorandum of Understanding (MoU): Explaining the programme, what will be required of schools, data protection and expected activities.
- School information sheet: Provides further information about all topics in the MoU and explains the study design, including the process for randomisation.

We will also share with schools a Parent Information Sheet and Consent Withdrawal form which explains the programme and trial to parents and gives them the opportunity to withdraw their children from the trial.

These documents provide clear explanations of the programme and evaluation and give schools the chance to make an informed decision about participating in the trial. These have been designed collaboratively by the trial partners following specific EEF guidance.

Trial registration

The trial will be registered on [Open Science Framework](#) (OSF). It will be registered by the BIT Evaluation Manager by the 31st of January 2024 after the Statistical Analysis Plan has

been approved. The Evaluation Manager will also be responsible for updating the registry with the trial results after the report is published in 2025.

Data protection

All pupil data will be processed in accordance with the General Data Protection Regulation (2018) and Data Protection Act (2018).

Data security

BIT takes steps to protect participants' personal information and prevent unauthorised access, alteration, loss or disclosure. We train those with access to personal data, provide access only on a need-to-know basis, and remove access when it's no longer needed. Devices are encrypted, confidential data is kept securely, and we have procedures in place to deal with data breaches and notify participants and regulators if necessary.

Data retention

The data collected during the Reciprocal Reading Evaluation Research Project will only be kept for as long as necessary and will be deleted securely when it is no longer needed. The appropriate retention period is determined by considering factors such as the amount and sensitivity of the data, potential risk of harm, and legal requirements. The anticipated deletion date for participants' personal data is September 2025, 3 months after the project's completion. However, names and work contact details of school staff may be kept for future similar projects. Research consent forms with personal information may be kept for a number of years for legal and statutory requirements and to meet the research funder's requirements. Personal data is needed to evaluate the impact of Reciprocal Reading in schools.

At the end of the project, a dataset will be archived with the EEF, following their guidelines. We will preserve anonymity and confidentiality in all publications, and no school or individual will be named or identifiable.

Data rights

To exercise their data rights, research subjects can contact the BIT's Data Protection Officer. The Data Protection Officer at BIT can be contacted via email at dpo@bi.team. It should be noted that the extent to which these rights apply to research may vary and may be restricted in some circumstances. Typically, there is no fee required to access personal data or exercise rights, but a reasonable fee may be charged if the request is unfounded, excessive, or repetitive. If the request is deemed as such, BIT may refuse to comply with the request. BIT may request specific information to confirm the identity of the requestor and ensure that the personal data is not disclosed to an unauthorised party. In some cases, additional information may be requested to expedite the response.

BIT aims to respond to legitimate requests within a month, but if the request is complex or multiple requests have been made, it may take longer. In such a scenario, the requestor will be notified and updated on the status. It should also be noted that BIT can only comply with requests to exercise rights for personal information that directly identifies the requestor. If the information is pseudonymised or has been irreversibly anonymized and is part of the research data set, BIT will not be able to comply with the request.

More information can be found in BIT's [Privacy Notice for Reciprocal Reading](#).

The following data will be collected from research participants to complete the trial.

Pupil data

We will collect personal data from pupils involved in the research including:

- Unique Pupil Number (UPN, identifies the pupil in the National Pupil Database)
- First name
- Last name
- Date of birth
- School name
- Year group
- New Group Reading Test (NGRT) pre-intervention and post-intervention scores
- Key Stage 2 SATs Reading scores (for Year 6 pupils at the start of the trial)
- Gender
- Free school meal (FSM) status
- Ethnicity*
- Whether the pupil speaks English as an Additional Language
- Special Educational Needs status*
- Type of special educational need*
- Focus group discussion data - qualitative data gathered in focus groups on attitudes to the Reciprocal Reading intervention
- Reciprocal Reading session attendance data - data on the number and length of Reciprocal Reading interventions attended

*Some of this data – ethnicity and data about health conditions – constitutes ‘special category data’ under data protection laws, and additional protections will apply to our collection and use of this data. This information is vital for our analysis, to assess the reach and possible differential effects of the programme. Reporting on these fields for the purpose of our research will be in an aggregated format only.

