



E
E
F
ducation
ndowment
oundation

REACH

Evaluation report and executive summary

February 2016

Independent evaluator:

Luke Sibieta (Institute of Fiscal Studies)



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 Trust (now part of Impetus - Private Equity Foundation) 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:

Danielle Mason

Head of Research and Publications
Education Endowment Foundation
9th Floor, Millbank Tower
21–24 Millbank
SW1P 4QP

p: 020 7802 1679

e: danielle.mason@eefoundation.org.uk

w: www.educationendowmentfoundation.org.uk

About the evaluator

The project was independently evaluated by a team from the Institute for Fiscal Studies. The lead evaluator was Luke Sibieta.

Contact details:
The Institute for Fiscal Studies
3rd Floor
7 Ridgmount Street
London
WC1E 7AE

p: 020 7291 4800
e: mailbox@ifs.org.uk

Contents

About the evaluator.....	2
Contents.....	3
Executive summary.....	4
Introduction	6
Methods	10
Impact evaluation	19
Process evaluation.....	33
Conclusion.....	38
References.....	41
Appendix A: Project team training plan	42
Appendix B: Consent forms and information packs.....	44
Appendix C: Teaching assistant survey	60
Appendix D: Robustness Checks	67
Appendix E: Cost rating	70
Appendix F: Security classification of trial findings	71

Executive summary

The project

REACH is a targeted reading support programme designed to improve reading accuracy and comprehension in pupils with reading difficulties in Years 7 and 8. It is based on research by the Centre for Reading and Language at York and is delivered by specially trained teaching assistants (TAs). This evaluation tested two REACH interventions, one based directly on the original 'Reading Intervention' developed by York, and one adapted from it with supplementary material on language comprehension. In both versions, pupils received three one to one 35 minute sessions each week for 20 weeks. Pupils were taken out of other lessons (typically not English lessons) for the sessions and so this evaluation assesses the effect of the interventions combined with more time focused on literacy, compared with standard provision.

The impact of the interventions on the reading skills of 287 pupils in 27 schools was tested using a randomised controlled trial. Schools in areas close to Leeds were recruited to the trial in 2013. Pupils identified as having relatively poor reading skills were randomly allocated to the original REACH reading intervention, the language comprehension version, or standard provision. In response to slow initial recruitment, the trial was implemented in two phases. A process evaluation was carried out involving a survey of teaching assistants and interviews with staff from participating schools.

Key conclusions

1. Both REACH interventions had a positive effect on the reading skills of the pupils in the trial. These effects are unlikely to have occurred by chance.
2. Pupils receiving the reading intervention with language comprehension experienced the equivalent of about six months of additional progress on average. For pupils receiving the standard reading intervention the figure was about four months.
3. The evaluation did not provide any evidence that the interventions improved reading comprehension in particular, as opposed to other skills such as word recognition.
4. Staff reported that the interventions improved literacy, reading ability, and confidence. Staff views were more positive in schools where the interventions were delivered by experienced teaching assistants, supported by senior staff, and allocated a dedicated space for delivery.
5. TAs sometimes found the interventions challenging to deliver. In particular, many said they were not confident delivering the one to one sessions even after training, and some found that the reading comprehension elements sometimes failed to hold pupils' attention.

Security rating

Security rating awarded as part of the EEF peer review process

Findings from this study have moderate to low security. The study was designed as a single randomised controlled trial which aimed to compare the progress of pupils who received the interventions with that of similar pupils who did not. However, the original design had to be changed because of delays in recruiting schools, meaning that the trial was run in two separate phases. The trial was also smaller in size than expected because not as many pupils were recruited as planned, and because 29.6% of the pupils did not complete all the tests at the end of the project.

The process evaluation also suggested that some participating TAs used some of the REACH techniques they had learned when teaching pupils from the comparison group. These pupils were not supposed to receive the REACH interventions, and the fact that they did makes it harder to estimate the size of the impact accurately.

Findings

Both REACH interventions had a positive effect on the reading skills and reading accuracy of the pupils in the trial. Pupils receiving the reading intervention with language comprehension experienced the equivalent of about six months of additional progress. For pupils receiving the standard reading intervention the figure was four months. These effects are unlikely to have occurred by chance.

However, the impact of the interventions on pupils' reading comprehension in particular, measured using a combination of reading comprehension tests, was much smaller. These effects are also more likely to have occurred by chance. It is therefore not possible to say with confidence that the REACH interventions improve reading comprehension. This is true even for the intervention which had greater focus on language comprehension. The process evaluation also revealed that TAs generally reported that the language comprehension component was the most difficult to deliver and that in some cases pupils became bored by it. It was suggested that it could be more varied and segmented into shorter pieces.

The process evaluation revealed a number of areas where schools felt the programme could be improved. Of the TAs interviewed for the process evaluation, many said they did not feel confident in delivering the intervention after the initial five days of training without ongoing support, and most agreed that more focus on the practical elements of delivering the interventions would have been helpful. In practical terms, the 35-minute long sessions were not well matched with standard one hour school lessons. The evaluation also suggests that some lead-in time for schools is valuable. Schools in the second phase, which had more notice of the interventions' introduction, were noticeably more prepared than those in the first phase.

Although the overall evaluation results are promising, it is important to note the concerns over the security of these findings. These include the phasing of the trial, the fact that not all pupils completed the tests, and some differences in the characteristics of pupils in the treatment and control groups.

How much does it cost?

The programme is relatively cheap to buy, but requires significant delivery time from teaching assistants. The cost of the materials for each intervention is £486 per TA. The cost of a trainer for five days was £2,500. This trainer could train a number of TAs and so this cost could be split between a number of schools. If the training were held at a hotel or training centre, as opposed to a school, there would be an estimated additional cost of £28 to £35 per day for each TA. Each TA could then deliver the intervention repeatedly. In terms of staff time, the intervention requires TAs to deliver three 35-minute one to one sessions with each pupil involved for 20 weeks.

Table 1: Summary of impact on reading skills

Group	Effect Size (95% confidence interval)	Estimated months' progress	Security rating	EEF cost rating
REACH reading intervention vs. standard provision	0.33 (0.14; 0.52)	4 months		£
REACH reading intervention with language comprehension vs. standard provision	0.51 (0.34; 0.68)	6 months		£

Note: See notes to Table 11 for more details on how the impact estimates were calculated. See the 'cost' section in the 'impact evaluation' chapter for more detail on the EEF cost rating.

Introduction

Intervention

REACH is a targeted reading support programme designed to improve the reading comprehension and accuracy of those with reading difficulties in Years 7 and 8.

The interventions targeted pupils who demonstrated poor reading skills, selected on the basis of a Single Word Reading Test (SWRT)¹ administered to all those who scored below level 4 at Key Stage 2. Trial participants were randomly assigned to three equally sized groups: two treatment groups and one waitlist control group. Both treatment interventions were delivered by trained teaching assistants working with selected pupils, one to one, for 35 minutes three times per week for 20 weeks. Pupils were taken out of classes to receive the intervention, with the survey of TAs suggesting this was mostly from subjects other than English. We are therefore assessing a different method or intensity of teaching rather than any additional hours of teaching, though there may have been a greater focus of time on literacy rather than other subjects.

The first treatment group received a reading intervention based directly on the original 'Reading Intervention' developed by York, and referred to throughout as REACH RI, that consisted of the following activities:

- reading books at an appropriate level and addressing any errors—
 - while the child reads out loud the TA keeps a 'running record', recording verbatim all reading errors;
 - the running record forms the basis of teaching points to be addressed in the session;
- instruction in phoneme awareness (learning to identify and manipulate phonemes in spoken words);
- instruction on taught letter sounds; and
- phonological linkage training—learning to apply knowledge of letter sounds and phoneme awareness to the task of identifying new printed words, and work that reinforces this understanding through spelling.

