

Updates to Protocol – February 2018

Following completion of the first phase of research, several updates to the protocol have been made:

- Page 2 – Pre-testing has been completed, and based on this data, the number of pupils has been changed to 5339 in the Evaluation Summary Table. The TIDieR checklist has been updated in terms of the tailoring and modifications sections, now that Phase 1 is completed
- Page 4 – Figure 1 has been updated to the new logic model, reflecting the implementation factors and secondary outcomes that have been described in the SAP. This logic model was developed in collaboration with FFT literacy at a research retreat in May 2017.
- Page 7 – The Randomisation section has been updated with the exact detail of how the pre-tested schools were randomly allocation to treatment condition.
- Page 9 – A recruitment update has been added. This also includes detail on the targeted intervention selection tool which was developed in Phase 1.
- Pages 12-13 – Updated effect size calculations have been included, to reflect recruited sample size and NGRT pre-test to post-test correlation.
- Pages 13-14 – The Outcomes section has been updated to include the final decision on having two primary outcomes. Further detail has also been added on secondary outcomes, which were developed in Phase 1.
- Pages 14-15 – The Analysis plan has been updated, based on the SAP which has now been developed, reviewed and published. Updated details are included for primary, secondary, additional and subgroup analyses, in line with the SAP.
- Page 17 – The Ethics and Registration section has been updated to include information on GDPR compliance and the link to the ISRCTN registration of this trial.
- Page 19 – The section on Substantial Attrition has been updated to include information on dealing with school concerns regarding GDPR compliance.

Evaluation Summary

Age range	Key Stage 2: Year 4, 5 & 6
Number of pupils	5339
Number of schools	100
Design	Logic Modelling, Implementation Study, Cluster RCT & Process Evaluation
Primary outcome	Reading Comprehension
Protocol version	2.0
Protocol version date	07/03/2018

Programme

The programme to be evaluated is called Reciprocal reading for a detailed description of the intervention being evaluated, please see TIDieR checklist in Table 1 below (Hoffman *et al.*, 2014). Furthermore, the logic model for the programme is displayed in Figure 1.

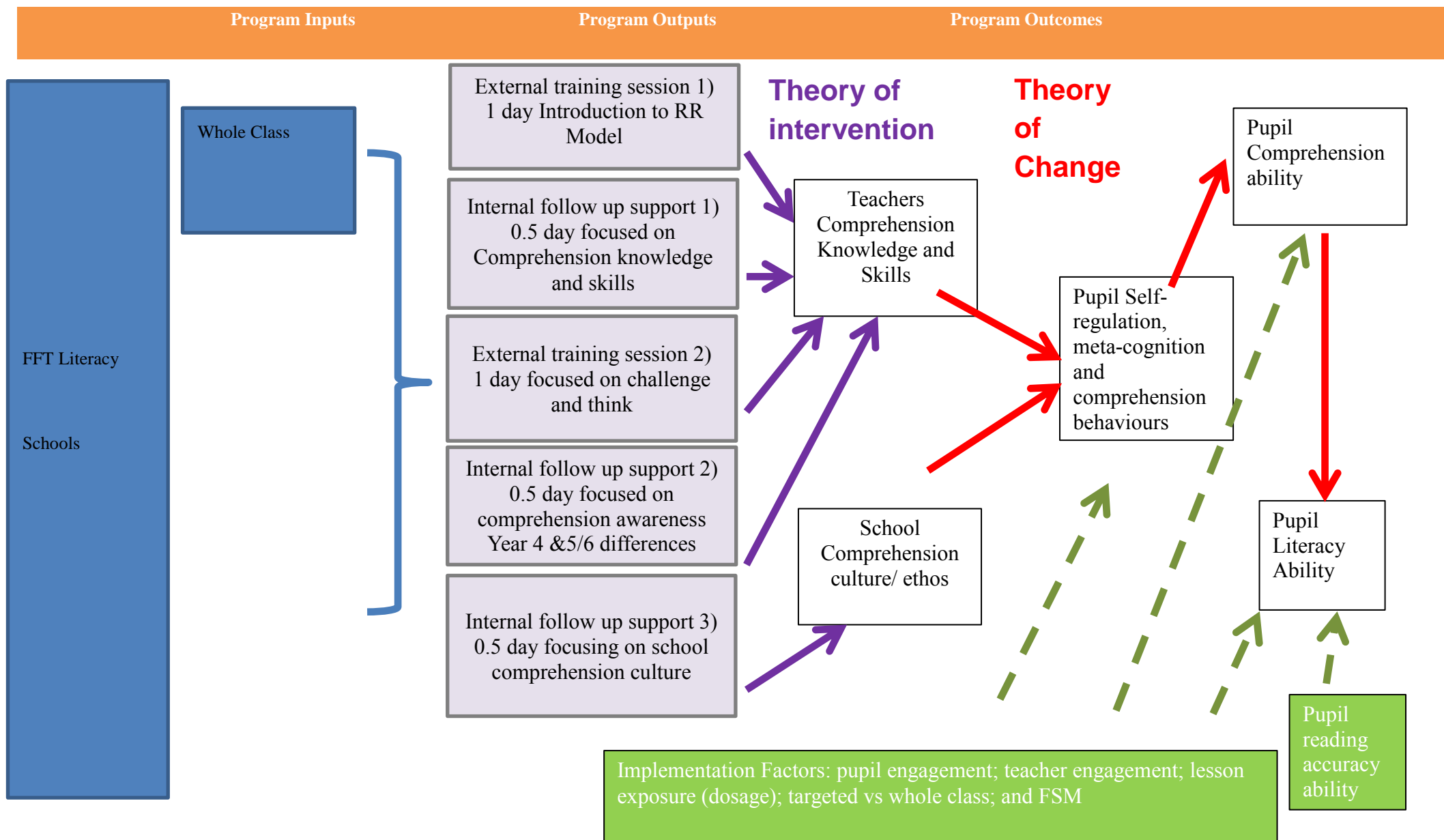
Table 1 Reciprocal Reading TIDieR checklist¹

ITEM No.	Item
Brief Name	
1	Reciprocal Reading (RR) (two versions: 1. Universal; & 2. Targeted)
Why	
2	Training programme for teachers aimed at improving comprehension ability of pupils in years 4 (universal) & years 5-6 (targeted)
What	
3	Materials: A teacher training programme delivered by the Fischer Family Trust Literacy, which includes external school training (for teachers and teaching assistants) interspersed with internal follow up support/training.
4	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.
Who Provided	
5	Reciprocal Reading Trainer provides teacher internal and external training. Teachers and teaching assistants provide reciprocal reading activities to pupils based on their training
How	
6	Initial training sessions provided to groups of teachers
Where	
7	External training provided out of school setting. Internal training provided in school setting including classroom.