Pupils’ parent data

We will collect personal data from parents of the pupils involved in the research including:

- First name
- Last name
- Email address
- Telephone number
- Data on socio-economic background (FSM status) of their children
- Interview data - qualitative data on attitudes to reading and self-reported reading behaviours in the home

School staff data

We will collect personal data from school staff involved in the research including:

- First name
- Last name
- Work email address
- Work telephone number
- Survey data - quantitative and qualitative data on attitudes to the Reciprocal Reading intervention
- Interview data - qualitative data on attitudes to the Reciprocal Reading intervention

BIT will also use pupils' UPNs to access the following data using the DfE's NPD:

- KS2 SATs Reading scores
- FSM - Free school meal status
- Gender
- Ethnicity
- EAL - whether the pupil speaks English as an Additional Language
- SEN - Special Educational Needs status
- SEN Type - Type of special educational need

BIT will share participants' data with the following organisations:

- Qa Research – An independent provider contracted to administer the New Group Reading Test (NGRT) in partner schools.
- GL Assessments – An independent provider which will provide the platform to run the NGRT tests.
- Schools participating in the research, which will receive pupil NGRT scores at the end of the study.*
- McGowan - BIT's transcription provider.
- SmartSurvey - Online platform to administer surveys.
- FFT, which will receive pupil NGRT scores at the end of the study.
- The Department for Education (DfE) – National Pupil Database (NPD) team
- The Office for National Statistics – Secure Research Service (ONS SRS) – We will use the ONS SRS to analyse the trial data securely when it is matched to NPD data.
- At the end of the evaluation, the pupil data will be shared with the EEF and FFT Education (EEF's data processor for their archive). All EEF trial data is stored in the EEF data archive, held within the ONS SRS. The archive does not contain direct identifiers like pupil name, contact details and date of birth, but does hold a Pupil Matching Reference (PMR). The PMR is used for further matching to the NPD and other administrative datasets that may be required as part of subsequent research. We will not use pupil names or school names in any report arising from the research. For information on how the EEF will use and protect participants' data, please see their Data protection statement regarding EEF evaluations.

*Schools that want to receive individual level NGRT results for their pupils will have to sign a data sharing agreement with BIT.

Data processing roles

Data processing roles during the evaluation up to the point of data being deleted from all locations by the evaluator and/ or delivery team:

- BIT: Independent data controller
- FFT: Independent data controller
- Participating schools: Controllers
- Department for Education: Controller
- EEF*: Processor

- Qa Research: sub-processor
- GL Assessments: sub-processor
- SmartSurvey: sub-processor
- McGowan: sub-processor

*The EEF becomes data controller for the datasets archived after the trial, once internal quality checks have been successfully completed by the archive manager.

Legal basis for processing data

For all information collected, BIT is relying on the lawful basis of Legitimate Interest.

BIT's lawful basis for processing personal data is based on legitimate interests according to Article 6 (1) (f) of the GDPR. The processing is necessary for conducting an evaluation of the Reciprocal Reading program commissioned by the EEF, which aligns with BIT's business aims of delivering social impact through research and evaluation.

The processing of personal data from pupils, school staff, and parents is required to understand the effects of the program, as well as to arrange observations and interviews. The processing of interview/focus group data cannot be made anonymous due to unavoidable information that may be revealed during the interviews.

The processing of pupil demographic data, participation data, and outcome data is necessary to ensure correct assessments and to match data for analysis. Anonymous data cannot be used for this purpose.

For special category data, we also rely on Scientific Research Purposes as a lawful basis.

The processing of personal data is necessary for scientific research purposes with the aim of serving the public interest. BIT has implemented appropriate measures to protect the rights and interests of the data subjects, in accordance with relevant laws. The collected data is limited to what is required for the research, and any direct identifying information, such as names or contact details, has been removed where feasible. The processing is not expected to cause harm or distress to the participants and is not intended to be used for making decisions regarding specific individuals.

Personnel

Table 9: Project personnel

Team member	Organisation	Project role and responsibilities
Programme Management		
Rachael Morris	EEF	Senior Evaluation Manager Key liaison with BIT
Katie Luxton	EEF	Programme Manager Key liaison with FFT
Evaluation Team		
Patrick Taylor	BIT	Principal Investigator - Key liaison with EEF especially on issues regarding contract and budget - Managing the team - Overseeing design and quality assurance of all research activities
Neus Torres	BIT	Evaluation Manager and Quantitative Researcher - Day-to-day management of evaluation - Managing subcontractors (Qa Research and GL Assessments) - Overseeing data collection for the impact evaluation - Quantitative analysis and reporting
Jemuwem Eno-Amooquaye	BIT	Qualitative Researcher - Managing data collection for the IPE - Qualitative data analysis and reporting - Conducting qualitative data collection
Andrés Cueto	BIT	Qualitative Researcher - Supporting quantitative data collection - Conducting qualitative data collection - Supporting qualitative data analysis and reporting
Delivery Team		
Andy Taylor	FFT	Education Director (Literacy) Education lead and overall Project Director
Paul Charman	FFT	CEO
Laura James	FFT	Head of FFT School Support Team Team Head of school support team
Katie Kielty	FFT	Education Product Manager