The second treatment group received a similar reading intervention programme which had a greater focus on language comprehension (referred to throughout as REACH LC). The language comprehension material was based on the approaches used in the York READING for MEaning Project,² and consisted of metacognitive strategies (building strategies for approaching text), reading comprehension, making inferences from text, written narrative (writing stories), and vocabulary training using a multiple context learning approach.³ The two programmes differed in content, but the time allocated to delivery was the same.

The impact of each treatment on a range of literacy measures was estimated by comparing outcomes to those in the control group. The delivery team took the decision that the control group would receive the second treatment arm (REACH LC) after the initial 20-week intervention had been completed.

Each school participating in the trial was asked to provide between one and three TAs to deliver the intervention depending on how many children were receiving it. The TAs received five days of training provided by the project team (principally Glynnis Smith, Paula Clarke and Charles Hulme), with support from their research assistants. Each training day ran from 9.30am to 3.30pm. The training plan is provided in Appendix A.

¹ Foster and NFER (2008).

² Clarke *et al.* (2010). For more information please see <http://readingformeaning.co.uk/>

³ For more details see <http://readingformeaning.co.uk/materials/>

All TAs were given a teaching pack that included session-by-session guidelines, general teaching principles, copiable resources, and progress monitoring sheets. The project team also provided email and telephone support throughout the trial. Each school was also provided with a book box of levelled books and a range of supplies including tactile letters, phoneme charts and figurative language cards. When requested the team would visit schools to provide on-site assistance.

Background evidence

Children with poor literacy skills when they start secondary school often make slow educational progress and therefore face poor labour market prospects (Crawford and Cribb, 2015). There is thus substantial interest in understanding how to improve literacy skills during early secondary school, and whether such improvements increase subsequent educational attainment.

Reading comprehension requires two skills: decoding (learning to translate written text into spoken language) and language comprehension (understanding the decoded words). Many children with poor reading comprehension can read accurately by decoding words, but have a limited understanding of what those words mean (Nation and Snowling, 1997).

The REACH RI programme has grown from a long established line of research carried out in the Centre for Reading and Language at York. A series of trials has shown that the ‘Reading Intervention’ approach is an effective and reliable method of improving reading skills in primary school pupils with reading difficulties. Estimated effect sizes have typically ranged from 0.4 to 0.6 standard deviations (Hatcher *et al.*, 2006). The REACH LC approach was developed following a successful randomised controlled trial (RCT) that compared three different interventions designed to improve language comprehension. This highlighted the importance of oral language training in improving text comprehension (Clarke *et al.*, 2010).

Following an RCT in North Yorkshire (Hatcher *et al.*, 2006) the Local Education Authority included the approach in TA training both for primary school TAs, and for those working with Years 7 and 8 in secondary schools. Results from extending the approach to secondary school pupils suggest it is just as effective as for younger pupils. On the basis of these positive results, the EEF commissioned a randomised controlled trial to provide robust evidence of the effectiveness of the Reading Intervention method in secondary school. This trial forms one of the various projects undertaken as part of the EEF’s ‘Literacy Catch Up Round’ which has sought to find programmes that are effective at supporting pupils during the transition from primary to secondary school.

Evaluation objectives

The main research questions for this evaluation are those published in the original evaluation protocol⁴ and are as follows:

- What is the impact of receiving REACH RI or REACH LC in Years 7 and 8 on reading skills after 20 weeks?
- Do the reading skills of children that receive the reading intervention plus language comprehension training improve more than those who receive the reading intervention alone?
- To what extent do differences in the effects [between the treatment and control groups] of the two programmes continue or fade out once the programmes have finished?

The process evaluation will provide qualitative evidence to improve understanding of programme implementation and highlight any issues that may be relevant to future roll-out of the programme.

⁴ The protocol is available at:
[http://educationendowmentfoundation.org.uk/uploads/pdf/Transitions_-_REACH_\(UCL\).pdf](http://educationendowmentfoundation.org.uk/uploads/pdf/Transitions_-_REACH_(UCL).pdf)

Ethical review

For the intervention, ethical approval was obtained from the University of Leeds AREA Faculty Research Ethics Committee; for the impact evaluation undertaken by IFS researchers, from the UCL ethics board. IFS researchers are also required to adhere to the Economic and Social Research Council's Ethics Framework, the Social Research Association's Ethical Guidelines, as well as the IFS Information Security guidelines and the IFS Information Classification and Handling Policy (both of which comply with the international standard for data security, ISO27001). Ipsos MORI carried out an internal review to ensure the evaluation adhered to the key principles outlined in the MRS Code of Conduct and Data Protection Act. Key considerations included ensuring that TAs had consented for their contact details to be passed to Ipsos MORI so they could be approached about the survey and case studies, and that participants in all elements of the primary research were able to make fully informed consent before participating.

Schools gave initial permission for pupils to be given a screening test. Pupils were eligible for the intervention if their scores were below a threshold. If there were fewer than 18 pupils falling below this threshold, all pupils were selected. If more than 18 pupils were eligible, the 18 pupils with the lowest scores were selected. Parents/carers of the selected pupils were then asked for informed written consent for pupils' participation in the research project. At this point, they were informed about the nature of the project, what data would be collected, linkage of data to the National Pupil Database (NPD), and how it would be used. Due to difficulties in obtaining written consent from parents, the University of Leeds AREA Faculty Research Ethics Committee agreed that verbal consent given over the phone from a parent to a member of staff at the child's school was sufficient. However, the ethics committee ruled that written consent was required for schools to release Unique Pupil Numbers (UPNs) for linkage to the NPD. There is therefore no NPD linkage for pupils whose parents only gave verbal consent. Parents were also regularly informed about the progress of the trial, contact details of the project team, and the option to withdraw their child at any point. For pupils whose parents/carers provided written consent, schools provided UPNs to enable linkage of pupils test scores to their NPD records. Not all schools provided UPNs to enable this data linkage. All consent forms are provided in Appendix B.

Evaluation Team

Lead researcher from IFS:

Luke Sibieta, Programme Director at IFS

Supported by:

Claire Crawford, Research Fellow at IFS

Elaine Kelly, Senior Research Economist at IFS

Agnes Norris Keiller, Research Associate at IFS

Annabelle Phillips, Research Director at Ipsos MORI

Fay Yorath, Research Manager at Ipsos MORI

Julia Pye, Research Director at Ipsos MORI

Rachael Emmett, Senior Research Executive at Ipsos MORI

Henning Sandberg, Research Executive at Ipsos MORI

Project and Delivery Team

Charles Hulme, University College London

Paula Clarke, University of Leeds

Glynnis Smith, Educational Consultant

Shirley-Anne Paul, University of Leeds

Trial registration

This randomised controlled trial was not registered by the evaluation or project team. However, the initial evaluation protocol is available on the EEF website:

[https://educationendowmentfoundation.org.uk/uploads/pdf/Transitions_-_REACH_\(UCL\).pdf](https://educationendowmentfoundation.org.uk/uploads/pdf/Transitions_-_REACH_(UCL).pdf)

Methods

Trial design

The trial was designed as a randomised controlled trial, with randomisation at the pupil level. The original trial design intended to identify and recruit children with reading difficulties and to randomly assign participating pupils to one of 3 groups:

- a group receiving REACH RI for 20 weeks;
- a group receiving REACH LC for 20 weeks; or
- a waitlist control group that would receive an intervention (whichever intervention appeared to have been most effective) after the initial 20-week trial had ended.