¹ More details of programme training and delivery will be developed after logic modelling phase of project is completed.

When and how much	
8	There are two external training sessions and three internal follow up sessions over a 16 week period. Teachers are utilising their training over the course of the year.
Tailoring	
9	The programme logic model was refined during the first Phase of the research and is included in Figure 1.
Modifications	
10	No programme modifications are being made during the trial. Suggestions for modifications based on process evaluation data will be made in the final report.
How well	
11	Planned: This will be assessed through the research process evaluation
12	Actual: This will be assessed through the programme Cluster Randomised Controlled Trial evaluation.

Figure 1 – Emerging Logic Model from Phase 1



Significance

Reciprocal teaching and more recently reciprocal reading 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 a way of explicitly teaching reading comprehension using four cognitive strategies: summarising, questioning, clarifying and predicting. Teachers coach their pupils' one skill at a time until pupils are themselves able to lead small groups or pairs using structured discussion around text. This process results in pupils being equipped to take turns as tutor, leading dialogue using the four key strategies.

Reciprocal Reading was developed in New Zealand in the 1980s as reciprocal teaching, and has not been much used in the UK until recently. It is a group approach to reading intended to boost both accuracy and comprehension using small groups including shared reading. A meta-analysis of 16 reciprocal teaching studies outside the UK (including some randomised trials), not of uniformly high quality, found an average effect size of $ES > 0.32$ on reading and $ES > 0.88$ on comprehension using an experimenter-developed test (Rosenshine and Meister, 1994). A further meta-analysis also reports a large positive effect size of $ES > 0.74$ with regards to reciprocal teaching (Hattie, 2009). Most recently, a single randomised control study of reciprocal teaching funded by the Education Endowment Foundation, *The LIT programme*, was undertaken in the UK involving 41 schools. 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 the characteristics of pupils in treatment and control schools were too different to yield an unbiased estimate of the impact of the programme (Crawford and Skipp, 2014).

Reciprocal reading can be understood as a form of peer learning when used as shared reading. Peer learning or peer tutoring is widely reported to be an effective approach to learning and teaching in primary schools. A meta-analytic review of peer-assisted learning interventions with primary school pupils suggests positive effect sizes of between $ES > 0.33-0.4$ indicating increases in achievement (Cohen et al., 1982). Another meta-analysis suggests a positive median effect size of $ES > 0.59$, with effect sizes being strongest for studies with urban, low income, ethnic minority students and where learners are provided with more autonomy (Rohrbeck *et al.*, 2003). Peer learning continues to be cited as providing high impact ($ES > 0.5$) for low cost in the Sutton Trust report on pupil premium (Higgins *et al.*, 2013), although most recent refinements including additional meta-analytic review evidence result in a median effect size of $ES > 0.43$ in the Education Endowment Foundation Toolkit (EEF, 2016).

Peer tutoring includes a range of approaches in which learners work in pairs or small groups to provide each other with explicit teaching support in various curriculum areas, particularly mathematics and reading. These approaches include: cross-age tutoring, peer assisted learning and reciprocal learning. The common characteristic between these approaches is that learners take on responsibility for aspects of teaching and for evaluating their success. Peer learning through shared reading, is characterised by

specific role taking as tutor or tutee, with clear procedures and support for interaction, in which participants receive training. When paired reading is implemented with reasonably high integrity, results have been typically good (Van Keer, & Verhaeghe, 2005; Topping, 1987). Fuchs, Fuchs, Mathes and Simmons (1997) indicated positive effects for Peer-Assisted Learning Strategies techniques using class wide reciprocal peer tutoring in reading amongst grade 2-6 pupils in elementary and middle schools over a 15-week period. A single study evaluation of peer learning implemented in primary schools across a whole local authority using Paired Reading (the Fife Peer Learning project) was also found to have a positive impact on pupils' outcomes with effect sizes on reading attainment of $ES > 0.25$ (Tymms *et al.*, 2011). Though overall the evidence base related to peer tutoring is relatively consistent, some recent studies of peer tutoring have found lower average effects, suggesting that monitoring the implementation and impact of peer learning is important (EEF, 2016).

Methods

Research questions

The evaluation is designed to answer the following research questions:

1. **Does the universal Reciprocal Reading approach impact pupil reading outcomes?** Improving reading outcomes is the ultimate goal of the programme. We will assess the impact of the programmes on reading levels or progress through the New Group Reading Test results for Year 4.
2. **Does the targeted Reciprocal Reading approach impact pupil reading outcomes?** Again, we will assess the impact of the programmes on reading levels or progress through the New Group Reading Test results for pupils with low comprehension and relatively good decoding ability in Years 5 & 6.
3. **How are the programmes being implemented?** In order to get a better picture of the effectiveness of the programmes on learning outcomes and potential implementation factors, we will conduct a "light touch" examination of how the programmes are being implemented. Process evaluation questions are likely to be refined during Phase 1, but are likely to include: Are the teachers teaching the programmes as envisioned by the developers or adapting in practice? How does training help support the teacher to implement the programme with fidelity? How do pupils manage their respective roles? How has the programme impacted teacher knowledge on how to teach reading? What the relationships between pupil comprehension behaviours, comprehension awareness and school comprehension culture on programme effectiveness.

Design

Phase 1: Logic Modelling & Implementation Pre-test

We aim to conduct a light-touch and formative evaluation of the programme during phase 1. This will involve close working with the Fischer Family Trust Literacy (in workshops) to clarify programme components, training schedules and materials, develop a theory of change for the approach, and make potential refinements to the programme logic model to inform delivery.

To further inform the logic modelling phase we will collect some primary data focused upon the development of programme implementation strategies. This data will be collected through supplementary questions added to the memorandum of understanding (MOU) between schools, evaluators and project team. These questions will focus on the current quality and sufficiency of comprehension related training, support, acceptability, practice adaptations, teacher comprehension behaviours, awareness and school comprehension culture etc.