Risks

Table 10: Risk assessment & mitigation

Risk	Risk level	Mitigation strategy
Under-powered study We risk not detecting the intervention's effect due to a small sample of pupils/schools as a result of: • Delivery partner failing to recruit enough schools • Missing pupils on testing days due to absences or other circumstances	Medium	As the Evaluators we will take a number of precautions to ensure that the study is not underpowered including: • Accounting for 15% attrition in our power calculations. • Using baseline outcome data to improve the precision of estimates and obtaining the rest of covariates from the NPD • Working with Qa Research to ensure that absent pupils on testing days are subsequently tested. At the pre-assessment stage, teachers/TAs will self-administer the tests, within one month from the testing date. At the post-assessments Qa Research will arrange in person visits to schools with high numbers of absent pupils and or coordinate remote mop-ups with schools with a low percentage of absent pupils. • Schools will receive email nudges from Qa Research encouraging them to administer the mop-ups to pupils who were absent on testing days, within a one month timeline from the initial testing session. • Monitoring the progress on data collection by Qa Research on weekly basis. The EEF will also monitor FFT's recruitment progress to support them to meet recruitment targets (100% of schools recruited by the end of the summer term 2023).
Spillovers between trial arms Schools in the control and treatment arms exchange information and training resources on Reading Comprehension strategies, which threatens identification of the treatment effect.	Low	The trial has been designed to minimise spillovers between arms by randomising and delivering treatment at the school level. FFT will mitigate this risk by: • Including a clause in the MOU forbidding schools in the same MAT to exchange information on reading comprehension strategies
Schools in control arm give additional reading support to pupils Control schools could alter their business-as-usual after identifying potential candidates for the intervention, which would alter the counterfactual and threaten causal identification.	Medium	We will mitigate this risk by: • Providing schools with baseline and endline outcome data only after the trial is completed • Monitoring the business-as-usual in schools in the IPE using teacher and school surveys and teacher interviews
Incomplete or inaccurate outcome data • Pupils making errors in completing the	High	We will ensure accurate data collection for the trial outcomes by: • Helping Qa Research to design effective training for their test administrators (e.g. providing mock tests

- NGRT tests - Pupil absences on test days		for them to practise). In addition Qa Research's team will have project managers who are experienced in administering NGRT tests. - Providing Qa Research with pupils' names to ensure they are logging onto the correct NGRT account (cross-validating entries). - Having Qa Research conduct mop-up visits and provide remote assistance to schools to conduct tests at schools where pupils are absent on the test day.
Inaccurate or incomplete data from schools and teachers - Schools submitting incomplete or inaccurate data to FFT which will create complications for accurate randomisation and NPD matching - Teachers submitting inaccurate attendance records	High	We will mitigate this risk by: - Working with FFT to see the process for requesting school data and ensuring it is clear and user friendly. Our Quantitative Researcher will request a demo of how schools will submit data and review a mock data output before we start requesting data to schools. - Ongoing monitoring of incoming school data (lead by our Quantitative Researcher).
Delays in administering the NGRT tests Delays in administering the NGRT tests due to - Poor responsiveness for schools in arranging Qa Research visits - Delays in receiving school data from FFT - Delays in receiving the delivery timeline, which will cause delays for the endline assessment	Medium	These risks are due to effective project management, and are linked to a timely delivery of Reciprocal Reading. These can be mitigated by: - an effective approach to communicating with schools and introducing Qa Research. - A timely approach to making introductions and an easy way to schedule test visits. - FFT complying with the agreed timelines; especially with regards to the submission of school data to the Evaluation Team (see timeline in this document page 53).
Delays accessing NPD data Delays accessing NPD data may cause delays in the impact analysis.	High	We will: - Submit our NPD application 3 months before the release of the unamended dataset. This is the earliest time that the application can be submitted. - Use our high-level contacts at the NPD (including the head of the NPD team, with whom we have monthly check-ins) to monitor progress with our application and troubleshoot issues. Even with these mitigations in place, this risk remains high however. The NPD has had very serious data protection and operational issues for a number of years. If it becomes apparent that substantial delays are possible, we will notify EEF and enter into good faith negotiations to try to find a solution acceptable to both parties, which may involve extending the project deadline and increasing the project budget.
Failure to recruit enough schools for IPE and cost evaluation	Medium	To promote effective recruitment the Evaluation Team will: Work with FFT to identify suitable and proactive schools.