The main reasons for adopting within-school rather than across-school randomisation were (a) the increase in statistical power resulting from within-school randomisation, and (b) that any school drop-out would result in a loss of equal numbers of treatment and control pupils (rather than whole schools of treatment or control pupils). The main downside of within-school randomisation is that spillover effects are more likely if, for example, TAs applied the intervention techniques to pupils in the control group with poor literacy skills. If pupils in the control group were impacted positively the impact estimates would *underestimate* the impact of the treatments, whereas if they were impacted negatively the impact estimates would *overestimate* the impact of the treatments. In light of this concern, the training emphasised the importance of maintaining treatment and control conditions and the process evaluation explicitly asked TAs about whether they used skills gained from the trial with other pupils.

A waitlist control group was chosen as the project team determined that it would not be ethical to identify a group of struggling pupils and then not provide them any additional support (beyond 'business as usual' support).

Following difficulties in recruiting schools and pupils, the trial was run in two phases. The first phase included 12 schools and both treatments began in July 2013 when the pupils were in Year 7 and ended in January 2014, when the pupils were in Year 8. The second phase included 15 schools and ran from November 2013 to April 2014. Pupils in the second phase were in Year 7 throughout. The phasing of the trial was a major departure from the initial project plan and evaluation protocol. We discuss in detail below how we seek to address this issue in our analysis.

A final and important point to note is that the trial design does not allow us to identify the impact of the content of the intervention separately from (a) the provision of one to one teaching time or (b) the increased time devoted to literacy as pupils were typically withdrawn from lessons other than English. This issue of interpretation is common to most trials of this type.

Participant selection

Schools:

The original intention was to recruit 27 schools that fulfilled the following criteria:

- The pupil roll was above 800 (in order that there was likely to be a sufficient number of children in the school eligible to take part in the project);
- The school was within a 1-hour journey from the University of Leeds;
- Recent GSCE results indicated there were levels of underachievement; and
- Sufficient numbers of pupils were entering the school with English below level 4.

These eligibility criteria were relaxed in the latter stages of recruitment in order to maximise take-up. Specifically: the distance of the schools from Leeds University was increased so that schools could be reached within 90 minutes; GCSE results ceased to determine eligibility; and schools with smaller rolls were also approached if there was an indication that they had sufficient numbers of pupils entering Year 7 with English below Level 4.

Due to the recent expansion of academies in England that are outside of local authority control, it was decided to approach schools directly. Schools that met the selection criteria were telephoned and the name and email address of either the Head of Year 7 or the school's Special Educational Needs Coordinator (SENCo)—or equivalent post—were requested. An initial contact email was then sent giving a brief overview of the project, asking the school to express its interest and to provide the number of Year 7 pupils eligible for screening (those scoring below level 4 in English).⁵ Schools that expressed interest were sent an information pack that had to be returned to Leeds University with the headteacher's signature.⁶ For those who did not respond to the initial email, the project and delivery team sent a follow-up email and attempted telephone contact.

Approximately 70 schools were initially contacted regarding phase one of the project, and 150 schools regarding phase two. However, there was some crossover in the schools approached across each phase. A total of 207 schools were approached for participation in one of the phases (70 in phase one and a further 137 in phase two). Approximately 50% of the schools contacted did not respond. Out of those who declined, reasons given were that:

- The school was already participating in at least one form of literacy intervention.
- The TA's workloads were already very demanding.

Some of the reasons for accepting the intervention were that:

- The school was keen to help pupils who were under-achieving in English.
- The school was keen to participate in a literacy intervention.

Phase one of recruitment was completed in June 2013, followed by the completion of phase two in November 2013. Over the two phases 27 schools were recruited (12 schools in phase one and 15 in phase two).

Pupils:

To identify eligible pupils, the Single Word Reading Test (SWRT)⁷ was administered to all Year 7 pupils in participating schools who had scored below level 4 at Key Stage 2. The 18 lowest-scoring pupils in each of the 27 schools would then be selected to participate in a trial, resulting in an intended sample of 486 pupils and 162 pupils per group.

The contacts at the school were asked to send out a parent's information pack⁸ to the parents of the eligible children. Parents were asked to read the project information sheet and then complete and return the parent/guardian consent form to the school.

For those parents from whom it was not possible to obtain written consent, a Verbal Consent Information Sheet and Form (a shortened version of the aforementioned form) was completed by the School Contact. Completion of this form involved the School Contact telephoning the parent and asking for verbal consent by talking through the consent statements. The date and time of the conversation was logged on the verbal consent form and the School Contact ensured that the parent

⁵ See Appendix B.1

⁶ See Appendix B.2 and B.3

⁷ Foster and NFER (2008).

⁸ See appendix B.4 and B.5.

had the opportunity to ask questions. If verbal consent was given, a copy of the consent form was signed on the parent's behalf and sent out to the parent.

For parents that gave verbal consent, subsequent efforts were made to gain written consent in order to ensure written consent for linkage to the NPD. We only linked pupils to the NPD where parents had given informed written consent, as per the requirements of the University of Leeds AREA Faculty Research Ethics Committee.

The screening process indicated that fewer pupils than anticipated were eligible for the intervention. As a result, fewer pupils per school were recruited to participate in the intervention. Instead of the expected sample of 486 pupils, the actual sample consisted of 287 pupils of whom 117 received the intervention during phase one and 170 received the intervention during phase two.⁹

Outcome measures

The primary outcome measure for this evaluation is the New Group Reading Test (NGRT) score. This test is used for all EEF projects funded through the EEF 'Literacy Catch Up' round of projects.¹⁰ The NGRT seeks to measure reading and comprehension for various age groups. The NGRT test appropriate for this age group (Years 7–8) comprises 20 sentence-completion items and various comprehension questions based on the reading of three passages.¹¹

The NGRT score was recorded at three points: just before the start of each intervention (pre-test); after 10 weeks; and after 20 weeks (after the end of the intervention). This represented a deviation from the protocol. The original intention was to record NGRT scores just twice (at the beginning and end of the intervention). However, miscommunication with the various parties involved led to the additional testing at 10 weeks. It is possible that pupils could have become more familiar with the NGRT test as a result of this deviation. However, as this would apply to both treatment and control pupils, it is unlikely to affect our estimates of the impact of the intervention.

The secondary outcomes are:

- a composite reading comprehension score based on the reading comprehension component of the York Assessment of Reading Comprehension test (YARC-RC),¹² and WIAT II reading comprehension test scores;¹³ and
- a composite reading accuracy score based on the YARC Single Word Reading Test (SWRT), and the TOWRE measure of word reading efficiency.¹⁴

These scores were recorded at pre-test, 20 weeks and nine months after the end of the intervention. As such, they represent the only outcomes considered for the follow-up analysis at nine months. However, it should be noted that the control group had received the intervention by the nine-month follow-up. We therefore can only meaningfully compare the scores for the two treatment groups.

The WIAT II reading comprehension and TOWRE measure of word reading efficiency were not included in the original evaluation protocol published on the EEF website. They represented additional outcomes collected by the project and delivery team and were added to an updated version of the protocol before the trial began (not published on the EEF website, but available on request).

⁹ Almost of all of this reduced sample size is attributable to lower eligibility rather than lack of parental willingness to consent.

¹⁰ Please see <https://educationendowmentfoundation.org.uk/news/literacy-catch-up-projects/> for a summary of other interventions funded in the same round.

¹¹ More information is available here (<http://www.gj-assessment.co.uk/products/new-group-reading-test/test-detail>).

¹² Snowling *et al.* (2009). For details of these tests see <http://www.yarcsupport.co.uk/about.html>

¹³ Wechsler (2005).

¹⁴ Rashotte *et al.* (1999).

Details on how the secondary composite outcomes were calculated can be found in the 'Outcomes and analysis' section of the impact evaluation below.

All tests were administered and scored by research assistants trained by the project and delivery team who were blind to the allocation of pupils to groups. At no point were grouping details shared with the testers. They were instead given a list of pupils to work with. The schools were also briefed at the training about the importance of blinding to group allocation and were advised not to reveal to the testers which children were in each group. The importance of this was also stressed at the training that was done with the testers.