Phase 2: RCT efficacy trial and Implementation Study

This stage of the evaluation will focus on understanding the impact of the reciprocal reading approach. We will test both the universal whole class and targeted small-group approaches to reciprocal reading within the same schools. This permits an assessment of both versions of the programme against a control group at minimal extra cost compared to testing only one version of the programme. Year 4 children in the intervention group would receive the whole class approach, with pupils low in comprehension and relatively good decoding ability in Year 5 & 6 receiving the targeted version of the programme.

This phase will also see additional implementation study including: an audit tool administered during FFT literacy follow up training on dosage, reach, fidelity and quality completed by all trainees once a term; a post- test survey will go to all teachers and TAs working with year 4-6 and include items on teacher level outcomes (comprehension behaviour and awareness), implementation issues (for intervention schools) and comprehension culture of the school; structured observations (in a set of 10 case study schools with interviews with senior leaders and teachers/ TAs and pupil focus groups).

Randomisation

Randomisation was carried out in July 2017 after pre-test 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 we conducted minimisation 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, and is especially useful when randomising a small number of cases (Torgerson and Torgerson, 2007). Minimisation was used to ensure the schools were as evenly matched as possible. A number of school level covariates was used in the matching process specifically: reading comprehension score (NGRT passage comprehension), % FSMEver, 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. This randomisation process resulted in 51 schools being assigned to the Intervention condition and 49 being assigned to the Control condition.

Participants

Schools would be recruited to the trial during the pilot phase, which we suggest should start from January 2017 as recruitment during the summer term is more difficult.

Pupils:

We are aiming that the universal programme should be restricted to all pupils in either Year 4 and to those who are assessed as needing it in Years 5 & 6 who will be given the targeted version of the programme. For the universal programme all children in the respective year would be given the intervention. The eligibility criteria for the targeted version are children who are poor at reading comprehension but relatively good at decoding. The children in this category were selected by teachers using guidance and materials provided by FFT literacy. Teachers will be provided with a document that will allow them to compare each child in their class to two sets of criteria. The first set of criteria will describe the reading skills possessed by a child who could be classified as having good decoding skills. The second set of criteria will describe difficulties that may be experienced by a child who struggles with reading comprehension. By comparing each child with these two sets of criteria, teachers will be able to identify children who are good at decoding, but struggle with reading comprehension. Children who match the provided criteria for having good decoding skills and poor reading comprehension skills will be deemed to be eligible for the targeted intervention.

Schools:

QUB will carry out all the recruitment with FFT literacy's support in reviewing the documents.

a. Eligibility criteria

Schools should:

- not have received the FFT literacy package of training (external and internal) to deliver Reciprocal Reading
- not be in another EEF literacy trial at KS2,
- the full school sample (N=100) should have an average everfsm percentage of 29% (i.e., the national average) or over if possible
- be prioritised if they are not a NE schools with an advocate
- be prioritised if they are not a first and middle school

Areas to target will include:

The North East, its surrounding regions, and any other regions in England where a group of 10 schools or more applies and FFT literacy is willing to deliver training.

b. Schools data

Once schools have expressed interest QUB will request that the schools complete a registration form/MoU. The information provided on the registration form will be used in the random assignment of the schools. Before being informed of their random assignment, Queens University Belfast will conduct the pre-test and collect the following information from schools about pupils in the current Year 3, Year 4 and Year 5: school name; school post-code; year group; class name; pupil name; pupil DOB; UPNS; and ethnicity (all other relevant data will be retrieved from the NPD). Also at school level the following items: Confirmation that consent letters have been completed and any opt-outs removed; average KS2 reading score; and everFSM.

c. Process and materials (consent, MoU)

QUB will collect a combined registration and MOUs from schools and send out opt- out parental consent letters (allowing approx. 2 weeks for parents to return forms). The MOU will include a few short questions on usual literacy practice, existing reading programmes, and what they do for struggling readers in KS2.

d. Retention

Dr Maria Cockerill will set up and oversee a system for achieving recruitment targets based on previous successful models of recruitment and retention used on projects funded by EEF (e.g., ongoing Read Write Inc. trial with 120 schools). This will include usage of existing extensive networks of schools in Northeast England.

Recruitment update

Recruitment for the study took place from March 2017 to June 2017 and 114 schools were recruited by QUB (and signed a Memorandum of Understanding). The universal intervention is being delivered to pupils in Year 4. The targeted intervention is being delivered to pupils in Years 5 and 6. For the universal programme all children in the respective year are being given the intervention by a RR trained instructor. The eligibility criteria for the targeted version are children who are 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. Teachers were provided with a document that allowed them to compare each child in their class to two sets of criteria. The first set of criteria describes the reading skills possessed by a child who could be classified as having good decoding skills. The second set of criteria describes 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 are good at decoding, but struggle with reading comprehension. Children who matched the provided criteria for having good decoding skills and poor reading comprehension skills will be deemed to be eligible for the targeted intervention. Teachers selected 6 pupils from each Year 5 and 6 class for the targeted intervention.

Sample size calculations

As the RCT will assess both the whole group approach and the more targeted version of the programme it would need to be adequately powered for both approaches (with the targeted approach requiring the highest number of schools as it has fewer pupils per school). So the following power calculations have been based on the targeted approach.

While some syntheses of reciprocal reading related approaches report large effect sizes, these are likely to be over-represented by small-scale non-randomised studies (EEF, 2016). Effect sizes for literacy interventions evaluated through a good quality RCT design would more typically be in the range of 0.2-0.3 (e.g. Biggart et al, 2013; Borman et al, 2007; Tymms et al, 2011) and we have opted for a more conservative effect size of 0.2 in our power calculation.

Figure 2 shows the power calculation for the targeted approach provided by Optimal Design software, using estimates of: $ES=0.2$; $p=0.05$; $ICC = .145$ (NE/NW KS2 reading score from EEF's ICC guidance [here](#)); $r^2=0.50$ (due to having a pre-test of NGRT); average class size $n=20$.

This calculation suggests a total sample size of 94 schools (clusters) to detect a significant effect if present with a power of .8

Figure 2. Study power according to number of clusters.