Issues recruiting schools for the case-studies, or for the cost-evaluation data-collection activities may create complications for these elements of the evaluation.		● Use specific delivery touch-points (e.g. training days) to recruit school staff for data collection activities. ● Provide cash incentives for data completion. ● Use behaviourally informed communications to encourage effective recruitment. ● Design user-friendly systems to collect data
Failure to recruit parents for interviews Parents may be unresponsive to our recruitment requests for interviews.	Medium	To meet the recruitment targets for parents interview BIT will do the following: ● Design a behaviourally-informed recruitment strategy with trial partners. ● Work with schools to recruit parents. ● Offer cash incentives.
Unanticipated school closure School closures due to unexpected events such as pandemic cannot be dismissed.	Low	If the delivery of Reciprocal Reading has taken place, we will work with GL Assessments/Qa Research and partner schools to administer the NGRT tests to pupils remotely.
Staff turnover in the Evaluation Team Key staff members in the Evaluation Team leaving the organisation.	Low	A number of measures are in place to ensure that this risk does not impact the trial: ● BIT uses a cloud-based approach to project work which would ensure a smooth handover (everything from meetings notes, documents is stored on our drive and easily transferable to other staff members). ● BIT has a list of ONS certified researchers who could conduct the quantitative analysis. ● BIT has a number of qualified quantitative and qualitative researchers who could be resourced on the project. ● Programme a handover period to ensure that the new member of staff is able to complete all the relevant duties.

Timeline

Table 11: Timeline

Dates	Activity	Staff responsible/leading
14/11/2022	Ethical Approval for the Trial	Head of Quantitative Research at BIT
01/02/2023	Finalise recruitment documents	EEF/BIT/FFT
06/02/2023	Start of school recruitment (FFT requests pupil level data)	FFT
13/02/2023-17/2/2023	Spring half term holidays	School dates
13/03/2023	Share trial protocol with trial partners	BIT Evaluation Manager / Principal Investigator
20/03/2023	Qa Research and testing process is introduced to schools during school leads webinars	FFT
03/04/2023	BIT starts receiving school data from FFT	FFT
03/04/2023 - 14/04/2023	Easter break	School dates
14/05/2023 to 15/10/2023	Evaluators enrol pupils on the GL Assessment platform	BIT Evaluation Manager
14/04/2023 - 01/09/2023	Evaluators book Qa Research school visits	Qa Research
09/05/2023 - 12/05/2023	KS2 SATS	School dates
22/05/2023 - 15/10/2023	FFT shares pupil data with Evaluators	FFT
29/05/2023 - 02/06/2023	Summer half term holidays	School dates
12/06/2023	Start administering NGRT tests in schools	BIT / Qa Research
12/06/2023 - 21/07/2023	Administering 80% of NGRT tests in schools (Qa Research visits schools)	BIT / Qa Research
07/07/2023	Publish trial protocol	EEF/BIT
21/07/2023	Trial registration on the OSF	BIT
21/07/2022	Randomisation of the 1st batch of schools	BIT
21/07/2022	FFT finishes recruiting schools	FFT
04/05/2023 - 21/07/2023	Online briefing with school coordinators	FFT