Sample size

The project team aimed to recruit a total of 487 pupils across 27 schools, with pupils spread equally across the two treatment groups and the control group in each school. Recruitment was not as successful as hoped for and had to be spread across two phases. In phase one, the project team recruited 12 schools and 117 pupils. A further 15 schools and 170 pupils were added in phase two. Across the two phases, the project team therefore managed to reach their target number of schools, but fell well short of the expected number of pupils (287 compared to a target of 487). This was because fewer pupils than expected met the eligibility criteria.

Our initial power calculations assumed a central scenario where 67% of the variation in post-test was explained by pre-test characteristics (also assuming 80% test power and 5% significance level). In this central scenario the sample size of 287 achieved would be sufficient to detect an effect size of 0.24 SDs.

In reality, pre-test characteristics explained almost exactly two thirds of the variation (as expected). However, sample sizes further dropped due to school attrition. As discussed in the participants section of the impact evaluation below, the net effect was to further increase the minimum detectable effect size to 0.28.

This is based on pooling phase one and phase two pupils together. However, the experimental conditions differed across phases (different age levels, delivery during different points in the school year, and different levels of preparation). This could imply different impacts of the intervention across phases. As detailed in the 'Impact evaluation' section, estimating the impact of the interventions by phase further increases the minimum detectable effect sizes (to 0.42 for phase one and 0.32 for phase two).

Randomisation

The randomisation was undertaken by the evaluation team. Randomisation occurred within school and explicitly sought to minimise differences across groups in terms of age, gender and screening test score (SWRT).

In phase one, the project team restricted the number of pupils per school recruited to the intervention to be multiples of three. Hence, if there were ten pupils eligible for the trial, only the nine with the lowest scores were recruited. During phase one there were an average of 9.75 pupils per school compared to a target of 18. In phase two, the within-school sample was not restricted to being a multiple of three, and the mean number of pupils per school rose to 11.3.

The randomisation process was performed separately for each phase but followed an identical procedure. Where pupil numbers were multiples of three, allocations were achieved using a single random number generator. This was the case with all phase one schools. Where the numbers of pupils per school were not multiples of three, these additional pupils were allocated using a

randomisation algorithm designed to ensure sample sizes were as even as possible across groups, both within schools and for the sample as a whole.

Random assignment will, on average, lead to small and statistically insignificant differences between each group in terms of gender, age and SWRT score. However, in any particular random draw it is possible that larger, significant differences can arise purely by chance—for example, one group might have a disproportionately large share of pupils born in the summer months.

We used an iterative procedure to identify an ‘optimal’ random assignment. The process outlined above was carried out, and then two diagnostic checks were performed. First, the three groups were compared to each other in terms of age, gender and SWRT score, and the number of statistically significant differences was recorded.¹⁵ Second, the difference in average characteristics between the three groups was calculated.¹⁶

For each iteration, these two numbers were stored. The randomisation process was repeated 1,000 times, resulting in 1,000 different allocations. To identify the optimal randomisation, we first restricted our attention to the random assignments that led to zero significant differences between groups in terms of age, gender and SWRT score. Among this set of assignments, we then selected the one that yielded the smallest value of the total differences in average characteristics. This was the final REACH treatment allocation that we shared with the project team.

Analysis

The original evaluation protocol expected the intervention to be run across a single phase. The fact that the intervention was run across two phases therefore represents a major departure from the protocol. The effect of the intervention could also differ across the two phases as a result of different experimental conditions (different age groups, different points in the school year, and different levels of preparation), and the two phases could be treated as separate experiments. However, the sample sizes are relatively small for individual phases, meaning that estimates of the effects of the intervention by phase are likely to be quite imprecise. In response to this issue, our main estimates of the effect of the REACH interventions combine pupils across phases (in keeping with the spirit of the original evaluation protocol). We then perform a prominent robustness check where we estimate the results separately by phase.

In our analysis, we present both raw comparisons and analysis that accounts for pupil characteristics and baseline test scores,¹⁷ with the latter representing our preferred estimates. Raw comparisons of pupil test scores between treatment and control groups should in principle provide unbiased estimates of the effect of the intervention if the randomisation has been successful. Methods that account for pupil characteristics will also yield unbiased estimates, but should produce more precise estimates as a greater amount of the variation in test scores can be accounted for.

The preferred method used to account for the pre-test and pupil characteristics is Fully-Interacted Linear Matching (FILM).¹⁸ FILM allows the effect of the treatment to vary linearly with the pre-test and pupil characteristics which means that it is more flexible than Ordinary Least Squares (OLS) regression. FILM is more restrictive than propensity score matching which implicitly allows the effect

¹⁵ This was assessed by regressing membership of a particular group relative to another on the baseline covariates of interest (gender, age, and SWRT score).

¹⁶ This was assessed by adding together the absolute difference in mean outcomes (rescaled by the standard deviation) between Group 1 and Group 2, Group 1 and Group 3, and Group 2 and Group 3.

¹⁷ In particular, we control for gender, age, and pre-treatment test scores. For the subsample of pupils with observed NPD records, additional characteristics that were controlled for are: whether pupils are recorded as having SEN (statement or school action plus), whether pupils are eligible for FSM, whether pupils have English as an Additional Language (EAL), deprivation of the pupil’s residential neighbourhood as measured by the IDACI percentile rank, and KS2 Maths and English fine point scores.

¹⁸ Blundell *et al.* (2005).

of the treatment to vary non-linearly with baseline characteristics. However, kernel matching is less precise than FILM and OLS as the standard errors must be estimated using the bootstrap method. FILM was chosen as it represents a compromise between precision and flexibility. Nevertheless, to ensure the robustness of the results, robustness checks were performed by comparing treatment effect estimates across four alternative methodologies (raw comparison, OLS, FILM, and propensity score matching, see Table D3). We will be depositing the evaluation dataset with the UK and EEF data archives; this will enable other researchers to further test the robustness of our findings by using alternative methodologies or assumptions.

All outcomes are standardised by the unadjusted standard deviation within the estimation sample (either pooled across phases or within phase depending on whether the analysis is pooled or separate by phase). As a result, we are able to estimate the effect size as the coefficient on a treatment dummy variable.

To account for the experimental design, robust standard errors are clustered at the school level to allow for correlation of pupil outcomes within schools. This approach is used across all methods presented in the paper. Another way to account for the experimental design in our analysis is to also allow pupil outcomes to explicitly depend on the school that they attend. This could take the form of a school effect that is assumed to be uncorrelated with all observable pupil characteristics (a random effects model) or one can explicitly estimate the individual effects of schools (a fixed effects model) (Wooldridge, 2010). Neither of these approaches should affect our estimates of the impact of the programme if the number of pupils in the treatment and control groups is equal across schools. However, estimating the treatment effects using these alternative methodologies represents another robustness check on our impact estimates (see Table D3). The random effects model is also identical to a hierarchical linear model with random intercepts (Raudenbush and Bryk, 2002).

The fact that the random assignment was re-run until balance was achieved has implications for the analysis which are still being debated. Both Bruhn and McKenzie (2009) and Scott *et al.* (2002) suggest that the most practical approach is to control for all covariates used in the randomisation, which we always do. Morgan and Rubin (2012) go further and show that standard errors calculated in the normal way are likely to be too conservative. They show that one can instead perform randomisation or permutation tests to perform inference which are likely to generate smaller confidence intervals. However, these methods are still relatively new and only valid where a specific criterion has been used to determine acceptable randomisations (we instead chose the randomisation with the 'best' level of balance). We therefore still use conventional standard errors, but accept these are likely to be too conservative.

