Reciprocal Reading

Evaluation Report

September 2019

Dr Liam O'Hare, Dr Patrick Stark, Dr Maria Cockerill, Dr Katrina Lloyd, Dr Sheila McConnellogue, Dr Aideen Gildea, Dr Andy Biggart, Professor Paul Connolly, Christine Bower

The Education Endowment Foundation (EEF) is an independent grant-making charity dedicated to breaking the link between family income and educational achievement, ensuring that children from all backgrounds can fulfil their potential and make the most of their talents.

The EEF aims to raise the attainment of children facing disadvantage by:

- identifying promising educational innovations that address the needs of disadvantaged children in primary and secondary schools in England;
- evaluating these innovations to extend and secure the evidence on what works and can be made to work at scale; and
- encouraging schools, government, charities, and others to apply evidence and adopt innovations found to be effective.

The EEF was established in 2011 by the Sutton Trust as lead charity in partnership with Impetus (formerly Impetus Trust) and received a founding £125m grant from the Department for Education.
Together, the EEF and Sutton Trust are the government-designated What Works Centre for improving education outcomes for school-aged children.

For more information about the EEF or this report please contact:

Jonathan Kay
Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

0207 802 1653

jonathan.kay@eefoundation.org.uk

www.educationendowmentfoundation.org.uk

Contents

About the evaluator

The project was independently evaluated by a team from the Centre for Evidence and Social Innovation (CESI) at Queen's University Belfast; Dr Liam O'Hare, Dr Patrick Stark, Dr Maria Cockerill, Dr Katrina Lloyd, Dr Sheila McConnellogue, Dr Aideen Gildea, Dr Andy Biggart, Professor Paul Connolly and Christine Bower.

The lead evaluator was Dr Liam O'Hare.

Contact details:

Queen's University Belfast,
Room 71.01,
004 69–71 University St,
Belfast
BT7 1LN,
Northern Ireland,
United Kingdom

Direct line: +44 (0)28 9097 5973

Email: l.ohare@qub.ac.uk

Executive summary

The project

The FFT Literacy Reciprocal Reading intervention aimed to improve reading comprehension and overall reading ability through two programmes: a whole-class ('universal') intervention for pupils in Year 4 (8–9 years old), and a targeted intervention for pupils with poor reading comprehension, but good decoding skills in Year 5 and Year 6 (9–11 years old). The programme, created by Fischer Family Trust Literacy (FFT Literacy), lasted for one academic year and was delivered by trained teachers and teaching assistants (TAs). The universal intervention was delivered to the whole class for 20 minutes at least once per week for at least twelve weeks over the year. The targeted intervention was delivered to groups of approximately six pupils, with a recommended delivery of at least two sessions of twenty minutes per week for twelve weeks.

This project was a randomised controlled trial. 98 schools from the North of England and the Midlands were randomised at school level to intervention or control groups. All pupils in Year 4 in each school participated in the universal intervention. Prior to randomisation, teachers of Years 5 and 6 were each asked to each select six pupils (total 12) for the targeted intervention, using guidance provided to them by FFT Literacy. Sessions consisted of collaborative reading of texts with the use of four evidence-based strategies: *predicting, clarifying, questioning, summarising.* The project measured pupils' overall reading, reading comprehension, reading accuracy and reading comprehension meta-cognition (the extent to which they recognised and used the four strategies). The process evaluation included observations, interviews and surveys with teachers, TAs and head teachers at the participating schools as well as focus groups with participating pupils. The intervention ran from September 2017 to June 2018.

Key conclusions	
1.	Children in the FFT Literacy Reciprocal Reading targeted intervention group made the equivalent of 2 additional months' progress in both primary outcomes (overall reading and reading comprehension), on average, compared to the equivalent children in the other schools. This result has a moderate to high security rating.
2.	There is no evidence that the FFT Literacy Reciprocal Reading universal intervention had an impact on pupils' overall reading or reading comprehension outcomes, on average. This result has a moderate to high security rating.
3.	Children in the targeted intervention made more progress in both secondary outcomes (reading accuracy and pupil comprehension meta-cognition), whilst the universal intervention showed effects for only pupil comprehension meta-cognition.
4.	Among children receiving free school meals (FSM) analysis showed an effect on reading comprehension for pupils in the targeted intervention, but not for the universal version. However, this initial analysis was conducted using information supplied directly from schools and was constrained by missing data. Subsequent analysis was able to match data for FSM eligible pupils within the trial using the National Pupil Database. These results found signs of promise for both the targeted and universal interventions on outcomes for children eligible for FSM. These results are summarised in full in appendix K.

EEF security rating

The findings from the both interventions have a moderate to high security rating. This trial was an efficacy trial, which tested whether the intervention worked under developer-led conditions in a number of schools. The trial was a well-designed and well-powered two-armed randomised controlled trial. However, the following factors reduced the security of the trial. For the universal intervention, 14.9% of the pupils who started the trial were not included in the final analysis of overall reading, with 15.78% missing for the analysis of reading comprehension and similar drop-out for the targeted intervention (13.54% and 14.21% respectively) because schools did not provide test data. Peer reviewers identified several threats to the validity of the findings, including concerns about the differences between pupils in intervention and control schools at the start of the intervention, and risks caused by the missing data. The evaluation team controlled for these differences in the analysis. Additional analysis indicated that the missing data did not impact the direction of the results.

Additional findings

Preliminary analysis suggested that the targeted version of FFT Literacy Reciprocal Reading found promising results for reading comprehension and overall reading ability. Initial analysis showed that pupils in receipt of FSM made increased progress with their reading comprehension but did not see the same improvement to their overall reading progress, though this was based on small numbers of pupils due to incomplete data and mean that these results should be treated with caution. Further analysis was undertaken following the trial to match pupils in receipt of FSM using information received from the National Pupil Database. These results indicate that the universal intervention had a small positive impact on overall reading for pupils in receipt of FSM. The analysis also indicates a positive impact for both overall reading and reading comprehension outcomes for the targeted intervention. This additional analysis and interpretation is included in full as an appendix to this report (appendix K).

The pupils in the universal intervention performed as well as pupils in comparison classes on most measures, implying that FFT Literacy Reciprocal Reading was as effective as usual practice in other schools. Pupils in this intervention performed significantly better than comparison pupils in one measure: pupil comprehension meta-cognition. However, exploratory analysis did not find that higher comprehension meta-cognition correlated with increased comprehension or overall reading attainment. Some teachers reported issues with pupils in the class accessing the programme, which could imply that the older age and appropriate targeting in the targeted programme led to the attainment gain.

Pupil-reported enjoyment of the programme was linked to higher reading comprehension at post-test. Teachers spent substantially more time delivering the universal and targeted programmes than the developer's minimum recommendation, delivering throughout the school year, well above the minimum of 12 weeks (though only 32 of 47 teachers submitted their time reports). This suggests that the programme is highly feasible to implement. Schools implementing the minimum may not see the same results, and even with this high intensity, the universal intervention did not show evidence of improved reading outcomes. The success of the targeted version of this programme is aligned with the wider research literature, indicating that comprehension can be improved by direct targeting.

Cost

The average cost of the FFT Literacy Reciprocal Reading universal intervention was around £2502, or £35 per pupil per year when averaged over 3 years, assuming delivery to one class of 25 pupils per year. The average cost of the targeted intervention was around £2436, or £135 per pupil per year averaged over 3 years, assuming the programme was delivered to 6 pupils per year.

Impact

Table 1: Summary of impact on primary outcomes

Outcome / Group	Effect size (95% confidence interval)	Estimated no. of months progress	No. of pupils analysed in the whole sample	*p*-value	EEF security rating	EEF cost rating
Universal intervention: Overall Reading (Y4)	0.00 (–0.06, 0.07)	0	3198	0.795	🔒🔒🔒🔓🔓	£ £ £ £ £
Universal intervention: Reading Comprehension (Y4)	–0.02 (–0.09, 0.06)	0	2958	0.494	🔒🔒🔒🔓🔓	£ £ £ £ £
Targeted intervention: Overall Reading (Y5 and Y6)	0.14 (0.04, 0.25)	2	1296	< 0.01	🔒🔒🔒🔓🔓	£ £ £ £ £
Targeted intervention: Reading Comprehension (Y5 and Y6)	0.18 (0.07, 0.29)	2	1270	<0.001	🔒🔒🔒🔓🔓	£ £ £ £ £

Introduction

Background evidence

In England, pupil progress in reading is recorded in independently marked tests at the end of Key Stage 2. In 2018, twenty-five per cent of students nationally failed to reach the expected standard of Level 4 in reading at the end of Key Stage 2 (Department for Education, 2018). There is, therefore, a need to find evidenced-based ways of raising literacy attainment for those failing to reach expected standards.

An influential UK-based meta-synthesis carried out by Brooks (2007) concluded that ordinary teaching (no intervention) does not enable children with literacy difficulties to catch up with peers whose reading development is in the normal range. There is evidence to support the view that the specific deficits of learners should be targeted in order to raise overall literacy attainment; for example, students who struggle with reading comprehension, as opposed to word-reading skills (Organisation for Economic Co-operation and Development, 2014). Brooks reported that it is difficult to produce positive outcomes for pupils with reading comprehension difficulties, necessitating targeted intervention approaches which tend to be expensive.

More recently, the Education Endowment Foundation (EEF) has commissioned 35 trials of language and literacy interventions in England (EEF, 2018a). There has been a wide range of results during these evaluations with the programmes having a median impact of two months' progress on attainment outcomes (EEF, 2018a). Some of the interventions which have demonstrated positive effects on reading outcomes in efficacy trials have explicitly targeted literacy skills. These include ABRA Online Reading Support (McNally, Ruiz-Valenzuela and Rolfe, 2016) delivered at Key Stage 1 (effect size (ES) = 0.23 for the non-online ICT version and 0.14 for the online ICT version) and Accelerated Reader (Gorard, Siddiqui and See, 2015) delivered at Key Stage 3 (ES = 0.24). Other interventions, which do not explicitly target literacy skills, but which focus on the development of thinking and meta-cognitive skills, have also shown positive effects for literacy attainments; for example, Philosophy for Children (Gorard, Siddiqui and See, 2015) delivered at Key Stage 2 (ES = 0.12), and Dialogic Teaching (Jay, Willis, Thomas, Taylor, Moore, Burnett, Merchant and Stevens, 2017) delivered at Key Stage 2 (ES = 0.15). This is interesting as the following study is of a specific literacy intervention which also has a thinking and meta-cognitive development focus.

The EEF toolkit evidence review on reading comprehension strategies (Education Endowment Foundation, 2018b) states that there is extensive evidence for reading comprehension strategies, defined as strategies which focus on the learners' understanding of written text and in which a range of techniques are explicitly taught to enable learners to comprehend the meaning of texts. The toolkit draws on evidence from over one hundred studies from between 1980 and 2011 and states that reading comprehension approaches are more effective than phonic or oral language approaches at Key Stage 2 for both short-term and long-term impact, leading to high impact on pupil learning, an average six months' progress, with very good strength of evidence (four out of five padlocks).

An example of a reading comprehension strategy intervention which is recommended by the EEF toolkit (Education Endowment Foundation, 2018b) is reciprocal teaching. This is informed by research and development in the United States that was led originally by Palincsar and Brown (Palincsar, 1982; Palincsar and Brown, 1984; Palincsar, 1986). Reciprocal teaching is now essentially a worldwide approach and it has been used with general school populations, those with learning difficulties and those with specific reading comprehension deficits. Although there are a range of interventions based on the principles of reciprocal teaching, the core factors are strategy instruction, modelling of strategies by an adult and student practice of the strategies.

Learners are explicitly taught four evidence-based strategies which effective readers are known to use, to help them derive meaning from texts. These are: summarising what has been read, asking questions about the text, predicting what will happen next, and clarifying new or unfamiliar words. The modelling is interactive in that, whilst an adult models strategic action with the text, there is a collaborative problem-solving approach to the text. Prior research on reciprocal reading has mostly been carried out outside of the UK, mainly in the US. Previous meta-analyses have claimed positive effects for reading strategy instruction (Sencibaugh, 2007; Swanson, 1999; Chiu, 1998) including reciprocal teaching (Rosenshine and Meister, 1994). In Rosenshine and Meister's (1994) meta-analysis of 16

reciprocal teaching studies, an average effect size of 0.32 was found for reading and 0.88 for reading comprehension using an experimenter-developed test rather than a standardised test. Evidence-based guidance which advocates the use of reading strategy instruction, such as reciprocal teaching, include that of The National Reading Panel (2000) in the US and that of Shanahan *et al.* (2010).

Reciprocal reading has not been widely implemented in the United Kingdom (UK) as a discrete intervention until recently, although the teaching of reading strategies has become a standard part of the reading curriculum in primary and secondary education over the last few decades both in the UK (Department for Education, 2013) and internationally (Pressley, 2002; Okkinga, 2018). Reciprocal teaching is now among the interventions recommended to improve reading comprehension in the EEF toolkit (EEF, 2018c), having been evaluated through a small efficacy trial (Crawford and Skipp, 2014). This randomised control study of the Literacy Intervention Toolkit examined the effectiveness of reciprocal teaching for students at the start of their post-primary education (Year 7) and involved 41 schools. It was a targeted intervention study focused on students who had scored below Level 4 at the end of primary school. The intervention was delivered in small groups with exposure of 3 to 4 hours per week. Although this study found a small positive effect size of ES>0.09, the evaluation could not conclude with certainty what impact the programme had on reading ability for those pupils who received the intervention because of issues with outcome testing, school attrition and balance between the intervention and control groups.

In spite of a large research base the EEF (2018b) stated that in recent UK-based evaluations of programmes using reading comprehension strategy instruction, the impact has not matched the extensive impact claimed in the international literature. There are some key issues that are worthy of note in the literature. The evidence base for reciprocal teaching generally centres on specific targeted populations (students with literacy or learning difficulties) and, in a recent meta-analysis of 52 studies (Okkinga *et al*., 2018), larger effect sizes have been shown for low-achievers in reading than for typically developing students. However, Okkinga *et al*. (2018) did report that reading strategy intervention is beneficial for universal populations, especially for students aged 8 to 14 (Key Stages 2 and 3). This is in spite of the fact that the authors acknowledged that the intervention is challenging to address in a whole-class setting and that it is harder to maintain intervention quality (Okkinga *et al*, 2018). In addition, Okkinga *et al*. (2018) reported that teachers are less successful than researchers at delivering reading strategy instruction. This may suggest that teachers require deeper understanding of reading comprehension itself and effective comprehension strategies in order to deliver the intervention.

Methodologically, smaller effect sizes (mean ES = 0.19) from studies using standardised tests compared to bespoke researcher developed reading comprehension tests (mean ES = 0.43) were found by Okkinga *et al*. (2018) and this is a common finding in the literature. There is also a lack of clarity about the optimal age at which reciprocal reading is beneficial. Some studies report that teachers find it hard to induce strategic thinking in students (Duffy, 1993) and that students showed poor application of reading strategies while collaborating (Hacker and Tenent, 2002). This may be due to the age and developmental stage of students receiving the intervention.

This report describes a cluster randomised controlled trial (CRCT) evaluation of a reciprocal teaching intervention delivered by FFT Literacy. This is an interesting study both methodologically and theoretically, as it uses pre/post standardised outcome measures, which include a specific measure of reading comprehension ability, allowing evaluation of the specific skills targeted by the intervention (i.e. the specificity matching principle: McWilliams *et al*., 2013). The current study also assesses a range of secondary outcomes, namely pupil comprehension strategies and reading accuracy as well as teacher attitudes to reading comprehension, which will help explore the theory of change (logic model: see Figure 1), for reciprocal reading approaches such as this. The study also incorporates a large sample (98 schools), which will improve study power and thus potential to identify effectiveness of the approach. It is also an interesting study in terms of implementation factors as it is delivered at Key Stage 2 rather than Key Stage 3, which should add to the research base on suitability of reciprocal reading approaches with this younger age group. In addition, the study looks at both universal and targeted implementations of the approach, which allows for comparison of who benefits most from the intervention, the general population or students who have been identified as having a specific comprehension deficit. Furthermore, the universal and targeted interventions are delivered to different age groups, allowing the potential to consider age as a factor, although this is not explicitly analysed as part of the study design. Finally, the intervention is based on the teacher continuing professional development model, whereby general teachers are trained to deliver the intervention, rather than specialists, which has potential benefits in terms of scaling and cost.

Intervention

FFT Literacy's training on reciprocal reading emphasises the importance of all readers applying the four strategies of predict, clarify, question, and summarise, rather than readers taking on roles within the group as in other reciprocal reading programmes. The training considers the difficulties involved in comprehension and how reciprocal reading addresses these, how to identify the children who will benefit most from reciprocal reading, as well as giving teachers opportunities to practise using the strategies, to build their understanding of the impact the approach can have. Additional sessions include considering how to build challenge into the lesson without losing the emphasis on children's talk and reviewing how best to enhance reciprocal reading through a rich and coherent comprehension curriculum. Training is supplemented by school visits which can include demonstration, review of developing practice and advice on embedding the approach into classroom practice.

FFT Literacy Reciprocal Reading is delivered by practising teachers/TAs in mainstream UK settings for pupils aged 5 to 11 years. Staff development is, therefore, an essential part of the programme including: (1) two days of teacher training; (2) two to three half days in-school support visits from the FFT Literacy team throughout the year; (3) teacher manuals; and (4) pupil books. The implementation of the FFT Literacy Reciprocal Reading programme includes teacher training delivered by the FFT Literacy, including external school training (for teachers and TAs) interspersed with internal follow-up support. Teachers who delivered the FFT Literacy Reciprocal Reading programme received two full-day external training sessions; one of these was in advance of programme delivery in October 2017 and the second training session took place in January 2018. Three internal follow-up sessions then took place during the implementation period. After the initial training session, teachers were provided with a set of core materials (including teacher manual, pupil book, journal activities and reading books) needed for delivering the programme aimed at improving comprehension ability of pupils in Year 4 (universal) and Years 5–6 (targeted).

The training covered the knowledge, skills and understanding that practitioners need to deliver the FFT Literacy Reciprocal Reading programme in either a universal or a targeted format. The training also covered an understanding of the nature of reading comprehension and an evidence-based package of strategies as well as instructional components, such as how to conduct FFT Literacy Reciprocal Reading sessions and associated issues such as choices of texts and the use of planning and recording sheets. Overall, the training is focused on:

1. increasing teachers' sense of the importance of comprehension
2. enhancing teachers' knowledge of key reading comprehension strategies
3. increasing the school comprehension ethos.

FFT Literacy Reciprocal Reading instruction is teacher-led with facilitation that involves collaborative reading of texts (where the teacher reads to the class; teacher and pupils read together; or pupils read the text independently and discuss in small groups). The activity during sessions is the use of evidence-based strategies – predicting, clarifying, questioning, summarising – modelled by the teacher and used collaboratively between teacher and students and between students, to derive meaning from the text. Hence both the teacher and the students use these strategies in their instruction and learning respectively. All teachers involved in the study had to make changes to their teaching focus and lesson timetable to make space to deliver FFT Literacy Reciprocal Reading sessions. The programme is recommended to be delivered during literacy time as a universal programme. When FFT Literacy Reciprocal Reading is delivered as a targeted programme it is recommended this be delivered to groups of 6 pupils as an additional session rather than replacing taught sessions.

The participants in this trial were students in mixed-ability Years 4, 5 and 6 classes (aged 8 to 11 years). Two variants of the programme were delivered: targeted and whole class. The trial assigned different year groups in the same schools to participate in different versions of the programme, with Year 4 pupils assigned to the whole class (universal version) and Years 5 and 6 pupils assigned to the targeted version. This assignment pattern allowed evaluation of the two different versions of the programme (universal and targeted) within one sample of schools.

The targeted intervention was delivered only to pupils who were selected by teachers as having reading comprehension difficulties but normal reading accuracy abilities. Teachers were given guidance on how to do this (see Appendix D). Teachers were asked to select 6 pupils per class. All teachers and TAs involved in delivering the

programme received two days' off-site training from FFT Literacy, a programme manual and pupil books, as well as on-site advisory support from the FFT Literacy team during delivery of the programme.

Universal (whole-class) version

Teachers involved in the universal version of the programme delivered FFT Literacy Reciprocal Reading sessions of 20–30 minutes in length, at least once a week, for a minimum of 12 weeks over two academic terms to students aged 8–9 years (Year 4 in primary schools in England). Over the course of the intervention, students encountered many texts, which were a set of short stories provided by the trainer. The key strategies (predict, clarify, question, summarise) were used flexibly and adapted to the texts. In the universal intervention the delivery was often with the whole class, although sometimes the children read in groups and the teacher supported one group at a time within the class. Whole class or whole class with small group reciprocal reading sessions within this were followed by individual tasks, which were based on the reading sessions, in the form of book journal activities. Whole-class reading sessions were led by the teacher with participation from the whole class, whereas whole-class with small-group reading sessions included whole-class input, with opportunities for groups of pupils to work together on elements of the reading session where appropriate (e.g. predict, clarify, question, summarise). The teacher then supported each small group in turn. This is contrasted with the targeted intervention where one teacher worked with a small group of approximately 6 pupils throughout each reading session.

The whole-class FFT Literacy Reciprocal Reading programme comprises:

a) a set of strategies (predict, clarify, question, summarise) used to strategically process text

b) an instructional dialogue

c) materials – texts

d) book journal activities (undertaken individually by pupils).

Targeted version

Teachers and TAs involved in the targeted version of the RR programme delivered guided-reading sessions of 20–30 minutes in length to small groups of 6 pupils aged 9 to 11 years (Year 5 and Year 6 in primary schools in England) who were identified as having reading comprehension difficulties but normal or above average levels of reading accuracy abilities. Teachers used guidelines (Appendix D) provided by FFT Literacy based on using the simple view of reading (Gough and Tunmer, 1986) to select the pupils who fit within the 'good decoders/poor comprehenders' quadrant. The frequency of the targeted version was twice a week for at least 12 weeks over two academic terms. Pupils were generally taken out of their normal classroom by their teacher or TA for the sessions, but there was flexibility in how this was done. The targeted version of the FFT Literacy Reciprocal Reading programme also comprised the four activities as described for the universal version above.

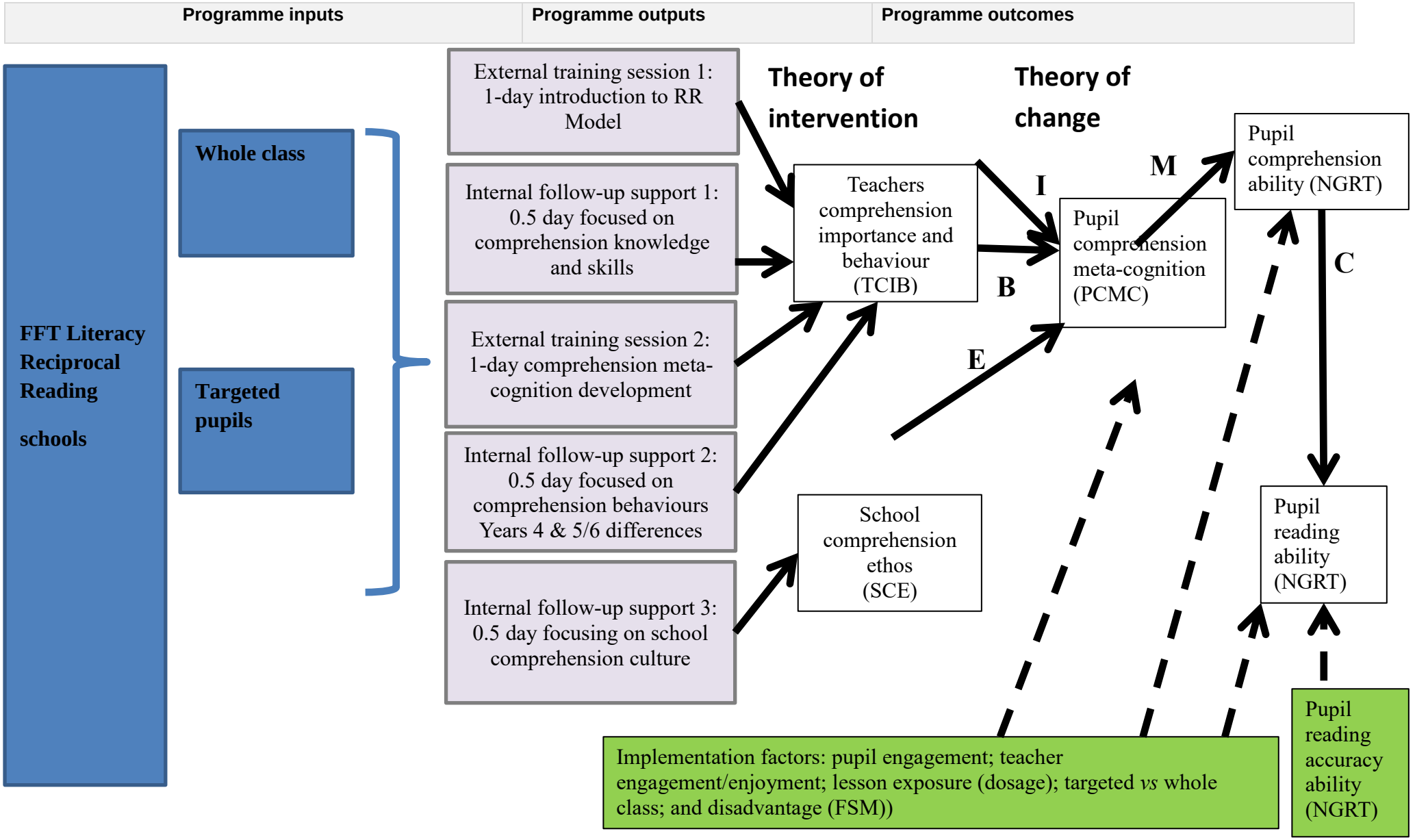
Control group

Control group schools proceeded with business as usual, comprising of their normal literacy and specific comprehension instruction practices for students aged 8 to 11 years.

Table 2: FFT Literacy Reciprocal Reading TIDieR checklist

Item	Content
Brief name	FFT Literacy Reciprocal Reading (two versions: 1. universal; and 2. targeted)
Why	Training programme for teachers to allow them to improve the reading comprehension ability of their pupils aged 8 to 9 years (universal) and pupils aged 9 to 11 years (targeted) and therefore raise literacy attainments.
What	Materials: A teacher training programme delivered by FFT Literacy, which includes external school training (for teachers and TAs) interspersed with internal follow-up support/training.
	Procedures: External school training for both universal and targeted versions is similar. Internal follow-up training/support sessions are tailored for the two different versions but have overarching themes of comprehension behaviours, awareness and school culture. Teachers and TAs provide reciprocal reading activities to pupils based on their training. This is in a whole-class context for the universal programme and a small-group withdrawal context for the targeted programme.
Who provided	FFT Literacy Reciprocal Reading Trainer provides teacher internal and external training. Teachers and TAs provide reciprocal reading activities to pupils based on their training.
How	Initial training sessions provided to groups of 20–30 teachers on average and consisted of the following 5 components: 1. External training session: 1-day introduction to RR model 2. Internal follow-up support: 0.5 day focused on comprehension knowledge and skills 3. External training session: 1-day comprehension meta-cognition development 4. Internal follow-up support: 0.5 day focused on comprehension behaviours; Years 4 and 5/6 differences 5. Optional: Internal follow-up support: 0.5 day focusing on school comprehension culture
Where	External training provided in out-of-school setting. Internal training provided in school setting including classroom.
When and how much	There are two external full-day training sessions and three internal half-day follow-up sessions over a 16-week period. Teachers deliver reciprocal reading sessions of 20–30 minutes' length. In the universal programme the sessions are delivered at a frequency of once a week for at least 16 weeks. In the targeted programme the sessions are delivered at a frequency of twice a week for at least 16 weeks over 2 academic terms.
Tailoring	There was no specific tailoring to the programme delivery beyond those specified between the universal and targeted versions. However, the programme logic model was tailored to fit the programmes activities and outcomes during the first Phase of the research and is included in Figure 1. This logic model specified that 5 specific outputs of the programme would impact on three outcomes: 1. increasing teachers' sense of comprehension importance (I) 2. enhancing teachers' comprehension behaviours (B) 3. increasing school comprehension ethos (E). As a result of changes in I, B and E would have a proximal impact on pupil comprehension meta-cognition (M), which would lead to distal changes in pupil reading comprehension ability (C) and overall reading (see programme theory description for more details).
Modifications	No programme modifications were made during the trial. Suggestions for modifications based on process evaluation data are included in this report.
How well	Planned: This has been assessed through the research process evaluation. Focus group interviews with students and individual interviews with school staff were carried out in 10 case-study schools and questionnaires administered.
	Actual: This was assessed through the programme cluster randomised controlled trial evaluation.

Figure 1: The FFT Literacy Reciprocal Reading logic model investigated in this evaluation



Programme theory

The team of independent evaluators worked with the intervention team (FFT Literacy) to design a logic model for the intervention for RCT evaluation purposes (see Connolly *et al.,* 2017 Chapter 2 for more details on using logic models in RCT evaluations). Figure 1 provides a graphical representation of the programme components (inputs, outputs and outcomes) and the theories tested by the trial are outlined briefly below. Specific details on these theories are described in more detail in O'Hare *et al*. (2018).

Theory of intervention

The theory of intervention describes the relationship between programme outputs and proximal (short-term) outcomes. The first component pathway of the FFT Literacy Reciprocal Reading theory of intervention is that programme training days aim to improve teachers' sense of comprehension importance and behaviours. Teachers are trained on both how to deliver the programme, and how to identify children who may have reading comprehension difficulties but have normal reading accuracy abilities. For this trial, teachers were given guidance on how to select the pupils who would be eligible for the targeted intervention, and selected the pupils pre-randomisation; and all teachers received the training post-randomisation. Notwithstanding, teachers were trained both how to deliver the programme, and how to identify children who may have reading comprehension difficulties but have normal reading accuracy abilities. To do this FFT Literacy Reciprocal Reading training explains the simple view of reading (Gough and Tunmer, 1986) to teachers and how to identify the pupils who may fit within the 'good decoders/poor comprehenders' quadrant (Nation and Snowling, 1997). The second pathway of the theory of intervention is that programme training aims to impact school comprehension culture and ethos by encouraging teachers and school leadership to improve awareness and prioritisation of reading comprehension in schools and not just focus on reading accuracy/decoding.

Theory of change

The theory of change is that teacher practice is improved by the earlier training, described above in the theory of intervention. There are four component pathways within the theory of change. Firstly, the programme aims to effect change on pupil comprehension meta-cognition (PCMC) via teacher comprehension importance (I in Figure 1: see secondary outcomes in methods for description) and behaviour (B in Figure 1: see secondary outcomes in methods for description). The second pathway of the theory of change is that an improvement in school comprehension ethos (E in Figure 1) improves pupil comprehension meta-cognition. The third pathway is that improvements in PCMC drive change (M in Figure 1) in pupil comprehension ability. The final pathway in the theory of change describes how change in reading comprehension ability (C in Figure 1) will have a subsequent effect on overall reading ability. If a pupil experiences improvement in reading comprehension score, as assessed by the passage comprehension sub-component of the New Group Reading Test (NGRT), this will naturally improve their overall reading score.