30/07/2023	Evaluators to share draft training surveys with trial partners	BIT
21/07/2023 - 04/09/2023	Summer holidays for schools	School dates
05/09/2023	FFT shares the final list of schools that will participate in the trial with Evaluators	FFT
04/09/2023 - 16/10/2023	Administer the remaining 20% of NGRT test results	BIT / Qa Research
15/09/2023	Randomisation (second batch)	BIT
01/10/2023-30/11/2023	Reciprocal Reading training delivered to schools (and training surveys pre- post- collected)	FFT
01/10/2023 - 22/07/24	Reciprocal Reading delivery in schools	FFT/Partner schools
23/10/2023 - 27/10/2023	Autumn half term holidays	School dates
01/12/2023 - 16/03/2024	Recruiting case-study schools	BIT
21/12/2023-08/01/2024	Christmas Holidays in school	School dates
12/02/2024	Evaluators share qualitative research instruments with project partners (observational frameworks and interview guides for pupils, school staff and parents)	BIT
12/02/2024-16/02/2024	Spring half term holidays	School dates
16/02/2024	Share Reciprocal Reading delivery timetable with Qa Research	FFT
26/02/2023 - 25/03/2024	Qa Research to book school visits for the retrospective NGRT assessments	Qa Research
05/03/2024	SAP submitted to the EEF	BIT Quantitative researcher
04/03/2024 - 08/07/2024	Retrospective NGRT tests administered in schools	BIT/Qa Research
11/03/2024	Most schools will have completed ≥ 12 weeks of intervention	FFT/ partner schools
21/03/2024 - 03/04/2023	Co-development of parent recruitment strategy for parent interviews	BIT / FFT
29/03/2024-12/04/2024	Easter Holiday in schools	School dates
04/03/2024-28/06/2024	Conduct qualitative research in case-study schools	BIT Qualitative Researcher
13/05/2024-19/05/2024	KS2 SATs	School dates

04/03/2024-08/07/2024	Administer post- NGRT tests in schools (Qa Research visits schools)	BIT Evaluation Manager / Qa Research
10/04/2024-06/05/2024	Conduct parent interviews (x20)	BIT Qualitative Researcher
15/05/2024	Evaluators share draft retrospective surveys for school staff (control and treatment arm)	BIT Evaluation Manager
TBC	Evaluators launch retrospective survey with school staff (treatment and control group)	BIT Quantitative Researcher
27/05/2024-31/05/2024	Summer half term holidays	School dates
05/04/2024-28/06/2024	Delivery team conducts second meeting with schools	FFT
01/06/2024	Submit NPD data access request	BIT Quantitative Researcher
03/06/2024-21/06/2024	Cost evaluation interviews with school coordinators	BIT Qualitative Researcher
24/07/2024-02/09/2024	Summer holidays	School dates
30/09/2024	NPD data release (unamended data-set)	DfE
01/01/2025	Analysis in ONS SRS begins	BIT Quantitative Researcher
31/03/2025	First draft of report submitted to EEF	BIT Evaluation Manager /Principal Investigator
21/04/2025	EEF returns 1st round feedback	BIT Evaluation Manager /Principal Investigator
5/05/2025	Second draft of report submitted to EEF	BIT Evaluation Manager /Principal Investigator
26/05/2025	EEF returns 2nd round feedback	BIT Evaluation Manager /Principal Investigator
09/06/2025	Third draft of report submitted to EEF	BIT Evaluation Manager /Principal Investigator
31/08/2025	EEF returns developer feedback	EEF
30/09/2025	Fourth draft of report submitted to EEF	BIT Evaluation Manager /Principal Investigator
31/10/2025	Trial results are updated on the OSF and data archived	BIT Evaluation Manager

References

Austin, P.C. (2009) Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research, *Communications in Statistics - Simulation and Computation*, 38:6, 1228-1234, DOI: 10.1080/03610910902859574.

Bradley, P., Nordheim, L., Harpe, D., Innvær, S., & Thompson, C. (2005). A systematic review of qualitative literature on educational interventions for evidence-based practice. *Learning in Health and Social Care*.

Clark, C. and de Zoyah, S. (2016) Mapping the interrelationships of reading enjoyment, behaviour and attainment. National Literacy Trust London.

Cockerill, M., Thurston, A., O'Keeffe, J., & Taylor, A. (2021). A phase 3 definitive RCT of reciprocal reading in high schools in England. *International Journal of Educational Research*, 109.

Cockerill, M., O'Keeffe, J., Thurston, A., & Taylor, A. (2022). Reciprocal Reading for struggling readers: An exemplar of evidence implementation in schools. *Review of Education*, 10, e3332.

Crawford, C., & Skipp, A. (2014). LIT Programme Evaluation Report and Executive Summary October 2014. London: Educational Endowment Foundation.

Department for Education. (2022). Academic Year 2021/22: Key stage 1 and phonics screening check attainment. Available at: <https://bit.ly/3rS8dJW>.

Department for Education. (2022). Key Stage 1 phonics and screening check attainment.

Department for Education. (2022). Outcome delivery plan.

Department for Education. (latest accessed on 15/02/2023). Understanding scaled scores at key stage 2.

Department for Education Data Explorer (latest accessed on 02/03/2023).