To allow us to account for a wider range of pupil characteristics, NPD records were obtained for pupils whose parents had provided written consent to link to the NPD. Because not all parents gave such consent, the analysis that considers these additional pupil characteristics is conducted on a subsample.

Eligibility for free school meals (FSM) is recorded in the NPD, and it is therefore theoretically possible to examine whether the intervention was more or less effective at improving the reading skills of these relatively deprived pupils. However, attrition and the lack of written consent from all parents mean that the sample sizes for such analysis are extremely small. Combining pupils across phases, the sample sizes for each group would be just under 20 pupils eligible for FSM. Analysis based on such small samples is unlikely to be secure. The small sample sizes mean there are likely to be very wide confidence intervals and any results could just be due to imbalance.

Follow-up analysis was conducted using the secondary composite outcomes that were used for the post-test analysis. These tests were conducted about nine months after the end of the interventions. By this time, pupils in the control group had already received the REACH LC intervention. A comparison between the original treatment groups and the control group after this follow-up period is

problematic, as any estimated differences could be the result of a combination of fade out, differences in the age of pupils when the treatment was administered, and TA learning improving the quality of delivery. However, the post-test does provide the opportunity to examine if differences across treatment groups persist to the follow-up test. To achieve this, we compare follow-up test scores for pupils receiving the REACH LC with those receiving the REACH RI (excluding the control group). The difference between the two will tell us whether differences between the two treatments persisted.

All analysis was conducted using Stata 13 and undertaken on an intention-to-treat basis which conforms with EEF guidance for interventions of this type. The syntax used is clearly documented and available to access from the UK data archive.

Implementation and process evaluation

The process evaluation used a mixed-method approach to provide context for the impact evaluation and an understanding of how the project was implemented. The approach had two main elements: (1) a survey of TAs delivering the programme that was combined with data collected as part of the intervention, and (2) in-depth interviews conducted with six schools.

The main difference from the protocol published on the EEF website is that was not possible to analyse session plans or delivery logs of TAs (the next section explains why and how the survey was amended to collect similar information).

TA survey and interviews

The survey was a self-completion questionnaire, and comprised 25 questions over seven pages. It consisted of predominantly closed questions with some open-ended questions. The survey is included in Appendix C.

The survey was sent to all TAs who gave permission to be re-contacted for follow-up during a questionnaire that was filled out when they agreed to participate in the intervention. The aim of the survey was to understand the TAs' perception of: how well the intervention had been administered and managed in their school; how well the training equipped them for the intervention; and how the school facilitated the sessions within the school.

Originally the process evaluation team had hoped to analyse TAs' session plans in order to understand (1) which lessons the pupils were withdrawn from for each session, (2) how the pupil engaged in specific sessions, and (3) how TAs felt about the sessions overall. However, due to logistical and financial reasons it was not practical to collect this information via session plans. Instead it was agreed that equivalent information could be captured through the TA survey retrospectively, and questions were added to this in order to supplement the information.

The surveys were sent out via email to the TAs from an Ipsos MORI email address.¹⁹ The accompanying text emphasised the confidentiality and anonymity of the responses so that TAs could feel confident in giving their full and frank opinions about the intervention. The text encouraged TAs to send the completed questionnaire back to the same email address, or to one of the process evaluation team members.

On the basis of the TA questionnaires, six schools were selected to act as case studies to explore questions raised in further detail. The selection of these case studies was made after the project surveys had been completed. In total, two case studies were conducted with phase one schools and four with phase two schools.

¹⁹ reachprocessevaluation@ipsos.com

The case studies were selected based on the following criteria:

- a spread of schools (in order to avoid selecting more than one TA from the same school);
- a spread in terms of TAs' number of years' experience and the number of students they had; and
- getting a mix of positive and negative views (although overall views of the intervention were very positive across the board, some expressed dissatisfaction with aspects of the organisation, the administration involved, and school support).

In each case, we aimed to interview the TA, a member of the school's senior leadership team such as the Deputy Headteacher, and the SENCo. However, this sometimes varied depending on how the intervention was run in schools—for example, in one school a teacher rather than a TA had been involved in delivering the intervention. Up to three members of staff were interviewed in each case-study school (all interviewed separately). The SENCo was often closely involved in the programme and could provide useful overview information about the implementation of the programme. Through these case studies we were able to gather a broader view on how the programme worked from the perspective of participating schools. We aimed to cover those expressing varying degrees of satisfaction with the programme, including those who were very satisfied and dissatisfied, and those who were neutral.

Interviews were conducted by telephone in order to, as much as possible, fit around the working day of the TAs. Each interview lasted about 30 minutes. In order to facilitate the interviews we used a discussion guide that was tailored to each discussion, depending on the survey answers. The interviews focused on sharing best practice, identifying sticking points, and making recommendations for improving the processes in future.

Timeline

The overall timeline for the intervention and evaluation is shown in Table 2 below. The project started with recruitment of schools by the REACH project team from January 2013 onwards. However, difficulties in recruiting schools led to the trial being split into two phases. Schools already recruited would commence the intervention in the summer term (with pupils towards the end of Year 7), while further efforts were made to recruit more schools for inclusion in a second phase to start at the beginning of the following academic year (for pupils at the beginning of Year 7).

For phase one schools, the training of TAs was conducted in the spring of 2013. Eligible pupils were then screened and randomised into one of the two treatment groups or the control group in June 2013. The pre-test data was collected from all eligible pupils and the intervention commenced soon afterwards. The intervention stopped at the end of the summer term, and recommenced when the pupils started Year 8 in September. The 20-week intervention ended in January 2014 and post-tests were administered in January or February of 2014.

For phase two schools, screening of pupils and TA training began in September or October 2013. In November 2013, the pupils were assigned to their treatment or control group and the intervention began. The 20-week intervention was completed in April 2014 and post-tests were completed in June and July of 2014.

Table 2: Project timeline

Activity	Phase 1 date	Phase 2 date
Recruitment of schools commences	January 2013	June 2013
Screening followed by pre-test (Test 1)	June 2013	October 2013
Pupil random assignment to groups (evaluation team)	June 2013	November 2013

Block 1–10 weeks	July–October 2013 (3 weeks in Summer term, 7 weeks in Autumn term)	December 2013–February 2014
Mid-test (Test 2)	October 2013	February 2014
Block 2–10 weeks	November 2013–January 2014	March–June 2014
Post-test (Test 3)	January–February 2014	June–July 2014
Survey of TAs	February 2014	June 2014
Control group: Block 1	March–May 2014	September–November 2014
Control group: Block 2	May–July 2014	November 2014–February 2015
Follow-up test (Test 4)	October–November 2014	February–March 2015
Final report to EEF	July 2015	July 2015

Costs

To calculate the cost of the intervention, we rely on information recorded by the project team on the monetary cost of the training and materials, as well as expected time commitments from staff involved. Monetary costs are presented in terms of each staff member able to deliver the programme, as well their expected time commitment. In addition, we indicate how many pupils this is expected to cover. Schools could deliver the interventions to smaller or larger number of pupils at their discretion, though there is no guarantee that the expected impact will be the same as that estimated in this evaluation.

This trial was commissioned before new guidance from the EEF on the systematic collection of cost data. All new trials are expected to collect such data in order to produce cost estimates in line with EEF guidance.

Impact evaluation

Participants

The original intention was to recruit 27 schools and 18 pupils per school (giving an overall sample of 486 pupils). Over the course of the experiment, six schools dropped out of the trial and no post-test data was collected for any of these pupils. Our estimates of the impact of the intervention are based on 202 pupils across the 21 schools that completed the intervention. In this section, we describe how the number of participants changed at different stages of the experiment, and the determination of the overall estimation. This is illustrated by Figure 1 which shows the flow of participants at different stages of the trial, and Table 3 which provides more detail on the test scores collected from pupils at different stages. Table 4 summarises the impact of the reductions in sample size at different stages on the minimum detectable effect size (through the use of power calculations).