Implementation theory

In addition to investigating the relationship between the intervention, outputs and outcomes, this study explored which implementation factors were related to outcome change for primary (reading comprehension and overall reading) and secondary outcomes (reading accuracy and pupil reading comprehension meta-cognition). These implementation factors cover: pupil disadvantage; dosage (exposure to lessons); and pupils' and teachers' levels of programme engagement (in terms of enjoyment, engagement, feelings of benefits of the programme measured at post-test). Due to the close relationship between reading accuracy and reading comprehension, reading accuracy was also analysed as an implementation factor. This was to investigate if the reading accuracy was a significant predictor of overall reading at post-test. Normal levels of reading accuracy are a recommended prerequisite for inclusion in the targeted intervention (i.e. 'good' reading accuracy, 'poor' reading comprehension). This prerequisite of the programme suggests that reading accuracy is important for the success of the programme. This implementation factor analysis of reading accuracy as a predictor of primary outcome, therefore, was to analyse if reading accuracy prior to the programme was actually important for later reading outcomes after receiving the programme.

Evaluation objectives

As documented in the EEF trial protocol,[1] and statistical analysis plan and independently published trial protocol (O'Hare *et al.*, 2018): the main aim of the evaluation was to determine the effect of targeted and universal versions of FFT Literacy Reciprocal Reading on primary outcomes of reading comprehension and overall reading ability in children in primary schools aged 8 to 11 years. This study also looks at FFT Literacy Reciprocal Reading effects on a series of secondary outcomes related to pupil and, teacher comprehension attitudes and behaviours as well as the overall implementation of the FFT Literacy Reciprocal Reading programme. Schools' ethos towards the prioritisation of reading comprehension is measured as a potential implementation factor for programme success.

The evaluation is designed to answer the following research questions:

1. What is the impact of the FFT Literacy Reciprocal Reading training programme at post-test on reading outcomes (primary: overall reading and reading comprehension and secondary: reading accuracy and reading comprehension meta-cognition) in pupils participating in a universal/whole-class version?

2. What is the impact of the FFT Literacy Reciprocal Reading training programme at post-test on reading outcomes (primary: overall reading and reading comprehension and secondary: reading accuracy and reading comprehension meta-cognition) in pupils participating in a targeted version which involves pupils identified with having relatively good reading accuracy with relatively weaker reading comprehension skills?

3. What evidence is there to support the pathways for change in both primary and secondary outcomes as proposed in the logic model for both the universal and targeted versions of the programme?

4. What is the relationship between outcomes and implementation factors for both the universal and targeted versions of the programme?

5. What does the implementation process data tell us about how the programme was implemented?

Ethics, data protection and trial registration

The ethics were approved through the ethics committee in the School of Social Sciences Education and Social Work at Queen's University Belfast and granted on 23 March 2017.

Information sheets were issued to the parents of all pupils potentially receiving the programme in May 2017. These information sheets also included a form for parents to return if they wished to remove their child's data from the data processing (June 2017[2]) i.e. QUB would not access their child's reading scores or NPD data. A further opportunity to remove their child's data from processing was provided in June 2018. This was to ensure that parents were aware of what data was being processed for their child before they made a decision to withdraw this data from processing. Schools were provided with a memorandum of understanding. All ethics documentation is available in Appendix E.

All pupil data was treated with the strictest confidence and is stored in accordance with the data protection legislation, including the General Data Protection Regulation (GDPR) which came into effect in May 2018. Research data will be held securely within Queen's University for a minimum of five years in line with the data retention policy in the School of Social Sciences, Education and Social Work. Personal data was processed as per condition 6(1)e of the GDPR under public interest purposes, because the research is considered to be a 'task carried out in the public interest'. The public benefit of this work will be determining if the FFT Literacy Reciprocal Reading intervention is successful in

[1] https://educationendowmentfoundation.org.uk/projects-and-evaluation/projects/reciprocal-reading/
[2] The legal bases for processing personal and special data changed during the course of this study. The recruitment documents were completed pre-GDPR, when the evaluation team relied on opt-out consent but subsequently the legal bases for processing personal and special data was processed in the public interest.

improving pupils' reading attainment. This reading intervention is currently being used in schools, and it is in the public interest to examine what effect it has on pupils' reading ability. The special data (FSM) is processed under the legal basis of it being necessary for archiving, scientific or historical research purposes (GDPR, 9 (2) j).

The trial has been registered on the ISRCTN website and summarised here:
http://www.isrctn.com/ISRCTN81582662

Project team

Evaluation team

The evaluation team is drawn from senior and experienced staff within the Centre for Evidence and Social Innovation (CESI) at Queen's University Belfast.[3] CESI has considerable experience in the conduct and analysis of randomised control trials and cluster randomised control trials in educational and community settings. CESI staff have undertaken over 50 randomised control trials in educational settings over the past 10 years. The centre has developed particular expertise in working with programme developers in the design and reporting of trials, while providing a thorough and robust independent evaluation of programme impacts. The team also has experience in accounting for the nested nature of the data when schools are randomised at the school level through the use of multilevel modelling.

Dr Liam O'Hare (Principal Investigator) had overall responsibility for project delivery, including the final report. He also led on the research design, refinement of the programme theory and logic modelling for the programme.

Dr Patrick Stark (Trial Manager) was the lead analyst and oversaw data collection, data management and data analysis for the project as well as writing aspects of the final report.

Dr Maria Cockerill (Programme Manager) had particular responsibility for the recruitment and retention of schools. She oversaw the school registration, MOU and consent[4] processes as well as some implementation data collection aspects of the trial and writing aspects of the final report.

Dr Katrina Lloyd (Survey Expert) supported the development of survey instruments in the project including the head teacher survey in the Memorandum of Understanding, teacher surveys and post-test implementation survey as well as writing aspects of the final report.

Dr Sheila McConnellogue (Educational Psychologist) acted as an advisor and contributed to literature review, data interpretation and writing the final report.

Dr Aideen Gildea (Process Evaluation Specialist) supported several aspects of the trial data collection, cleaning and analysis including the process evaluation.

Dr Andy Biggart (Educational RCT Expert) contributed to the research design and data analysis in the project as well as writing aspects of the final report.

Professor Paul Connolly (Educational RCT Expert) acted as an expert consultant for the project and played a key role in quality control, interpretation of findings, final write-up of the report and research dissemination.

The team also drew upon other experienced research staff within CESI to conduct fieldwork for the evaluation and would like to acknowledge Dr Nicole Craig for her valuable comments on the analysis.

[3] https://www.qub.ac.uk/research-centres/cesi/
[4] The legal bases for processing personal and special data changed during the course of this study. The recruitment documents were completed pre-GDPR when the evaluation team relied on opt-out consent, but subsequently the legal bases for processing personal and special data were processed in the public interest.

Christine Bower (Evaluation Support) supported trial data collection and analysis of the process evaluation data.

Project team

Andy Taylor (Project Team Lead) from the Fisher Family Trust Literacy reviewed documents produced by the evaluation team, co-designed the programme logic model and was responsible for the implementation of FFT Literacy Reciprocal Reading programme, and provided some of the external training and follow-up visits.

John Catron (Training Associate) from the Fisher Family Trust Literacy provided the remainder of the external training and follow-up visits.

Methods

Trial design

Universal version FFT Literacy Reciprocal Reading

Trial type and number of arms		Efficacy trial: 2 arms (intervention and control)
Unit of randomisation		School
Stratification variable(s) (if applicable)		Minimisation variables: reading comprehension at baseline (NGRT Passage Comprehension Score), reading accuracy at baseline (NGRT sentence completion Score) and School FSM %
Primary outcomes	variable	Overall Reading Ability and Reading Comprehension Ability (both primary outcomes)
	measure (instrument, scale)	NGRT Overall Reading Score and NGRT Passage Comprehension Score
Secondary outcome(s)	variable(s)	Reading Accuracy and Comprehension Meta-cognition
	measure(s) (instrument, scale)	NGRT sentence completion Score and pupil comprehension meta-cognition (PCMC[5] – bespoke measure)

Targeted version FFT Literacy Reciprocal Reading

Trial type and number of arms		Efficacy trial: 2 arms (intervention and control)
Unit of randomisation		School
Stratification variable(s) (if applicable)		Minimisation variables: reading comprehension at baseline (NGRT Passage Comprehension Score), reading accuracy at baseline (NGRT sentence completion Score) and School FSM %
Primary outcomes	variable	Overall Reading Ability and Reading Comprehension Ability (dual primary outcome)
	measure (instrument, scale)	NGRT Overall Reading Score and NGRT Passage Comprehension Score
Secondary outcome(s)	variable(s)	Reading Accuracy and Comprehension Meta-Cognition
	measure(s) (instrument, scale)	NGRT sentence completion Score and pupil comprehension meta-cognition (PCMC – bespoke measure)

[5] The PCMC measure was developed by the evaluation team based on the four key concepts of the RR programme: predicting, clarifying, questioning and summarising. This was to ensure that the measure provided specificity matching to the intervention which is essential for construct validity. The responses to the items were added together to create a scale which was tested for internal consistency reliability and factor structure.

1. Cluster RCT evaluation: The main outcomes of the FFT Literacy Reciprocal Reading training were evaluated using an intention to treat cluster RCT (school clusters) with control-group analysis. The RCT tests for changes in pupils' primary reading outcomes (reading comprehension, overall reading ability) as well as a range of secondary outcomes (reading accuracy and reading comprehension meta-cognition). Both the universal whole-class version and the targeted small-group version of the implementation were tested. Any changes in the intervention group receiving the FFT Literacy Reciprocal Reading training were measured against the control group who did not receive the treatment. Control schools proceeded within a 'business as usual condition' and agreed not to participate in any other EEF literacy trial at KS2 during implementation of FFT Literacy Reciprocal Reading. Control schools were provided with £1000 as compensation for their time to complete testing and intervention schools paid a small contribution of £500 to participate, thereafter to receive the programme at no further cost, as compensation for their participation. Pre-test NGRT data was collected in June–July 2017. Post-test NGRT data was collected in June–July 2018.

2. Process evaluation. A process evaluation supplemented the RCT to measure implementation factors, assessing dosage, reach, fidelity and quality. To help assess this, all teachers and TAs delivering the programme completed a questionnaire, an audit tool was administered by FFT and observations and interviews were carried out in ten case-study schools.

Participant selection

Recruitment for the study took place from March 2017 to June 2017 and 119 schools were recruited by the Centre for Evidence and Social Innovation (CESI) and signed a Memorandum of Understanding (MOU: see appendix E). Of these, 14 schools withdrew pre-randomisation owing to changing circumstances including emerging capacity issues. In addition, 35 schools who signed the MOU were excluded pre-randomisation, as there were not enough schools from this geographic area, and 6 were excluded because they were already participating in part of another on-going EEF literacy trial. This left 100 schools to be pre-tested and randomised. A further 2 schools were excluded post-randomisation due to ineligibility as explained below.

Therefore, 98 primary schools in England with particular focus on the North East of England were in the ITT model of this study. Outside this area, recruitment was targeted to include schools in geographic clusters from any one area. Areas of high disadvantage were given priority and invited to participate with others in their geographic region. In addition, schools of all types (not targeted) expressed an interest to participate through the EEF website. There were no other specific location criteria. In total, 31 schools from a total of 98 schools were from the North East of England (32%). The schools were selected to ensure that there is was an overall mean in the implementation group (98 schools) of at least 29% EverFSM, which is the national average. Schools with higher than average levels of Year 6 pupils eligible for FSM from the North East were contacted first and priority was given to these schools when they expressed an interest to participate during selection. Further eligibility criteria were that schools had not received the FFT literacy package of training to deliver FFT Literacy Reciprocal Reading and were not involved in another on-going EEF trial for pupils aged 8 to 11 years. Two schools who signed up to the trial failed to disclose that they had already signed up to a similar EEF literacy trial. These two schools were therefore ineligible to continue with the reciprocal reading trial. Once this had been discovered, after pre-testing and randomisation, these schools were excluded from the reciprocal reading trial.

Head teachers from each of the schools were asked to read and sign a MOU, which explained their commitments to the programme and the research. Schools also completed a short questionnaire on usual literacy practice within the school with a focus on reading comprehension practices. Parental information about the trial and data collection were provided with the option to withdraw their child's data from the study. Schools allocated to the control group did not receive the training and continued as normal with regular curriculum and usual classroom activity.

The universal intervention was delivered to pupils in Year 4. The targeted intervention was delivered to pupils in Years 5 and 6. For the universal programme all children in the respective year were given the intervention by an FFT Literacy Reciprocal Reading trained instructor. The eligibility criteria for the targeted version were children who were poor at reading comprehension but relatively good at decoding. The children in this category were selected by teachers in June 2017 using guidance and materials provided by FFT literacy (see Appendix D) pre-baseline testing and pre-randomisation. Teachers were provided with a document that allowed them to compare each child in their

class with two sets of criteria. The first set of criteria described the reading skills possessed by a child who could be classified as having good decoding skills. The second set of criteria described difficulties that may be experienced by a child who struggles with reading comprehension. By comparing each child with these two sets of criteria, teachers identified children who were relatively good at decoding, but struggled with reading comprehension. Children who matched the provided criteria for having good decoding skills and poor reading comprehension skills were deemed to be eligible for the targeted intervention. Teachers selected six pupils from each Year 5 and Year 6 class for the targeted intervention. Although there may have been more than six pupils who met the criteria for receiving the targeted intervention, teachers were asked to select only six. A caveat to consider here is that the sample selected for the targeted intervention may have been susceptible to teacher bias. However, teachers were given standardised instruction on how to select the pupils based on the simple view of reading (see Appendix D)

Assigning different year groups in the same schools to participate in different versions of the programme allowed evaluation of the two different versions of the programme (universal and targeted) within one sample of schools.

Outcome measures

The following section describes the primary and secondary outcome measures used in the study. It also describes measures that were used for additional analysis and implementation analysis. We understand that EEF requires previously standardised measures for *primary* outcomes, but also that the need for specificity matching of secondary outcomes and implementation factors requires the need to use bespoke measures where appropriate and has been the norm with numerous previous trials, e.g. development of bespoke pupil/teacher engagement measures for programmes (O'Hare *et al*., 2017, O'Hare *et al*., 2015). Also, secondary outcome measures are used to explore theory of change. Therefore, it is important to match secondary outcomes even more closely than primary outcomes. There was no pilot phase for secondary and implementation measures in the protocol for this RCT as time and funding did not allow for this. However, the measures were developed through careful consideration of content validity and *post hoc* reliability and validity tests were conducted. If reliability and validity tests showed issues with the measures they were not included in the analysis. Essentially, secondary outcome measures are used to explore the programme logic model and the primary standardised outcomes are the main test of programme effectiveness.

Primary outcomes

The primary outcomes of the trial are two attainment indicators from the digital New Group Reading Test (NGRT): 1. overall reading score; and 2. comprehension subscale score: see Table 3). Although current EEF practice is to prioritise one primary outcome, dual primary outcomes were agreed for both interventions in the set-up phase for the following reasons. The NGRT is an adaptive test which has high reliability (see Table 3 for reliability of outcome measures) and measures both reading accuracy and comprehension. The paper version of the NGRT (recommended for pupils aged 6 to 16 years) was standardised in 2010 using data from 11,716 students, then further paper versions were standardised in 2014 using 4960 students (GL Assessment, 2018). When the paper questions were converted to a digital version, a study of 1721 pupils in 2010 found a floor effect for pupils in Year 1, and the digital version is therefore recommended only for Year 2 (pupils aged 7) onwards. All pupils who take the NGRT test receive two subscales. The NGRT has three possible subscales: sentence completion, passage comprehension and phonics. Sentence completion is the first subscale and was used as the Reading Accuracy secondary outcome. Pupils then progress to the passage comprehension subscale (used as the Reading Comprehension primary outcome), unless they score at an extremely low level for the sentence completion, in which case they progress to the more basic Phonics subscale instead. The number of items for each subscale is adaptive, as the difficulty of the next item is chosen based on their performance on items already administered. The maximum number of items depends on pupil performance. The aggregate Overall Reading score is calculated by GL before the researcher downloads the scores from the GL software, and is based on the two subscales each pupil received. All Y4, Y5 and Y6 pupils were provided with the NGRT digital test. The adaptive system means that different test forms for different ages are not necessary and therefore not offered by GL. One caveat for the adaptive testing system used by the digital NGRT test is that Reading Comprehension (passage comprehension) outcome scores were not available for pupils who had scored at an extremely low level on the Reading Accuracy (sentence completion) subscale; these pupils received the phonics test instead. Overall Reading outcome data was still available for these pupils, as the GL system still generates an overall score, based on sentence completion and Phonics. There were, therefore, fewer pupils with reading comprehension outcome data than there were pupils with Overall Reading data. Overall reading is analysed as a

primary outcome as it may be seen by teachers as an accessible and easily interpretable outcome of an intervention. Reading comprehension is analysed as a primary outcome as it is the specified target of the programme and, therefore, arguably the most appropriate way to evaluate programme success (see logic model Figure 1). Considering the benefits of each of these reading outcomes, it was agreed with EEF that both would be analysed as primary outcomes. At pre-test the NGRT was delivered (digitally) by schools under exam conditions. At post-test the NGRT was administered (digitally) by the evaluation team under exam conditions with a field worker in attendance to ensure objective assessment conditions. Field workers were research staff from CESI. Due to the scale of the project and the geographical spread of schools, it was necessary to use this team of fieldworkers to ensure schools were visited to oversee the delivery of post-testing. All fieldworkers were trained by the evaluation team. This increased the chance of objective assessment conditions as the fieldworkers were instructed to not provide any additional help to pupils, outside of the required test instructions, that may impact their scores. The NGRT provides two versions of the digital tests, A and B, and a different test was used at pre-test and post-test to avoid practice effects. The digital NGRT form A was delivered at pre-test. Digital NGRT form B was delivered at post-test. Overall reading scale score is used as the score for Overall reading ability. Passage Comprehension scale score is used as the score for reading comprehension.

Table 3: Primary outcome measures

Primary outcome	Measure	Level of measurement	Number of items	Reliability (Cronbach's Alpha)	Range
Reading Comprehension	New Group Reading Test: passage comprehension	Pupil	27	0.9	Minimum possible score = 0, maximum = 500
Overall reading ability	New Group Reading Test: total score	Pupil	47	0.9	Minimum possible score = 0, maximum = 500

Multiple testing

There are two primary outcomes for this trial and there is much debate in the literature over cut-offs and criteria for correcting for multiple comparisons. There are similarly harmful aspects to correcting when unnecessary as to not correcting when necessary (Wason, Stecher and Mander (2014). The claimed advantage of correction for multiple comparisons is that it can reduce the chance of a Type 1 error, by enforcing a stricter alpha cut-off for significance, dependent on the number of tests in the analysis (the greater the number of tests, the stricter the alpha cut-off). This, however, inherently increases the chance of a Type 2 error, as an extremely strict and small alpha value for significance may result in failing to reject the null hypothesis when it should have been rejected (Gelman, Hill and Yajima, 2013; Perneger, 1998). The analysis in this study meets the criteria to not require correction for multiple testing due to a small number of planned comparisons (Armstrong, 2014).

The relationship between the two primary outcomes should also be considered here. The overall reading score is an aggregate score which includes the other primary outcome, reading comprehension score. As such, if a Type 1 error, i.e. a 'false positive' result is found for one, the chance of a similar situation for the other is increased. Concern with reducing Type 1 error is based on the fact that identifying a relationship which occurred by chance, when in fact the relationship between the factors is zero, would be a major problem for both theory and practice. However, the likelihood of a true null hypothesis, i.e. *zero* relationship between factors in social science is small, unlike a field such as genetics where such a relationship is highly likely (Gelman *et al*., 2013). It should also be considered that for students with the lowest reading ability, this aggregate score comprises accuracy and phonics, rather than accuracy and comprehension. This means that change in their overall reading outcome may be based on different types of reading skill than for the rest of the sample. Despite this, change in overall reading for these pupils would still be representative of improved reading ability, as if they progress to receiving the reading comprehension subtest at post-test, this indicates improved reading ability, even if it is not directly comparable with the phonics questions they received at pre-test.

It should also be noted that the trial tested the two interventions in the same schools. This suggests that any bias or imbalance that exists in the sample of schools exists for both interventions. However, randomising the two interventions into two separate randomisations of schools would have involved either: 1. Recruiting 200 schools, and delivering to 100 schools, which was beyond the current capacity of the programme developer; or 2. Conducting two

separate randomisations in the one sample of 100 schools, which would have resulted in contamination effects, i.e. if school A was randomised into being an intervention school for targeted but not universal, contamination may still have happened for the control children in the universal analysis. All targeted pupils were identified pre-randomisation which eliminated the chance of bias or contamination during the selection process. A caveat to consider here is that contamination from one intervention may have changed practice in the classes receiving the other intervention. For example, practice from the targeted intervention may have influenced teachers in the universal intervention to use an additional targeted approach with their lower-ability pupils.

The sample-size calculations did not account for two primary outcomes within the trial. Sample size outcomes were based on overall NGRT outcome. The process of sample-size calculations at the protocol stage began before dual outcomes had been agreed on for the protocol, and to maintain consistency of sample-size calculations at each stage of the trial, this approach was maintained for the duration of the research.

Current EEF guidance advises correction for multiple comparisons. Due to there also being disadvantages to correcting for multiple comparisons (increased Type 2 errors), this trial did not apply such a correction. This means that the chance of this analysis failing to reject the null hypothesis (i.e. reject the hypothesis that there is zero relationship between the factors in the analysis) is reduced compared to the analysis if we had applied a correction. This also means there is a higher chance of this analysis rejecting the null hypothesis when it should not have been. The findings should be interpreted in full acknowledgement of both of these considerations. Future research in this area, using multiple comparisons to investigate multiple primary outcomes may benefit from using a Bayesian 'New Statistics' approach which does not depend on decisions based on the chance of false positives and error rates (Kruschke and Liddell, 2017).

Secondary outcomes

The secondary outcomes of the trial are reading accuracy score from the sentence completion subtest of the NGRT and comprehension meta-cognition (Table 4).

The pupil comprehension meta-cognition measure was developed by the evaluation team. This bespoke measure was designed to closely reflect the principles of FFT Literacy Reciprocal Reading and was developed to provide specificity matching to the intervention as no previously used standardised measures would be appropriate for this. Originally, two variables were planned: meta-cognition and pupil comprehension behaviour, as detailed in the SAP. However, during reliability and validity analysis, these were found to load onto one factor, so it would not have been appropriate to analyse as two (as in the SAP). This is why there are now only a total of two secondary outcomes in the analysis. As the children were also completing the NGRT and a measure of engagement with the RR programme (intervention group), the meta-cognition questionnaire needed to be as short as possible. Therefore, the evaluation team developed six questions that best matched the four key aspects of the programme: predicting, clarifying, questioning and summarising. The six items related directly to the two overarching concepts of meta-cognition and comprehension. The children completed the measure at post-test which was delivered online (using Questback) and completed in the classroom. The NGRT was completed first, then the secondary measures. It should be considered here that if pupil responses to the secondary measure were influenced by completing the primary measure first, the possibility of this was consistent across the whole sample. Furthermore, it is possible that completing a reading meta-cognition measure soon after performing a reading task is appropriate, as the skills and approaches used by the pupil could be particularly salient, thus allowing an accurate representation of their meta-cognition. These secondary pupil measures were completed during the same testing session as the primary outcome measure and took approximately 10 minutes to complete. Combined with the NGRT which takes approximately 20–30 minutes, the post-test session for pupils was approximately 30–40 minutes. They were asked to 'think about what kinds of things you can do to help you to understand a story better when you read it' and the items related specifically to the strategies they were taught in the FFT Literacy Reciprocal Reading sessions (e.g. 'Before I read a text I ask myself what I already know about the subject of the story' and 'I find ways to help me understand words I am not sure about such as looking them up in a dictionary.' The responses ranged from 1 (Always) to 5 (Never). Internal consistency reliability was measured using Cronbach's alpha and was deemed acceptable (0.68) given that the conventional cut-off is 0.70. The questions were designed by members of the evaluation team with expertise in reading programmes to ensure the measure had construct validity. Factor structure was assessed using Maximum Likelihood as the data were normally distributed.

One factor was extracted (eigenvalue 2.30) accounting for 26.46% of the variance. All items loaded 0.30 or above. Therefore, the measure was deemed to be a valid measure with acceptable reliability.

Table 4: Secondary outcome measures

Secondary outcome	Measure	Level of Measurement	Number of Items	Reliability (Cronbach's Alpha)
Reading accuracy	New group reading test: sentence completion	Pupil	20	0.90
Reading comprehension meta-cognition	Pupil comprehension meta-cognition	Pupil	6	0.68

There were also a number of measures used for additional analysis to help explore the programme logic model and associated theoretical pathways (see Figure 1). These were: teacher comprehension importance; teacher comprehension behaviour and school comprehension ethos. These bespoke measures were developed by the evaluation team, in conjunction with the programme developer, to ensure they closely matched the training that was delivered to the teachers.

Teacher comprehension importance and behaviour

Teachers completed a twenty-four-item post-test comprehension importance and behaviour questionnaire (TCIB) delivered online (again using Questback). This was not completed at pre-test as it was developed during the evaluation. Section 1 (11 items) of the TCIB questionnaire assessed their views on the importance of teaching reading strategies which reflected the four key aspects of the FFT Literacy Reciprocal Reading programme: predicting (e.g. 'it is important to teach children to anticipate what may happen next in a text to aid understanding.'); clarifying (e.g. 'it is important to teach children how to identify words and phrases they are unsure about when reading text.'); questioning (e.g. 'it is important to teach children how to ask questions about the text to help them understand what they are reading.') and summarising (e.g. 'It is important to teach children how to sum up what they have read to check how well they have understood the text.'). The responses ranged from 1 (Strongly agree) to 5 (Strongly disagree). The questions were designed by members of the evaluation team with expertise in reading programmes to ensure the measure had construct validity. As noted above, they were chosen to reflect the four main aspects of the RR programme covered in the training. The scale had excellent internal consistency reliability as assessed by Cronbach's alpha (0.93). Factor structure was assessed using principal axis factoring as the data were not normally distributed. One factor was extracted (eigenvalue 6.57) accounting for 55.74% of the variance. All items loaded 0.60 or above. Therefore, the measure was deemed to be valid and reliable.

Section 2 (10 items) of the TCIB questionnaire assessed the teachers' behaviours in relation to these four strategies (e.g. 'I explicitly teach children to anticipate what may happen next to aid understanding'; 'I explicitly teach children how to identify words and phrases they are unsure about when reading text'; I explicitly teach children how to ask questions about the text to help them understand what they are reading', and 'I give children the opportunity to consolidate their understanding of what they have read in individual post-reading tasks'). The responses ranged from 1 (Always) to 5 (Never). The questions were designed by members of the evaluation team with expertise in reading programmes to ensure the measure had construct validity. As noted above, they were chosen to reflect the four main aspects of the RR programme covered in the training. The scale had excellent internal consistency reliability as assessed by Cronbach's alpha (0.87). Factor structure was assessed using principal axis factoring as the data were not normally distributed. Two factors were extracted (eigenvalues 4.89 and 1.06 respectively) accounting for 50.41% of the variance. Factors were rotated using the Direct Oblimin method (assuming the underlying factors are correlated) (Field, 2013). Six items loaded on factor 1 and four on factor 2.

The final three questions assessed the teachers' behaviour in relation to choosing and preparing texts for teaching reading (e.g. 'I read and prepare the texts I use prior to leading sessions with the children.') The scale had an acceptable internal consistency as assessed by Cronbach's alpha (0.69), particularly given the small number of items (3). Factor structure was assessed using principal axis factoring as the data were not normally distributed. One factor was extracted (eigenvalue 1.85) accounting for 43.96% of the variance. All items loaded 0.50 or above. Therefore, the measure was deemed to be valid and reliable.

School comprehension ethos

The school-level comprehension ethos questionnaire (SCE) was completed by head teachers and is designed to measure the prioritisation and focus of the teaching of reading comprehension in the school, e.g. if there is an ethos within the school that reading comprehension is an aspect of development to be strongly emphasised within teaching. This was completed at pre-test only. The measure consisted of twelve items. The head teachers were asked how much they agreed or disagreed with a range of statements including 'Current teaching in Years 4, 5 and 6 focuses more on comprehension skills than on decoding skills'; 'Teachers are aware of instruction strategies that target reading comprehension difficulties', and 'There is a lot of support and/or training available to teachers for improving reading comprehension in my school'. The internal consistency of the scale as assessed by Cronbach's alpha was low at 0.62. Removing one of the items i.e. 'Current teaching in Years 4, 5 and 6 places as much emphasis on reading comprehension as on decoding' increases the scale's reliability to 0.68. However, given that the removal of this one item would not increase the scale's reliability to 0.70 or above, the item was retained. Overall, the results of analysis using this scale should be treated with some caution.

Table 5: Implementation factor instruments: level of measurement and reliability

Implementation factor	Measure	Number of items	Level of measurement	Reliability (Cronbach's alpha)	Source
Teacher comprehension importance	TCIB (section 1)	11	Teacher	0.93	Schools
Teacher comprehension behaviour	TCIB (section 2)	10	Teacher	0.87	Schools
School comprehension ethos (SCE)	SCE	12	Head teacher	0.62	Schools
Teacher engagement	TE	11	Teacher	0.94	Schools
Pupil engagement	PE	8	Pupil	0.86	Schools
Dosage	24-week implementation record	N/A	Teacher	N/A	Schools
FSM	EverFSM	N/A	Pupil	N/A	Schools/ NPD
Gender	Gender	N/A	Pupil	N/A	NPD
KS1 Literacy	KS1 Literacy Score	N/A	Pupil	N/A	NPD

The additional analysis explored the role of implementation factors in primary and secondary outcome change. Regression analyses are conducted using each implementation factor as a predictor of primary and secondary outcomes. These implementation factors and sources of measurement are detailed in Table 5: Implementation factor instruments: level of measurement and reliability. Pupil and teacher measures in this table are described above in the sections on Secondary outcomes and Teacher comprehension importance and behaviour.