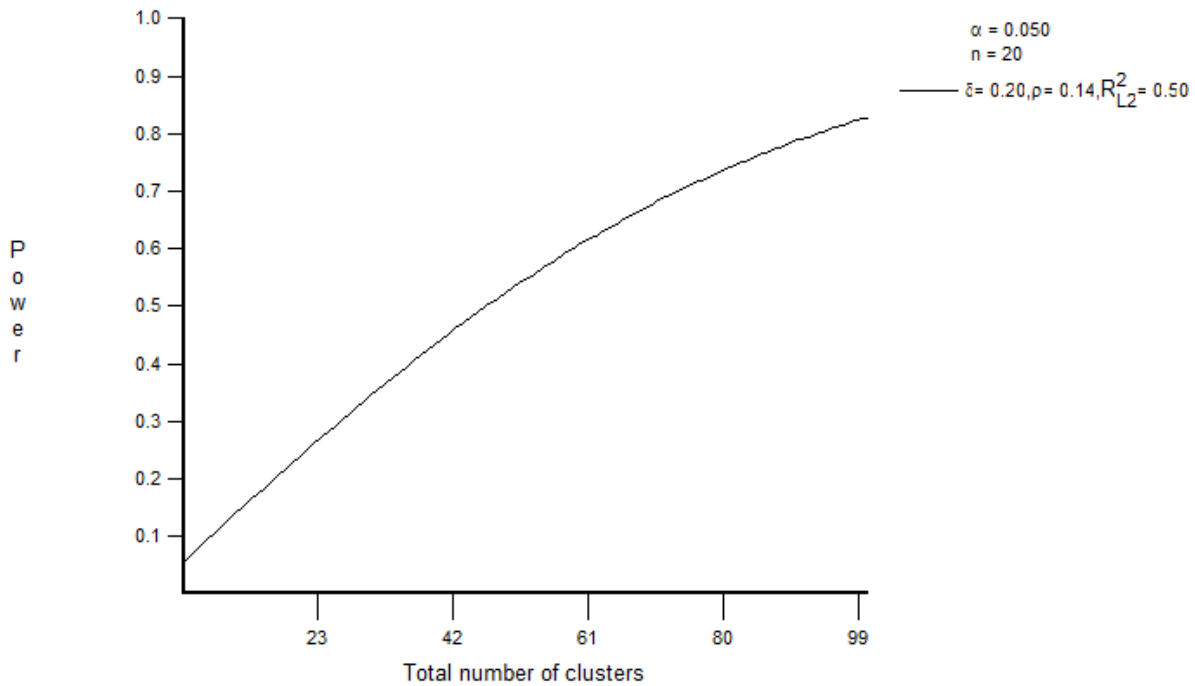
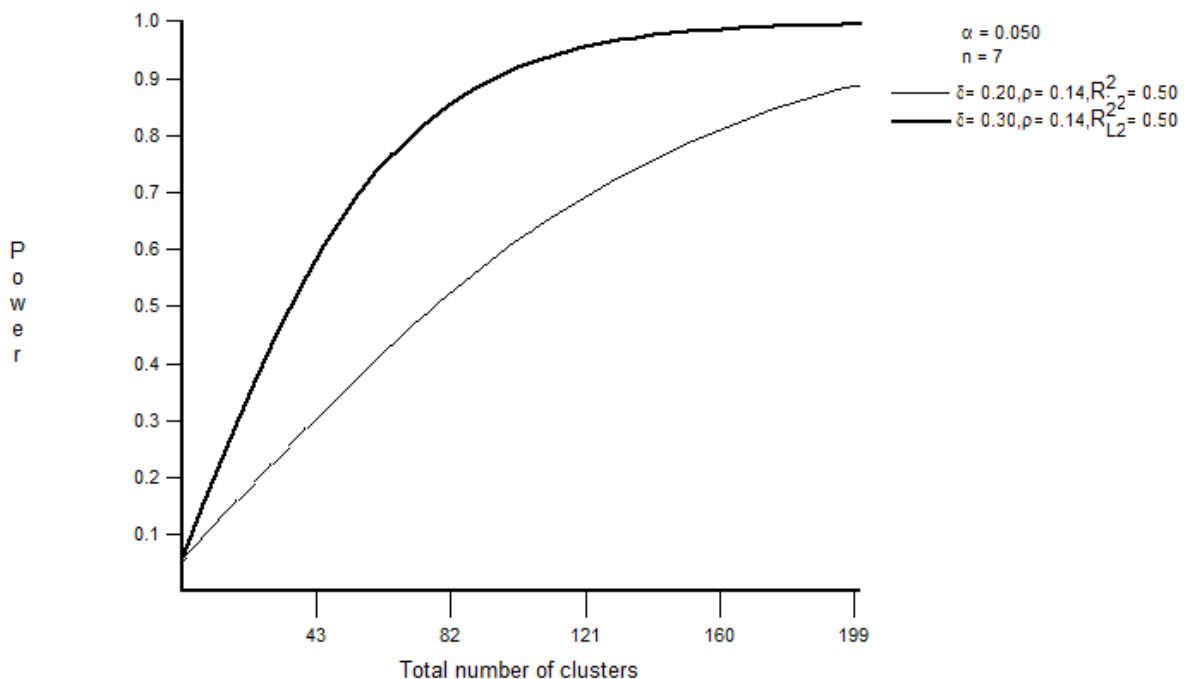


Figure 3 shows a repeat power calculation for everFSM students provided by Optimal Design software, using estimates of: ES=0.2 & .03; $\rho=0.05$; ICC = .089 (FSM quintile 3 KS2 reading score from EEF's ICC guidance [here](#)); $r^2=0.50$ (due to having a pre-test of NGRT); average class size $n=7$ (assuming recruitment of classes with above national average FSM students of 29%).

Figure 3. Study power according to number of clusters (for everFSM students)



This calculation suggests a total sample size of 152 schools (clusters) to detect a significant effect of ES= 0.2 if present with a power of .8. Or a total sample size of 72 schools (clusters) to detect a significant

effect of ES= 0.3 if present with a power of .8.