Schools were recruited close to the project team in Leeds for practical reasons. Schools in disadvantaged areas were targeted in line with the EEF focus on disadvantaged pupils. Recruitment was split across two phases: 12 schools were recruited in phase one and 15 in phase two.

In each of the 27 participating schools, the parents of the pupils that scored under 90 in the screening SWRT (or the 18 lowest in the one school that had more than 18 eligible) were invited to give consent for their child to participate in the trial. The original intention was to recruit 27 schools and 18 pupils per school. However, although the project team reached their target number of schools, fewer pupils were eligible than expected, reducing the initial sample to 287 pupils at randomisation (117 in phase one and 170 in phase two). One of these schools dropped out shortly after the randomisation and so recorded no pre-test scores either (except for the SWRT used for screening). The net result of this dropout was a reduction in sample size of 57 pupils (from 287 to 230).

The process evaluation team evaluated the reasons for withdrawing through analysing the motives these schools gave and through a case study with one of the withdrawn schools during the process evaluation. Many of the reasons given for withdrawing were related to difficulties in communications between the schools and the project team. For example:

- Many schools changed the contact person responsible for liaising with the REACH team which made communication more difficult.
- Communications from the delivery team did not always take into account the way the intervention was treated in schools: in some schools the project was seen as part of literacy provision, and in others part of special needs provision.
- The lack of communication between the REACH contact person and senior management in some schools created issues.
- There was also a lack of information given to TAs who attended the training: many were unaware of the scope of the project.

The case study that was conducted with a school that withdrew from the intervention highlighted many of the same issues that arose in the schools that had participated in the process evaluation. The two TAs the school sent for training believed that the time commitment exceeded the time that had been allocated for them to deliver the intervention. The SENCo of the school believed that much of the hesitation about the time commitment required to deliver the trial was down to the experience and confidence level of the TAs. The TAs and senior members of staff at the school felt that too much time was involved to justify using full-time TAs to deliver REACH: it was neither feasible given their workload, nor proportionate given the needs of the school. There were also concerns about the lessons pupils should miss in order to participate, and which lessons the TAs should miss to deliver the trial. It was deemed that pupils could not be taken out of certain classes as they would fall behind.

A member of the Senior Leadership Team in this case-study school believed that senior members of staff at the school, including the SENCo, had not engaged enough with the material and information at the early stages of the intervention. Therefore they did not realise the complexity of the intervention, or the time commitment that it would involve. Ultimately, even though extra support was offered by the University of Leeds, the headteacher made the decision to withdraw, citing 'significant impact on staffing' as the ultimate reason.

In addition to school and pupil dropout, the potential estimation sample was reduced further by missing pre- and post-test data. Table 3 provides further detail about how the sample changed at different stages of the trial. Collecting the YARC-RC scores proved particularly problematic as the reading skills of some pupils were too low to record any score and other pupils were only able to complete one of the two passages.²⁰ It was therefore decided to remove the YARC-RC score from the baseline measures. Excluding YARC-RC, 17 pupils did not record any post-data which further reduced the sample size to 213 pupils with some post-test data. In addition, pre- and post-test outcomes were missing on some dimensions for the remaining pupils. We decided to exclude pupils who had incomplete pre- or post-test records. Incomplete post-test records led us to exclude five pupils and incomplete pre-test records led us to exclude a further five. The developers reported that these incomplete records are largely attributable to pupils' choosing not to complete all assessments. One pupil recorded pre-test and post-test NGRT outcomes that implied a gain score that was over six standard deviations above the average. Given that pilot studies had found an effect size of 0.4–0.6, this observation was deemed an outlier and has also been excluded.

This gives a final estimation sample of 202 pupils: 70 in the REACH RI group, 69 in the REACH LC group, and 63 in the control group.

At this point, it is important to state that we are estimating the effects of the interventions at schools that continued with the intervention and recorded post-test outcomes. This is estimated on an intention-to-treat basis as we do not take account of the amount or dosage of the intervention received by individual pupils. This could be biased if the effect of the interventions differs across schools—and schools that dropped out differed in this respect. It is impossible to know the size or direction of any bias. However, we would speculate that the effect of the intervention would have been smaller among schools that dropped out as the expected benefit is likely to have been one factor determining continued participation.

²⁰ Pupils that made more than 16 accuracy errors on the first passage of the test recorded no score. Where children were able to read the first passage with 16 errors or fewer, accuracy scores were recorded and the comprehension questions were administered. Pupils were then invited to attempt the second passage. The second passage is more complex than the first and was again discontinued if the pupils made more than 16 errors. This meant that some pupils recorded no score on either passage, some scored on one passage, and some scored on two.