Dosage was collected using data from FFT as recorded by teachers in the weekly lesson frequency recording tool given to teachers by FFT on their initial training day. Dosage was defined as the time spent delivering the programme to pupils. This was calculated by multiplying the average weekly minutes of delivery by the number of weeks of delivery. This was reported at the teacher level. This survey asked teachers to record the number of FFT Literacy Reciprocal Reading lessons and the average length of these lessons on a weekly basis. The agreed minimum compliant dosage as stated by the programme developer was 240 minutes for the universal intervention and 480 minutes for the targeted intervention. The progress in completion of this recording tool was to be checked during the school follow-up training visit by FFT. This survey was to be returned to QUB or QUB fieldworkers were to collect this survey from teachers on the day of the post-test. In practice, many teachers did not complete and return the lesson record logs, generally due to lack of time (discussed further under Process evaluation findings) and other teachers reporting they had not been made aware they had to do this by FFT. In total 68% of teachers (32 of 47 schools who

stayed in the trial for post-testing) in the intervention group returned the lesson record logs. An implication of this is that the estimate of dosage derived from teacher dosage records is subject to bias and should be considered in interpretation of this data. It would not have been feasible to record individual pupil dosage, but in future research of the targeted intervention, individual pupil dosage could be recorded for each targeted pupil. Individual pupil dosage records were not requested as part of this trial to avoid over-burdening teachers. Considering the reported lack of time to complete even the lesson records, there would likely have been a lot of missing data for individual pupil records. Future research should emphasise the importance of recording pupil dosage, and this could be incorporated into programme training. A caveat for the teacher-reported dosage of lessons is that it does not take into account absenteeism, and individual pupil dosage will have varied.

As part of the online survey administered to children at post-test, those who had taken part in RR were asked a series of questions about their engagement with, and enjoyment of, the sessions. The measure was designed by the evaluation team and consisted of eight questions (e.g. 'I found the reciprocal reading sessions interesting'; 'I enjoyed the reciprocal reading sessions'; 'The things I learned in the sessions help me to understand better what I am reading' and 'The things I learned in the sessions make reading easier'. The children were also asked if they thought they should keep doing RR in their class. The responses to all the items ranged from 1 (A lot) through to 4 (Not at all). The scale had an excellent internal consistency as assessed by Cronbach's alpha (0.86).

Teachers in the intervention group were asked a series of questions about their engagement with the RR programme. This was part of the online survey administered at post-test. The measure was designed by the evaluation team and consisted of ten questions (e.g. 'I felt engaged when I was delivering the reciprocal reading sessions'; 'The sessions were straightforward to implement'; 'There is sufficient support from the school in the delivery of the sessions' and 'I would be happy to keep doing the reciprocal reading sessions'. The responses to the ten questions ranged from 1 (Strongly agree) to 5 (Strongly disagree). The scale had an excellent internal consistency as assessed by Cronbach's alpha (0.94). The teachers were also asked one further question, 'Would you recommend the reciprocal reading intervention to other schools?' with the responses 'Yes/No/I don't know'.

Schools provided data on pupil names, DOB, UPN and FSM, with the intention of matching with the National Pupil Database to ensure accurate data on these variables and reduce missing data.

Sample size

Sample-size calculations were first carried out prior to school recruitment. (Please see effect size calculations and assumptions summarised in Tables 8–11 below.) Effect sizes for literacy interventions evaluated through a good quality RCT design would tend typically be in the range of 0.2–0.3 (Biggart, Kerr, O'Hare and Connolly, 2013; Borman, Slavin, Cheung, Chamberlin, Madder and Chambers, 2007; Tymms, Merrell, Thurston, Andor, Topping and Miller, 2011). Borman *et al*.'s study used approximately 1600 pupils per intervention/control arm, Tymms *et al*. used approximately 1900 per arm.

At the initial protocol stage, a power calculation for the universal intervention provided by Optimal Design software using estimates of ES = 0.2, p = 0.05; Intra-cluster correlation coefficient (ICC) = 0.14 (FSM quintile 3 KS2 reading score from EEF's ICC guidance); r^2 = 0.50 (due to having a pre-test of NGRT and correlation used was for overall NRGT); average cluster size n = 20 showed that 94 schools (clusters) would be needed to detect a significant effect of ES = 0.2 if present with a power of 0.8. ICC was calculated at the school level. Analysis was not conducted at the class level.

Also at the initial protocol stage, a power calculation for the NGRT overall reading outcome for the targeted intervention using estimates of ES = 0.2 and ES = 0.3, p = 0.05, ICC = 0.14, r^2 = 0.5, ICC = 0.14, average cluster size = 7 showed that 152 schools would be needed to detect an ES = 0.2 if present with a power of 0.8, or 72 schools would be needed to detect the programme effect if it had an ES = 0.3 if present with a power of 0.8. Cluster size of 7 was used prior to agreeing with the programme developers that they would recommend teachers select n = 6 pupils for the targeted intervention.

A maximum sample size of 100 schools (intervention and control) was a limiting factor as the programme delivery team had capacity to deliver the intervention to a maximum of 50 schools. At the point of writing the protocol, it was

concluded that this sample size (*n* = 100 schools) would be adequate to detect a universal intervention effect size of 0.2 and a targeted intervention effect size of 0.3.

After pre-test, the minimum detectable effect size (MDES) calculation was repeated using a school sample size of *N* = 100 schools. The cluster size for universal intervention was updated to 30 and the cluster size for the targeted intervention was updated to 6. After the initial-protocol phase, this cluster size of 6 had been agreed with the programme developers as the number of pupils a teacher should select for the targeted intervention. Using a correlation between pre-test and post-test of 0.85 for the NGRT, ICC = 0.14 and *p* = 0.05, the MDES for the universal intervention was found to be 0.15, and the MDES for the targeted intervention was found to be 0.24. These figures were updated in the amended evaluation protocol. This was carried out post-randomisation. The updated protocol also includes the original sample size calculations and assumptions. At the protocol stage, the published protocol included an MDES for targeted intervention with a conservative estimate of 72 schools for an effect size of 0.3. This was later updated to reflect the actual recruited sample size.

At the analysis stage, the sample size had been affected by 2 schools being excluded in consultation with EEF due to the schools having violated the terms of the memorandum of understanding (having already signed up to another EEF literacy trial) and 3 schools being lost to follow-up. The pre-test to post-test correlation was found to be 0.73 for the sample in the universal intervention analysis and 0.56 for the sample in the targeted intervention analysis, with ICC of 0.06 and 0.15 respectively. Repeating the MDES calculations for this sample gave MDES = 0.14 for universal and MDES = 0.24 for targeted.

MDES calculations for FSM pupils were carried out at the analysis stage. This was not possible at the earlier stages of MDES as schools had not returned FSM data on pupils. The cluster size for FSM pupils for the universal intervention was *n* = 10. The cluster size for FSM pupils for the targeted intervention was *n* = 5. The pre-test to post-test correlation for FSM pupils in the universal intervention sample was 0.77 with ICC = 0.04. The pre-test to post-test correlation for FSM pupils in the targeted intervention sample was 0.58 with ICC = 0.18. The MDES calculation for FSM pupils in the universal intervention sample gave MDES = 0.19. The MDES calculation for FSM pupils in the targeted intervention sample gave MDES = 0.3. These calculations will be repeated in the forthcoming NPD addendum, using FSM data received from the NPD. The FSM data used in the present calculations in this report was EverFSM supplied by schools.

The MDES at each stage of the study is summarised in Table 8 and Table 9 in the Impact evaluation section.

Randomisation

Randomisation was carried out in July 2017, after the pre-test had been administered in June 2017.

Stratification was used as part of the randomisation process. Stratification improves the precision of the estimates by helping to ensure that the treatment indicator is orthogonal to the other covariates (Cox and Reid, 2000).

In this case minimisation was conducted through the QMinim software package. Minimisation is a well-recognised approach that uses algorithms to ensure a balance on certain covariates between the control and intervention schools at baseline (Torgerson and Torgerson, 2007). Minimisation was used to ensure the schools were as evenly matched as possible. A number of school-level covariates were used in the matching process; specifically: reading comprehension score (NGRT passage comprehension), % EverFSM reading accuracy score (NGRT sentence completion). Median values were calculated for each of these characteristics to determine a mid-cut point and the creation of dichotomous variables, coding schools as ‘High’ or ‘Low’ for each of these characteristics. These variables were then entered into QMinim for each school; and all variables given a weight of one with the exception of reading comprehension which was double weighted as an important predictor of the outcome of interest. Analysis was performed by the evaluation team who had also performed recruitment and who were not blind to condition. A random number sequence is not generated during minimisation. Schools were entered into the QMinim software alongside their ‘High/Low’ scoring for FSM and reading baseline scores. The program then randomly assigns each school to a condition while simultaneously minimising the difference between each group on these variables. This randomisation process resulted in 51 schools being assigned to the intervention condition and 49 being assigned to the control condition. After randomisation, it was discovered that two intervention schools were ineligible to participate in the trial

due to violating the terms of their MOU (already participating in another EEF literacy trial). These schools had not made this violation of eligibility known to the evaluation team prior to randomisation, and this only became known when EEF were informed by the evaluation team of the other trial in which the schools were participating in. The two schools were excluded from the RR trial in agreement with EEF, and are not included in the ITT model. This resulted in 49 schools in each condition.

Targeted pupils were selected pre-randomisation for the targeted intervention. Baseline balance is fully described in the Pupil and school characteristics section. At baseline, pre-exclusion schools did not differ statistically significantly on any of the minimisation variables.

Statistical analysis

Analysis was conducted on an intention-to-treat basis and was carried out using STATA version IC15.1. The main effects of the intervention were estimated using multilevel mixed effects regression modelling to take account of the clustered nature of the data (clustered at the school level) and a series of models were estimated for each outcome (where pupil is level 1 and school is level 2). Mixed effects models account for both fixed (clustered at the school level) and random effects. Analysis was not conducted at the pupils' individual class level.

Firstly, two models were conducted for the universal intervention with the NGRT comprehension score and overall reading score forming the dependent variable of each model and the independent variables including a dummy variable representing whether the child was a member of the intervention or control group (coded '1' and '0' respectively) and pupils' baseline scores at pre-test. This same analysis was repeated for the targeted intervention. Each primary analysis model was conducted using standardised pre-test scores (*z*-scores) to control for variation in pre-test scores between groups (Connolly et al., 2017). These were calculated using the egen std function in STATA, to have a mean of 0.

ICCs were calculated after the primary analysis models were conducted using the estat ICC function in STATA and these are reported in the MDES calculations.

Schools that dropped out of delivering the intervention were encouraged to allow post-testing; however, they did not agree to do so. The pre-test data from drop-out schools was still included in raw means and the maximum number of participants was used for each analysis model, i.e. pupils were included in analysis Model 1 even if they were missing from analysis Model 2. The SAP specified that if missing data was higher than 5%, a 'missing at random' analysis would be carried out to determine if multiple imputation was required. Levels of missing data for each outcome are reported in the 'Sensitivity analysis: multiple imputation' section of the analysis. This was carried out by running chi-squared analyses of group allocation and missingness (missingness was a dummy variable, coding '1' for missing outcome data and '0' for not missing). Multiple imputation was performed using chained iterations to fill in missing values in multiple variables using univariate imputation with fully conditional specification of prediction equations. Twenty imputations were carried out in order to reduce the risk of the 'Monte Carlo' error (simulation error). Twenty imputations are recommended for between 10% to 30% missing data (Graham *et al*., 2007). A total of 200 iterations (with a burn-in of 10) were carried out, and estimates were combined using Rubin's pooling rules. This then allowed the primary analysis to be repeated using the imputed data, and the difference in primary effects was compared with the complete case analysis to investigate if the presence of missing data impacted the results.

Dosage is included as an independent variable in the exploratory analysis of programme implementation factors. The SAP stated that compliance analysis (CACE analysis) would be performed to investigate if a binary variable of compliant/non-compliant influenced programme impact. Schools were coded as compliant/non-compliant based on comparing their teachers' reported dosage with the minimum required dosage. FFT have reported that minimum dosage should be two sessions of twenty minutes per week for 12 weeks for the targeted intervention, and one session of 20 minutes per week for 12 weeks for the whole-class intervention. Year 4 (whole-class intervention) groups with a reported dosage of under 240 minutes, and Year 5 and 6 groups (targeted intervention) with a reported dosage of under 480 minutes will be coded as non-compliant. No dosage scores for schools were under the required minimum for compliance. The CACE analysis could not be performed. Reported dosage scores are included in the Dosage statistics table in Appendix G. It should be noted that only 32 out of the 49 intervention schools returned reported dosage data.

Secondary analyses investigated differences between the control and intervention groups in post-test reading accuracy and teachers comprehension importance and behaviour controlling for pre-test reading score. This analysis was carried out for the universal intervention and for the targeted intervention. These analyses were also conducted as mixed effects multi-level models, accounting for the clustered nature of the data (school level).

Additional analysis was conducted to explore the pathways for change in both primary and secondary outcomes as proposed in the logic model (Figure 1). This was carried out for both the universal intervention and the targeted intervention. The original protocol did not fully specify implementation factors. These were developed and updated since the original protocol and since the SAP. The SAP also specified two pupil implementation factors of pupil meta-cognition and pupil comprehension behaviour: this scale was found to load on to a single factor, 'Pupil comprehension meta-cognition' and has been analysed as such.

The second part of the additional analysis is the implementation factor and outcome change analysis. These models were carried out for the intervention group only, to explore programme implementation of both the universal intervention and the targeted intervention. The implementation factors are described above. These were theoretically determined during the refinement of the logic model for the updated protocol. They were analysed for the intervention group only. This was to investigate if any factors resulted in higher reading outcomes at post-test, thus allowing such a factor to be deemed important for implementation of the programme.

Further subgroup analyses repeated the primary outcome analysis for only pupils in receipt of FSM (both control and intervention). This change from the original protocol, which also stated SEN as a variable to be analysed here, was stated in the amended protocol. The SAP also stated that a correlation analysis between KS1 data from the NPD and NGRT Overall literacy would be conducted. This has been postponed until an addendum is submitted to EEF with the NPD data, as changes in data protection legislation have delayed data sharing agreements with the NPD. The FSM subgroup analysis will also be repeated in this addendum, using EverFSM data received from the NPD and compared with the FSM subgroup analysis conducted using the data received from schools.

Effect size was calculated for the primary outcomes: overall reading score and reading comprehension score. Effect size (Hedges' *g*) was calculated as the standardised mean difference between the control and intervention groups, using the standard deviation accounting for clustering. This is an update since the SAP, as the previous formula did not take into account the clustered nature of the data. The standard deviation was calculated for each group (control and intervention) using the formula (Connolly *et al*., 2017):

$$\text{standard deviation} = \text{standard error} \times \sqrt{\text{sample size of group in analysis model}}$$

The standard error for each group was calculated as the standard error of the coefficient in a 'null' mixed-effects model using only the post-test outcome with no covariates and including only the participants in each group who were included in the analysis. The effect size for each primary outcome was calculated using the formula for Hedge's *g*, using the adjusted post-test mean scores, the *N* of each group in that primary analysis and the SD as described above.

Adjusted post-test mean scores to be used in the calculation of the effect size for each primary outcome were calculated using the constant, group coefficient and pre-test coefficients from the multi-level model of the analysis of that outcome.

$$\text{adjusted post} - \text{test mean}$$

$$= \text{constant} + (\text{group coefficient} \times \text{group}) + (\text{pretest coefficient} \times \text{pretest mean})$$

Hedge's *g* was calculated using the formula:

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

EEF analysis guidance recommends that minimisation or stratification variables should be included in the primary analysis model. Minimisation was conducted using reading comprehension, reading accuracy and school level %

FSM. Minimisation did not include the primary outcome of overall reading, as randomisation occurred prior to the agreement with EEF on two primary outcomes (which was finalised at the updated protocol stage). Considering that reading accuracy is a secondary outcome, this minimisation variable was not included in the primary analysis. The other minimisation variable, school-level %FSM would not be appropriate to include in the primary analysis as it is not measured at the pupil level. The implications of FSM on primary outcome are explored in the subgroup analysis. Considering that the subgroup analysis of FSM should repeat the primary analysis, it is not appropriate for the primary analysis to also include FSM, as this could not be repeated in the subgroup analysis. The SAP stated that the primary models would be repeated for the FSM-pupils-only subgroup. It would not appropriate to repeat the primary model for an FSM-pupils-only subgroup if the primary model also included FSM as a predictor (all clusters would be 100% FSM). This would mean using two different models for primary analysis and FSM subgroup analysis, and that the analysis for FSM-pupils only would not be comparable with the primary analysis. Comparison between the effects for FSM and the overall effects can be made by comparing the primary analysis with the FSM subgroup analysis.

Implementation and process evaluation

The implementation and process evaluation (IPE) was undertaken alongside the trial to support understanding of the trial findings and to understand control group activity, using a mixed-methods (quantitative and qualitative) approach. Data sources were used to assess several issues including: the centrality of the teacher role in programme delivery, i.e. teaching of FFT Literacy Reciprocal Reading and pupil paired/group reading time; how well pupils use the different strategies (predicting, clarifying, questioning, summarising); facilitators and barriers to programme implementation. In addition, case-study data were generated in ten schools via interviews, focus groups and observations, drawing upon the views of a range of informants (e.g. pupils, teachers and head teacher).

Data collection took place between May 2017 and July 2018 using the data collection methods described below. The IPE data was collected directly by the evaluator.

Observation of training and lesson sessions

Structured observations were carried out by the evaluators for both targeted and universal approaches in a set of ten case-study schools using an EEF-funded instrument developed for the evaluation of previous Paired Reading studies. This involved observations of teachers' implementation of the FFT Literacy Reciprocal Reading lessons, focusing on fidelity (e.g. to what extent did the teachers include focused elements of prescribed procedures when implementing the programme?), participant responsiveness (e.g. to what extent did the children engage with FFT Literacy Reciprocal Reading?) and quality (e.g. how well did teachers use modelling strategies to deliver the components of FFT Literacy Reciprocal Reading?). Schools were selected primarily from the two geographic areas: the North of England and the Midlands, with 12 schools from other regions in the South of England, with five schools selected with high above average FSM average and five schools selected with below average FSM, respectively. The evaluator observed a one-day training session for teachers delivered by the programme delivery team in October 2017. Teachers delivering the universal and targeted versions of FFT Literacy Reciprocal Reading received the same training and were directed to aspects of the manual on how to use the approach in different ways. For example, Year 4 teachers and TAs using reciprocal reading with their whole class were encouraged to use reciprocal reading as a form of shared, whole-class reading. Year 5 and Year 6 teachers and TAs using reciprocal reading as a targeted strategy having already identified a group of children were directed to run it as a form of group or guided reading and that could take place in the class or as a withdrawal activity.

Lesson record logs

Lesson record logs (designed by the evaluation team with input from the programme delivery team) were designed to be completed by all intervention teachers with the aim of recording information on FFT Literacy Reciprocal Reading programme delivery and fidelity. This included:

- attendance at training
- date and duration of each session delivered
- start and end dates of FFT Literacy Reciprocal Reading programme delivery

These forms were supplied to teachers by FFT at the first training session, to be completed at the end of every week of term about delivery of sessions. These were collected by the evaluators on the final research visit in May, June and July 2018. Lesson record logs were intended as the basis for measuring dosage. A caveat for the teacher-reported dosage of lessons is that it does not take into account absenteeism, and individual pupil dosage will have varied.

Pre- and post-data on usual practice

The evaluators collected pre- and post-data on usual practice. Data on usual practice was gathered at pre-test using a survey attached to the MOUs signed by each head teacher. Usual practice data was gathered at post-test using the teacher survey (this was administered at the same time as the pupil post-test to maximise response). The survey at post-test was delivered to all (control and intervention) teachers and TAs working with Years 4–6 and included items on teacher-level outcomes (e.g. comprehension behaviours and importance), implementation issues (for intervention schools) and comprehension culture of the school.

Interviews and focus groups

As is the norm in case-study research, we made use of a range of methods (e.g. interviews, focus groups and observations), and informants (e.g. pupils, teachers, head teacher) to cover both universal and targeted approaches. This data contributed to the implementation factors section of the logic model as it provided qualitative reports of key stakeholders engagement with the programme. Interview schedules were developed by the evaluation team to target the process evaluation factors identified in the protocol. All such meetings were audio-recorded (with participants' consent) and transcribed verbatim. We undertook a hybrid inductive–deductive approach to thematic identification. The inductive aspect of the analysis drew directly from the data itself using the principles and processes outlined by Braun and Clarke (2006) (e.g. familiarisation of data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, producing the report). The deductive aspect drew upon key sources in the implementation science literature (e.g. Durlak and DuPre's 2008 and Humphrey *et al*., 2016) (e.g. fidelity, quality, dosage, adaptation, participant engagement, programme reach, programme differentiation, and monitoring of control or comparison conditions). Additional insights were probed through additional orientating concepts of implementation (specifically factors affecting implementation and perceived attractiveness of FFT Literacy Reciprocal Reading to stakeholders, and perceptions of impact (both congruent with programme theory and wider/unexpected benefits). Emergent coding allowed for the emergence of unanticipated themes specific to this project/context.

A total of nine out of ten invited teachers participated in semi-structured interviews with the evaluator. These took place over the course of the intervention delivery period during observation visits. Data was collected on teacher perceptions of the programme's impact: what was working well (or less well) in terms of delivery; the value added to the school (in terms of wider benefits such as school ethos or teacher–pupil relations); and any suggestion for improvements to the programme.

Semi-structured interviews took place with seven head teachers during the intervention delivery period. These focused on the value added to the school (in terms of wider benefits such as school ethos) and leadership support and barriers to implementation.

Eight focus groups took place with a total of 48 pupils from across eight of the ten case-study schools over the course of the intervention period during site visits to schools. Each focus group session lasted approximately 40 minutes. During these, pupils were asked to describe FFT Literacy Reciprocal Reading and what they had been learning during sessions. They were asked what they liked or disliked about the programme and, if it had helped them, they were asked for examples of how it had helped. This data contributed to the implementation factors section of the logic model, as it provided qualitative reports of pupil engagement with the programme.

Costs

Data on costs were collected directly from the programme provider, FFT Literacy, and no data on costs was collected from schools. Financial costs for the programme include: (1) Two days of staff training (includes training for 5 staff), (2) two to three half days in-school support visits from the FFT Literacy team throughout the year (schools are offered up to three but most only needed two), (3) teacher manuals and resources (£2386); and (4) pupil books (£210 if universal programme and £50 if targeted programme).

Overall, the total financial cost to deliver this project across one school over three academic years, including training and resources provided in the first year, is £2596 for the universal programme and £2436 for the targeted programme. This equates to roughly £35 per pupil per year over three years based on 25 pupils for the universal intervention. If delivered as a targeted intervention only in schools with one group of six pupils receiving the intervention, the total cost per pupil is £135. This is therefore a very low-cost intervention delivered as a universal programme and a low-cost intervention delivered as a targeted programme. Schools would make substantial per-pupil financial savings by delivering the interventions to more than one group.

In addition, the programme imposes some costs in terms of teacher time. Staff-cover time is required for the FFT Literacy Reciprocal Reading teacher/TA to attend two days of training and two half days for the in-school support visits. In addition, there is some teacher preparation time (approximately 20 minutes planning time per session).

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
1. Universal delivery:	Teacher resources Pupil books	£2386 £210	£2596	£35 (assuming programme delivered over 3 years to 25 pupils per year)
2. Targeted delivery:	Teacher resources Pupil books	£2386 £50	£2436	£135 (assuming programme delivered over 3 years to 6 pupils per year)

Timeline

Table 7: Timeline

Date	Activity
Set up meetings Literature review Logical model/theory of development and review Establish methods of data collection Ethical approval	Jan 17–Jun 17
Recruit schools to trial and completion of Memorandum of Understanding and baseline school survey Parent information letters sent[6] Unique pupil data collected from schools Teachers identified pupils who met the criteria for inclusion in the targeted intervention	Jan 17–Jun 17
Pre-test data collected in schools	Jun 17
Randomisation to condition	July 17
Refinement of methods of data collection including surveys and observation tool Agree naturally occurring data source (Dosage records) with FFT [7]	Apr 17–Sep 17
Teacher and TA training	Sep 17–Dec 17
Develop school interviews and school survey	Oct 17–Dec 17
Observation data collection	Jan 18–Mar 18
School interview	Apr 18–Jun 18
Teacher and TA survey	Jun 18–Jul 18
Collection of naturally occurring data (dosage records from teachers)	Jun 18–Sep 18
Post-test data collected in schools	Apr 18–Jul 18
Data linkage and cleaning Error rate checking Analyse data of outcomes Conduct process evaluation	Jul 18–Sep 18
First draft of final report written and submitted	Oct 18–Jan 19
Submission of appendix corroborating data collected from schools with NPD data for specific analyses, including correlations of primary outcomes with KS1 literacy scores and FSM subgroup analyses.	June 19

[6] Parent letters were sent home, giving parents the opportunity to withdraw their child's data from the study.
[7] It was planned in the protocol that lesson records would be collected as naturally occurring data, i.e. FFT would provide these to teachers and the evaluation team would get access to them at the end of the trial. However, FFT did not have an existing lesson record form. The evaluation team designed one, in agreement with FFT, and FFT provided this to schools as part of their programme training.

Impact evaluation

Participant flow including losses and exclusions (attrition)

Participant flow is detailed below in Figure 2. One hundred schools were pretested and randomised (51 intervention, 49 control). Following discussions with the EEF, two of these intervention schools were excluded after pretesting and randomisation as they were ineligible for the trial but had not disclosed this on their signed MOU (randomisation error). This meant that 98 eligible schools had been pretested and randomised. In the universal intervention sample, *n* = 3699 pupils (*n* = 98 schools) were pretested. The exclusion of two schools did not impact the minimisation procedure, as the exclusions occurred post-randomisation. Balance for the remaining 98 eligible schools is described in Table 10. At the point of analysis, there were *n* = 3198 pupils. This is an overall attrition rate of 13.5% for the universal intervention sample. In the targeted intervention sample, *n* = 1523 pupils (*n* = 98 schools) were pretested. At the point of analysis there were *n* = 1296 pupils. This is an overall attrition rate of 14.9% for the targeted intervention sample. Attrition for each primary outcome for each of the universal and targeted samples is shown in Table. Any pupils missing from post-test who were from schools that remained active in the trial were not present for post-test, either due to absence or leaving the school. The analysed *N* is lower than the baseline *N*, as pupils needed both pre-test and post-test scores to be included in the primary analysis models (both these variables are included in the primary models). Data withdrawal forms were issued before pretesting, and schools were instructed not to pre-test pupils who had withdrawn from the data collection. The pre-test data collection was the point at which the evaluation team first received any pupil-level data from schools. The number of pupil withdrawals is, therefore, not calculated. The analysed *N* refers to the number of pupils who the evaluation team received both pre-test and post-tests for (i.e. who sat both NGRT pre-tests and post-tests resulting in an overall reading score for both). The overall attrition is representative of the lost number of children from pre-test to post-test.

Participant flow is described from recruitment to randomisation to pre-test to post-test (i.e. pupils who received the digital NGRT at post-test, rather than splitting between the two outcomes). Missing data for individual outcomes is shown in Appendix H.

Minimum detectable effect sizes are shown for the universal intervention in Table 8. Minimum detectable effect sizes are shown for the targeted intervention in Table 9. The protocol had previously reported MDES. The MDES at randomisation figures include an updated NGRT correlation from pre-test to post-test of 0.85. This is a test–retest correlation published by Granada Learning (2013). At the earlier protocol stage, the NGRT had not been decided as the outcome measure, so an estimate was used for the correlation value. These figures have also been updated using the final data at the analysis stage. NGRT pre-test and post-test correlation was updated at the analysis stage using the data. For universal intervention: *N* per cluster was included as 34, based on dividing 3198 observations in pre-test to post-test correlation by 94 schools remaining in the trial after attrition and school exclusion. Pre-test to post-test correlation for universal intervention was 0.73 at the analysis stage. ICC was calculated using the Stata estat ICC function after running the primary analysis model of overall reading for the universal intervention. For targeted intervention, *N* per cluster was included as 14, based on dividing 1296 observations in the pre-test to post-test correlation by 94 remaining schools. Pre-test to post-test correlation was 0.56 at the analysis stage. MDES for FSM pupils have also been included for each intervention. These calculations were not conducted at either protocol or randomisation stage. The ICC was found to be considerably lower at the analysis stage. This is because the original ICC values were estimates based on other research, with the ICC at analysis taken directly from the analysis model for overall reading. The cluster size for FSM pupils for the universal intervention was *n* = 10. The cluster size for FSM pupils for the targeted intervention was *n* = 5. The pre-test to post-test correlation for FSM pupils in the universal intervention sample was 0.77 with ICC = 0.04. The pre-test to post-test correlation for FSM pupils in the targeted intervention sample was 0.58 with ICC = 0.18.

Baseline comparison between the two groups is fully described in the pupil and school characteristics section later in the report.