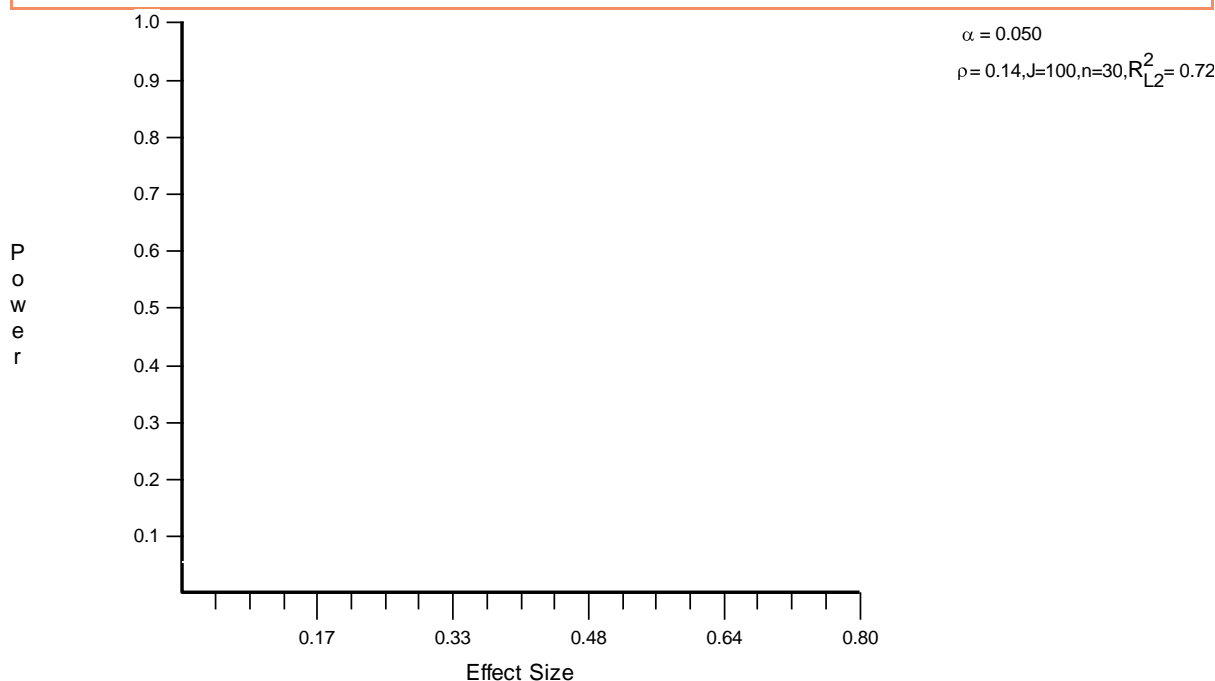
In conclusion, the nature of the trial with 100 schools (which is a limiting factor due to project team capacity to deliver training) would permit the well powered assessment of outcomes of the programme in comparison to control schools if an effect of 0.2 is present and a subgroup analysis of everFSM pupils if an effect of 0.3 is present.

Updated effect size calculations

Pre-test data was collected in June/July 2017 and an updated set of effect size calculations has been carried out based on the number of pupils who sat pre-test and an updated NGRT correlation of .85.

Minimum detectable effect size (MDES) at different stages of trial for whole school intervention

Stage	N (schools)	Correlation between pre-test and post-test	ICC	Blocking or pair matching	Power	Alpha	MDES
Protocol	94	.7	.14	Pair matching	.8	.05	.2
Pre-test ²	100	.85	.14	Pair matching	.8	.05	.15



² NGRT pre-test and post-test correlation was updated after recruitment stage to reflect the recruited sample size and the reported test-retest correlation for NGRT (Granada Learning, 2013). N per cluster was included in whole class calculation as 30 for one class, and 6 for the targeted intervention as teachers were asked to select 6 targeted pupils.

Figure 4 – Power and effect size for NGRT reading for whole school intervention as calculated after recruitment

Figure 4 is a plot of power and effect size with alpha = .05, pre-test post-test correlation = .85, ICC = .14, number of clusters = 100, n per cluster = 30. Number of pupils per cluster was increased from n=20 to n=30 to reflect the number of pupils recruited across the 100 schools (n=5339); twenty pupils was a low estimate of pupils per cluster, given the number of pupils now recruited. The same ICC estimate as in the original protocol has been used. This will be recalculated once post-test data has been collected.

Stage	N (schools)	Correlation between pre-test and post-test	ICC	Blocking or pair matching	Power	Alpha	MDES
Protocol	72	.7	.14	Pair matching	.8	.05	.3
Pre-test	100	.85	.14	Pair matching	.8	.05	.24

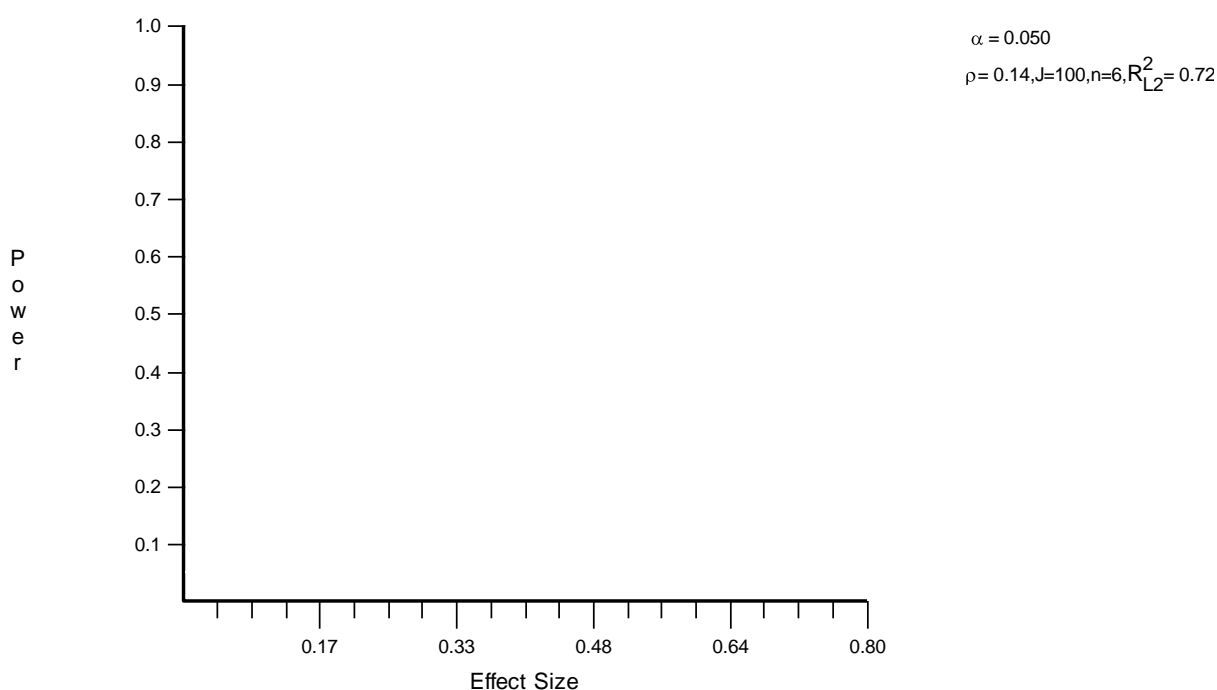


Figure 5 – Power and effect size for NGRT reading for targeted intervention as calculated after recruitment