Figure 1: Flow of participants at different stages of the trial

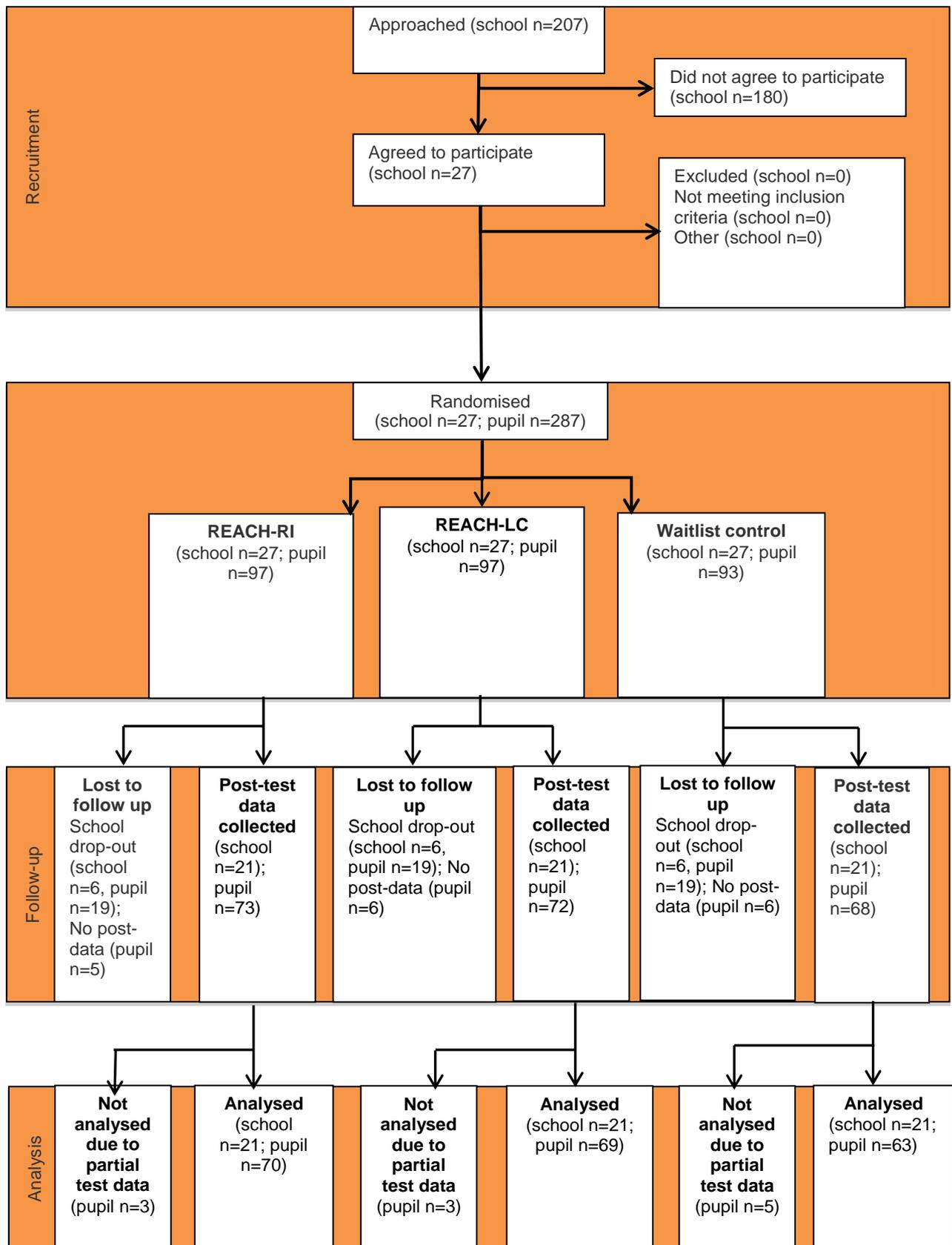


Table 3: Number of schools and pupils at randomisation and in final sample

	At randomisation	Baseline outcomes observed (exc. YARC-RC)	Baseline outcomes observed (inc. YARC-RC)	Post-test outcomes observed (exc. YARC-RC)	Post-test outcomes observed (inc. YARC-RC)	Final estimation sample
Schools	27	26	26	21	21	21
All pupils	287	235	165	203	140	202
Treatment 1	97	79	59	70	52	70
Treatment 2	97	80	54	69	46	69
Control	93	76	52	64	42	63

The reduction in sample size compared with the original plan—and compared with the randomisation—has some predictable consequences for the minimum detectable effect size. Table 4 shows the implications of power calculations for the minimum detectable effect size expected at different stages of the trial (assuming power of 80%, a significance level of 5%, and that baseline characteristics can explain two thirds of the variation in the final outcome). The initial plans implied a minimum detectable effect size of around 0.18. This quickly increased to 0.234 given the lower number of pupils at the randomisation stage. At the analysis stage, baseline characteristics and pre-test outcomes explained about two thirds of the variation in the primary outcome (as expected). The net result is a small increase in the minimum detectable effect size to 0.280.

It is also informative to calculate the minimum detectable effect sizes if we estimate the effects separately for each phase. As the sample size in each phase is lower, the minimum detectable effect sizes are larger. This increase in minimum detectable effect sizes is, however, dampened to some extent as we are able to explain a slightly larger share of the variance in the post-test outcomes by phase, increasing the degree of precision.

Table 4: Minimum detectable effect sizes at different stages

Stage	Number of pupils (T1, T2, C)	Correlation between baseline characteristics & post-test	ICC	Randomisation Method	Power	Alpha	Minimum detectable effect size (MDES)
Planned	486 (162, 162, 162)	0.67	n/a	Within School	80%	0.05	0.179
Randomisation	287 (97, 97, 93)	0.67	n/a	Within School	80%	0.05	0.234
Analysis	202 (70, 69, 63)	0.67	n/a	Within School	80%	0.05	0.280
Analysis by phase							
Phase 1	77 (26, 26, 25)	0.71	n/a	Within School	80%	0.05	0.423
Phase 2	125 (44, 43, 38)	0.73	n/a	Within School	80%	0.05	0.322

Pupil characteristics

The randomisation method ensured there were no statistically significant differences in gender, age or SWRT score across groups at the time of randomisation. However, reductions in pupil numbers that occurred post-randomisation could have resulted in some imbalances across groups. Characteristics that weren't considered during the randomisation may also vary between groups.

Table 5 presents the average characteristics of pupils in the treatment and control groups in the final estimation sample, with data combined across both phases. As well as gender, age, and SWRT baseline scores for the NGRT, WIAT comprehension, YARC-RC and TOWRE word efficiency tests are also included. The table also presents the differences between the treatment and control groups as effect sizes (the standardisation took place within the estimation sample).

For pupils where (written) permission was given to link to the NPD, we also show the proportions of pupils: eligible for FSM; with English as an additional language; with special educational needs (statement or school action plus); and who are not White-British. We also show the average KS2 points scores, where these are available. These variables are only available for a sub-sample as not all parents/carers provided written permission for linkage to the NPD, so the data might not be representative of the full cohort. The sample sizes are also smaller for the YARC-RC assessment as not all pupils recorded results for this measure (see earlier section on outcome measures for more details).

The figures in Table 5 also allow us to illustrate the characteristics of pupils in the trial and whether there are any imbalances in pupil characteristics across treatment and control groups. We now organise our discussion of these results into three parts: characteristics considered at randomisation; baseline or pre-test scores; and finally, additional characteristics from the NPD.

The final sample remains balanced in terms of the characteristics considered at randomisation (age, gender, and SWRT scores). Other baseline or pre-test outcomes were not considered during the randomisation process that allocated pupils to the different intervention groups. As a result, differences between groups are more likely for these variables. Looking at Table 5, we indeed see some differences across treatment and control groups, though none are statistically significant. For example, treatment 1 is about 0.12 standard deviations behind the control group on the NGRT (primary pre-test outcome), while treatment 2 is about 0.16 standard deviation ahead. We see differences of a similar scale for the WIAT, but slightly larger differences for the YARC-RC (with treatment 1 being over 0.3 standard deviations behind the control group). For the TOWRE and SWRT, differences are extremely small. The overall impression is that the groups are generally well-balanced, though there are some notable absolute differences in the pre-test scores as might be expected given the small sample sizes and the multiple dimensions captured by the pre-test scores.

We are also able to consider the characteristics of pupils collected from the NPD (where parents gave written consent). This illustrates some of the key characteristics of pupils in the experiment. First, they are relatively deprived, with around 30% or over eligible for FSM across groups. They are highly likely to have special educational needs, with over 60% having a statement of special educational needs or recorded as School Action Plus across all groups. They are also more likely to have English as an Additional Language (over 20% across groups), than the average for England. Unsurprisingly given this finding, a relatively large proportion of pupils are not from White-British backgrounds (around 30%). As one might expect given the design, pupils also have relatively low KS2 English scores (an average of around three or below) and low KS2 Maths scores.

There are also some notable differences in these characteristics across groups. For example, pupils in treatment 1 are seven percentage points less likely to be eligible for FSM as compared with the control group, both treatment groups are around six to seven percentage points less likely to have

EAL and seven to eight percentage points less likely to be not White-British. Although these differences are sizeable, none are statistically significant.

Table 5: Comparison of baseline characteristics

Variable	REACH-RI (T1)		REACH-LC (T2)		Control	Total N
	Mean (sd)	T1 - C effect size	Mean (sd)	T2 - C effect size	Mean (sd)	
Characteristics considered at randomisation						
% female	0.43 (0.50)	-0.03	0.43 (0.50)	-0.02	0.44 (0.50)	202
Age (months)	142.54 (5.19)	0.1	142.72 (5.24)	0.14	142.03 (4.97)	202
SWRT at baseline	32.21 (8.24)	0.01	32.10 (7.95)	0	32.10 (8.70)	202
Additional baseline tests						
NGRT at baseline	230.10 (57.20)	-0.12	245.74 (49.93)	0.16	236.89 (61.17)	202
WIAT at baseline	102.87 (16.50)	-0.13	101.49 (17.88)	-0.21	105.13 (18.21)	202
YARC-RC at baseline	6.92 (2.16)	-0.35	7.19 (2.25)	-0.23	7.74 (2.61)	141
TOWRE baseline	74.64 (9.84)	-0.03	75.01 (10.53)	0	74.98 (12.26)	202
Additional characteristics from the National Pupil Database						
% Eligible for Free School Meals	0.27 (0.45)	-0.15	0.33 (0.48)	-0.02	0.34 (0.48)	184
% English as an Additional Language	0.21 (0.41)	-0.15	0.22 (0.42)	-0.14	0.28 (0.45)	184
% SEN (Statement or School Action Plus)	0.64 (0.48)	0.07	0.62 (0.49)	0.03	0.60 (0.49)	184
% Not White British	0.29 (0.46)	-0.16	0.28 (0.45)	-0.17	0.36 (0.48)	184
KS2 English Points	2.98 (0.83)	-0.09	2.88 (0.67)	-0.23	3.05 (0.76)	181
KS2 Maths Points	3.54 (0.78)	-0.05	3.58 (0.67)	0	3.58 (0.81)	180
Total Sample Size	70		69		63	202
Total Sample with NPD data	66		60		58	184

Note: * indicates that the difference in means (TX - C) is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in brackets.