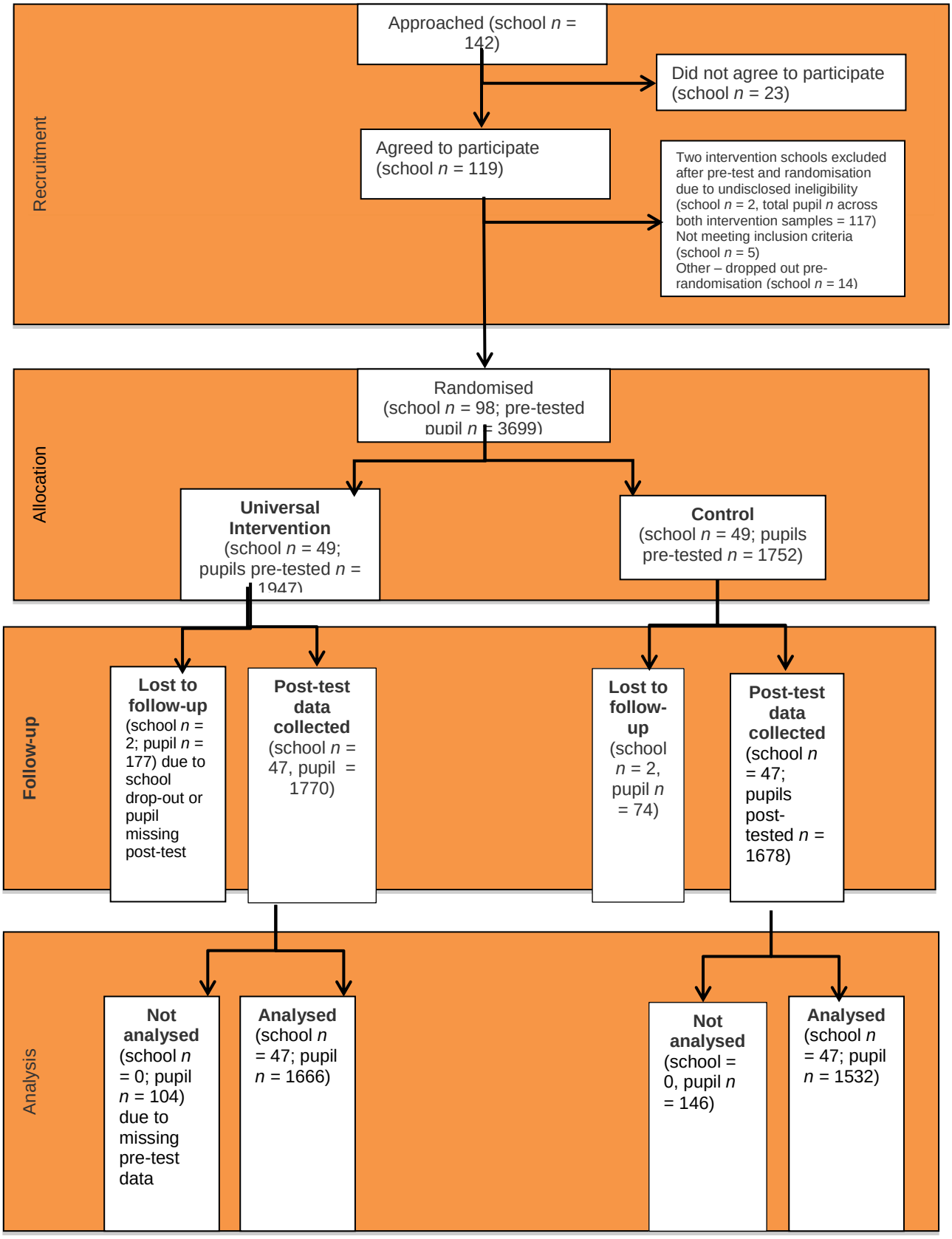
Table 8: Minimum detectable effect size at different stages for universal intervention

		Protocol	Randomisation	Analysis	
		Overall	Overall	Overall	FSM pupils
MDES		0.20	0.14	0.14	0.20
Pre-test/ post-test correlations	level 1 (pupil)	0.70	0.85	0.73	0.77
Intracluster correlations (ICCs)	level 2 (school)	0.14	0.14	0.06	0.04
Alpha		0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8
One-sided or two-sided?		2	2	2	2
Average cluster size		20	38	34	10
Number of schools	intervention	47	49	47	47
	control	47	49	47	47
	total	94	98	94	94
Number of pupils	intervention	940	1947	1666	470
	control	940	1752	1532	470
	total	1880	3699	3198	940

Table 9: Minimum detectable effect size (MDES) at different stages of trial for targeted intervention

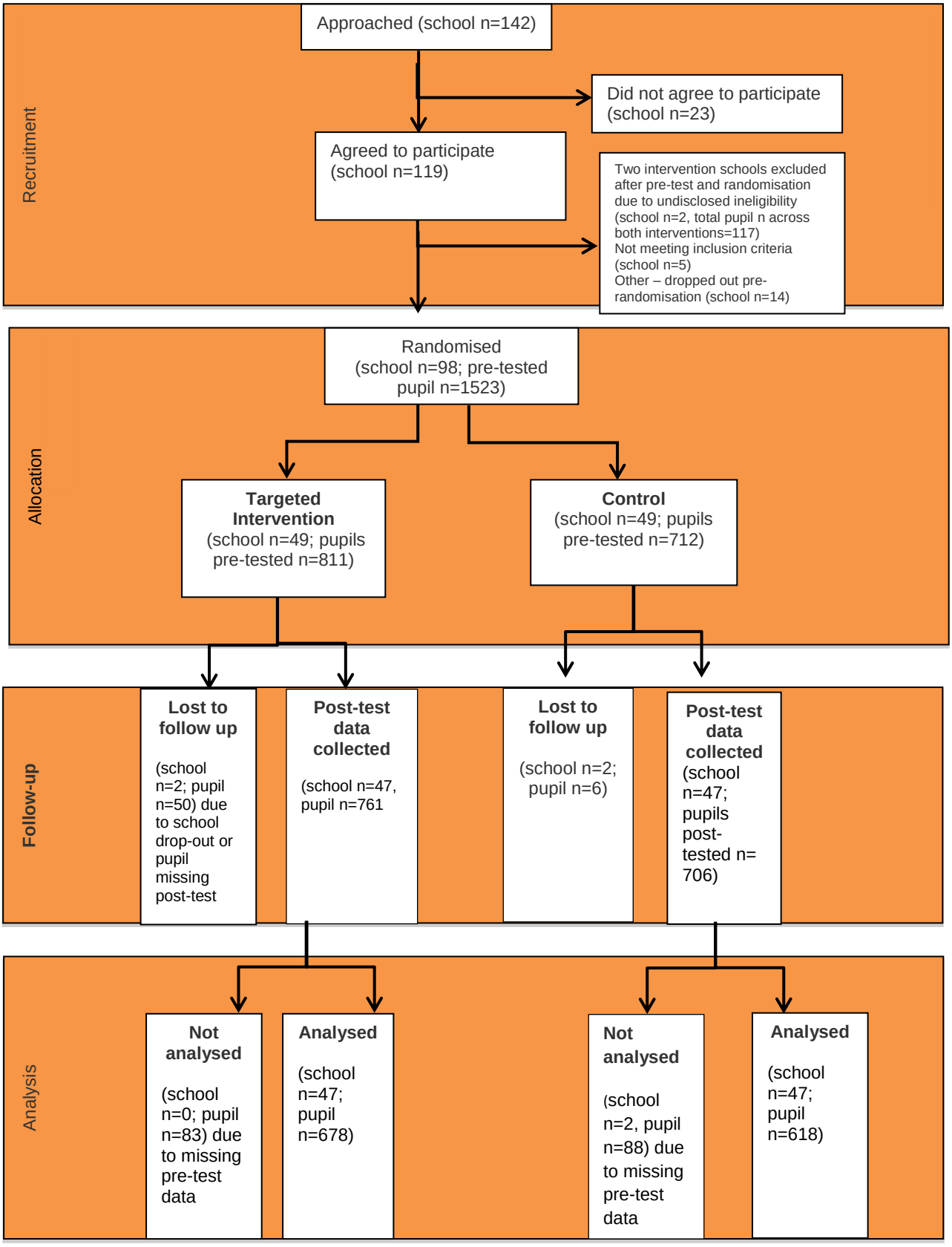
		Protocol	Randomisation	Analysis	
		Overall	Overall	Overall	FSM pupils
MDES		0.30	0.17	0.24	0.30
Pre-test/ post-test correlations	level 1 (pupil)	0.70	0.85	0.56	0.58
Intracluster correlations (ICCs)	level 2 (school)	0.14	0.14	0.15	0.18
Alpha		0.05	0.05	0.05	0.05
Power		0.8	0.8	0.8	0.8
One-sided or two-sided?		2	2	2	2
Average cluster size		7	16	14	5
Number of schools	intervention	36	49	47	47
	control	36	49	47	47
	total	72	98	94	94
Number of pupils	intervention	252	811	678	235
	control	252	712	618	235
	total	504	1523	1296	470

Figure 2a: Participant Flow: universal intervention[8]



[8] Allocation, Follow-up and Analysis: figures for the first primary outcome (Overall Reading). Please see Table 10 for the second primary outcome (Reading Comprehension).

Figure 2b: Participant flow: targeted intervention9



[9] Allocation, Follow-up and Analysis: figures for the first primary outcome (Overall Reading). Please see Table 10 for the second primary outcome (Reading Comprehension).

Table 10: Pupil-level attrition for each primary outcome for universal and targeted interventions

Intervention	Outcome	*N* at randomisation[10]	*N* at analysis	Attrition	
				Count	%
Universal	Overall reading	3699	3198	501	13.54
	Reading comprehension	3448	2958	490	14.21
Targeted	Overall reading	1523	1296	227	14.90
	Reading comprehension	1508	1270	238	15.78

Pupil and school characteristics

Pupil-level differences on the primary and secondary reading outcome measures at baseline are shown in Table 11. Numbers of pupils with FSM are also reported for each arm, as supplied by schools. This will be updated when data are received from the NPD and analysed in an addendum. School-level differences in Ofsted rating are shown in Table 12. Pre-test distribution histograms are included in Appendix F. All of the histograms show normal distribution of the primary outcomes except for the universal overall reading which shows a slight negative skew, i.e. more pupils were scoring towards the higher half of possible scores This, however, does not mean a ceiling effect is present. The maximum possible score is 500, and the pre-test distribution shows that there was no problem of pupils scoring at this maximum level and therefore being unable to progress at post-test.

No data is reported for the two schools that were excluded post-randomisation as all data from those two schools were destroyed once they were discovered to be ineligible and it was agreed with EEF that they would not be included in the ITT model or any analysis.

Effect sizes for baseline imbalances were only calculated for the pupil-level outcomes and not for school-level categorical data. Effect sizes were chosen as the measure of baseline imbalance as they account for the size of difference and are appropriate for interpretation of a simple group difference.

Table 11 shows that the effect sizes are very small for difference between control and intervention schools and pupils in the universal intervention sample for the pre-test on NGRT primary outcomes. This means that the groups did not differ strongly for reading scores before the intervention was delivered.

Table 11: Balance at baseline in intervention and control groups: universal intervention

School-level (categorical)	Intervention group		Control group	
	n/*N* (missing)	Count (%)	*n*/*N* (missing)	Count (%)
Ofsted rating: Outstanding Good Satisfactory Inadequate Requires improvement	49(0)	 14.3% 69.4% 0% 0% 16.3%	49(0)	 8.2% 83.7% 0% 2% 6.1%
Academy (Converter) Academy (Sponsor led) Community school Faith school (Academy converter) Faith school (Foundation school) Faith school (Voluntary aided) Faith school (Voluntary controlled) Foundation school Free school	4 4 25 1 0 7 3 4 1	8.2% 8.2% 51% 2% 0.00% 14.3% 6.1% 8.2% 2%	8 2 23 1 1 7 3 4 0	16.33% 4.08% 46.94% 2.04% 2.04% 14.29% 6.12% 8.16% 0.00%

[10] Note this is *N* at randomisation after excluding two ineligible schools.

School-level (continuous)	Mean		Mean		
Y4 Pupil *N* per school at baseline	39		36		
Pupil-level (categorical)	*N*	Count (%)	*N*	Count (%)	
Eligible for FSM	348	18.34%	418	20.38%	
Pupil-level (continuous)	*n* (missing)	Mean (SD)	*n* (missing)	Mean (SD)	Effect size
Pre-test NGRT overall reading scale (primary outcome)	1947 (104)	248.54 (62.72)	1752 (146)	246.10 (64.64)	0.04 (CI: –0.03–0.10)
Pre-test NGRT reading comprehension (primary outcome)	1832 (219)	248.61 (61.39)	1619 (279)	247 (62.24)	0.03 (CI:–0.04–0.09)

Table 12 shows that there are small to moderate effect sizes for the difference between control and intervention schools in the targeted sample at baseline on the NGRT pre-test for primary outcomes. This means that prior to the intervention being delivered, the intervention schools were already performing at a slightly higher level on the reading outcomes than the control schools. However, the effect sizes calculated for the primary analysis later in the report are calculated in a way that controls for imbalance at pre-test. The imbalance at baseline, therefore, may have meant that the intervention was being delivered in a group of schools who had higher reading attainment, but it does not strongly impact the interpretation of the results at post-test, i.e. that the intervention had a positive impact on reading attainment.

Table 12: Balance at baseline in intervention and control groups: targeted intervention

School-level (categorical)	Intervention group		Control group		
	N (missing)	Count (%)	*N* (missing)	Count (%)	
Ofsted rating: Outstanding Good Satisfactory Inadequate Requires improvement	49(0)	 14.3% 69.4% 0% 0% 16.3%	49(0)	 8.2% 83.7% 0% 2% 6.1%	
Academy (Converter) Academy (Sponsor led) Community school Faith school (Academy converter) Faith school (Foundation school) Faith school (Voluntary aided) Faith school (Voluntary controlled) Foundation school Free school	4 4 25 1 0 7 3 4 1	8.2% 8.2% 51% 2% 0% 14.3% 6.1% 8.2% 2%	8 2 23 1 1 7 3 4 0	16.33% 4.08% 46.94% 2.04% 2.04% 14.29% 6.12% 8.16% 0.00%	
School-level (continuous)	Mean		Mean		
Y5 & Y6 Pupil *N* per school at baseline	16		15		

Pupil-level (categorical)	*N*[11]	Count (%)	*N*	Count (%)	
Eligible for FSM	158	17.67%	182	22.75%	
Pupil-level (continuous)	*n* (missing)	Mean (SD)	*n* (missing)	Mean (SD)	Effect Size
Pre-test NGRT overall reading Scale (primary outcome)	811 (83)	277.05 (45.30)	712 (88)	268.57 (47.30)	0.18 (CI:0.08–0.28)
Pre-test NGRT reading comprehension (primary outcome)	804 (90)	270.57 (53.59)	704 (96)	262.70 (52.33)	0.15 (CI:0.05–0.25)

Outcomes and analysis

Primary analysis

Raw means are reported in the analysis tables, to allow comparison of mean scores between groups. Standardised pre-test scores, which have a mean of zero, were used in the analysis as they control for variation at pre-test between groups. As detailed in the description of the NGRT testing procedure, number of pupils (N) for sub-scores (reading comprehension) can be lower than for overall reading, as not all pupils who completed the NGRT took the reading comprehension test (if they scored extremely low on the reading accuracy test, they did not receive the reading comprehension, and sat a phonics sub-test instead).

Overall, no statistically significant effects of the universal intervention were found on either primary outcome.

Significant effects were found for the targeted intervention on both primary outcomes. This means pupils who received the targeted intervention scored more highly on reading at the end of the trial than pupils who did not receive the intervention. Pupils who were chosen to receive the targeted intervention were chosen prior to pre-testing and randomisation.

Table 13: Primary analysis for universal intervention

		Raw means					Effect size	
Outcome	Covariate	Intervention group		Control group		*N* in model (intervention; control)	Hedges *g* [95% CI]	*p*
Primary outcome								
		n (missing)	Mean [SD]	*n* (missing)	Mean [SD]			
NGRT overall reading ability	Pre-test NGRT overall reading ability	1770 (281)	284.98 [57.28]	1678 (220)	281.72 [61.81]	3198 (1666, 1532)	0.00 [–0.06, 0.07]	0.795
NGRT reading comprehension	Pre-test NGRT reading comprehension	1722 (329)	288.06 [57.64]	1605 (293)	288.52 [57.13]	2958 (1557, 1401)	–0.02 [–0.09, 0.06]	0.494

No significant effect of the universal intervention was found for either the overall reading primary outcome or the reading comprehension primary outcome.

[11] Note it is not possible to calculate missing data for FSM data from the data provided by schools as some schools left the cell blank for pupils unless they receive FSM, and others entered ‘no’ – hence a blank cell for FSM cannot be interpreted as missing data.

Table 14: Primary analysis for targeted intervention

		Raw means					Effect size	
Outcome	Covariate	Intervention group		Control group		*N* in model (intervention; control)	Hedges *g* [95% CI]	*p*
Primary outcome								
		n (missing)	Mean [SD])	*n* (missing)	Mean [SD]			
NGRT overall reading score	Pre-test NGRT overall reading score	761 (133)	309.71 [47.16]	706 (94)	288.46[49.98]	1296 (678, 618)	0.14 [0.04, 0.25]	0.001
NGRT reading comprehension	Pre-test NGRT reading comprehension	755 (139)	308.51 [51.38]	693 (107)	285.71 [52.06]	1270(668, 602)	0.18 [0.07, 0.29]	<0.001

A significant effect (ES = 0.14) was found for the targeted intervention on the overall reading primary outcome (p = 0.001). A significant effect (ES = 0.18) was also found for the targeted intervention on the reading comprehension primary outcome (p < 0.001). This means that the pupils who received the targeted RR intervention scored significantly higher than pupils on both the overall reading test and the reading comprehension test than pupils who did not receive the intervention (equivalent of 2 months additional progress based on EEF guidelines). Pupils who were chosen to receive the targeted intervention were chosen prior to pre-testing and randomisation.

Sensitivity analysis: multiple imputation

The proportion of missing data at pre-test and post-test was greater than 5% for both primary outcomes (see Appendix H)

The relationship between group allocation and missingness on each of the outcome measures was analysed using chi-square analyses. ‘Missingness’ refers to the pattern of missing data, i.e. this analysis measured the association between group allocation and whether or not a pupil would have missing data: was one group more likely than the other to have missing data? Control showed higher numbers of missing data for pre-test overall reading (p = 0.001) and for pre-test reading comprehension (p < 0.001). Intervention showed higher numbers of missing data for post-test overall reading (p = 0.007). This pattern in missingness means that data was presumed to be missing at random, rather than missing completely at random (where there is no pattern in missingness) or missing not at random (where missingness cannot be explained by observed variables). Multiple imputation was carried out (full details given in Appendix H). Using imputed data, the overall pattern of results was very similar to that obtained by the primary analysis using complete cases. The analysis of the data using multiple imputation suggests that missing data do not negatively impact the strength of the evidence for efficacy of the targeted intervention or the evidence of a lack of efficacy for the universal intervention.

No significant effect of the universal intervention on overall reading or reading comprehension was found using imputed data. A significant positive effect was found for the targeted intervention for overall reading and reading comprehension using imputed data. This means that positive result for the targeted intervention was not influenced by the missing data in the analysis. Even after filling in the gaps of the missing data, the result remained the same.

Non-compliance with intervention

The statistical analysis plan specified that year groups of participants would be coded as compliant or non-compliant based on teacher-reported dosage and that a regression analysis would be conducted on the primary outcome data including compliance level as a predictor. No year groups, however, were coded as non-compliant based on school-reported dosage. No schools reported universal intervention dosage below the compliance limit of 240 minutes (minimum reported by schools = 340 minutes) or targeted intervention dosage below the compliance limit of 480 minutes (minimum reported by schools = 510 minutes). Although some schools delivered a greater dosage of

universal than targeted, this does not categorise them as non-compliant, as they still delivered greater than the minimum recommended dose for each intervention. The programme developers specified a minimum dose for each intervention, not a recommended relationship between the two interventions. The relationship, in terms of time, between the two interventions in each school is not a factor for compliance. For example, if a school was able to spend more time delivering the universal intervention than the targeted intervention, it does not mean that they were non-compliant, as both interventions still received above the minimum dose. A high dose of the universal intervention does not mean that there is a negative impact on the targeted intervention. No regression analysis was, therefore, conducted to measure the effect of non-compliance on the efficacy of the intervention. Compliant dosage limits were defined using total number of minutes delivered. This, however, does not consider the nuances of recommended delivery, i.e. the importance of minutes per week and the importance of number of weeks of delivery. Future research into this intervention could define compliance as adhering to both the recommended number of weeks, and the recommended number of minutes per week. This would allow investigation into whether time spent during the year is important, or if a shorter more intensive intervention can also work. These dosage scores are, however, investigated as a predictor of outcome change in the additional analysis section below. Descriptive statistics for dosage scores for targeted and universal interventions are included in Appendix G. Mean school dosage scores schools were slightly higher for the universal intervention (1767 minutes) than for the targeted intervention (1707 minutes). These means are considerably higher than the recommended minimum. The programme developers informed the evaluation team of what they thought the minimum amount of dosage was that a school should deliver for each intervention. However, this was not the value that schools were informed of, i.e. they were not told that they could deliver only 240 minutes of the universal intervention and still be regarded as having delivered the programme with compliance. The schools were told during training to embed the programme in their literacy teaching, not that they should implement it for a finite number of weeks then stop.

Secondary outcome analyses

Secondary analyses were carried out using multi-level models to investigate the differences between control and intervention groups for reading accuracy and pupil comprehension meta-cognition. This was carried out for the universal intervention and targeted intervention.

No effect of the universal intervention was found for reading accuracy. A significant effect of the universal intervention was found for pupil comprehension meta-cognition. Significant effects were found for both reading accuracy and pupil comprehension meta-cognition in the targeted intervention analysis. This means that pupils who received the targeted intervention scored more highly than control pupils at the end of the trial at both reading accuracy and meta-cognition of reading comprehension.

Note that the variable ‘group’ refers to the membership of either control or intervention group, i.e. a significance value for this row in a results table refers to a significant effect of group allocation.

Table 15: Secondary analysis of reading accuracy for universal intervention

NGRT reading accuracy	Coef.	Standard error	*z*	Significance	95% CI	
Group	2.21	2.26	0.98	0.33	–2.22	6.65
Pre-test NGRT reading accuracy	36.43	0.76	47.68	<0.001	34.93	37.92
Constant	289.32	1.63	177.90	<0.001	286.13	292.50

No significant impact of the intervention on reading accuracy was found for the universal intervention.

Table 16: Secondary analysis of pupil comprehension meta-cognition for universal intervention

Pupil comprehension meta-cognition	Coef.	Standard error	*z*	Significance	95% CI	
Group	1.19	0.31	3.84	<0.001	0.58	1.79
Pre-test NGRT reading comprehension	0.06	0.09	0.61	0.55	–0.13	0.24
Constant	18.18	0.22	82.39	<0.001	17.74	18.61

The intervention group showed a significantly higher score for pupil comprehension meta-cognition (raw mean = 19.326, SD = 4.72) than controls (raw mean = 18.07, SD = 4.56) for the universal intervention. This means there was a significant effect of the intervention on pupil comprehension meta-cognition, i.e. pupils in the intervention group showed higher comprehension meta-cognition at post-test.

Table 17: Secondary analysis of Reading Accuracy for targeted intervention

NGRT Reading Accuracy	Coef.	Standard error	*z*	Significance	95% CI	
Group	8.57	4.21	2.04	0.04	0.32	16.82
Pre-test NGRT reading accuracy	23.62	1.61	14.70	<0.001	20.47	26.77
Constant	293.48	3.06	96.02	<0.001	287.49	299.47

The intervention group showed a significantly higher post-test score for reading accuracy (raw mean = 314.83, SD = 45.62) than controls (raw mean = 299.18. SD = 50.5), for the targeted intervention.

Table 18: Secondary analysis of Pupil comprehension meta-cognition for targeted intervention

Pupil comprehension meta-cognition	Coef.	Standard error	*z*	Significance	95% CI	
Group	1.13	0.37	3.08	0.00	0.41	1.85
Pre-test NGRT reading comprehension	–0.25	0.17	–1.50	0.13	–0.58	0.08
Constant	18.43	0.26	71.12	<0.001	17.92	18.94

The intervention group showed a significantly higher score for pupil comprehension meta-cognition (raw mean = 19.39, SD = 14.47) than controls (raw mean = 18.11, SD = 4.53) for the targeted intervention.

Additional analyses

Full results tables are available for these analyses in Appendix C

The exploratory analyses for universal and targeted interventions investigate the relationship between implementation factors and primary outcomes and the relationship between pupil comprehension meta-cognition and overall reading for pupils who received the programme. They were conducted using intervention pupil data only, no control group. They measure the impact of implementation factors on post-test outcomes for intervention pupils. These models are exploratory as they investigate potential pathways for the logic model. The analysis of the effect of reading comprehension meta-cognition on overall reading was conducted to investigate if meta-cognition is predictive of reading ability. Considering the impact of the programme on meta-cognition for both the universal and targeted interventions, this analysis is important as it tells us if this factor is actually important for reading ability (otherwise the fact that the programme improved it would not be important).

Exploratory analyses for universal intervention

No significant effects were found of teacher comprehension importance, teacher comprehension behaviour or school comprehension ethos on pupil comprehension meta-cognition for the universal intervention.

No significant effect of pupil comprehension meta-cognition was found on reading comprehension at post-test for the universal intervention.

There was a significant effect of pre-test reading comprehension on overall reading at post-test for the universal intervention (Table 35).

There was a significant effect of pre-test reading accuracy on overall reading at post-test for the universal intervention (Table 36).

This means that pre-test comprehension is significantly related to later attainment in overall reading, i.e. early comprehension is significantly predictive of later overall ability at reading. Furthermore, pre-test reading accuracy is significantly predictive of later attainment in overall reading. This analysis explores the likelihood that both aspects of reading development (accuracy and comprehension) are important for later reading attainment. Although this is a comprehension intervention, the importance of reading accuracy for later overall reading performance is clear.

Exploratory analyses for targeted intervention

A significant negative effect of teacher comprehension importance was found on pupil comprehension meta-cognition for the targeted intervention.

No significant effect of pupil comprehension meta-cognition was found on reading comprehension at post-test for the targeted intervention.

There was a significant effect of pre-test reading comprehension on overall reading at post-test for the targeted intervention (Table 39).

There was a significant effect of pre-test reading accuracy on overall reading at post-test for the targeted intervention (Table 40).

This means that, in the targeted intervention sample, both comprehension and accuracy are important for later overall reading attainment. A pupil's early reading accuracy is predictive of their later overall reading ability, even though this programme directly targeted comprehension.

Analyses of implementation factors for universal intervention

The implementation factors of pupil engagement, teacher engagement (mean at the school level) and dosage were analysed for their effect on pupil comprehension meta-cognition, reading comprehension and overall reading using mixed effects multi-level models. These analyses were conducted for only the intervention group.

A significant effect of pupil programme engagement on pupil comprehension meta-cognition was found for the universal intervention.

Analyses of implementation factors for targeted intervention

A significant effect of pupil engagement on pupil comprehension meta-cognition was found for the targeted intervention.

No other implementation factors showed a significant effect on pupil outcomes for either the universal or targeted interventions (see Appendix C for full results of additional analyses).

Subgroup analyses

The primary analyses were repeated for only pupils in receipt of free school meals, using FSM data provided by the schools. (These analyses will be repeated and published in an appendix using FSM data from the NPD.[12])

Subgroup analyses for universal intervention

The primary analyses were repeated using only the pupils who were eligible for FSM. This is to measure the impact of the universal intervention on pupils who are eligible for FSM. This is the analysis planned in the SAP for an FSM subgroup. Additional effect sizes for FSM will be calculated in the NPD addendum. The variable 'group' in the tables below is the effect of allocation to either control or intervention on post-test reading outcomes. The significance value for each table represents if being in the intervention group resulted in a statistically significantly improved reading outcome, compared with the control group.

Table 19: Subgroup analysis for FSM pupils – Primary outcome of overall reading for universal intervention

NGRT Overall Reading	Coef.	Standard error	*z*	Significance	95% CI	
Group	3.95	3.58	1.10	0.27	–3.07	10.96
Pre-test NGRT overall reading	45.62	1.44	31.66	<0.001	42.79627	48.44
Constant	281.60	2.68	104.90	<0.001	276.34	286.86

There was no significant effect of the universal intervention on overall reading for FSM pupils.

Table 20: Subgroup analysis for FSM pupils – Primary outcome of reading comprehension for universal intervention

NGRT reading comprehension	Coef.	Standard error	*z*	Significance	95% CI	
Group	–3.31	3.79	–0.87	0.38	–10.74	4.13
Pre-test NGRT reading comprehension	33.31	1.87	17.83	<0.001	29.65	36.97
Constant	288.57	2.91	99.26	<0.001	282.87	294.27

There was no significant effect of the universal intervention on reading comprehension for FSM pupils.

Subgroup analyses for targeted intervention

The primary analyses were repeated using only the pupils who were eligible for FSM. This is to measure the impact of the targeted intervention for pupils who are eligible for FSM.

Table 21: Subgroup analysis for FSM pupils – Primary outcome of overall reading for targeted intervention

NGRT overall reading	Coef.	Standard error	*z*	Significance	95% CI	
Group	10.18	6.14	1.66	0.10	–1.86	22.23
Pre-test NGRT overall reading	34.88	2.94	11.85	<0.001	29.11	40.65
Constant	280.87	4.37	64.30	<0.001	272.31	289.44

[12] At the time of analysis and writing up, new procedures for accessing the NPD were being developed by the Department for Education. To prevent a significant delay to the publication of this report, and given that the primary outcome analysis is based on independent commercial tests and not NPD data, it was agreed that this section would be reported using FSM data provided by the schools. The subsequent appendix will verify these analyses against FSM status from the NPD.

There was no significant effect of the targeted intervention on overall reading for FSM pupils.

Table 22: Subgroup analysis for FSM pupils – Primary outcome of reading comprehension for targeted intervention

NGRT reading comprehension	Coef.	Standard error	*z*	Significance	95% CI	
Group	15.96	5.52	2.89	0.00	5.14	26.78
Pre-test NGRT reading comprehension	30.77	2.86	10.75	<0.001	25.16	36.38
Constant	279.35	3.91	71.46	<0.001	271.69	287.01

There was a significant effect of the targeted intervention on reading comprehension for FSM pupils

This shows that the targeted intervention is not only effective for the sample of pupils as a whole, but also specifically for pupils who receive FSM. The significant effect was found only for the reading comprehension outcome for pupils who receive FSM. This means that FSM pupils did not seem to benefit in terms of their overall reading ability, but did benefit in terms of their reading comprehension. The targeted intervention therefore shows promise in two ways. Firstly, it shows promise as an intervention for targeted pupils (good accuracy/poor comprehension) on both overall reading and reading comprehension. Secondly, it shows promise as an intervention for targeted pupils who receive FSM on reading comprehension.

Cost

Data on costs were collected directly from the programme provider, FFT Literacy, and no data on costs was collected from schools. Financial costs for the programme include: (1) two days of staff training (includes training for 5 staff), (2) two half days in-school support visits from the FFT Literacy team throughout the year, (3) teacher manuals and resources (£2,386); and (4) pupil books (£210 if universal programme and £50 if targeted programme). Overall, the total financial cost to deliver this project across one school as a universal programme, averaged over three academic years, with training and resources provided in the first year is £2596 for the universal programme and £2436 for the targeted programme. This equates to roughly £35 per pupil per year over 3 years based on 25 pupils, for the universal intervention. If delivered as a targeted intervention only in schools with groups of 6 pupils receiving the intervention, across 5 classes over 3 years, the total cost per pupil is £135. This is therefore a very low-cost intervention delivered as a universal programme and a low-cost intervention delivered as a targeted programme.

In addition, the programme imposes some costs in terms of teacher time. Staff-cover time is required for the FFT Literacy Reciprocal Reading teacher/TA to attend two days of training and two half days of in-school support visits. In addition, there is some teacher preparation time (approximately 20 minutes planning time per session).

Table 23: Cost analysis of intervention

Item	Type of cost	Cost	Total cost over 3 years	Total cost per pupil per year over 3 years
1. Universal delivery: Two-day teacher training and two ½ day in-school support visits. 2. Targeted intervention: Two-day teacher training and two ½ day in-school support visits.	Start-up cost per school Start-up cost per school	£2386 £2386	£2386 £2386	£32 £133
Pupil reading books: 1. Universal delivery: 2. Targeted intervention:	Running cost per pupil	 £210 £50	 1. (£8.4 × 25 pupils) = £210 2. (£8.4 × 6 pupils) = £50	 £3 £3
Total: 1. Universal delivery: 2. Targeted version:	Cost over average of 3 years	 £2596 £2436	 £2596 £2436	 £2596/25 pupils/ 3 years = £35 £2436/6 pupils/ 3 years = £135

Table 24: Cumulative cost of delivering FFT Literacy Reciprocal Reading over three years (assuming the same staff delivered the programme over the 3-year period)

	Year 1	Year 2	Year 3
Universal	£2596	£0	£0
Targeted	£2436	£0	£0

All the costs for delivering the programme in respect of training, in-school support, staff manuals and resources, and pupil books are assumed in Year 1. Additional costs to schools will be in respect of staff planning time to deliver the programme (this is not costed here).