Figure 5 is a plot of power and effect size with alpha = .05, pre-test post-test correlation = .85, ICC = .14, number of clusters = 100, n per cluster = 6. Number of pupils per cluster was updated from n=7 to n=6 to reflect the number of targeted intervention pupils identified by teachers. The same ICC estimate as in the original protocol has been used. This will be recalculated once post-test data has been collected.

Outcome Measures

Primary Outcomes

The SAP has been reviewed by EEF and external peer reviewers and the primary outcomes of the trial are now confirmed. The primary outcomes of the trial are two attainment indicators from the New Group Reading Test (NGRT 1. Overall reading score and 2. Comprehension subscale score). The NGRT is an adaptive test which has high reliability and measures both reading accuracy and comprehension. Overall reading is being 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 being analysed as a primary outcome as it is the specified target of the programme and the most valid way to evaluate programme success. At pre-test the NGRT was delivered (digitally) by schools under exam conditions. At post-test the NGRT will be administered (digitally) by the evaluation team under exam conditions. The NGRT provides two tests A & B, and a different test will be 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 will be delivered online at post-test. Overall reading scale score will be used as the score for Overall reading ability. Paragraph Comprehension scale score will be used as the score for reading comprehension. These scores provide a valid measure of each of the primary outcomes. 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 & Mander (2014). The analysis in this study meets the criteria to not require correction for multiple testing due a small number of planned comparisons (Armstrong, 2014).

Secondary Outcomes

Exact measures for secondary outcomes were also decided during the development of the SAP. The secondary outcomes of the trial are Reading accuracy score from the Sentence Completion subtest of the NGRT, pupil meta-cognition and pupil comprehension behaviour. The pupil meta-cognition and pupil comprehension behaviour measures were developed by the evaluation team. These bespoke measures were developed to provide specificity matching to the intervention as no previously used standardised measures would be appropriate for this. These measures have been designed to closely reflect the principles of the Reciprocal Reading intervention. The school level comprehension ethos questionnaire will be 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.

No pilot phase for secondary measures was in the protocol for this RCT. 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. Also, secondary outcome measures are used to explore theory of change. Therefore it is important to match programme outcomes even more closely than primary outcomes. Overall, the programme is judged on primary outcomes and theory of change explained by secondary outcomes

Analysis plan

Analysis will be conducted on an intention-to-treat basis. The initial characteristics of the intervention and control groups will be compared at baseline in relation to their core characteristics: e.g. gender, FSM eligibility, and mean scores on the pre-test.

Analysis will be carried out using STATA. Analysis will be conducted on an intention-to-treat basis. The main effects of the intervention will be estimated using multilevel random-effects linear regression modelling to take account of the clustered nature of the data and a series of models will be estimated for each outcome (where pupil is level 1 and school is level 2). Multilevel random effects models allow for estimates of variance of both levels.

Firstly, analysis will be conducted for the universal intervention with the NGRT comprehension score, overall reading score forming the dependent variable 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.

Secondly, this analysis will be repeated for the targeted intervention.

Please refer to the SAP for further detail on multilevel models to be included in the analysis.

The analysis used to help determine whether attrition has led to imbalance at baseline will be multi-level regression models. These will test for differences in reading comprehension, FSMEver % and reading accuracy between the Control and Intervention groups. Imbalance at baseline will be reported as effect size. Effect size will be calculated using the equation described in "Effect Size Calculation" in the SAP.

Schools who drop out of delivering the intervention will be encouraged to allow post-testing and will still be included in an intention-to-treat analysis. If the proportion of missing data is greater than 5%, a 'missing at random' data analysis will tell us whether imputation is required. If so, data will be imputed using multiple imputation which will be presented as a sensitivity analysis. The missing at random analysis will be carried out in Stata and will determine if the pattern of missing data is related to the primary outcome. Schools that were pre-tested but then dropped out of the trial will be included as ITT.

Non-compliance with the intervention will be analysed using CACE analysis. FFT have reported that minimum dosage should be two sessions of twenty minutes per week for twelve weeks for the Targeted Intervention, and one session of twenty minutes per week for twelve 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. A regression analysis will then be conducted on the same scores as the primary analysis, but also including compliance level as a predictor.

Secondary analyses will investigate differences between the control and intervention groups in post-test reading accuracy, pupil meta-cognition and pupil comprehension behaviour, controlling for pre-test reading score. This analysis will be carried out for the universal intervention and for the targeted intervention.

Additional analyses will be conducted to explore the pathways for change in both primary and secondary outcomes as proposed in the logic model (see logic model in Figure 1). This will be carried out for both the universal intervention and the targeted intervention. The original protocol had not fully specified

implementation factors. These have been developed and updated since the original protocol. The additional analyses will investigate the relationships between pupil meta-cognition and comprehension behaviour, teacher comprehension and behaviour, school comprehension ethos and the primary outcomes.

A subgroup analysis will repeat the primary outcome analysis for only pupils in receipt of FSM. We have since updated the protocol to include analysis of only FSM as a subgroup and this will be included in the amended protocol to be published in the near future.

Please see the SAP for further detail on the secondary, additional and subgroup analyses.

Implementation and process evaluation methods

We propose a mixed methods (quantitative and qualitative) approach to measure implementation factors and would envisage the programme to be delivered in the intervention schools from September 2017 and expect the intervention to run for between 12 and 16 weeks (to be firmed up after logic modelling, which is part of phase 1).