EEF (2022) Cost Evaluation Guidelines. London.

EEF. (2019). Implementation and process evaluation guidance for EEF evaluations.

EEF. (2019). Reciprocal Reading: A structured approach to teach reading comprehension strategies.

EEF. (2020). Classification of the security of findings from EEF evaluations.

GL Assessment. (latest accessed on 02/03/2023). Using NGRT in primary schools.

GL Assessment. (latest accessed on 20/02/2023). NGRT Technical guidance.

Franklin, M., Lomas, J., Walker, S. & Young, T., 2019. An Educational Review About Using Cost Data for the Purpose of Cost-Effectiveness Analysis. *PharmacoEconomics*, Volume 37, pp. 631-643.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, 7(1), 6-10.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D., Macdonald, H., Johnston, M., Lamb, S.E., Dixon-Woods, M., McCulloch, P., Wyatt, J.C., Chan, A-W., & Michie, S. (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, 348.

Levin, H. M. et al., 2018. *Economic Evaluation in Education. Cost-Effectiveness and Benefit-Cost Analysis (Third Edition)*. London: Sage Publishing.

Montgomery, J.M., Nyhan, B. and Torres, M., 2018. How conditioning on post-treatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3), pp.760-775.

Oakhill, J., Cain, K., & Elbro, C. (2014). *Understanding and teaching reading comprehension: A handbook*. Routledge.

O'Hare, L., Stark, P., Cockerill, M., Lloyd, K., McConnellogue, S., Gildea, A., Biggart, A., Connolly, P., & Bower, C. (2019). *Reciprocal Reading Evaluation Report*. EEF.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). *Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research*. *Adm Policy Ment Health*.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research*, Vol. 3, pp. 545–561). Lawrence Erlbaum Associates Publishers.

Pressley, M. & Allington, R. L. (2015). *Reading instruction that works: the case for balanced teaching*, 4th edition. Guilford Press.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479–530.

Schaffer, J.L. (1999). Multiple imputation: a primer. *Statistical Methods in Medical Research*. 9(1):3-15. doi:10.1177/096228029900800102

Schulz, K. F., & Grimes, D. A. (2002). Sample size slippages in randomized trials: exclusions and the lost and wayward. *The Lancet*, 359(9308), 784.

Vassall, A. et al., 2017. *Reference Case for Estimating the Costs of Global Health Services and Interventions*, NA: GHCC.

Appendix A: Changes since the previous EEF evaluation

Appendix Table 1: Changes since the previous evaluation

Feature		Efficacy to effectiveness stage
Intervention	Intervention content	There are no key intervention content changes from the efficacy trial.
	Delivery model	Change from developer-led delivery to train-the-trainers delivery: Training for teachers was led by the developer in the efficacy trial, now it will be led by FFT trainers that have been trained by the developer before the intervention. The support sessions that are part of the training package were face-to-face in the school, and will be online in the effectiveness trial. The efficacy trial had two delivery models: a universal intervention where delivery was with the whole class, and a targeted intervention for small groups of pupils. The effectiveness trial will only have targeted delivery.
	Intervention duration	The previous efficacy trial the baseline results were administered in March 2017 and the endline results in April to June 2018. For the effectiveness trial, pupils will complete the pre- NGRT results between June and September 2023 and the final tests between March to July 2024.
Evaluation	Eligibility criteria	Targeted schools: The efficacy trial specified the same eligibility criteria for schools as the effectiveness trial the DfE has stipulated that 50% of the schools recruited for the trial must be from the DfE's 55 Education Investment Areas, areas facing 'entrenched deprivation', which includes the 12 areas which were previously classed as DfE Opportunity Areas. Targeted pupils: The efficacy trial tested a universal intervention (to the whole class) with Year 4 pupils and a targeted intervention (with a pre-selected group) with Year 5 and 6 pupils. The effectiveness trial will be a targeted intervention for Year 5 and 6 exclusively.
	Level of randomisation	No change (both trials clustered randomisation at the school level) The efficacy trial minimised randomisation with school-level covariates: %FSM and NGRT baseline scores. The effectiveness trial will randomise in two batches without stratification.
	Outcomes and baseline	The efficacy trial had dual primary outcomes: NGRT overall score and passage comprehension subscale. The effectiveness trial only has NGRT overall score as primary outcome. The passage comprehension subscale will be considered a secondary outcome. KS2 SATs have been added as another secondary outcome in the effectiveness trial.
	Control condition	No changes in the control condition