Implementation and process evaluation

This section presents findings from the implementation and process evaluation of the FFT Literacy Reciprocal Reading programme. Participants included teachers, head teachers and pupils involved in the project. Data was collected through interviews with a sample of Year 4, 5 and 6 teachers, head teachers, and focus groups with pupils, observations of programme delivery, an implementation survey, and lesson record logs across both the universal and targeted approaches. The structured observations of teachers' implementation of the FFT Literacy Reciprocal Reading lessons, focusing on fidelity (e.g. to what extent did the teachers include focused elements of prescribed procedures when implementing FFT Literacy Reciprocal Reading?), participant responsiveness (e.g. to what extent did the children engage with FFT Literacy Reciprocal Reading?), and quality (e.g. how well did teachers use modelling strategies to deliver the components of FFT Literacy Reciprocal Reading?). Table 25 summarises the case-study school data collection. Findings from these components have been integrated and are organised under the main headings of Implementation, Fidelity, Outcomes, Formative findings and Control group activity. A selection of verbatim quotes from across all the interviews were selected to support the findings and interpretations that arose from the data.

Table 25. Summary of qualitative data collected from case-study schools. Year 4: universal intervention, Year 5 and 6: targeted intervention

School	Observations	Focus group with pupils	Teacher interviews	Head teacher interviews
1	1	1 focus group with 8 pupils Year 4	1	1
2	1	1 focus group with 8 pupils Year 4	1	1
3	1	1 focus group with 8 pupils Year 4	1	1
4	1	1 focus group with 10 pupils Year 4	1	1
5	1	1 focus group with 6 pupils Year 5 and 6	1	1
6	1	1 focus group with 6 pupils Year 5 and 6	1	0
7	1	1 focus group with 8 pupils Year 4	1	0
8	1	0	0	1
9	1	0	1	0
10	1	1 focus group with 6 pupils Year 5 and 6	1	1
Total	10 observations	8 focus groups (58 pupils)	9 teacher interviews	7 head teacher interviews

Table 26: Survey responses from teachers and pupils across both universal and targeted interventions

	Comprehension surveys		Programme engagement surveys
	Intervention group (*N*)	Control group (*N*)	Intervention group (*N*)
Teachers	100	77	86*
Pupils	2105	2073	1811*

* 14 teachers and 294 pupils in the intervention groups selected 'no' as a response to the question 'Did you have reciprocal reading sessions? Please ask your teacher if you are unsure.' Pupils only progressed to the programme engagement section of the online survey if they selected 'yes' as a response to this question. This either means that the pupils were absent for the entire programme, which is unlikely considering the high delivery dosage, or that they clicked the wrong option and had actually received the programme. If the reason was the latter, this could be due to pupils being unaware of the name of the programme or random error. This is a weakness of the instrument. A future survey instrument would use two versions of the post-test pupil survey, one for control and one for intervention schools to avoid this problem.

Implementation

Training and support for delivery

Overall, teachers and TAs were confident about delivering the programme following the training, with 98% (of the 86 implementation survey respondents from 36 schools) agreeing or strongly agreeing with the statement 'I am confident in the delivery of the FFT Literacy Reciprocal Reading programme.' This is also reflected in the interview data, where the majority of teachers were positive that the external training they received prepared them for implementing the FFT Literacy Reciprocal Reading programme with their class. Teachers reported that the training was very practical, not complicated, and enabled them to integrate the four key concepts of the RR programme (i.e. predicting, clarifying, questioning and summarising) into their reading instruction. In addition, a small number of teachers commented that they had found teaching the subject of reading difficult and the training provided them with techniques for effective and interesting ways to teach reading. This is reflected in quotations from two teachers below:

> *It [the training] was amazing actually. It is very much a process and yeah, a cycle that you could apply to anything really, when it comes to reading.*

> *The training was incredibly practical and it just made sense. It wasn't complicated because reading's always been something that I think is incredibly difficult to teach, to actually improve on.*

One teacher mentioned the gap between the 2 external training days was a disadvantage. It was suggested that they were not confident enough after the first training event to implement the programme and this in turn hindered implementation of the programme due to gaps in teachers' knowledge.

> *The second training day, I think was in February, so there was quite a big gap where we were sort of, what do we do? What are we doing? And so for that reason it didn't really kick off in the school this year until the beginning of this year.*

Whilst most teachers felt that the two days training were very useful, the majority of teachers interviewed felt that the second day of training was the more beneficial as it provided staff with practical examples of activities and texts that were suitable to use within the FFT Literacy Reciprocal Reading lessons. In addition, some teachers felt that the second day of training rectified any uncertainties they were having whilst utilising their learning from the original training day:

> *I think particularly the second session that we did which was more about how we could use it with regard to poetry and a few other bits and pieces.*

> *I think, that second one ironed out any gaps. I think it was just the ambiguity of what texts do we use, are we supposed to be, is it to be used as a whole class or can we use it in smaller groups and do we have to stick to that strict, that's what we do, that's what we do next, that's what you do after sort of thing. Getting the flexibility, we were a bit unsure whether we were doing the right thing by being flexible with it.*

In addition to the training, participating schools were provided with support in the form of visits, observations and feedback. This support was perceived as a positive addition to the training by the majority of staff who participated in the interviews.

> *[Trainer X] that came to visit, it was really positive to have that external person come in and have a look at what we were doing. I think that was really valuable.*

> *So, kind of, gave you the initial information, let you have a go at it, have a couple of visits from people from Fisher Family Trust. I don't know if that was just our school, but that was kind of good as well, whereby they could come in and see what you were doing with it, discuss, meet with the children with you or without you and observe some sessions and things like that and then obviously, you guys came in as well, and then the other session.*

It would appear that some teachers who had to implement the FFT Literacy Reciprocal Reading programme did not attend the training and had to rely on the feedback of those who attended the training (not in line with the recommendation of the RR programme where all delivery staff are recommended to attend training):

> *I think they felt they'd missed out a little bit and it meant that there wasn't quite as much continuity…I felt as if really, there was just that little bit lack of knowledge and the two staff not fully understanding the benefits…*

Facilitators to implementation

The majority of teachers who were interviewed considered that FFT Literacy Reciprocal Reading was successful within their schools because of the flexibility the programme offered. The flexibility to adapt the programme to meet the needs of each class or pupil empowered the teachers to make judgements on how it was best delivered in their classrooms. This allowed teachers to take ownership of the programme:

> *But I think it definitely, the flexibility that we've been allowed with it has been good, there are some programmes where you strictly have to do this and it's kind of like, well I'm going to lose him if I do this, I'm going to lose her if I do this but we've been able to get all the children involved and their work has improved as a result of it. (Teacher: universal intervention)*

> *For me, whether, I don't think there needs to be a change in the programme, I think teachers as professionals need to interpret that and say what's right for their children. (Head teacher)*

The flexibility of the programme was also acknowledged to provide a framework for teachers to work within their own preferred teaching style rather than making major changes to their professional practice. Further, the programme does not require teachers to assess their understandings of conceptions of pupil ability and provides teachers with an opportunity to continue to use ability grouping if they so desire:

> *Yes, it's worked really well. They've been doing it in their English sets because when we set for English and that's worked really well.' (Teacher: universal intervention)*

> *We started off doing it as a whole class and then it was difficult to get the challenge in with that so then we split into smaller groups, I sort of grouped them in ability. (Teacher: universal intervention)*

A number of teachers took some time to establish their role and the children's role in FFT Literacy Reciprocal Reading sessions. The flexibility of the programme provided a space for the teachers to try out new teaching methods that may not have been so familiar to them, and allowed them a space to be successful and to also fail:

> *I've tried different texts. I've worked with different and things like that and tried to bring in different levels of reading and things like that. I've tried focusing solely on certain aspects of it. (Teacher: universal intervention).*

> *It seems more of a learning curve to see what works for individual circumstances. (Teacher: targeted intervention).*

Two teachers commented that the implementation of FFT Literacy Reciprocal Reading was successful within their school because of commitment, collaboration and support shown by their colleagues and management. This is reflected in the comments below:

> *Yes, definitely, because I think then people really value its importance. Not everybody was initially valuing the importance because they didn't understand it as fully, whereas now, people do and …everybody's talking about it, so everyone who's done it talks about it and everyone's like a reciprocal reading thing? It's definitely become a bit of a buzz word within the school. (Teacher: universal intervention)*

> *The only few tweaks and changes we've had to make was the through lesson dropping where staff needed some support on the logistics of the children that were working independently whilst you were with a group. (Teacher: targeted intervention)*

One teacher implementing the targeted version of FFT Literacy Reciprocal Reading recalled that to be successful it needed to be implemented in a way that considered and met the educational needs of the pupils who were involved. The teacher considered this to be a strength of the programme, referring to the fact that both of their schools were located within high deprivation catchment areas and many of children were not as exposed to reading in the home. This is reflected in the comments below:

> *Yeah, I think a school like ours – you know, we're in an area of high deprivation and kids lack experience of the wider world, so choosing the right text, it, kind of, engages the child and allows them to almost become an expert in an area of something that they haven't had a real-life experience of. Because you're breaking it down into small sections, you're not expecting them to read for hours on end, you're reading small pieces at a time and you're really taking it apart in terms of vocabulary.*

Teachers implementing both the targeted and universal versions of the programme highlighted that the scaffolding approach of incorporating the four main strategies of predicting, clarifying, questioning and summarising was one of the main strengths of the programme and it helped to improve teaching approaches to reading and increased many of the pupil's confidence levels. For example, one teacher who was implementing the programme to the whole class cited the use of the four main strategies in each session as helping her to be more analytical about her approach to teaching. Another teacher using the targeted approach stated:

> *Those key steps you know, are the key elements of my lesson plan for reading in all my lessons with a few bullet pointed steps within each, predicting, clarifying, questioning, summarising, with maybe a few images to support what's going on, It just seems to be a more logical and inquisitive way to teach reading... it's great. (Teacher: universal approach).*

On the same note, a teacher using the targeted approach cited the repeated use of the four main strategies as being particularly beneficial for those pupils who were less assertive and normally less participatory in the reading class.

> *Upper Key Stage 2-wise, the children in the past… a lot of them would be quite passive when it came to reading and now they're much happier to engage in the class and not afraid to take part.*

Two teachers implementing the targeted approach highlighted that pupils in their classes had made progress, changed their practices and were more confident to independently use techniques such as decoding or using a dictionary reading comprehension. For example, one teacher recalled that in the past when a pupil in her class could not decode words independently; they would just ignore the words and move on in the text, but now they are use reciprocal reading strategies and are more confident in reading independently.

> *So if they came across a word that they didn't know they would just read it and just forget about it. Whereas now, I think they're a lot more confident in wanting to, kind of, find out what something means, and also looking at strategies to look at the word within a sentence and try and work it out for themselves. And they know that they can go to a dictionary. There's certain strategies and things that they would follow now to find an answer, whereas they weren't necessarily doing that before.*

Reciprocal reading instruction is teacher-led with facilitation which involves collaborative reading of texts (where teacher and students take turns to read, or students themselves take turns to read the text in small groups). The task is the use of the four FFT Literacy Reciprocal Reading evidence-based strategies – predicting, clarifying, questioning and summarising – modelled by the teacher and used collaboratively between teacher and students and between students, to derive meaning from the text. Hence both the teacher and the students use these strategies, in their instruction and learning respectively. All teachers involved in the study had to make changes to their teaching focus and lesson timetable to make space to deliver reciprocal reading sessions. The FFT Literacy Reciprocal Reading

programme is recommended to be delivered during literacy time as a universal programme. When Reciprocal Reading is delivered as a targeted programme it is recommended this be delivered to groups of 6 pupils as an additional session rather than replacing taught sessions.

Barriers to facilitation

All schools reported that the biggest barrier to implementing the programme was fitting it into the timetable both at the outset of implementation and throughout the year of delivering reciprocal reading. The issue of timetabling the programme was a problem faced by both universal and target groups. However, the timetabling of the sessions for the older children in the target group proved to be the most challenging, especially when schools did not feel it was appropriate that the children selected for reciprocal reading should miss out on other curriculum areas. FFT Literacy recommends, when reciprocal reading is delivered as a targeted programme, it should be as an additional session rather than replacing other taught literacy sessions, so could be taught outside of lesson time (pre-school or lunchtime) or could replace other lessons such as topic work.

> *I think in Year Six they found it tricky just because it's Year Six and trying to fit everything in…There's a lot, so I think they just struggle to fit it in, especially in the run up to SATs which I expected anyway. (Teacher: targeted intervention)*

> *…is just sometimes, our mornings are relentless. It's bam, bam, bam into the next thing to try and fit everything in. So just that really. Just sometimes having to stop the previous lesson quite abruptly in order to fit in the guided reading. Sometimes I find that I really feel I can't fit it in the morning because they've not finished other things, but the kids are just past it in the afternoons. (Teacher: universal intervention)*

In addition, staff across both the universal and targeted classes experienced pressure with the structure and length of the session. Whilst responsive teaching was part of the training programme, the unpredictability of reciprocal reading sessions made it difficult for teachers to plan for the appropriate amount of time that should have been dedicated to each element of reciprocal reading (i.e. predicting, clarifying, questioning and summarising). This was particularly an issue for teachers who had to deliver the programme within a limited and restricted amount of time. Across the group of teachers interviewed it was reported that timing was an issue in relation to finding an appropriate level and length text for the lesson, often choosing a text that was too long, the inability to complete a text, spending too long on one technique, spending too long on vocabulary, lack of understanding from the children and having to provide more support to the pupils than originally anticipated. Some of these challenges are reflected in the comments below.

> *It's quite a tight half hour. I mean, sometimes the kids could talk all day about the text but we've got that half hour, so I think that staff were saying they spent so long at the start with an unknown text unpicking vocabulary and that sort of thing they missed the discussion, which was really the most beneficial part. (Teacher: universal intervention)*

In the universal group several teachers mentioned that the ability and confidence level of the children was a challenge when teaching the programme to the whole class. One teacher reported having to remove children from the session, as they were unable to access the content of the programme. In relation to the low confidence levels of other children, a few teachers reported that as the programme became more familiar to the children this was less of an issue:

> *I think there are definitely some children in my class who cannot access it at all and it's done nothing really but confuse them. So, I do actually take them out, this isn't going to help because, like I said, my class goes from being very extreme lower children where they are going to be Year One all the way to children with a GCSE level.*

In the main it was the teachers and TAs delivering the reciprocal reading to small groups of children (targeted) that experienced barriers around implementing the programme. There were two main challenges. Firstly, the majority of teachers teaching this group reported issues around staffing levels for adequate delivery of reciprocal reading. This appeared to be particularly challenging for schools that expected the class teacher to guide the small group of targeted Years 5 and 6 pupils. It was reported that teachers struggled to ensure the rest of the class continued to learn with an appropriate, independent activity. Some teachers suggested that delivery of reciprocal reading to the target group would be less challenging with a second member of staff:

> *The thing that's quite difficult is being freed up to do the groups. There's sometimes I'm having to do the group while I've got the rest of the class and then it's not as good quality as if I am able to come out of class. At one point in the year I had*

a PGCE student, so she could take my class and I could take the group away and then it would be real quality. But that's quite tricky. (Teacher: targeted intervention)

I think, sort of, managing the rest of the class, I think it's been a little bit of a challenge. Our children aren't quite so ready to perform a typical reading independently, so they can't, kind of, do their own session while I'm leading a different session. Keeping them working independently has been a bit of a challenge because the sessions are so focused on the group that you're working with. (Teacher: targeted intervention)

Secondly, many of the teachers from the targeted groups suggest that fitting in two sessions of reciprocal reading a week would impact on teaching the rest of the curriculum for this age group (Years 5 and 6) particularly when preparing them for the impact on SATs. This is interesting, considering the high level of dosage reported by all schools. It may be the case that despite the difficulty reported qualitatively, schools still managed to fit in an adequate level of dosage for all groups. Alternatively, it may be the case that some teachers over-reported dosage or that the low response to the dosage survey is not representative of these struggles. The difficulty with fitting in delivery is reflected in the following comments made by teachers:

I think it was trying to fit in two sessions where it wasn't going to disrupt the curriculum too much. For children who were in Year Six, especially, it was really difficult. Even though we knew it was going to be really, really valuable, it was making sure that it was timetabled when it was gonna be up. Kind of, have least impact on other areas that they needed to cover.

I think in Year Six this is tricky just because it's Year Six and trying to fit everything in. I know that it's been really good because it's complementing all the skills they need for their SATs but in Year Six it is a different ball game.

Teacher buy-in

Overall, 177 teachers from 63 schools (36 intervention and 27 control) participated in the survey. Of these, 100 teachers were from intervention schools and 77 were from control schools. Of the 100 teachers who were eligible to take part in the section asking about engagement with RR, 86 said they were part of the intervention and were therefore directed to the section. However, as 14 teachers from the intervention schools said they had not taken part in the RR programme, they were not asked the questions on programme engagement. There were no missing values to individual questions: all questions were compulsory.

The majority of survey respondents (90% of 86 respondents from 36 intervention schools) felt that the FFT Literacy Reciprocal Reading programme was necessary, with a similarly high proportion of respondents (95%) reporting that they were 'engaged when delivering the programme'. The commitment of the school staff involved with implementing the programme emerged from the interviews. Some schools expressed the view that they chose to participate in the evaluation of FFT Literacy Reciprocal Reading because they felt the programme and the structure of the lessons used techniques for teaching reading effectively. This is reflected in the comments below:

I really like, I see it as a methodology for teaching reading rather than the discreet intervention role I'd say. I think it's, for me it's essentially quality first teaching and then those skills are sort of, instilling those skills in the children.

I feel like that they, I don't know, been given a bit of a magic wand because reading, if you like, it's one of those subjects that they've always felt a bit wobbly about teaching. I think writing, because it's got a lot more structure, reading can be such an awesome thing that they find the structure of reciprocal makes teaching reading easier to them.

As Table 27 shows, the teachers who took part in the intervention were overwhelmingly positive about it. Reflecting the comments in the teacher interviews, the survey suggests that the training was perceived positively: the respondents reported that they were confident and engaged when delivering the programme. The vast majority of respondents to the survey felt their school provided a suitable environment (92% agreement) and sufficient support (91% agreement) to enable the delivery of the programme. Once again, this reflects the views of the teachers who were interviewed, many of whom said that the supportive school environment contributed to the programme's success. The survey respondents were least positive about the paperwork; nonetheless, around three quarters (73%) strongly agreed or agreed that the paperwork to be completed did not take too long to do. There was no missing data within the sample who returned a teacher programme engagement survey, so all $N = 86$ for Tables 31–33.

Table 27: Teacher engagement with the programme (*N* = 86)

Response	% saying strongly agree or agree
I am confident in my delivery of the reciprocal reading programme.	98
I followed the guidance on how to deliver reciprocal reading closely.	96
I felt engaged when I was delivering the reciprocal reading sessions.	95
The sessions were straightforward to implement.	95
The working environment in this school is suitable for the delivery of reciprocal reading.	92
There is sufficient support from the school in the delivery of the sessions.	91
I feel reciprocal reading is necessary in this school.	90
The paperwork to be completed did not take too long to do.	73

The vast majority of survey respondents enjoyed doing the reciprocal reading sessions and would be happy to keep doing them (Table 28). Furthermore, 96 per cent said they would recommend the programme to other schools (Table 29).

Table 28: Teachers' attitudes to the programme (*N* = 86)

Response	% saying strongly agree or agree
I would be happy to keep doing the reciprocal reading sessions.	98
I enjoyed doing the reciprocal reading sessions with the children.	96
Overall, I was happy with reciprocal reading.	98

Table 29: Whether respondents would recommend the programme to other schools (*N* = 86)

Response	Yes %	No %	I don't know %
Would you recommend the reciprocal reading intervention to other schools?	96	0	4

Pupil engagement

Some of the pupils who were in the intervention schools selected 'no' as a response to the question, 'Did you have reciprocal reading sessions? Please ask your teacher if you are unsure' (*N* = 294) and did not access the engagement section of the questionnaire. As described earlier, it is possible that this means some pupils were absent for the whole programme, but there are other possibilities that are potential flaws of the measure (not knowing the name of the programme, or selecting 'no' erroneously). Some who were in the control schools selected 'yes' as a response to this question (*N* = 337). This may be interpreted as a crossover effect, although the programme team have stated that there is no possibility that the control schools had access to the FFT Literacy Reciprocal Reading programme. It is more likely, therefore, that pupils in the control schools were told that the project activities (post-test and questionnaire) were part of the 'reciprocal reading' project, and selected 'yes', thinking this was appropriate. They may also have selected 'yes' if they didn't know what reciprocal reading was, and chosen 'yes' assuming it referred to them receiving reading lessons in general. The FFT Literacy Reciprocal Reading programme cannot be accessed outside of receiving training from the FFT Literacy team, and we think a crossover effect is unlikely.

Future research would use a separate questionnaire for control schools, ensuring they could not access the engagement part of the questionnaire. Only one version of the pupil questionnaire was used, and pupils accessed the programme engagement section if they answered 'yes' to the above question. The analysis of pupil engagement data was, therefore, carried out using the 1811 eligible respondents (those who had taken part in the FFT Literacy Reciprocal Reading sessions). Their responses to individual items from the FFT Literacy Reciprocal Reading engagement questionnaire, and the numbers responding to each item, are presented in Table 30 and Table 31. The survey results generally reflect the focus group findings in demonstrating that the children found the FFT Literacy Reciprocal Reading sessions helped their reading and increased their interest. Just over half (29% a lot and 24% a little) said they had read more outside school since doing reciprocal reading, something that was also reported by the children taking part in the focus groups. See Appendix J for pupil survey.

Table 30: Children's perceptions of the effects of taking part in reciprocal reading

Response	A lot %	Quite a lot %	A little %	Not at all %	*N*
The things I learned in the sessions help me to understand better what I am reading.	47	32	18	3	1775
The things I learned in the sessions make reading easier.	39	33	23	5	1785
Reciprocal reading has made me more interested in reading.	36	30	25	9	1787
I read more outside school since doing reciprocal reading.	29	24	30	16	1783
Total					1811

The children were generally very positive about the reciprocal reading sessions, with the majority saying they enjoyed them, found them interesting and were happy to keep doing them Table 31). Only four per cent of children did not enjoy the sessions at all, five per cent did not find the sessions interesting at all and six per cent would not want to keep doing reciprocal reading at all. Overall, 84 per cent of the survey respondents said they were happy with reciprocal reading 'a lot' or 'quite a lot'.

Table 31: Children's attitudes towards the programme

Response	A lot %	Quite a lot %	A little %	Not at all %	*N*
I think we should keep doing reciprocal reading sessions in our class.	50	24	20	6	1763
I enjoyed the reciprocal reading sessions.	43	30	23	4	1796
I found the reciprocal reading sessions interesting.	34	35	27	5	1780
Overall, I am happy with reciprocal reading.	60	24	13	3	1743
Total					1811

Influence of school ethos on implementation

School ethos was measured quantitatively at pre-test for all schools. Therefore, quantitative results are not discussed here as an outcome, but are discussed later in the context of comparison with control group usual practice for reading comprehension. Reciprocal reading was considered to align well with the ethos of all schools involved. The majority reported that the ability to read lay at the core of their ethos as being able to read provides children with access to knowledge. Therefore, the reciprocal reading programme was a tool to be used by teachers to aid the development of their pupils' reading ability:

> *So we've kind of, put enjoyment and the love of reading at the centre of everything that we do, because we know that that's the foundation for all the other curriculum areas. (Teacher: universal intervention)*

> *So reciprocal reading was something that we felt was fitting to a need in school to look at doing something with reading that wasn't just a kind of unintelligible and something that would become embedded within the school. (Teacher: target intervention)*

> *You know, the reading policy and the reading ethos is obviously … you know, we want the kids to be lifelong learners and develop a love of reading. (Head teacher)*

Fidelity

Fidelity of implementation was examined quantitatively through the lesson record forms and qualitatively through interviews with teachers.

Delivery of FFT Literacy Reciprocal Reading

All teachers involved in the intervention group had to make changes to their teaching focus and lesson timetable to make space to deliver reciprocal reading sessions. Whilst some teachers found this challenging, below are ways they suggested the programme was perceived to be adaptable to accommodate this:

Yeah, so what we've done is, we've changed our timetable so that we have a slightly later lunch, which allows us to do half an hour of guided reading every day from 11.20 to 12.10.

We've moved away from the kind of rotation of activities that we've done in the past and made reciprocal reading the focus. So, we've fitted in the timetable pretty much first thing in the morning…

All schools that completed the lesson logs in the trial exceeded the required minimum recommended dosage for both universal and targeted interventions. The mean dosage was 60 minutes higher for universal than for targeted. The maximum reported dosage was the same for each intervention (4830 mins) and was ten times higher than the recommended minimum dosage for targeted (480 mins) and twenty times higher than the recommended minimum dosage for universal (240 mins), This appears to have been influenced by schools running the programme for a lot longer (mean = 26 weeks of delivery) than the recommended minimum number of weeks (12 weeks). Descriptive statistics are further detailed in Appendix G. Although some teachers reported that they found it difficult to find enough time for the programme, the quantitative statistics show that they all delivered at least the minimum recommended dosage. This may mean that teacher-reported records of dosage are less accurate than asking teachers directly in an interview, or it may simply mean that teachers over-estimated how much of the programme they should be delivering, and therefore felt like they were under-dosing. The teachers may not have been informed by the programme developers about how much was the minimum recommended dosage, and were unaware that they had already delivered an adequate amount. Furthermore, although some schools delivered a higher dosage of the universal than of the targeted intervention, this does not mean that they were non-compliant. As discussed earlier, as long as schools delivered at least the recommended dosage, they were compliant. The relationship between the two dosages within a school is not important for fidelity. More simply, the standard for minimum dosage is higher for the targeted version, but this is not negatively impacted if a school also delivers a high dose of universal (even if it exceeds the targeted dose).

It should be considered that the minimum dosage was likely a very conservative estimate by the programme developers and not a minimum amount that was stressed to the schools as what they should deliver to remain compliant. This raises the question 'Should the minimum dosage be raised by the developers?' Dosage was not found to be predictive of primary outcomes: therefore, the lower end of dosage found in this sample was not found to be associated with lower reading outcomes. Similarly, a very high dosage was not associated with higher reading outcomes. This suggests that the lowest dosages in this sample were still effective, and the point at which the programme is negatively impacted may be lower than any of these reported dosages. The positive effects of the programme were found with dosages consistently (and noticeably) higher than the recommended minimum. A caveat for programme success, therefore, is that if a school implemented lower than the dosages in this sample, the likelihood of a positive effect may be less certain. However, dosage above the minimum was consistently achieved, so even if a low dosage may result in a less positive effect, the programme is highly feasible and this study found that all schools were able to implement the dosage with success.

The delivery of FFT Literacy Reciprocal Reading within each school relied heavily on the teachers' own perceptions, understanding and knowledge of the educational needs of each pupil within their class and the flexibility within the programme to being able to put the children into reading groups. All but one teacher from the universal groups referred to the diverse pupil abilities contained within their classes and how whole-class delivery of reciprocal reading was challenging to effectively teach and challenge all pupils together. This resulted in many teachers teaching FFT Literacy Reciprocal Reading in ability grouping style where children were placed together with children of a similar ability in small groups. Each small reciprocal reading ability group would be given a different text and follow-up activities that matched their overall perceived ability. Teachers made the decision to use grouping to help FFT Literacy Reciprocal Reading meet the whole class needs in a whole class session. This is reflected in the comments below:

We started off doing it as a whole class and then it was difficult to get the challenge in with that so then we split into smaller groups, I sort of grouped them in ability…

So, it didn't work as a whole class because the needs in my class are so massive…There's no way to challenge the higher while keeping the lowers in play so…

They've been doing it in their English sets because when we set for English. That's worked really well.

> *…how I deliver will depend on the children I get in. I think the children coming up to me are going to be lower attainment than what I've got currently. And I think their learning behaviours are more challenging. So I might choose to put that programme into smaller groups.*

In all schools, teachers led the implementation and delivery of reciprocal reading within their own classrooms for the Year 4 whole-class intervention. However, the role of the teacher within the implementation and delivery of reciprocal reading for the target groups in Years 5 and 6 was not as prominent. The teacher interviews suggest that around half of the schools made use of TAs to assist with reciprocal reading as an intervention for children with poor comprehension skills:

> *I think because we've got such a strong group of teaching assistants who were doing the interventions, we had an initial meeting where we talked about what was gonna happen and planned initially and then from there, they've just, kind of gone away and done their planning and got on with it because they really believe in it.*

In the few schools that did not make use of the TAs in the same way, the teacher's role in the delivery of reciprocal reading for the target groups was hindered as the teacher was faced with the challenge of managing their classroom while trying to focus on the small target group:

> *The thing that's quite difficult is being freed up to do the groups. There's sometimes I'm having to do the group while I've got the rest of the class and then it's not as good quality as if I am able to come out of the class.*

> *Because I feel like, if it's only going to work with some identified children, like a small group of children for it to work, at times you probably need two members of staff.*

No data was collected on if children were further divided into pairs or smaller groups within the targeted groups.

The successful implementation of the reciprocal reading programme within schools does not only depend on the teachers or school staff but also on the children themselves.

> *…so the challenge for children of eight years old is they have to predict, interpret what information they don't understand, lots of vocab, they have to read it and they have to be able to summarise it, synthesise it really quickly, to then be able to generate questions.*

A small number of teachers were concerned that the children in their Year 4 class were too young to be able to engage fully with the strategies. This meant that the children were less independent with their learning as they needed more teacher input throughout the year:

> *Ideally, I know that the programme is designed around handing the reins over to the children. I've found in Year Four that they still need quite a lot of scaffold and guidance.*

For pupils who were not familiar with the style of learning and independence promoted within reciprocal reading it took some time before they become comfortable in their new roles:

Although the majority of teachers assumed the children enjoyed the respective roles in reciprocal reading, some thought that the children found some roles easier than others. Many teachers had concerns about the effectiveness of the questioning role as children found it difficult to generate higher level and challenging questions:

> *Yeah definitely, out of the four key skills there's ones that they find easier I think. I think summarising, I think today, you saw that it was one that some of them are still working on and yeah, questions, although they are asking questions, it's having the time to challenge them and try to get them to get a better question. (Teacher: universal intervention).*

> *The one thing that's challenging and I don't feel I've cracked it at all is the questioning. So, I, I've no problem generating questions to discuss with the children but it's the children generating the questions, that's been challenging for Year Four children. So then coming up with questions that they would want to explore themselves, that's been quite a challenge. (Teacher: universal intervention)*

Outcomes

Perceived benefits of the programme

Perceived benefits of attending the reciprocal reading sessions include increased confidence, concentration, greater enjoyment of reading, development of greater reading comprehension, fluency and listening skills in addition to improved group work such as turn taking and helping each other.