FFT literacy and QUB will develop an audit tool which can be used with each teacher before the follow-up sessions. This will include information on dosage (e.g. number and time period of RR sessions conducted), fidelity (e.g. how close the RR sessions adhere to the training criteria and if key principles were included) and engagement (both teacher and pupil). FFT literacy will administer these at each of the training sessions and collect them by email and pass them on to QUB. Every teacher/ TA taking part in the sessions will need to complete one of these every term. They will help to inform the follow-up sessions by outlining how far they have got with the work and what their issues are.

In addition, QUB will collect pre and post data on usual practice. This will be through the MOU at pre-test and a survey at post- test (this could be administered at the same time as the pupil post- test to maximise response). The survey at post- test will go to all teachers and TAs working with year 4-6 and include items on teacher level outcomes (e.g. comprehension behaviours and awareness), implementation issues (for intervention schools) and comprehension culture of the school.

We will also undertake structured observations in a set of 10 case study schools using an instrument developed for the evaluation of previous Paired Reading studies interviews with senior leaders and teachers/ TAs (case studies will also include pupil focus groups and teacher interviews). We will use all these data sources to assess several issues including: the centrality of the teacher role in programme delivery i.e., teaching of Reciprocal Reading and pupil paired/group reading time; how pupils adapt to their respective roles (summarising, questioning, clarifying and predicting); facilitators and barriers to programme implementation; exploration of programme theory of change and intervention; and the perceptions over the scalability of the programme. Naturally occurring data will also be collected from FFT literacy at the end of the programme (including their records and training notes).

It is also increasingly acknowledged in both the health and educational trial literature that in the case of complex interventions there is a need to take account of the way in which context interacts with an intervention to produce outcomes (Bonell, et al., 2012; Craig, et al, 2008; Jamal. et al., 2015). This requires a greater integration of the process evaluation and impact analysis through implementation factor analyses than is traditionally the case in RCTs. In this respect, we could envisage revisions to the initial protocol through emerging findings from Phase 1 and the wider process evaluation (clarifying the important implementation factors and essential components of programme delivery are, i.e., ‘on-treatment’), however, initial hypothesised implementation factors likely to be included in the analysis are dosage, pupil and teacher engagement, comprehension culture of the school and leadership support. Initial hypotheses will be firmed up through dialogue with FFT literacy as part of the evaluation set-up process, but others may also emerge over the course of Phase 1.

Costs

QUB will collect information from the FFT literacy (and schools) to assess the estimated cost of Reciprocal Reading as it was delivered. This will include both the direct and marginal costs of the intervention. Cost estimates are likely to include intervention materials, training costs, substitute teaching cover etc. A cost per pupil will be derived for both the whole school and targeted approach.

Ethics and registration

Ethics will be applied for through the ethics committee in the School of Social Sciences Education and Social Work at Queen’s University Belfast.

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 at both pre-test (June 2017) and post-test (June 2018), i.e. QUB would not access their child’s reading scores or NPD data. Schools will be provided with an information sheet regarding compliance with the General Data Protection Regulation (GDPR). All pupil data will be treated with the strictest confidence and will be stored in accordance with the data protection legislation, including the General Data Protection Regulation (GDPR) which comes into effect in May 2018. Personal data will be 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 trial has been registered on the ISRCTN website and summarised here: <http://www.isrctn.com/ISRCTN81582662>

Personnel

Evaluation Team

The evaluation team will be drawn from senior and experienced staff within the Centre for Evidence and Social Innovation (CESI) at Queen’s University Belfast. The 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 has also experience in accounting for the nested nature of the data when schools are randomised at the school level through the use of multi-level modelling.

Dr Andy Biggart will be the lead analyst of the project data. He will be responsible for overseeing the statistical analysis, qualitative analysis and mixed methods analysis.

Dr Maria Cockerill will have particular responsibility for the recruitment and retention of schools. She will also oversee the school registration, MOU and consent processes as well as some implementation data collection aspects of the trial.

Dr Katrina Lloyd will support the development of survey instruments in the project including in the MOU, teacher surveys and post-test implementation survey.

Dr Sheila McConnellogue will act as an educational psychology advisor and contribute to literature review, data interpretation and writing the final report.

Dr Liam O'Hare will have overall responsibility for project delivery, including the final report. He will also lead the development of implementation strategies, and refinement of theory of change and logic modelling for the programme.

Dr Patrick Stark will act as the trial manager and oversee data collection and data management for the project.

Professor Paul Connolly will act as an overall expert consultant for the project and play a key role in quality control, interpretation of findings, final write up of the report and research dissemination.

The team will also draw upon other research staff within the Centre that have experience of conducting fieldwork for EEF evaluations.

Project Team

Andy Taylor from the Fisher Family Trust Literacy will act as the project team lead. He will review documents produced by the evaluation team and be responsible for the design and implementation of the programme.

John Catron from the Fisher Family Trust Literacy who is a training associate and will provide some of the external training and follow up visits.

Risks

One of the major benefits of EEF funding this proposed evaluation are the extensive experience, strong controls and contingency measures that Queen's University of Belfast will be able to provide. This adds

security to the funding body and peace of mind that the proposal will be delivered on specification and on-time.

No.	Risks	Current Controls	Net Risk*		Contingency Plans
			I	L	
1	Failure to recruit required number of schools	Maria Cockerill is experienced in recruiting large numbers of school in short time periods through her established networks in English schools.	5	2	Additional experienced CESI staff will be assigned to recruitment
2	Substantial attrition	A MOU will be completed by schools ensuring their awareness of their commitments. Also, Maria Cockerill will maintain ongoing relationships with schools encouraging continued engagement. Schools will be provided with an information sheet regarding GDPR compliance to minimise the risk of drop-out due to data protection issues. Maria will also be available to schools to answer any queries regarding GDPR and research participation in line with QUB ethics guidelines.	4	2	Re-engagement of schools of schools will be sought if substantial overall attrition occurs.
3	Failure to collect the required attainment data	Research team has extensive experience of data collection including: standardised measures like NGRT, NPD etc.	5	1	Team can administer online or pencil paper versions of tests if required
5	Fire or other damage to electronically-stored data in CESI	Daily back-up of servers is carried out in QUB. Full back-up for off-site storage carried out every week Data is double backed up from key CEE personnel on a weekly basis also.	2	1	Files would be restored and any lost electronic data would be retrieved directly from schools and/or email systems. Files can be restored from one of the two back-ups.
6	Temporary loss of key project staff due to illness	CESI employs 17 full-time and dozens of part-time 'field worker' staff including administrators, technical staff who are familiar with supporting schools to administer assessments. There are several senior staff in the project. In the unlikely event of illness these staff could fulfil the role of the other. The School of	1	2	Other members of the CESI team who are not working on this project would be recruited along with others in their respective departments with the necessary skills and experience