> *I know from data that I've looked at with Year Four in terms of year group across schools that our year group has made a lot of progress compared to other year groups; for example, their reading fluency and the pace they read. (Teacher)*

> *Their fluency and expression have improved because they're reading for other people. (Teacher)*

Teachers reported that techniques and strategies learned through reciprocal reading lessons such as the four key strategies of the FFT Literacy Reciprocal Reading programme (predicting, clarifying, questioning and summarising) were beneficial for the pupils. They also thought these strategies were empowering for the children, as once the skills were secure the children were able to use them in their independent reading. These perceived positive outcomes were particularly prevalent for the targeted group and were attributed by the teachers who were interviewed to being part of a small group rather than as a whole class. This is reflected in the comments below from three of the teachers who facilitated the targeted intervention.

> *I think that it's just they feel like it's a safe space to get things wrong in, not that the classroom isn't but it's just, when it's smaller they have come on a lot.*

> *I think, when it came to reading, and I think being in smaller groups where he's been able to take a bit more time, focusing on a text. You know, reading in a secure environment and then really being able to take the text apart and discuss it.*

> *Whereas now, I think they're a lot more confident in wanting to, kind of, find out what something means, and also looking at strategies to look at the word within a sentence and try and work it out for themselves…There's certain strategies and things that they would follow now to find an answer, whereas they weren't necessarily doing that before.*

Around half of the teachers interviewed found that reciprocal reading sessions were beneficial for the teacher–pupil relationship as it provided a space that enabled the teachers to hear their pupils reading more often and allowed them to assess their reading progress on a weekly basis:

> *I feel that I'm listening to the children read more often, I've got a better understanding of where they are. (Teacher: targeted group)*

> *I have found in the mornings that they know that it's there to read with me, they do get excited on the morning because they know they can have a chat about it, the text, and talk about stories. (Teacher: universal group)*

The quality time teachers were spending with the pupils during the lessons provided teachers with a reminder that they could not assume pupil knowledge. This recognition reminded some teachers to slow their teaching down to ensure pupils were equipped with the knowledge they needed to understand a text:

> *And it also means teachers slow down so that they're not, kind of, skipping ahead and presuming that children know things, cause [sic] they actually don't…for example, not knowing what a seagull is or something like that, which totally throws a curveball in the session where you might end up spending the majority of the session trying to define one noun… (Teacher: universal intervention).*

The majority of teachers from the targeted groups felt that reciprocal reading gave them a way to teach children to read in an inclusive way as the chosen texts were more accessible and the small group nature of the lessons provided lower ability children with a safe space to share their ideas. This important point is reflected in the comments below:

> *…and you get the majority of the group taking part, rather than just the characters who always get involved, if that makes sense?*

> *I think it's the fact that, that you don't have to read that much in a session is good because it means that the children don't get freaked out by the length of a text, which sometimes happens.*

Only one school did not feel that the reciprocal reading worked as well for the target groups as it did for the universal groups. However, this was attributed to the approach taken by staff rather than an issue with the programme:

> *I don't think that the reciprocal reading has been, has made as much impact in Year Five and Year Six as it has in Year Four. And I think that's probably to do with staffing and how they've approached reading (Head teacher)*

Some schools thought the programme would be an attractive intervention as it was considered to be an affordable and cost-effective way to teach children reading strategies:

> *To me, it's one of those, it's cost effective which is a massive tick with budgets being so tight.*

> *Financially, the schools are getting less and less money. But reciprocal reading, that wasn't, I don't think it costs that much…*

Across the focus groups, many children reported that the techniques and strategies taught through the programme gave them more confidence and made reading and spelling appear less of a chore. As a result, many children reported reading outside of school more frequently. This important point is reflected in the comments below:

> *It has helped me with my words and reading, a lot. (Pupil: Year 4)*

> *Normally in English, we read paragraphs out of the book or something and reading helps me to figure out all those hard words and because I didn't used to figure out all those hard words and I got a bit nervous saying them in front of the class, but now I don't. (Pupil: Year 6)*

> *Break it down, because when I was reading my book every day, the word was 'psychic' - I didn't know because it was a funny spelling so I broke it down.' (Pupil: Year 5)*

> *I think it helps me more at home because I love reading and when I read at home and I find something tricky, instead of just asking my mum, I can look in a dictionary and I can spell the word off the top of my head. (Pupil: Year 6)*

Negative effects

The majority of teachers did not feel that reciprocal reading led to any negative or adverse effects for themselves, the school or their pupils. One school, however, disclosed that by doing reciprocal reading with one of their classes it resulted in the exclusion of some pupils. The teachers believed these pupils were not academically able enough and, therefore, could not access it. The reuse of this quote below reflects this point.

> *I think there are definitely some children in my class who cannot access it at all and it's done nothing really but confuse them. So, I do actually take them out, this isn't going to help because, like I said, my class goes from being very extreme lower children where they are going to be Year one all the way to children with a GCSE level. (Teacher: universal intervention)*

Overall, the majority of the pupils did not feel that reciprocal reading had any negative effects. However, one pupil was concerned that some pupils dominate the small group sessions, meaning that some pupils' voices were silenced:

> *Like when you're stuck on a word, that is the difficult part because everyone else is speaking over you. You can't sound it out.' (Pupil: Year 6)*

Formative findings

Improvements

As indicated above, overall satisfaction with the FFT Literacy Reciprocal Reading programme was very high. Over half of the schools who participated in the interviews had no suggestions for improvements. There were two main recommendations for improving the delivery of the programme that requires more direction from the programme

writers. The first was a desire to have more direction on the texts that would be suitable to be used with reciprocal reading.

> *I would quite like some examples of further reading that we could do…*
>
> *Just maybe a bit more information around the texts. You know, and just think about the different ages and abilities of the children as well.*

The second was a desire to have more guidance about how to adapt reciprocal reading for the needs and ages of different pupils:

> *…whether the programme needs to be tailored slightly depending on the Key Stage, lower Key Stage Two, needs to be more emphasis on vocabulary…It's maybe about you guys or the programme developers looking at how do we implement this in Key Stage One, what elements do we want to focus on? How does that link into the beginning of Key Stage Two?*

Another point to consider are teachers' concerns about timetabling. This was perceived as a barrier to delivery rather than an area for improvement and is discussed at length in barriers section above. However, there were still high levels of (reported) implementation despite this barrier. So one suggestion for improvement could be that FFT Literacy Reciprocal Reading recalibrate their training and guidance on how to avoid timetabling issues for teachers given that there were generally high levels of implementation anyway.

Control group activity

No control schools received training from FFT literacy during the trial. No control schools were eligible to participate in other EEF literacy trials during the trial. Only one control school dropped out of the trial after allocation to the control group and refused to facilitate post-testing. This suggests that demoralisation or compensation rivalry were not a problem for this trial. It should be considered that school comprehension ethos survey was delivered at post-test and this means that if control group schools changed practice during the course of the trial, this would not have been captured. The ethos survey does include more than just perceived attitudes, it also includes measures of current teaching (at the time of the survey), and is therefore still appropriate as a measure of practice in the control schools, with the caveat that it was at the beginning of the trial.

School comprehension ethos

The results from the school comprehension ethos questionnaire, completed at pre-test by head teachers from 49 control and 46 intervention schools, showed that there were no statistically significant differences in the mean scores on the overall scale (control mean = 45.86 and intervention mean = 45.28; t = 0.646, df = 80, p>0.05). This suggests that the prioritisation and focus of the teaching of reading comprehension in the school was similar at pre-test in both control and intervention schools which is also evident from the post-programme findings presented in Table 32. As can be seen, the percentages of principals in control and intervention schools who said they strongly agreed or agreed with each statement was similar for most items on the questionnaire. The total number of participants who responded to each question is shown in the '*N*' column for each item. The largest difference between control and intervention schools was for the item: 'Teachers are aware of instruction strategies that target reading comprehension difficulties' with 49 per cent of the former compared with 29 per cent of the latter agreeing or strongly agreeing with the statement. This was followed by 'Teachers can identify children who are good decoders but struggle with reading comprehension' (87% and 68% respectively). The final item for which a fairly large difference in agreement was found between control and intervention schools was for 'Current teaching in Years 4, 5 and 6 focuses more on comprehension skills than on decoding skills' with 82 per cent of the principals in the control group agreeing with this statement compared to 98 per cent of principals in the intervention schools. However, none of the differences between teachers from control and intervention schools were statistically significant.

Table 32: School comprehension ethos completed by head teachers at pre-testing

Response	% saying strongly agree or agree			
	Control	Intervention	*N*	*p*[13]
Current teaching in Years 4, 5 and 6 includes a lot of work to target reading comprehension.	84	83	86	0.926
Current teaching in Years 4, 5 and 6 focuses more on comprehension skills than on decoding skills.	82	98	86	0.129
Current teaching in Years 4, 5 and 6 makes use of small group activities for the teaching of reading comprehension.	76	78	86	0.827
Current teaching in Years 4, 5 and 6 places as much emphasis on reading comprehension as on decoding.	36	38	85	0.855
It is usual for teachers to identify children who are falling behind in reading comprehension.	93	88	86	0.388
Teachers can identify children who are good decoders but struggle with reading comprehension.	87	68	86	0.148
Teachers are aware of instruction strategies that target reading comprehension difficulties.	49	29	86	0.260
Many children in my school would benefit from more teaching focusing on reading comprehension.	93	98	86	0.443
Reading comprehension for KS2 is vital to literacy development planning in my school.	98	95	85	0.181
There is a lot of support and/or training available to teachers for improving reading comprehension in my school.	36	42	86	0.131
Discussion about texts is integral to reading comprehension in my school.	89	88	86	0.972
There is a lot of support and/or training available to teachers in my school for teaching reading comprehension.	42	38	83	0.097
Total			95	

Teacher attitudes and behaviour

This section presents the descriptive data for the responses to the TCIB questionnaire which was completed at post-test by teachers in control and intervention schools. Table 33 shows the percentages for the two groups to the questions on the importance of teaching reading strategies to pupils in their classrooms. As the vast majority of teachers in both groups strongly agreed or agreed with each statement, the percentages are presented for strongly agree. The percentages of teachers in control and intervention schools who said they strongly agreed with each statement were similar for most items on the questionnaire. The largest differences between control and intervention schools were for the items: 'It is important to teach children how to sum up what they have read to check how well they have understood the text' (69% and 82% respectively) and 'It is important to explicitly show children how to check their understanding when reading text' (65% for control and 78% intervention schools). This was followed by 'It is important to give children the opportunity to consolidate their understanding of what they have read in individual post-reading tasks' with 57 per cent of the control group compared with 67 per cent of the intervention group strongly agreeing with the statement. However, none of the differences between teachers from control and intervention schools were statistically significant.

[13] *p*-value of chi-squared analyses

Table 33: Teacher views on importance of teaching reading strategies to children at post-test

Response	% saying 'strongly agree'		*N*	*p*[14]
	Control	Intervention		
It is important to teach children to draw on what they already know about the subject of a text.	69	70	177	0.834
It is important to teach children to anticipate what may happen next in a text to aid understanding.	65	68	177	0.735
It is important to teach children how to identify words and phrases they are unsure about when reading text.	88	93	177	0.357
It is important to teach children how to seek out the meaning of words and phrases they are unsure about when reading text.	83	87	177	0.306
It is important to teach children to ask questions about things that are unclear or confusing to help them understand what they are reading.	88	88	177	0.674
It is important to teach children how to ask questions about the text to help them understand what they are reading.	77	79	177	0.534
It is important to teach children how to sum up what they have read to check how well they have understood the text.	69	82	177	0.063
It is important to explicitly show children how to check their understanding when reading text.	65	78	177	0.133
It is important to provide opportunities for children to talk about the processes of reading text.	77	69	177	0.380
It is important to give children the opportunity to consolidate their understanding of what they have read in individual post-reading tasks.	57	67	177	0.217
It is important to carefully select suitable texts as the context to teach active reading.	73	78	177	0.176

While there were no statistically significant differences between teachers in the control and intervention schools in their views on the importance of teaching reading strategies there were some notable, statistically significant, differences in relation to their actual behaviour. As Table 34 shows, overall, teachers in the intervention group were more likely than their counterparts in the control group to say they 'always' used the reading strategies presented in the questionnaire items. These differences were statistically significant for three items 'I explicitly teach children to anticipate what may happen next to aid understanding'; 'I explicitly teach children how to identify words and phrases they are unsure about when reading text' and 'I explicitly teach children how to seek out the meaning of words and phrases they are unsure about when reading text.'

[14] *p*-value of chi-squared analyses

Table 34: Teachers' reading strategy behaviours at post-test

Response	% saying 'always'		*N*	*p*
	Control	**Intervention**		
I explicitly teach children to draw on what they already know about the subject of a text.	40	50	177	0.262
I explicitly teach children to anticipate what may happen next to aid understanding.	49	66	177	0.006** (Cramer's *V* = 0.26)
I explicitly teach children how to identify words and phrases they are unsure about when reading text.	61	82	177	0.012* (Cramer's *V* = 0.25)
I explicitly teach children how to seek out the meaning of words and phrases they are unsure about when reading text.	56	81	177	0.001** (Cramer's *V* = 0.30)
I explicitly teach children to ask questions about things that are unclear or confusing to help them understand what they are reading.	58	71	177	0.111
I explicitly teach children how to ask questions about the text to help them understand what they are reading.	48	62	177	0.054
I explicitly teach children how to sum up what they have read to check how well they have understood the text.	47	57	177	0.056
I explicitly show children how to check their understanding when reading text.	47	55	177	0.326
I provide opportunities for children to talk about the processes of reading text.	34	46	177	0.368
I give children the opportunity to consolidate their understanding of what they have read in individual post-reading tasks.	39	46	177	0.572

*$p < 0.05$; **$p < 0.01$ effect sizes (Cramer's *V*): 0.25 'I explicitly teach children how to identify words and phrases they are unsure about when reading text'; 0.26 'I explicitly teach children to anticipate what may happen next to aid understanding'; 0.30 'I explicitly teach children how to seek out the meaning of words and phrases they are unsure about when reading text.'

This shows that intervention teachers scored more highly than controls on items focused on teaching of predicting within a text and identifying and clarifying unknown words.

Conclusions

Key conclusions	
1.	Children in the FFT Literacy Reciprocal Reading targeted intervention group made the equivalent of 2 additional months' progress in both primary outcomes (overall reading and reading comprehension), on average, compared to the equivalent children in the other schools. This result has a moderate to high security rating.
2.	There is no evidence that the FFT Literacy Reciprocal Reading universal intervention had an impact on pupils' overall reading or reading comprehension outcomes, on average. This result has a moderate to high security rating.
3.	Children in the targeted intervention made more progress in both secondary outcomes (reading accuracy and pupil comprehension meta-cognition), whilst the universal intervention showed effects for only pupil comprehension meta-cognition.
4.	Among children receiving free school meals (FSM) analysis showed an effect on reading comprehension for pupils in the targeted intervention, but not for the universal version. However, this initial analysis was conducted using information supplied directly from schools and was constrained by missing data. Subsequent analysis was able to match data for FSM eligible pupils within the trial using the National Pupil Database. These results found signs of promise for both the targeted and universal interventions on outcomes for children eligible for FSM. These results are summarised in full in appendix K.

Interpretation

This CRCT evaluation indicated that the targeted version of FFT Literacy Reciprocal Reading programme had significant positive effects on both primary and secondary reading outcomes. Specifically, the targeted version of the programme improved both primary outcomes of overall reading (ES = 0.14) and reading comprehension (ES = 0.18), which is the equivalent of 2 months additional progress based on EEF guidelines. It also improved the secondary outcomes of reading accuracy and pupil comprehension meta-cognition (PCMC).

However, the evaluation did not support the universal version of the programme having an effect on primary outcomes with no significant effects on either overall pupil reading or pupil reading comprehension. There was one significant positive effect of the universal version of FFT Literacy Reciprocal Reading on secondary reading outcomes of pupil comprehension meta-cognition but no significant effect on pupil reading accuracy. However, the exploratory analysis did not find PCMC to be significantly associated with higher reading comprehension in the intervention groups. Therefore, additional analysis and research would be required to investigate any potential significance of this finding

These findings should be considered in the context that all schools reported that they were compliant with the intervention. They all reported delivery higher than the recommended minimum dosages for both the universal and targeted interventions (particularly for the universal version). However, this was self-report and only 32 out of 47 returned this information. A caveat here is that if a school implemented the minimum dosage, we cannot be certain they would achieve a successful improvement, as no school in the sample reported the minimum level. The effectiveness of the minimum effective dose has, therefore, not been tested. Furthermore, even with extremely high dosage, the universal intervention did not show a positive effect and this should be considered when evaluating its promise.

This leads to the overarching question: Why did the targeted version have better effects than the universal version? The data can be interpreted in several ways to answer this question. Firstly, the pupils in the targeted version were specifically chosen as those who best aligned with the programme theory of change. That is, it helps develop comprehension ability in those who are typically developing in reading accuracy but below average in reading comprehension ability. This notion is also supported by the fact that the significant effect on comprehension was slightly larger (ES = 0.18) than on overall reading (ES = 0.14) within the targeted group.

Secondly, the implementation process data suggested that pupils in the universal intervention sometimes struggled to access the programme. One interpretation could be that the older age of the readers (Year 5 and Year 6) in the targeted intervention allowed them to better understand and apply the four core concepts of the programme (predicting, clarifying, questioning and summarising) than the younger readers in the universal version (Year 4). This

would concur with the previous research on younger students being less able to engage with strategic thinking in regard to reading (Duffy, 1993; Hacker and Tenent, 2002)

Thirdly, the implementation process data also suggests that the context of the targeted version was more appropriate for delivery. Specifically, the smaller groups allowed the pupils to understand and apply the four key concepts of the programme. Some teachers reported excluding pupils from the universal intervention, which may have undermined impact of this intervention, as some pupils who were analysed in the ITT model as intervention may not have received the programme. However, this was only reported qualitatively and the implications for the quantitative effects are therefore limited. Also, teachers and TAs delivering the programme were able to engage the pupils in all four elements of the programme and give them an opportunity to practise them all. It was reported the universal version did not allow time and opportunity for going through the full cycle of the four elements and in particular the 'questioning' element. It was also suggested that the context of smaller groups during the targeted version was less intimidating for the pupils to practise all the elements with the tutor, especially for those who were deemed to be weaker at comprehension-type activities.

Finally, one further reason that is often given for some reading interventions being more effective than others is intervention dosage or exposure. However, this does not appear to be a factor in this study as the pupil dosage for the universal version was slightly higher than the targeted version and overall dosage was a not significant predictor of outcome change. Although the recommended dosage of the targeted intervention is higher than the recommended dosage of the universal intervention, this does not mean that any deviation from this makes a school non-compliant with the intervention, as discussed above. The higher than minimum dosage found in all intervention schools underlines how highly feasible and successfully implemented the programme was in this study. Again, we cannot be sure that the minimum effective dosage declared by the developers is in fact 'effective', and the minimum effective dosage remains untested due to the high dosage in this study.

The exploratory analyses for the targeted intervention showed a significant negative relationship between teacher comprehension importance and pupil comprehension meta-cognition. One hypothesis is that the teachers in the intervention groups gained greater insight into the importance of comprehension and re-calibrated their responses. However, additional investigation across control and intervention groups would be required to interpret this relationship further. The implementation analysis showed a significant relationship between pupil programme enjoyment/engagement and pupil reading comprehension for the universal intervention and a significant effect of pupil enjoyment/engagement on pupil comprehension meta-cognition for the targeted intervention. However, there were no associations between dosage and teacher engagement on programme outcomes. Again, further analysis would be required to fully interpret these findings but generally it shows pupil engagement/enjoyment of the programme was the most influential implementation factor across both universal and targeted versions. Interestingly, pupil enjoyment/engagement in evidence-informed educational programmes has been found to be a significant implementation factor in several other implementation and RCT studies by the authors (O'Hare, 2014; O'Hare, *et al*., 2017, O'Hare *et al*., 2018), but this relationship has yet to be widely or systematically explored in the literature. In addition, with regard to additional quantitative analysis, the sub-group analyses repeated the main analysis for only pupils in receipt of FSM and showed a significant main effect on reading comprehension for pupils in the targeted intervention, but not for the universal version. However, this initial analysis was conducted using information supplied directly from schools and was constrained by missing data. Subsequent analysis was able to match data for FSM eligible pupils within the trial using the National Pupil Database. These results found signs of promise (i.e., significant effects) for both the targeted and universal interventions on outcomes for children eligible for FSM. These results are summarised in full in appendix K.

There were a number of other interesting findings in the implementation process evaluation in addition to those mentioned above, regarding age suitability and contextual issues (in the classroom) with the intervention. One of the key things to emerge across all these factors was how much the teachers enjoyed the flexibility within the FFT Literacy Reciprocal Reading programme to allow them to adapt it to their pupils' requirements. This was particularly the case for the targeted version.

With the pattern of effects on the primary outcomes in mind, it is useful to consider all other data analyses and identify what this can tell us about the overall logic model for the programme. Despite comprehensive measurement of pre-identified proximal outcomes (i.e. a range of teacher and pupil comprehension dispositions) there was generally

inconclusive evidence on the nuanced theory of change for this programme. This may be a result of FFT Literacy Reciprocal Reading having a holistic but small impact across a wide range of these proximal variables resulting in the identified distal benefits on reading ability. There is some evidence that the pupil comprehension meta-cognition may be one of the proximal routes to the changes in the distal primary reading outcomes as significant improvements were seen in this for both the universal and targeted versions of the programme. This could also suggest that the universal version of the programme was powerful enough to change this proximal secondary outcome but not translate this into further distal changes in the primary outcome. This is a working hypothesis and, again, further analysis could be helpful to unlock this theory of change. The post-evaluation logic model is presented in Appendix I with the significant pathways highlighted. Essentially, the strongest conclusion that can be made is about the direct effect of the targeted version of FFT Literacy Reciprocal Reading on comprehension ability (see green highlighted arrows) and its mediating effects on reading accuracy and overall reading ability (C in Figure 1). Furthermore, the logic model also highlights the effects both versions of the programme had on pupil comprehension meta-cognition (PCMC), but leaves a question mark over whether changes in PCMC are driving the theory of change on the primary outcome (M – highlighted in amber in Figure 1). Finally, the post-evaluation logic model also highlights the anomalous effect of teacher comprehension importance on PCMC (I – highlighted in amber in Figure 1) and the implementation factors found to be significant (highlighted by dashed green arrows) i.e. FSM, pupil enjoyment/engagement and pupil reading accuracy.

In conclusion, this evaluation provides strong support for further implementation of the targeted version of FFT Literacy Reciprocal Reading intervention. This finding should be considered in recognition of the fact that all intervention schools for whom compliance data is available reported compliant dosage of the targeted intervention. Despite qualitative reports that finding time for this dosage was a challenge including 14 teachers reporting they had not delivered the intervention, there was a high average overall dose/exposure to the programme provided. The teachers in the intervention schools not reporting this could be due to the programme being delivered by TAs, especially for the targeted intervention. If the link to the online survey was forwarded to the teachers of those targeted intervention pupils, but not the TAs, this could account for the 14 intervention teachers reporting in the post-test survey that they did not deliver the programme. The intervention pupils selecting ‘no’, and the control pupils selecting ‘yes’, to the question ‘did you have reciprocal reading lessons? Please ask your teacher if you are not sure’, could be due to numerous reasons, as discussed in the results section. It is likely that this was a flaw in the measure, as control pupils could access the programme engagement questionnaire even if they were in a control school, and intervention pupils could click ‘no’ even if they had received the programme. The programme delivery team are confident that no control schools could have had access to their programme, and a crossover effect is, therefore, unlikely. In addition, it was not possible to test statistically the effect of differential compliance. The main reasons for recommending further implementation of the targeted version are: the evidence this evaluation provides of its effect on reading outcomes; its focus on the pupils who will benefit the most through the targeting process, its low cost because it uses a teacher professional development model; and the generally high levels of teacher and pupil engagement with the programme. Furthermore, these positive features are enhanced in the context of a substantial proportion of language and literacy focused interventions rigorously evaluated by EFF not showing significant effects on reading outcomes (EEF, 2018a). Also, with the National Curriculum (2013) advocating a universal focus on comprehension instruction, the targeted version of FFT Literacy Reciprocal Reading can offer additional support to those who may be unable to initially benefit from the universal approach being delivered through the curriculum. Finally, with regard to using RCTs in education, this study demonstrates the importance of preliminary examination of a programme’s theory to ensure that programme outcomes are specifically matched to evaluation outcomes and that the programme is being delivered to the most appropriate groups of participants.

Limitations

The generalisability of these findings is quite good for pupils of similar characteristics to the sample tested, i.e. Year 4, Year 5 and Year 6 pupils in English primary schools, due to the power of this study. However, generalisability to other age groups may be less certain, particularly since there was some qualitative evidence that the younger children struggled with some of the concepts within the programme.

Another potential limitation is that there was some potential for teacher selection bias in their identification of pupils in receipt of the targeted version of the programme. However, both control and intervention teachers did make their

selections before randomisation and their choices were based on standardised guidance based on the simple view of reading (see Appendix D).

One potential issue with this evaluation was that the NGRT test uses adaptive testing. Therefore, pupils who perform very poorly on accuracy questions may not get access to the comprehension questions, although this may not have impacted upon the pupils in the targeted group as they were selected on above average accuracy ability. No ceiling effects were evident on the NGRT measures as the curve shows distribution well under the maximum score of 500. Distributions of neither the pre-test nor post-test of primary outcomes showed a ceiling effect. Furthermore, reading comprehension questions in standardised tests are posed with limited time for thinking and using questions. It is known within the literature that it is hard to identify effects on reading strategy instruction interventions through standardised tests (Paris et al., 1984) because of the limited time allowed for thinking. Therefore, the test may not have captured the full extent of comprehension development seen in the targeted group.

One further issue with regard to measures is that the secondary outcome measures were developed in bespoke manner as no existing tests had specificity to measures these outcomes (e.g. pupil comprehension meta-cognition). However, these tests were designed with careful scrutiny of content validity and *post hoc* tests of reliability and validity were conducted. In addition, these secondary measures are mainly to explore the theory of change. Effectiveness of the programme should mainly rest on the performance of the intervention on the primary standardised outcome measures. However, the primary outcome measures which provide the main insights of this evaluation on the effects of FFT Literacy Reciprocal Reading are reliable, robust and standardised measures.

Future research and publications

It would be beneficial to explore the sustainability and size of the programme's identified effects on pupils through their national Standardised Achievement Tests (SATs) at the end of Key Stage 2. In addition, future research would be useful in terms of further exploration of FFT Literacy Reciprocal Reading programme theory by exploring the relationships between the various primary outcomes, secondary outcomes and implementation factors measured in this study (particularly between PCMC and reading comprehension). Finally, it would be useful to further explore the selection and dosage aspects of the programme. For example, is a minimum dosage required or is there a point where level of exposure plateaus in its effect? Also, is six the optimum number of pupils for selection for the targeted version or can effects be enhanced/maintained if more or fewer pupils are selected? Also, are there ways to improve the selection process by more accurately identifying pupils who are good at decoding and struggling with comprehension?

The authors intend to publish in international open-access journals and present the findings at international conferences. To do this they will conduct additional analysis on the data set.

References

Armstrong, R.A., 2014. When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, 34(5), 502–508.

Biggart, A., Kerr, K., O'Hare, L. and Connolly, P., 2013. A randomised control trial evaluation of a literacy after-school programme for struggling beginning readers. *International Journal of Educational Research*, 62, 129–140.

Borman, G., Slavin, R.E., Cheung, A., Chamberlain, A., Madden, N.A., and Chambers, B. 2007. Final reading outcomes of the national randomized field trial of Success for All. *American Educational Research Journal*, 44(3), 701–731.

Braun, V. and Clarke, V. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3, 77–101.

Brooks, G. 2007. *What works for children with literacy difficulties? The effectiveness of intervention schemes*. London: Department for Children, Schools and Families.

Chiu, C. W. T. 1998. Synthesizing metacognitive interventions: What training characteristics can improve reading performance? Paper presented at the annual meeting of the American Educational Research Association, San Diego.

Connolly, P., Biggart, A., Miller, S., O'Hare, L., and Thurston, A. 2017. *Using randomised controlled trials in education*. SAGE.

Cox, D.R. and Reid, N., 2000. *The theory of the design of experiments*. CRC Press.

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

Department for Education 2013. English programmes of study: Key Stages 1 and 2 National curriculum in England. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads

Department for Education 2018. National curriculum assessments at Key Stage 2 in England, 2018 (provisional). Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment

Duffy, G.G. 1993. Teachers' progress toward becoming expert strategy teachers. *The Elementary School Journal*, 94(2), 109–120 Retrieved from: http://www.jstor.org/stable/1001963.

Durlak, J.A. 2015. Studying program implementation is not easy but it is essential. *Prevention Science*, 16(8), 1123–1127.

Durlak, J.A., and DuPre, E.P. 2008. Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41, 327–350.

Education Endowment Foundation 2018a. Completed Projects. Retrieved from: https://educationendowmentfoundation.org.uk/projects-and-evaluation/reports/

Education Endowment Foundation 2018b. Reading comprehension strategies. Retrieved from https://educationendowmentfoundation.org.uk/pdf/generate/?u = https://educationendowmentfoundation.org.uk/pdf/toolkit/?id = 160&t = Teaching%20and%20Learning%20Toolkit&e = 160&s =

Education Endowment Foundation 2018c. Sutton Trust–Education Endowment Foundation Teaching and Learning Toolkit. London: Education Endowment Foundation. Retrieved from: https://educationendowmentfoundation.org.uk/evidence-summaries/teaching-learning-toolkit

Gorard, S., Siddiqui, N. and See, B. H. 2015. Accelerated Reader: Evaluation Report and Executive Summary. Education Endowment Foundation.

Gorard, S., Siddiqui, N. and See, B. H. 2015. Philosophy for children: Evaluation report and executive summary. Education Endowment Foundation.

Gough, P.B. and Tunmer, W.E. 1986. Decoding, reading, and reading disability. *Remedial and Special Education*, 7, 6–10.

GL Assessment 2018. Technical Information. New Group Reading Test (NGRT) Digital Edition. [online] Available at: https://www.gl-assessment.co.uk/media/294737/ngrt-technical-final-proof.pdf [Accessed 02/08/18]

Graham, J., Olchowski, A.E. and Gilreath, T. 2007. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science: The Official Journal of the Society for Prevention Research*, 8, 206–213.

Hacker, D.J., and Tenent, A. 2002. Implementing reciprocal teaching in the classroom: overcoming obstacles and making modifications. Journal of Educational Psychology, 94, 699–718.

Higgins, S., Katsipataki, M., Kokotsaki, D., Coleman, R., Major, L.E. and Coe, R., 2013. The Sutton Trust–Education Endowment Foundation Teaching and Learning Toolkit.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. 2016. Implementation and process evaluation (IPE) for interventions in education settings: An introductory handbook. *Education Endowment Foundation*, *1*.