		Education at QUB employs a similar number of staff and has very experienced researchers who could provide temporary cover for staff.			
7	Permanent loss of key project staff	Whilst a temporary solution would be put in place, a recruitment process would follow.	1	2	Other members of the CESI team who are not working on this project would be recruited along with others in their respective departments with the necessary skills and experience
8	Failure to obtain informed consent from schools	The bid will receive ethical permission from the School of Social Sciences, Education and Social Work Ethics Committee, which will scrutinise practices. The nature and purpose of data collection will be explained to schools in advance of their recruitment to the project. Clear participant information sheets will be provided early on in the project to all potential participants. The data collection methods have been utilised in a number of previous projects without the materialisation of this potential risk as an actual problem.	5	1	Meetings with schools would be held to reassure them of the nature and purpose of data collection.
9	Differential student attrition between conditions.	The sample is significantly robust enough to deal with such attrition.	3	2	Thought may be given to how to deal with missing data in the final model.

*I=importance L = Likelihood

Timeline

<u>TASK/DATE</u>	<u>JAN- MAR 17</u>	<u>APR- JUN 17</u>	<u>JUL- SEP 17</u>	<u>OCT- DEC 17</u>	<u>JAN- MAR 18</u>	<u>APR- JUN 18</u>	<u>JUL- SEP 18</u>	<u>OCT- DEC 18</u>
<u>SET UP AND MONITORING</u>								
<u>SET-UP MEETINGS</u>								
<u>PROGRESS MEETINGS WITH PROJECT TEAM</u>								
<u>PHASE 1 REFINEMENT, PILOTING AND RECRUITMENT</u>								
<u>RECRUITMENT OF SCHOOLS</u>								
<u>ETHICAL APPROVAL</u>								
<u>LITERATURE REVIEW</u>								
<u>LOGIC MODELLING/THEORY OF CHANGE/OUTCOME MEASURES PREP</u>								
<u>DEVELOPMENT OF RECRUITMENT MATERIALS/REGISTRATION TEMPLATES AND MOU</u>								
<u>RECRUITMENT/SCHOOL REGISTRATION/MOUs SIGNED</u>								
<u>RANDOMISATION</u>								
<u>PUBLISH TRIAL PROTOCOL</u>								
<u>REFINEMENT AND PILOTING OF PROGRAMME</u>								
<u>REFINEMENT/PILOT OF STANDARDISED RESOURCES FOR TRIAL – WORKING WITH PROJECT TEAM</u>								
<u>IDENTIFY/AGREE NATURALLY OCCURRING DATA</u>								
<u>PROCESS EVALUATION</u>								
<u>BASELINE SURVEY QUESTIONS COLLECTED VIA MOUs</u>								
<u>OBSERVATION METHODS- A)DEVELOPMENT, B) PILOT</u>								
<u>OBSERVATION DATA COLLECTION</u>								
<u>SCHOOL INTERVIEWS- A) DEVELOPMENT, B) COLLECTION</u>								
<u>ONLINE SURVEY OF INTERVENTION SCHOOLS</u>								
<u>COLLECTION OF NATURALLY OCCURRING DATA FROM FFT LITERACY</u>								
<u>POST PROGRAMME SCHOOL SURVEY (DIGITAL) – A) DEVELOPMENT, B) DATA COLLECTION</u>								
<u>BASELINE DATA COLLECTION</u>								
<u>UNIQUE PUPIL DATA COLLECTED FROM SCHOOLS</u>								
<u>PRE-TEST GL ASSESMENT</u>								
<u>POST DATA COLLECTION</u>								
<u>POST-TEST GL ASSESMENT NGRT</u>								
<u>DATA ANALYSIS</u>								
<u>DATA LINKAGE</u>								
<u>FULL DATA CLEANING, ERROR-RATE CHECKING</u>								
<u>ANALYSES OF DATA (PROCESS)</u>								

<u>ANALYSES OF TRIAL DATA (OUTCOMES)</u>								
<u>WRITE FINAL REPORT</u>								
<u>PRESENT FINDINGS & DATA ARCHIVING</u>								

References

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, pp.129-140.

Bonell, C., Fletcher, A., Morton, M., Lorenc, T. and Moore, L., 2012. Realist randomised controlled trials: a new approach to evaluating complex public health interventions. *Social science & medicine*, 75(12), pp.2299-2306.

Borman, G.D., Slavin, R.E., Cheung, A.C., Chamberlain, A.M., 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), pp.701-731.

Cohen, P.A., Kulik, J.A. and Kulik, C.L.C., 1982. Educational outcomes of tutoring: A meta-analysis of findings. *American educational research journal*, 19(2), pp.237-248.

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

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. and Petticrew, M., 2008. Developing and evaluating complex interventions: the new Medical Research Council guidance. *Bmj*, 337, p.a1655.

Fuchs, D., Fuchs, L.S., Mathes, P.G. and Simmons, D.C., 1997. Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, 34(1), pp.174-206.

Hattie, J.A.C., 2009. *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon: Routledge.

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*.

Hoffmann, T.C., Glasziou, P.P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D.G., Barbour, V., Macdonald, H., Johnston, M. and Lamb, S.E., 2014. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *Bmj*, 348, p.g1687.

Jamal, F., Fletcher, A., Shackleton, N., Elbourne, D., Viner, R., & Bonell, C. (2015). The three stages of building and testing mid-level theories in a realist RCT: a theoretical and methodological case-example. *Trials*, 16(1), 1.

Palincsar, A.S., 1986. *Reciprocal teaching: Teaching reading as thinking*. Oak Brook, IL: North Central Regional Educational Laboratory.

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

Rohrbeck, C.A., Ginsburg-Block, M.D., Fantuzzo, J.W. and Miller, T.R., 2003. Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology*, 95(2), p.240.

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

Torgerson, C.J. & Torgerson, D.J. (2007). The need for Pragmatic Experimentation in Educational Research. *Economics of Innovation and New Technology* 16(5): 323-330

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), pp.265-289.

Van Keer, H. and Verhaeghe, J.P., 2005. Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions. *The Journal of Experimental Education*, 73(4), pp.291-329.