Jay, T., Willis, B., Thomas, P., Taylor, R., Moore, N., Burnett C., Merchant, G. and Stevens, A. 2017. Dialogic teaching: talk of the town: evaluation report and executive summary. Education Endowment Foundation.

McNally, S., Ruiz-Valenzuela, J., Rolfe, H. 2016. ABRA: Online reading support: evaluation report and executive summary. Education Endowment Foundation.

McWilliams, M.A., Nier, J.A. and Singer, J.A., 2013. The implicit self and the specificity-matching principle: implicit self-concept predicts domain-specific outcomes. *Personality and Individual Differences*, 54(4), 474–478.

Nation, K. and Snowling, M. 1998. Semantic processing and the development of word-recognition skills: evidence from children with reading comprehension difficulties. *Journal of Memory and Language*, 39, 85–101

National Reading Panel. 2000. *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Washington: National Institute of Child Health and Human Development.

O'Hare, L., 2014. Did Children's Perceptions of an after-school social learning program predict change in their behavior? *Procedia – Social and Behavioral Sciences*, 116, 3786–3792.

O'Hare, L., Biggart, A., Kerr, K. and Connolly, P., 2015. A randomized controlled trial evaluation of an after-school prosocial behavior program in an area of socioeconomic disadvantage. *The Elementary School Journal*, 116(1), 1–29.

O'Hare, L., Stark, P. McConnellogue, S., Lloyd, K., Cockerill, M. Biggart, A. 2018. Protocol: A cluster randomised controlled trial of reciprocal reading: A teacher training comprehension programme. *International Journal of Educational Research*, 92, 30–42

O'Hare, L., Stark, P., McGuinness, C., Biggart, A., and Thurston, A. 2017. Spaced learning: the design, feasibility and optimisation of SMART spaces. Evaluation Report and Executive Summary. Education Endowment Foundation.

O'Hare, L., Stark, P., Orr, K., Biggart, A. and Bonnell, C., 2018. Positive action. Evaluation report and executive summary. Education Endowment Foundation.

Organisation for Economic Co-operation and Development [OECD] 2014. Profile of student performance in reading, in PISA 2012 Results: what students know and can do (Volume I, Revised edition, February 2014): Student performance in mathematics, reading and science, OECD Publishing.

Organisation for Economic Co-operation and Development [OECD] 2017*., PISA 2015 Assessment and analytical framework: science, reading, mathematic, financial literacy and collaborative problem solving*, revised edition, PISA, OECD Publishing, Paris.

Okkinga, M., van Steensel, A., van Gelderen, A. and van Schooten, E. 2018. Effectiveness of reading strategy intervention in whole classrooms: a meta-analysis. *Educational Psychology Review*, 30 (4), 1215–1239.

Palincsar, A.S. 1982. Improving the reading comprehension of junior high school students through the reciprocal teaching of comprehension-monitoring strategies. Unpublished Doctoral Dissertation, University of Illinois

Palinscar, A.S. and Brown, A.L., 1984. Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1(2), 117–175.

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

Sencibaugh, J. M. 2007. Meta-analysis of reading comprehension interventions for students with learning disabilities: Strategies and implications. *Reading Improvement*, 44(10), 6–22.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., and Torgesen, J. 2010. Improving reading comprehension in kindergarten through 3rd grade: IES Practice Guide. NCEE 2010-4038. What Works Clearinghouse. http://files.eric.ed.gov/fulltext/ED512029.pdf

Swanson, H. 1999. Reading research for students with LD: a meta-analysis of intervention outcomes. *Journal of Learning Disabilities*, 32(6), 504–532. https://doi.org/10.1177/002221949903200605

Thurston, A., Roseth, C., O'Hare, L., Davison, J. and Stark, P. 2016. Talk of the town: evaluation report and executive summary. Education Endowment Foundation.

Torgerson, C.J. and Torgerson, D.J., 2007. The use of minimization to form comparison groups in educational research. *Educational Studies*, 33(3), 333–337.

Tymms, P., Merrell, C., Thurston, A., Andor, J., Topping, K. and Miller, D., 2011. Improving attainment across a whole district: school reform through peer tutoring in a randomized controlled trial. *School Effectiveness and School Improvement*, *22*(3), 265–289.

Wason, J.M., Stecher, L., and Mander, A.P. 2014. Correcting for multiple-testing in multi-arm trials: is it necessary and is it done?. *Trials*, 15(1), 364.

Appendix A: EEF cost rating

Cost ratings are based on the approximate cost per pupil per year of implementing the intervention over three years. More information about the EEF's approach to cost evaluation can be found **here**. Cost ratings are awarded as follows:

Cost rating	Description
£ £ £ £ £	*Very low:* less than £80 per pupil per year.
£ £ £ £ £	*Low:* up to about £200 per pupil per year.
£ £ £ £ £	*Moderate:* up to about £700 per pupil per year.
£ £ £ £ £	*High:* up to £1,200 per pupil per year.
£ £ £ £ £	*Very high:* over £1,200 per pupil per year.

Appendix B: Security classification of trial findings

OUTCOME: *Overall reading and reading comprehension: universal intervention*

Please use this template to assign a separate security rating for each primary outcome.

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	4				
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **[-1]**		3
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	High	Confounding should be controlled by the evaluation method (RCT). However: -The Balance on observable characteristics was only assessed descriptively, but not statistically tested (although later controlled by the model). For instance, it seems to be a difference in the quality rating of schools across the control and treatment group. Even if pre-treatment characteristics were tested, there is no recognition that this does not rule out imbalances in unobservable characteristics. - Sensitivity analysis was only carried out in terms of missingness, but not pre-intervention imbalance. -Analysis was performed by the evaluation team who had also performed recruitment and who were not blind to condition.
Threat 2: Concurrent Interventions	Low	No relevant treats.
Threat 3: Experimental effects	Low	Contamination from one intervention (e.g. targeted) may have changed practice in the classes receiving the other intervention (e.g. universal).
Threat 4: Implementation fidelity	Low	-Pupils in the universal intervention sometimes struggled to access the programme. - Some teachers reported excluding pupils from the universal intervention, which may have undermined impact of this intervention
Threat 5: Missing Data	Moderate	Missingness was moderate (13%). Nevertheless, sensitivity analysis shows no changes in evaluation results.
Threat 6: Measurement of Outcomes	Low	All of the histograms show normal distribution of the primary outcomes except for the universal overall reading which shows a slight negative skew, i.e. more pupils were scoring towards the higher half of possible scores.
Threat 7: Selective reporting	Low	Imbalance may have an impact on the universal intervention results (size and direction). The report could do more to highlight the high dosage of the intervention

- **Initial padlock score:** 4 Padlocks – Random Control Trial, small MDE (<0.2) and 13.5% attrition at pupil level.
- **Reason for adjustment for threats to validity**: 1 Padlocks – Concerns about imbalance and missingness.
- **Final padlock score:** initial score adjusted for threats to validity = 3 Padlocks

OUTCOME: *Overall reading and reading comprehension: targeted intervention*

Please use this template to assign a separate security rating for each primary outcome.

Rating	Criteria for rating			Initial score		Adjust		Final score
	Design	**MDES**	**Attrition**					
5	Randomised design	<= 0.2	0-10%					
4	Design for comparison that considers some type of selection on unobservable characteristics (e.g. RDD, Diff-in-Diffs, Matched Diff-in-Diffs)	0.21 - 0.29	11-20%	4				
3	Design for comparison that considers selection on all relevant observable confounders (e.g. Matching or Regression Analysis with variables descriptive of the selection mechanism)	0.30 - 0.39	21-30%			Adjustment for threats to internal validity **[-1]**		3
2	Design for comparison that considers selection only on some relevant confounders	0.40 - 0.49	31-40%					
1	Design for comparison that does not consider selection on any relevant confounders	0.50 - 0.59	41-50%					
0	No comparator	>=0.6	>50%					

Threats to validity	Threat to internal validity?	Comments
Threat 1: Confounding	High	Confounding should be controlled by the evaluation method (RCT). However: -The Balance on observable characteristics was only assessed descriptively, but not statistically tested (although later controlled by the model). For instance, there is a difference in the quality rating of schools across the control and treatment group. Even if pre-treatment characteristics were tested, this does not rule out imbalances in unobservable characteristics. - Sensitivity analysis was only carried out in terms of missingness, but not pre-intervention imbalance. -Analysis was performed by the evaluation team who had also performed recruitment and who were not blind to condition. -The difference between pre-intervention outcomes was >0.1 SD (Targeted intervention)
Threat 2: Concurrent Interventions	Low	No relevant treats.
Threat 3: Experimental effects	Low	Contamination from one intervention (e.g. universal) may have changed practice in the classes receiving the other intervention (e.g. targeted)
Threat 4: Implementation fidelity	Low	No relevant threats.
Threat 5: Missing Data	Moderate	Missingness was moderate (14%). Nevertheless, sensitivity analysis shows no changes in evaluation results.
Threat 6: Measurement of Outcomes	Low	No relevant treats.
Threat 7: Selective reporting	Low	While it is unlikely that the imbalance affects the direction of the results for the Targeted intervention, it may have an impact on the size of the effect of it. The report could do more to highlight the high dosage of the intervention.

- **Initial padlock score:** 4 Padlocks – Random Control Trial, small MDE (<0.2) and 15% attrition at pupil level.
- **Reason for adjustment for threats to validity**: 1 Padlocks – Concerns about imbalance and missingness.
- **Final padlock score:** initial score adjusted for threats to validity = 3 Padlocks

Appendix C: Additional Analyses

Exploratory Analyses for universal intervention

Table 35: Exploratory analysis for effect of teacher comprehension importance, teacher comprehension behaviour and school comprehension ethos on pupil reading comprehension meta-cognition – universal intervention

Pupil Comprehension Meta-Cognition	Coef.	Std. Err.	z	Significance	95% CI	
Group	1.73	0.40	4.30	<0.001	0.94	2.52
Teacher Comprehension Importance	–0.04	0.06	–0.73	0.46	–0.16	0.07
Teacher Comprehension Behaviour	–0.06	0.07	–0.89	0.37	–0.19	0.07
School Comprehension Ethos	0.03	0.05	0.55	0.58	–0.07	0.12
_cons	21.40	3.99	5.37	<0.001	13.58	29.21

No significant effects were found of teacher comprehension importance, teacher comprehension behaviour or school comprehension ethos on pupil reading comprehension meta-cognition for universal intervention.

Table 36: Exploratory analysis for effect of pupil reading comprehension meta-cognition on reading comprehension at post-test: universal intervention

NGRT Reading Comprehension	Coef.	Std. Err.	z	Significance	95% CI	
Group	0.84	4.58	0.18	0.86	–8.14	9.82
Pupil comprehension meta-cognition	0.11	0.24	0.46	0.64	–0.36	0.58
constant	289.46	5.44	53.23	<0.001	278.80	300.12

No significant effect of pupil reading comprehension meta-cognition was found on reading comprehension at post-test for the universal intervention.

Table 37: Exploratory analysis for effect of pre-test comprehension on post-test overall reading: universal intervention

NGRT Overall Reading	Coef.	Std. Err.	z	Significance	95% CI	
Group	–0.83	2.62	–0.32	0.75	–5.97	4.30
Pre-test NGRT Reading Comprehension	32.14	0.74	43.72	<0.001	30.70	33.58
constant	293.28	1.88	156.30	<0.001	289.61	296.96

There was a significant effect of pre-test reading comprehension on overall reading at post-test for the universal intervention.

Results Table 38: Exploratory analysis for effect of pre-test reading accuracy on overall reading at post-test: universal intervention

NGRT Overall Reading	Coef.	Std. Err.	z	Significance	95% CI	
Group	1.43	2.83	0.51	0.61	–4.12	6.98
Pre-test NGRT Reading Accuracy	38.30	0.76	50.42	<0.001	36.81	39.79
constant	287.32	2.02	142.10	<0.001	283.36	291.28

There was a significant effect of pre-test reading accuracy on overall reading at post-test for the universal intervention.

Exploratory Analyses for targeted intervention

Results Table 39: Exploratory analysis for effect of teacher comprehension importance, teacher comprehension behaviour and school comprehension ethos on pupil reading comprehension meta-cognition – targeted intervention

Pupil comprehension meta-cognition	Coef.	Std. Err.	z	Significance	95% CI	
Group	1.15	0.49	2.35	0.02	0.19	2.11
Teacher Comprehension Importance	–0.14	0.07	–2.02	0.04	–0.29	0.00
Teacher comprehension Behaviour	0.00	0.08	–0.02	0.98	–0.16	0.15
School Comprehension Ethos	0.03	0.06	0.56	0.58	–0.08	0.15
constant	24.25	4.96	4.89	<0.001	14.53	33.97

A significant negative effect of teacher comprehension importance was found on pupil reading comprehension meta-cognition for the targeted intervention.

Results Table 40: Exploratory analysis for effect of pupil reading comprehension meta-cognition on reading comprehension at post-test: targeted intervention

NGRT Reading Comprehension Coef.	Std. Err.	z	Significance	95% CI	Interval]
Group	21.45	4.73	4.53	<0.001	12.17	30.74
Pupil comprehension meta-cognition	0.16	0.33	0.50	0.62	–0.48	0.80
constant	285.64	6.77	42.22	<0.001	272.38	298.91

No significant effect of pupil reading comprehension meta-cognition was found on reading comprehension at post-test for the targeted intervention.

Table 41: Exploratory analysis for effect of pre-test comprehension on post-test overall reading: universal intervention

NGRT Overall Reading	Coef.	Std. Err.	z	Significance	95% CI	
Group	13.02	3.85	3.38	0.00	5.48	20.57
Pre-test NGRT Reading Comprehension	27.32	1.28	21.36	<0.001	24.81	29.83
constant	285.43	2.77	103.05	<0.001	280.00	290.86

There was a significant effect of pre-test reading comprehension on overall reading at post-test for the targeted intervention.

Table 42: Exploratory analysis for effect of pre-test reading accuracy on overall reading at post-test

NGRT Overall Reading	Coef.	Std. Err.	z	Significance	95% CI	
Group	13.56	4.37	3.10	0.00	4.99	22.12
Pre-test NGRT Reading Accuracy	26.87	1.56	17.24	<0.001	23.82	29.93
constant	281.77	3.16	89.20	<0.001	275.58	287.96

There was a significant effect of pre-test reading accuracy on overall reading at post-test for the targeted intervention.

Analyses of implementation factors for universal intervention

Table 43: Analysis of effect of pupil programme engagement, teacher programme engagement and dosage on pupil reading comprehension meta-cognition for universal intervention

Pupil comprehension meta-cognition	Coef.	Std. Err.	z	Significance	95% CI	
Pupil programme engagement	0.49	0.03	16.74	<0.001	0.44	0.55
Teacher programme engagement	–0.07	0.06	–1.14	0.26	–0.18	0.05
Universal total dosage	0.00	0.00	–1.34	0.18	0.00	0.00
constant	11.20	2.95	3.80	<0.001	5.42	16.97

A significant effect of pupil programme engagement on pupil reading comprehension behaviour was found for the universal intervention.

Table 44: Analysis of effect of pupil programme engagement, teacher programme engagement and dosage on reading comprehension at post-test for universal intervention

NGRT Reading Comprehension	Coef.	Std. Err.	z	Significance	95% CI	
Pupil programme engagement	0.16	0.37	0.43	0.67	–0.57	0.89
Teacher programme engagement	0.86	0.81	1.06	0.29	–0.73	2.44

Universal total dosage	0.00	0.00	0.17	0.86	–0.01	0.01
constant	247.91	40.64	6.10	<0.001	168.26	327.56

No significant effects of pupil engagement, teacher programme engagement or dosage on reading comprehension at post-test were found for the universal intervention.

Table 45: Analysis of effect of pupil programme engagement, teacher programme engagement and dosage on overall reading at post-test for universal intervention

NGRT Overall Reading	Coef.	Std. Err.	z	Significance	95% CI	
Pupil programme engagement	0.39	0.38	1.03	0.31	–0.36	1.14
Teacher programme engagement	0.73	0.84	0.87	0.38	–0.91	2.37
Universal total dosage	0.00	0.00	0.30	0.77	–0.01	0.01
constant	243.07	41.93	5.80	<0.001	160.88	325.25

No significant effects of pupil engagement, teacher programme engagement or dosage on overall reading at post-test were found for the universal intervention.

Analyses of Implementation factors for targeted intervention

Table 46: Analysis of effect of pupil programme engagement, teacher programme engagement and dosage on pupil reading strategy comprehension meta-cognition for targeted intervention

Pupil comprehension meta-cognition	Coef.	Std. Err.	z	Significance	95% CI	
Pupil programme engagement	0.32	0.05	6.96	<0.001	0.23	0.41
Teacher programme engagement	0.01	0.03	0.33	0.74	–0.05	0.07
Targeted total dosage	0.00	0.00	–1.43	0.15	0.00	0.00
constant	12.07	2.04	5.92	<0.001	8.07	16.07

A significant effect of pupil engagement on pupil reading comprehension meta-cognition was found for the targeted intervention.

Table 47: Analysis of effect of pupil programme engagement, teacher programme engagement and dosage on reading comprehension at post-test for targeted intervention

NGRT Reading Comprehension	Coef.	Std. Err.	z	Significance	95% CI	
Pupil Engagement	0.10	0.49	0.21	0.84	–0.86	1.07
Teacher programme engagement	0.31	0.39	0.79	0.43	–0.46	1.08

Targeted total dosage	0.00	0.00	0.98	0.33	0.00	0.01
constant	283.73	23.53	12.06	<0.001	237.62	329.85

No significant effects of pupil engagement, teacher programme engagement or dosage on reading comprehension at post-test were found for the targeted intervention.

Table 48: Analysis of effect of pupil programme engagement, teacher programme engagement and dosage on overall reading at post-test for targeted intervention

NGRT Overall Reading	Coef.	Std. Err.	z	Significance	95% CI	
Pupil programme engagement	–0.08	0.40	–0.21	0.83	–0.86	0.70
Teacher programme engagement	0.31	0.29	1.04	0.30	–0.27	0.88
Targeted total dosage	0.00	0.00	1.12	0.27	0.00	0.01
constant	291.49	18.04	16.16	<0.001	256.14	326.84

No significant effects of pupil engagement, teacher programme engagement or dosage on overall reading at post-test were found for the targeted intervention.

Appendix D: Guide for identifying suitable pupils for targeted intervention of FFT Literacy Reciprocal Reading[15]

Identifying good decoders but poor comprehenders

The purpose of this task

This task is for identifying children who are good at reading accurately, but less good at understanding what they read. This information will be used as part of the research.

Instructions for carrying out this task

Step 1. Please read all the information on page 2 about the characteristics of pupils who fall into this group. You will then select six pupils from your class who fit this profile using steps 2-5.

Step 2. Review the reading decoding skills of children in your class using the checklist on page 2 section A.

Step 3. Review the reading comprehension skills of the children in your class using the checklist on page 2 Section B.

Step 4. Now Identify **six children** in your class who read accurately, but less good at understanding what they read.. These are the children who are successfully decoding (page 2 Section section A) but who have limited reading comprehension skills (page 2 section Bi) and/or don't use or misapply important comprehension strategies (page 2 section Bii below).

Step 5. All teachers when you have selected the six names, email them to **Patrick Stark**

Children who can decode but struggle to understand, can read age appropriate texts aloud but often fail to take much sense from them, particularly struggling to understand anything which is not explicitly stated in the text.

Section A. Reading decoding

When reading an age appropriate text, children who read accurately can usually:

- Read at 90 – 95% accuracy on a text at their level
- Apply phonic knowledge and appropriate strategies to reading unfamiliar words
- Decode accurately without overt sounding out
- Read aloud at a reasonable speed and fluency
- Pay attention to punctuation marks to break reading into meaningful chunks
- Read with some appropriate expression e.g. when reading dialogue
- Read independently

Section B. Reading comprehension

When reading an age appropriate text, children who read accurately but struggle to understand generally::

- Retrieve information which is directly or literally stated in the text
- Recount events or give items of information directly stated in the text
- Make use of strategies such as reading picture clues and sounding out when trying to make sense of text

However, these children often don't use or misapply important comprehension strategies:

- They may not link what they've read to previous experience or knowledge of the world, nor to their previous reading
- When using prior knowledge, they often over-generalise from own experience rather than linking experience to what is in the text
- They find it difficult to identify main points or key ideas in a text if asked to summarise what has been read
- They find it difficult to infer from the hints and clues that an author uses to imply rather than state
- They rarely link new information to what they have read previously in the text
- They don't identify how a text may be similar to others
- They tend to ignore the meaning of unfamiliar words even when their disrupt understanding
- They rarely notice when they don't understand and do not apply 'fix it' strategies to try and make sense of their reading (e.g. they do not re-read, or stop to consider what they have misunderstood etc.)

Appendix E: Ethics Documentation

Reciprocal Reading Randomised Controlled Trial Study

MEMORANDUM OF UNDERSTANDING

Aims of the evaluation
The aim of this project is to evaluate the impact of reciprocal reading on children's attainment in reading and comprehension.

The project
Two versions of the reciprocal reading teacher training comprehension programme are to be tested to see if they have an impact on reading and comprehension. The first version of the programme is designed to train teachers and teaching assistants of Year 4 pupils to provide whole class instruction on comprehension. The second is designed to train teachers and teaching assistants to work with smaller targeted groups of Years 5 and 6 with lower comprehension skills, particularly those with a disparity between their ability to decode and to comprehend. The impact of this tutoring programme will be evaluated and compared with the 'business as usual' control group, i.e. usual teaching, using a randomised controlled trial (RCT). During this project, you will be contacted by both researchers from Queen's University Belfast and by the reciprocal reading delivery team from Fisher Family Trust, hereafter referred to as QUB and FFT Literacy respectively.

This memorandum of understanding (MoU) explains what your school's participation in the study will entail. If you agree to take part and accept the terms and conditions outlined, please sign a copy of this form, complete the questions at the end and return by email to maria.cockerill@qub.ac.uk.

Structure of the evaluation
As this is a randomised controlled trial schools will be randomly assigned to one of two school groups for the duration of the whole project:

1. *Reciprocal Reading (Intervention group):* Schools in the intervention group will deliver the reciprocal reading programme in Year 4 (whole class), and Year 5 and 6 (targeted programme version) in 2017/18.
2. *'Business as usual' (Control group): Schools in the control group will continue with usual teaching during academic year 2017–18 and not access the reciprocal reading programme from FFT Literacy. Furthermore, control schools will not access the reciprocal reading programme from FFT Literacy for universal or targeted teaching for the Year 4, 5 and 6 pupils in the current study until the final trial results are published.*

1. Reciprocal Reading (Intervention group): As indicated above all Year 4 pupils in these schools will receive reciprocal reading as a whole class approach. Only those Year 5 and 6 pupils who are identified as requiring reciprocal reading will receive it as a targeted programme.
Schools who are allocated to deliver reciprocal reading in academic year 2017/2018 will contribute £500 (25% of the normal cost of the programme at £2000), which includes two days training for up to 5 school staff, programme manuals and in-school support from FFT

Literacy. In addition, these schools will receive book resources from FFT Literacy worth £200.

School responsibilities:

a. All Year 3 pupils will complete a standardised reading pre-test in June 2017 (GL Assessment New Group Reading Test).
b. All Year 4 and 5 pupils will be assessed by their class teacher on their comprehension and decoding skills in Apr/May 2017. These scores will be used to select a group of 20 pupils across both year groups who will also complete a standardised reading pre-test in June 2017 (GL Assessment New Group Reading Test).
c. All pupils who completed a pre-test (as outlined in b.) will also be required to complete a post-test (GL Assessment New Group Reading Test in June 2018).
d. All teachers and teaching assistants will complete questionnaires after their training sessions with FFT Literacy and a survey at the end of the project (June 2018).
e. Some schools will be asked to become case-study schools and participate in classroom observations, staff interviews and focus groups.

2. 'Business as usual' (Control group): Schools in the control group are required to continue with usual teaching during academic year 2017–18.

These schools in the control group will receive £1000 following the completion of all evaluation requirements with staff/school and with the required pupils in 2017 and 2018. After the evaluation has finished, the school may purchase the reciprocal reading programme from FFT Literacy for use from September 2018.

School responsibilities:

a. All Year 3 pupils will complete a standardised reading pre-test in June 2017 (GL Assessment New Group Reading Test).
b. All year 4 and 5 pupils will be rated out of 10 by their class teacher on their comprehension and decoding skills in Apr/May 2017. These scores will be used to select a group of 20 pupils across both year groups who will also complete a standardised reading pre-test in June 2017 (GL Assessment New Group Reading Test).
c. All teachers and teaching assistants will complete a survey at the end of the project (June 2018).

Random allocation is essential to the evaluation as it is the best way of investigating what effect reciprocal reading had on pupils' reading. It is important that schools understand and consent[16] to this process.

The Evaluation Team will use school and pupil information provided by schools including KS2 results, and information from the Nation Pupil Database to assess any impact of reciprocal reading on attainment.

Use of Data

All pupil data will be treated with the strictest confidence and will be stored in accordance with the Data Protection Act (1998). Named data will be matched with the National Pupil

[16] The legal bases for processing personal and special data changed during the course of this study. The recruitment documents were completed pre-GDPR, when the evaluation team relied on opt-out consent but subsequently the legal bases for processing personal and special data was processed in the public interest.

Database and shared with the Evaluation Team: QUB, the Department for Education, the Education Endowment Foundation (EEF), FFT Education and in an anonymised form to the UK Data Archive.
All results will be anonymised so that no schools or individual pupils will be identified in any report arising from the research.

Responsibilities of FFT Literacy Reciprocal Reading Project Team:

- To provide two days training for up to 5 staff from every school in the intervention group
- Provide three half day on-going support sessions to the intervention schools
- To work closely with the Evaluation Team

Responsibilities of the Evaluation Team from QUB:

- Act as the first point of contact for any questions about the evaluation
- Conduct the random allocation of schools to either the intervention group or control group
- Provide information sheets and opt-out consent[17] forms for schools to send to parents
- Collect class and pupil level data (including name, date of birth, UPN)
- Request NPD data using pupil details
- Analyse the data from the project
- Disseminate the research findings
- To work closely with FFT Literacy Reciprocal Reading Project Team

Requirement for Schools

- The school will not participate in another EEF literacy randomised trial that would interfere with implementation of the intervention with Year 4, 5 and 6 pupils during 2017/18 academic year.
- The school will provide the Evaluation team with pupil and school data requested (see above) and will ensure all pupils participating in the trial complete all required standardised tests.
- Schools will deliver letters to parents giving them information about the study and an opportunity to opt their child out of the data gathering process. They will inform the Evaluation Team of any responses arising, and permit the publication of anonymised data collected.
- The school agrees to the Evaluation Team obtaining the evaluation cohorts' data from the National Pupil Database, and will provide the UPNs to enable this to be achieved.
- Teachers will, at the earliest opportunity, notify the Evaluation Team if there are any issues which could prevent the effective implementation of reciprocal reading.
- **If the school has to withdraw from the project for operational or other unavoidable reasons, it will notify the Evaluation Team (QUB) straight away and wherever possible still provide test data for the evaluation.**

[17] The legal bases for processing personal and special data changed during the course of this study. The recruitment documents were completed pre-GDPR, when the evaluation team relied on opt-out consent but subsequently the legal bases for processing personal and special data was processed in the public interest.

Head teacher agreement

I agree for my school to take part in the reciprocal reading study and I accept the eligibility terms and conditions.

School Name:

_____

Head Teacher Name:

_____

Head Teacher Signature: _____ Date: ____/____/_____

Head Teacher Email Address:

_____

School Contact (if not Head Teacher):

_____

School Contact Email Address (if not Head Teacher):

_____

School Telephone Number:

_____

Thank you for agreeing to take part in this research.

Please complete the information below and return this form to: maria.cockerill@qub.ac.uk

Parent Information Letter and Data Processing Withdrawal Form[18]

Dear Parent / Carer,

Your child's school is currently involved in a research project which is evaluating the reciprocal reading programme. This programme aims to improve reading and is a partnership between your child's school and Fischer Family Trust. The programme is being evaluated by a team from the Centre for Evidence and Social Innovation at Queen's University Belfast, Northern Ireland.

As part of measuring the success of this reading comprehension programme, your child took part in a reading test last summer (2017). This reading test is currently being repeated in schools at the moment and your child may already have taken part in this. The last letter we sent at the beginning of the project in June 2017 gave you the option to withdraw your child's reading test data from the data processing for this project.

The last letter you were sent about this programme also informed you that we would be accessing your child's Key Stage scores from the National Pupil Database (held by the Department for Education, part of the UK government) and from your child's school, and gave you the option to withdraw your child's data from this data processing.

In addition to these scores, we will now be accessing other information about pupils taking part in this project from your child's school and the National Pupil Database. This includes data about SEN statements, whether or not your child has received Free School Meals in the last 6 years and their UPN (Unique Pupil Number) to allow us to retrieve information from the National Pupil Database.

We will be sharing this data with the Department for Education, the Education Endowment Foundation (EEF, who funded the trial), EEF's data contractor Fischer Family Trust Education and storing the data in an anonymised form in the UK Data Archive.

This research has been reviewed and approved by the research ethics committee of Queen's University Belfast. The head teacher of your child's school agreed for the school to take part in the research programme.

If you have any questions please don't hesitate to contact Patrick Stark, 6 College Green, School of Education, Queen's University Belfast, Belfast. Telephone: 028 9097 5294, email: p.stark@qub.ac.uk.

If you have any other concerns about the conduct of the research then contact Liam O'Hare at Queen's University Belfast. Telephone: 028 9097 5973, email: l.ohare@qub.ac.uk

Because we are doing this research to improve understanding about what works in improving pupils' education, **if you are happy for information about your child to be used in the reciprocal reading research project you do not need to do anything.** Thank you for your help with this research, your support is much appreciated.

If you **DO NOT** want the above information from your child's school and the National Pupil Database about your child (Key Stage scores, Free School Meals, SEN and UPN) to be used to understand whether the reciprocal reading programme can help reading comprehension, please complete the enclosed form and return it to your child's school by 16/7/18 If you do this, then this information about your child will be withdrawn from the project and will not be shared with anyone.

[18] This was delivered to schools in June 2018 in preparation for the NPD application being submitted in the new GDPR system.

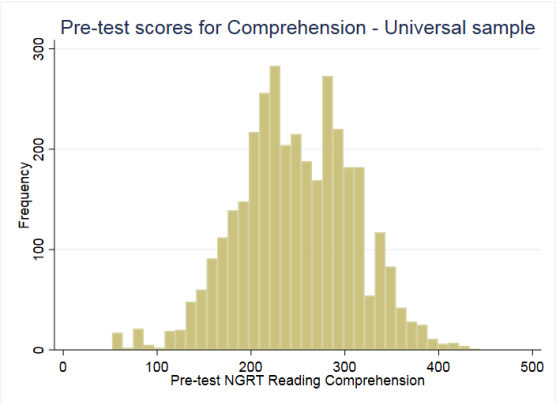
Reciprocal Reading Research Programme
(If you are happy for your child's Free School Meals, Key Stage Scores, Special Educational Needs and UPN data to be used in this project you DO NOT need to return this form.)
I **DO NOT** wish Free School Meals, Key Stage Scores, Special Educational Needs and UPN data about my child to be used as part of this research.

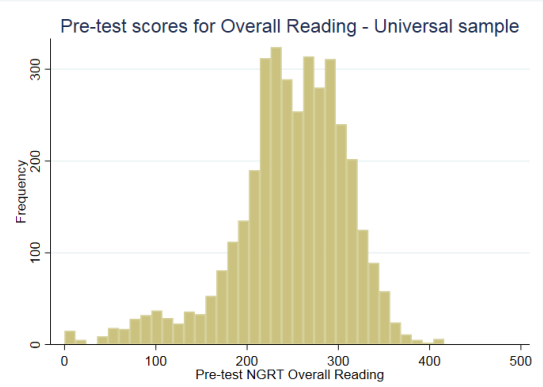
Child's name: ……Date of birth: ……
Child's class Teacher: ……..
School: ……

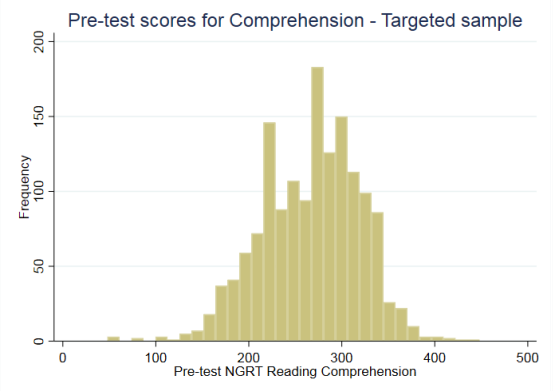
Parent name (BLOCK CAPITALS) ……
Parent signature: ……
Date ……
(Please return the completed form to your child's teacher.)

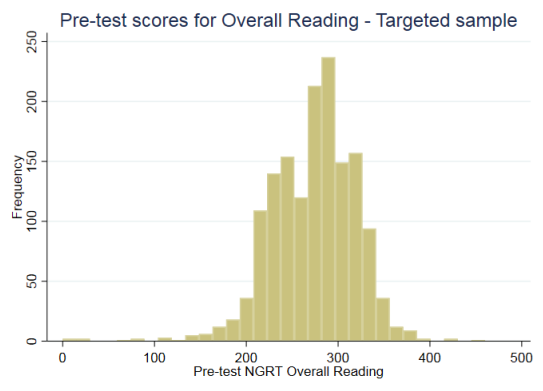
Appendix F: Pre-test and post-test distributions for Primary outcomes – universal and targeted samples

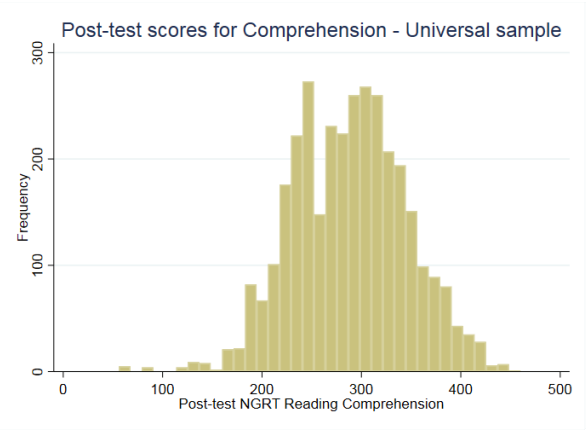
Distributions for pre-test and post-test scores for both primary outcomes are presented below. The left-hand tail extends to a zero score for the Overall Reading outcome at both pre-test and post-test. These outliers received a score of zero. Therefore, even with the adaptive nature of the digital NGRT, some pupils still scored at zero. This means that even after regressing to a phonics test instead of progressing to a reading comprehension test, some pupils still scored zero. These pupils showed extremely weak reading accuracy and are therefore not the intended pupils for FFT Literacy Reciprocal Reading, as the programme is described as being dependent on already having good reading accuracy by the programme developers. The digital NGRT remains an appropriate test for measuring the impact of FFT Literacy Reciprocal Reading.

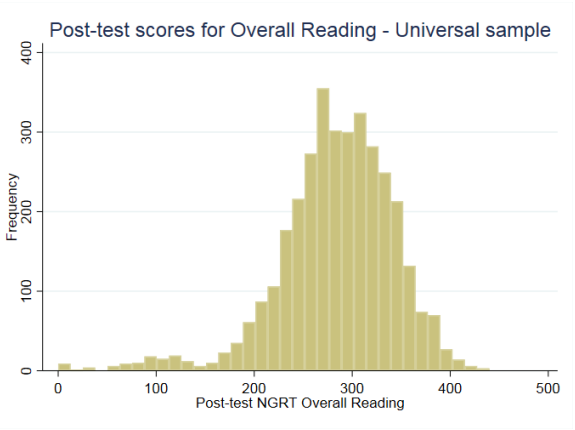


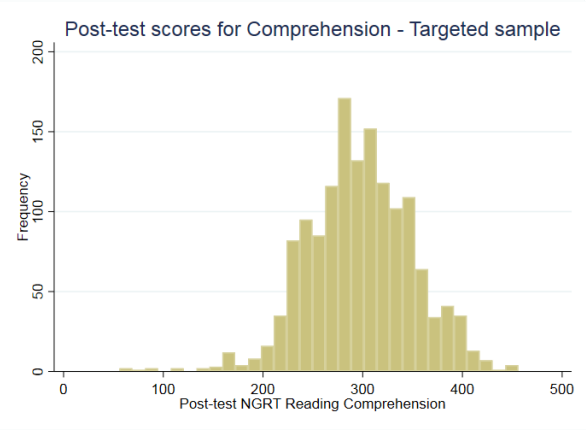


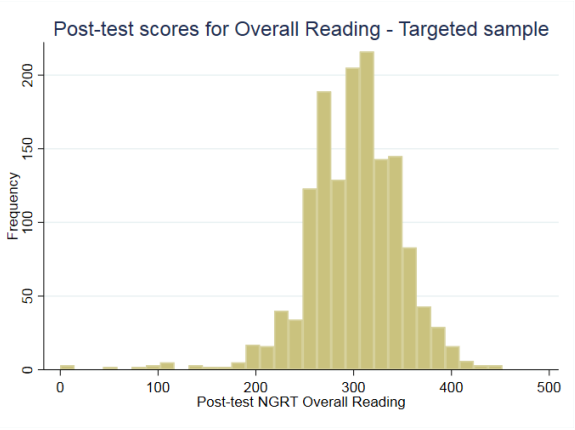





Post-test scores for Comprehension - Universal sample
Frequency
0
100
200
300
0
100
200
300
400
500
Post-test NGRT Reading Comprehension


Post-test scores for Overall Reading - Universal sample
Frequency
0
100
200
300
400
0
100
200
300
400
500
Post-test NGRT Overall Reading


Post-test scores for Comprehension - Targeted sample
Frequency
0
50
100
150
200
0
100
200
300
400
500
Post-test NGRT Reading Comprehension


Post-test scores for Overall Reading - Targeted sample
Frequency
0
50
100
150
200
0
100
200
300
400
500
Post-test NGRT Overall Reading

Appendix G: Dosage scores per intervention school for Universal and targeted Intervention

Intervention	Number of schools who reported dosage (School N)	Minimum reported by intervention schools	Maximum reported by intervention schools	Mean of reported values	SD of reported values
Total universal dosage as reported by teachers (minutes)	32	340.00	4829.95	1767.40	926.87.8
Total targeted dosage as reported by teachers (minutes)	32	509.96	4829.95	1707.40	855.54
Total number of weeks the universal intervention was delivered in intervention schools	32	17	34	26.41	5.06
Mean delivery of the universal Intervention per week (minutes)	32	20	167	66.21	30.13
Total number of weeks the targeted intervention was delivered in intervention schools	32	11	34	26.03	6.31
Mean delivery of the targeted intervention per week (minutes)	32	30	167	65.77	28.16

Total weeks of delivery of the interventions in a school were reported through the teacher survey at post-test. Universal dosage total and targeted dosage total were calculated for each school by multiplying the total weeks of delivery in that school by the mean weekly dosage (minutes) as reported in the survey.

Universal groups received, on average, 60 minutes more of the intervention than targeted groups in terms of total dosage. This does not mean that the universal dosage was non-compliant, as it is the dosage within an intervention that is important, not the relationship between the two interventions within a school.

Appendix H: Multiple imputation

Multiple imputation was carried out to allow the data to be analysed with imputed data and compare the results with the complete case analysis. The imputation model included all variables used in the primary analysis. The imputation was performed using chained iterations to fill in missing values in multiple variables using univariate imputation with fully conditional specification of prediction equations. Twenty imputations were carried out in order to lessen the risk of the 'Monte Carlo' error (simulation error). A total of two hundred iterations (with a burn-in of 10) were carried out, and estimates were combined using Rubin's pooling rules. As shown in Appendix H, both primary outcomes showed a non-significant result for the Universal intervention when using imputation. This is the same overall pattern of results as seen in the primary analysis for the universal intervention using complete cases. Both primary outcomes showed a significant result in the multi-level models for the targeted intervention. This is the same overall pattern of results as seen in the primary analysis for the targeted intervention using complete cases. This suggests that the pattern of results is highly likely to have been the same had there not been missing data.

More missing data at pre-test was found for the targeted sample than for the universal sample. It may have been the case that the small numbers required for the targeted sample in each school meant that schools did not prioritise testing for these children, in comparison with needing to organise whole-class testing for the universal sample. This difference may have been exacerbated because it was at the beginning of the trial and schools had not developed a relationship with either the evaluation team or the programme developers so engagement/commitment to the research was lower than at the end of the trial.

Missing data for each primary outcome scale for Universal and targeted samples

Sample	Outcome scale	% Missing
Universal	Pre-test overall reading	6.33%
	Post-test overall reading	12.68%
	Pre-test comprehension	12.61%
	Post-test comprehension	16.03%
Targeted	Pre-test overall reading	10.09%
	Post-test overall reading	13.40%
	Pre-test comprehension	10.98%
	Post-test comprehension	14.52%

Multiple Imputation: Analysis of Overall Reading for Universal intervention using imputed data

NGRT Overall Reading	Coef.	Std. Err.	t	Significance	95% CI	[Interval]
Group	1.58	2.19	0.72	0.47	–2.72	5.88
Pre-test NGRT Overall Reading	41.42	0.71	58.02	<0.001	40.01	42.82
Constant	286.38	1.57	182.95	<0.001	283.31	289.45

Multiple Imputation Analysis of Reading Comprehension for Universal intervention using imputed data

NGRT Reading Comprehension	Coef.	Std. Err.	t	significance	95% CI	Interval]
Group	0.57	2.93	0.20	0.85	–5.17	6.32
Pre-test NGRT Reading Comprehension	34.66	0.79	44.15	<0.001	33.11	36.21
Constant	289.67	2.11	137.36	<0.001	285.53	293.80

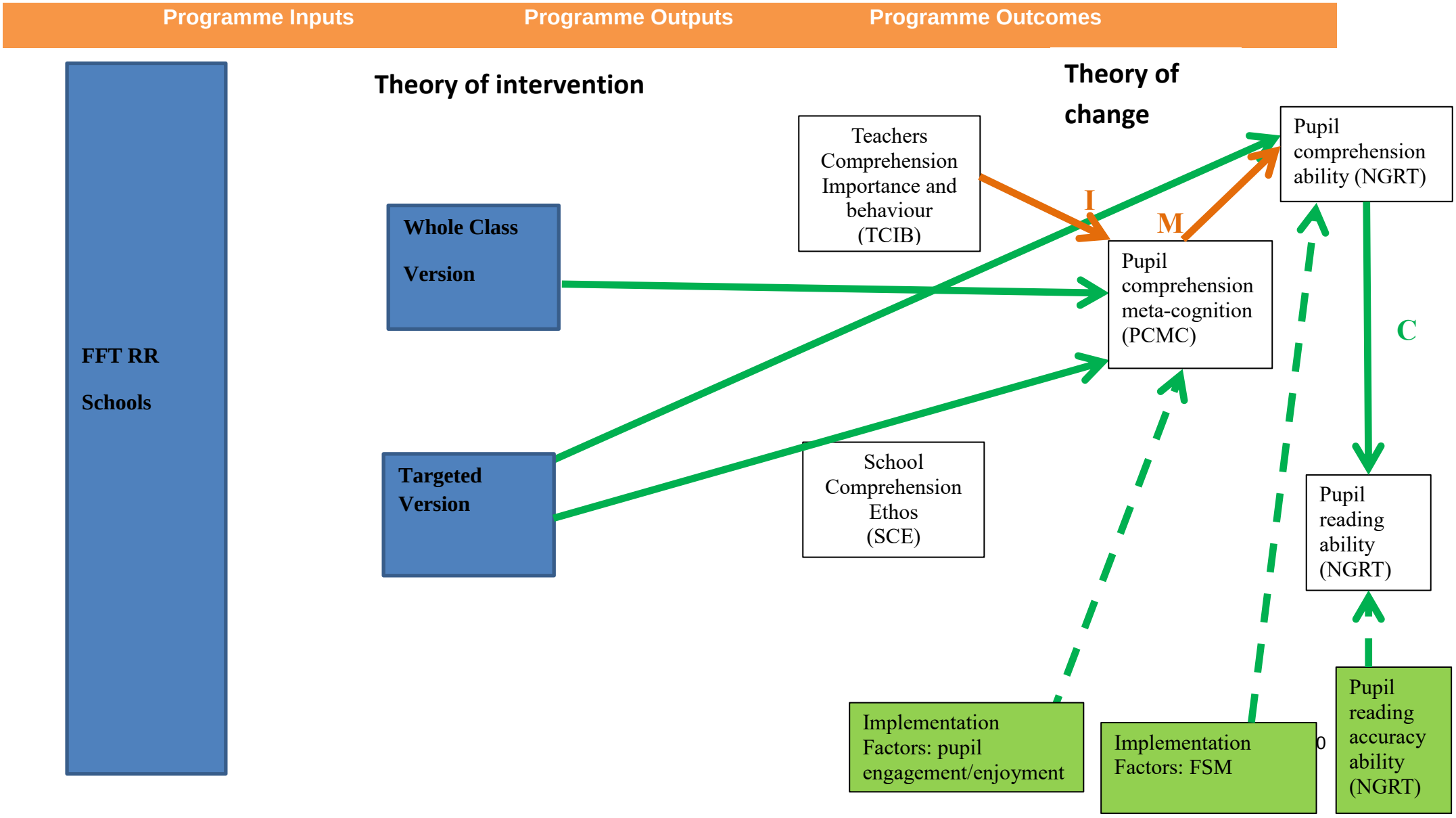
Multiple Imputation: Analysis of Overall Reading for targeted intervention using imputed data

NGRT Overall Reading	Coef.	Std. Err.	t	Significance	95% CI	[Interval]
Group	11.66	3.46	3.37	0.00	4.88	18.45
Pre-test NGRT Overall Reading	33.49	1.35	24.74	<0.001	30.84	36.15
Constant	282.70	2.53	111.66	<0.001	277.73	287.66

Multiple Imputation: Analysis of Reading Comprehension for targeted intervention using imputed data

NGRT Reading Comprehension	Coef.	Std. Err.	t	Significance	95% CI	[Interval]
Group	15.07	4.23	3.56	<0.001	6.76	23.37
Pre-test NGRT Reading Comprehension	28.23	1.39	20.24	<0.001	25.49	30.97
Constant	281.22	3.08	91.29	<0.001	275.18	287.26

Appendix I: Post evaluation version of the FFT Literacy Reciprocal Reading logic model



Appendix J: Pupil questionnaire – Pupil reading comprehension metacognition and programme engagement

Welcome to our survey and thank you very much for agreeing to take part!

1) The first questions are about you. If you don't know your school address, please ask your teacher.

* First name:
* Last name:
* School name:
* School address
* Class:

The next questions ask for your date of birth.

2) * **Day**

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17

- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31

3) * Month

- January
- February
- March
- April
- May
- June
- July
- August
- September
- October
- November
- December

4) * Year

- 2006
- 2007
- 2008
- 2009
- 2010

2011

Think about what kinds of things you can do to help you understand a story better when you read it. Read each of the statements carefully and decide which one of them would help you the most. *There are no right or wrong answers.* It's just what **you** think would help the most.

Please click on one answer for each statement.

5) Before I read a text I ask myself what I already know about the subject of the story.

Always Very often Quite often Seldom Never

6) I make some guesses about what I think will happen next in the story.

Always Very often Quite often Seldom Never

7) I try and work out the meaning of words and phrases that I am not sure about.

Always Very often Quite often Seldom Never

8) I find ways to help me understand words I am not sure about such as looking them up in a dictionary.

Always Very often Quite often Seldom Never

9) I stop my reading to ask questions to help me understand what the story is about.

Always Very often Quite often Seldom Never

10) I stop my reading to sum up what I have read to check how well I have understood it.

Always Very often Quite often Seldom Never

11) * Did you have reciprocal reading sessions? Please ask your teacher if you are not sure.

Yes No

The next questions ask you about the Reciprocal Reading sessions. Please click on one answer on each line.

12) I enjoyed the Reciprocal Reading sessions.

A lot Quite a lot A little Not at all

13) I found the Reciprocal Reading sessions interesting.

A lot Quite a lot A little Not at all

14) The things I learned in the sessions help me to understand better what I am reading.

A lot Quite a lot A little Not at all

15) The things I learned in the sessions make reading easier.

A lot Quite a lot A little Not at all

16) Reciprocal reading has made me more interested in reading.

A lot Quite a lot A little Not at all

17) I read more outside school since doing Reciprocal Reading.

A lot Quite a lot A little Not at all

18) I think we should keep doing Reciprocal Reading sessions in our class.

A lot Quite a lot A little Not at all

Last question!

When you have answered it, please don't forget to press SEND to submit your answers!

19) Overall, I am happy with Reciprocal Reading.

A lot Quite a lot A little Not at all

Appendix K: Additional analysis for the Reciprocal Reading Evaluation using NPD data

Introduction

The main report for the Reciprocal Reading evaluation detailed that updated analysis would be conducted once data had been received from the National Pupil Database (NPD) which was initially delayed by the changes in data protection legislation due to GDPR. The main report used free school meal (FSM) data supplied by schools, which included a significant amount of missing data. This appendix, therefore, includes updated figures for MDES for FSM pupils and balance at baseline for FSM (using the EverFSM 6 variable supplied by the NPD). Updated subgroup analyses for both the Universal intervention and Targeted intervention are also included. Finally, a correlation analysis between KS1 literacy data and pre-test NGRT literacy is also included.

MDES calculations

MDES calculations for FSM pupils were carried out at the analysis stage in the main report. These gave results of MDES of .2 for the Universal Intervention and .3 for the Targeted Intervention. These calculations have been repeated using the Everfsm_6 data collected from the NPD. The updated column is for FSM pupils and is in bold text. The cluster size for FSM pupils for the universal intervention was n= 15. The cluster size for FSM pupils for the targeted intervention was n=7. The pre-test to post-test correlation for FSM pupils in the universal intervention sample was .72 with ICC=.06. The pre-test to post-test correlation for FSM pupils in the targeted intervention sample was .58 with ICC=.17. The MDES calculation for FSM pupils in the universal intervention sample gave MDES=0.18. This is a smaller MDES for FSM pupils than was obtained using the FSM data from schools in the main report, which gave a value of MDES=.2. The MDES calculation for FSM pupils in the targeted intervention sample gave MDES=.28. Again, this is smaller than was obtained using the FSM data in the main report (MDES=.3).

Table 1: Minimum detectable effect size at different stages for Universal Intervention

		Protocol	Randomisation	Analysis		
		Overall	Overall	Overall	FSM pupils	
					Based on school data	Based on NPD data
MDES		.20	.14	.14	0.2	.18
Pre-test/ post-test correlations	level 1 (pupil)	.70	.85	.73	.77	.72
Intracluster correlations (ICCs)	level 2 (school)	.14	.14	.06	.04	.06
Alpha		.05	.05	.05	.05	.05
Power		.8	.8	.8	.8	0.8
One-sided or two-sided?		2	2	2	2	2
Average cluster size		20	38	34	10	15
Number of schools	intervention	47	49	47	47	47
	control	47	49	47	47	47
	total	94	98	94	94	94
Number of pupils	intervention	940	1947	1666	470	745
	control	940	1752	1532	470	705
	total	1880	3699	3198	940	1450

It should be noted that the smaller values for pupils eligible for FSM reported by schools (N=940) compared to those identified by the NPD (N=1450) may have largely been due to the amount of missing data provided by schools. The discrepancy between the NPD and school data is most likely due to a lack of schools reporting that pupils were FSM eligible and not due to inaccurate reporting of FSM eligibility by schools.

Table 2: Minimum detectable effect size (MDES) at different stages of trial for targeted intervention

		Protocol	Randomisation	Analysis		
					FSM Pupils	
		Overall	Overall	Overall	Based on school data	Based on NPD data
MDES		.30	.17	.24	.30	.28
Pre-test/ post-test correlations	level 1 (pupil)	.70	.85	.56	.58	.58
Intracluster correlations (ICCs)	level 2 (school)	.14	.14	.15	.18	.17
Alpha		.05	.05	.05	.05	.05
Power		.8	.8	.8	.8	.8
One-sided or two-sided?		2	2	2	2	2
Average cluster size		7	16	14	5	7
Number of schools	intervention	36	49	47	47	47
	control	36	49	47	47	47
	total	72	98	94	94	94
Number of pupils	intervention	252	811	678	235	342
	control	252	712	618	235	359
	total	504	1523	1296	470	701

FSM balance at baseline

Table 3 shows the percentage of pupils in each group who were coded as EverFSM 6 for the Universal Intervention. The overall levels of FSM are higher than originally estimated – the earlier values were 18.34% and 20.38% for the Intervention and Control groups respectively.

Table 3: EverFSM 6 balance at baseline in Intervention and Control groups – Universal intervention

	Intervention Group		Control Group		
Pupil-level (categorical)	N	Count (%)	N	Count (%)	
Eligible for FSM	745	36.77%	705	37.22%	

Table 4 shows the percentage of pupils in each group who were coded as EverFSM 6 for the Targeted Intervention. The overall levels of FSM are higher than originally estimated – the earlier values were 17.67% and 22.75% for the Intervention and Control groups respectively.

Table 4: EverFSM 6 balance at baseline in Intervention and Control groups – Targeted intervention

	Intervention Group		Control Group		
Pupil-level (categorical)	N	Count (%)	N	Count (%)	
Eligible for FSM	342	38.78%	359	45.73%	

Subgroup analysis for EverFSM 6 pupils

The primary analyses were repeated for only pupils in receipt of free school meals, using Ever FSM 6 data received from the NPD. These analyses were preliminarily conducted in the main report using FSM data supplied by schools. The effect sizes for each primary outcome for the two interventions have now been calculated. Raw means are reported in the analysis tables, to allow comparison of mean scores between groups. Standardised pre-test scores, which have a mean of zero, were used in the analysis as they control for variation at pre-test between groups. Missing data is reported alongside the n for each group. This missing data statistic represents the number of pupils who are missing from the post-test data from the total number of FSM pupils in that group. As detailed in the description of the NGRT testing procedure in the main report, number of pupils (N) for sub-scores (Reading Comprehension) can be lower than for Overall Reading, as not all pupils who completed the NGRT took the Reading Comprehension test (if they scored extremely low on the Reading accuracy test, they did not receive the Reading Comprehension, and sat a phonics sub-test instead).

Table 5 shows a significant effect of the universal intervention was found for Overall Reading Ability for everFSM 6 pupils (ES=.08). This means that, although there was no effect of the Universal intervention for the overall sample (see Main Report) there was a small positive effect of the universal intervention for everFSM 6 pupils in the intervention group when overall reading was assessed at post-test, in comparison with the control group. Disadvantaged pupils in this study showed a small benefit from the Universal Intervention for their overall reading. This effect was not observed in the FSM subgroup analysis in the main report, which was conducted using school supplied FSM data, prior to the NPD data collection, which had significant levels of missing data.

Table 5 shows no significant effect of the universal intervention was found for Reading Comprehension for everfsm 6 pupils.

Table 5: Primary analysis for FSM subgroup in Universal Intervention sample

		Raw means					Effect size	
Outcome	Covariate	Intervention group		Control group		N in model (intervention; control)	Hedges g [95% CI]	p
Primary outcome								
		n (missing)	Mean [SD]	n (missing)	Mean [SD]			
NGRT Overall reading ability	Pre-test NGRT Overall reading ability	618 (127)	268.42 [56.39]	593 (112)	259.96[66.23]	1120 (572; 548)	.08 [-.04,.2]	.05
NGRT Reading Comprehension	Pre-test NGRT Reading Comprehension	594 (151)	270.76 [57.06]	547 (158)	270.76[52.09]	989 (512; 477)	-.05 [-.17, .08]	.37

Table 6 shows significant positive effects of the Targeted Intervention were found for both Overall Reading (ES=0.2) and Reading Comprehension (ES=0.25) for everFSM 6 pupils. This shows that disadvantaged pupils in this study benefitted from the targeted intervention in terms of both overall

reading and reading comprehension. In the FSM subgroup analysis in the main report, only the reading comprehension effect was observed for the targeted intervention.

Table 6: Primary FSM analysis for Targeted Intervention

		Raw means					Effect size	
Outcome	Covariate	Intervention group		Control group		N in model (intervention; control)	Hedges g [95% CI]	p
Primary outcome								
		n (missing)	Mean [SD]	n (missing)	Mean [SD]			
NGRT Overall reading score	Pre-test NGRT Overall reading score	290 (52)	302.72 [47.10]	305 (54)	282.54 [49.70]	543 (280; 263)	0.2 [.03, .37]	.006
NGRT Reading Comprehension	Pre-test NGRT Reading Comprehension	286 (56)	302.37 [48.43]	299 (60)	279.68 [50.93]	529 (273; 256)	0.25 [.08, .42]	<.001

Correlation of KS1 Literacy with NGRT overall reading pre-test

The correlation of KS1 reading (points) with Overall reading at pre-test was r=.37 (n=1408) for the targeted intervention sample. The same KS1 measure correlated with overall reading at post-test was 0.39 (n=1378). This correlation analysis was not possible for the universal sample as for academic year 2015-2016, the NPD data for KS1 changes, and point scores are not available. These correlations represent a moderate (leaning towards weak) correlation of the pre-test reading measure with prior attainment data for literacy. This may represent a discrepancy between the types of literacy ability being measured in the NGRT test and the Key Stage 1 literacy test (outcome specificity matching). It may also represent a change in literacy ability between Key Stage 1 and the point of taking the pre-test measure. Either way, the Key Stage 1 measure of literacy was at most a moderate indicator of the pupils' reading ability at the beginning of the trial.

This analysis is interesting because it shows the usefulness of using KS1 as a pre-test measure compared to using a specifically matched pre-test/post-test instrument (i.e., the NGRT in this case) in an RCT evaluation. The closer pre-test instrument is matched to the post-test instrument allows for more sensitive analysis of any effects caused by the intervention in an RCT. Thus overall this analysis would suggest some caution in the use of KS1 as a pre-test measure when using NGRT as an outcome measure.

Conclusions

Only n=19 pupils were unable to be matched for everFSM 6 data with the NPD. This significantly reduced the missing data for pupil disadvantage. As a result, the subgroup analysis looks more promising, particularly for the targeted intervention.

The FSM sub-group analysis in the main report for the Universal Intervention, based on FSM status collected from schools, showed no effect for either Overall Reading or Reading Comprehension, yet this new FSM subgroup analysis, based on NPD data, shows the universal intervention had a small, but significant, effect for Overall Reading.

This new subgroup analysis suggests that the targeted intervention is beneficial for reading comprehension for disadvantaged pupils for both Overall Reading and Reading comprehension outcomes (whereas it was only the comprehension outcome in the preliminary analysis in the main report). Thus, the Targeted intervention shows effects in both the overall targeted cohort and the disadvantaged pupil subgroup of the targeted cohort for both primary outcomes.

The sample sizes for the subgroup analyses are considerably smaller than the primary analyses in the main report, and interpretation of the present results should take this into consideration. Accepting this limitation, the subgroup analyses suggest an overall pattern of FFT Reciprocal Reading showing promise as a targeted intervention: targeting both poor reading comprehenders and those in socio-economically disadvantaged circumstances. Thus the general evidence of efficacy of the Targeted version of the intervention seems robust. There was a significant improvement at post-test for both the sample as a whole and for the FSM subgroup, for both primary outcomes. The Universal intervention showed promise for the FSM subgroup. This may suggest that when delivered to a group with either form of disadvantage, be it reading comprehension or socio-economic disadvantage, the FFT Reciprocal Reading programme can improve reading outcomes.

These findings suggest that socio-economic disadvantage could be considered in the selection process for the targeted intervention. In addition to selecting pupils based on teacher's perceptions of pupils' poor reading comprehension ability, teachers could also use EverFSM as a variable for selecting pupils to receive the targeted intervention. Inclusion of this additional variable could improve the reliability of the targeting process as pupils selected solely on teacher perceptions of pupil reading accuracy and comprehension is open to teacher perception bias or error. However as already concluded in the main report, further research is required to optimise the selection process for the Targeted version of Reciprocal Reading, and these findings suggest that the EverFSM variable should also be considered in these investigations.

This additional analysis has highlighted the advantage of receiving data from the NPD. There was a huge reduction in the quantity of missing data for FSM and this allowed a more accurate subgroup analysis to be conducted. This analysis was delayed considerably in comparison with the analysis that was conducted using school-supplied data. This delay was caused by the changing legal protocols around GDPR, and the need to conduct the analysis at a secure ONS facility, not envisaged in the original timeline in the published protocol for the project. The timeline for acquiring NPD data would have fitted into the original timeline, however, had GDPR not come into effect late into the project. This timing issue, therefore, should not be considered discouraging for future NPD data collection.

NPD data collection may be especially advantageous for data such as FSM, which has numerous methods of coding (EverFSM variants, FSM for that year, etc) and requires more formatting or administrative work by schools to supply to evaluators (compared with simpler items such as DOB or Sex) leaving it prone to missing data.

However, there should still remain some caution in the use of NPD data as pre and post outcome data in a RCT if available attainment constructs are not closely matched to the primary outcome of a particular intervention. For example, in this case KS1 data does not provide specific information on comprehension ability and this would be problematic as a pre-test measure for an intervention specifically focused on comprehension development.

The Education Endowment Foundation
5th Floor, Millbank Tower
21–24 Millbank
London
SW1P 4QP

https://educationendowmentfoundation.org.uk

@EducEndowFoundn

Facebook.com/EducEndowFoundn