As indicated in the methodology section, we also seek to estimate the effects of the intervention separately by phase. An important test for the credibility of such analysis is whether the groups are well-balanced within individual phases as well. Appendix Tables D1 and D2 therefore repeat the analysis separately by phase. This shows that the treatment and control groups remain balanced in terms of gender, age, and the SWRT (as we might expect given that the randomisation explicitly targeted such balance). However, there are some very sizeable differences across groups in terms of the pre-test scores and NPD characteristics that were not considered at randomisation.

For phase one, T1 (treatment 1) is over 0.3 standard deviations behind the control group in terms of the NGRT (primary pre-test), and T2 (treatment 2) is over 0.2 standard deviations behind the control group on the WIAT and TOWRE. For phase two, the differences are generally smaller. However, T2 is

over 0.3 standard deviations ahead on the NGRT and over 0.3 standard deviations behind on the YARC-RC. Although none of these differences are statistically significant, these are large differences and therefore lead us to doubt whether the groups are well-balanced within each phase.

What is also clear is that there are some large differences in NPD characteristics across groups as well. For example, in phase one, pupils in T1 and T2 are more likely to be deprived (based on FSM eligibility) and have special educational needs. KS2 English scores are also lower for both treatment groups as compared with the control group in phase one (statistically significant at the 10% level). In phase two, the differences are again smaller.

Given that pupils were randomised within individual phases, how could these differences have occurred? There are two main potential explanations. First, some pupils and schools dropped out. If this was non-random, differences could emerge within the final sample. Second, the sample sizes involved are small and chance differences in characteristics can easily emerge, particularly as many of the pre-test measures are seeking to capture different dimensions of reading ability. Given that groups remain well-balanced in terms of the variables used in the randomisation (age, gender, and SWRT), we suspect that the second explanation is the more likely of the two. See the ‘randomisation’ section for more details.

In summary, the groups are generally well-balanced at baseline when pupils are combined across both phases of the experiment, though there are some notable differences of around 0.1–0.2 standard deviations in pre-test outcomes. However, the groups are not well-balanced at baseline within individual phases. Although there are almost no statistically significant differences, the differences between treatment and control groups are large in absolute value in terms of pre-test outcomes and pupil characteristics.

This set of conclusions has a number of important consequences for the interpretation of our impact analysis. First, it is important to say that the phasing of the trial was not intended at the outset or at the time the main evaluation protocol was published. It resulted from a combination of slow recruitment and conditions on the funding of the trial. Whatever the cause, the unintended phasing of the trial means that the phases could be interpreted as separate experiments. The randomisation was performed separately by phase and the conditions were quite different (slightly different age groups, points in the school year, level of preparedness of schools, and the follow-up test was conducted with slightly different delays). In principle, the groups should thus be balanced within each phase. The fact that they are not is a cause for concern and means that analysis by phase is unlikely to be secure. The greater balance achieved when combining the phases means that this analysis is more secure than analysis split across phases. However, the fact that we are only able to achieve balance by combining two experiments that were not well-balanced themselves is not entirely satisfactory. It is fortuitous that imbalances were in opposite directions in each phase and are thus masked by combining phases. They could easily have gone in different directions and got worse as a result of combining phases. The lack of balance in baseline outcomes also has the unfortunate consequence that estimates of the impact of the intervention are likely to depend on whether, and which, baseline covariates one accounts for in a regression setting.

Given these concerns and issues, we present our main impact analysis with the phases combined, include analysis by phase as a prominent robustness check, and emphasise that the lack of balance within individual phases represents an important limitation of the analysis. Unfortunately, we cannot say how large a limitation this represents.

Outcomes and analysis

Table 6 shows descriptive statistics about the primary outcome (NGRT) and the four individual components that make up the two composite outcomes (reading accuracy and reading comprehension). This is shown for the pre-test and post-test stages. A number of key features about

the different tests become apparent. Here, we do not restrict the sample to the final estimation sample, but instead show the full range of scores for pupils with non-missing scores on each outcome. This is done to illustrate the full variation in outcomes and the different sample sizes for each outcome.

First, the sample sizes are clearly reduced at the post-test compared with the baseline. This is the result of school and pupil attrition. Second, not all pupils recorded results for every single test. In most cases, the difference in the sample size for each assessment is very slight, with only four fewer pupils taking the WIAT II comprehension test at post-test as compared with the NGRT, TOWRE word reading efficiency, and SWRT. In the case of the primary outcome and secondary reading accuracy outcome, we thus restrict the estimation sample to cases where baseline and post-test outcome are all non-missing. This ensures that any differences in estimated impact across different outcomes are not driven by sample differences.

The problem of missing data is more severe for the YARC-RC with a drop in sample size of 30 compared with the NGRT. This is because not all pupils recorded scores on the YARC-RC. If we restricted the estimation sample to cases where the YARC-RC is non-missing, then there would be a much more severe drop in the sample size. As the sample sizes are already relatively low and this represents a secondary outcome, we have therefore chosen not to include the YARC-RC as a baseline covariate because doing so would involve a large reduction in sample size. Instead, we only restrict the sample size to cases where the YARC-RC is non-missing for the secondary reading comprehension outcome.

The YARC-RC assessment actually comprises two reading passages. However, pupils were only given the second passage if they scored sufficiently highly on the first passage (about 50% of pupils). As a result, we only consider scores based on the first passage as incorporating those from the second would imply an even larger drop in sample size. We will be depositing the evaluation dataset at the UK and EEF data archives. In future, other researchers could seek to apply other methods that do not imply a drop in sample size, such as imputing missing scores or item response theory.

Third, the outcomes all have different scales. In order to ensure that results can be easily interpreted, we thus standardise all outcomes to have a mean of zero and a standard deviation of one within the estimation sample. To create the composite outcomes, we add the two components together and then re-standardise. However, we include the individual components as baseline covariates (NGRT, WIAT, SWRT, and TOWRE, but not YARC-RC as discussed above). This approach ensures that the components of the composite outcomes count equally, and that results are not scale-dependent but allow for the differential effects of the individual components at baseline.

Table 6: Descriptive statistics of raw outcome components

Variable	Baseline					Post-test				
	mean	min	max	s.d.	N	mean	min	max	s.d.	N
Baseline										
New Group Reading Test (NGRT)	233.7	0	352	60.8	239	258.6	72	364	50.8	212
WIAT II reading comprehension	102.9	38	151	17.9	275	106.3	46	141	17.8	208
YARC-RC	7.3	0	13	2.3	202	7.1	1	13	2.6	182
SWRT	32.5	3	44	8.2	278	36.0	6	56	9.0	212
TOWRE word reading efficiency	75.0	53	100	10.7	276	77.4	53	112	10.9	212

Note: YARC-RC is based on comprehension of first passage only.

Impact Analysis

Table 7 illustrates our main estimates of the impact of the reading intervention (REACH RI) and reading intervention with a focus on language comprehension (REACH LC) on reading test scores,

pooling the pupils across phases (and including an indicator for which phase pupils were part of, in case this affected outcomes). We show both the raw outcomes and our preferred impact estimates. This analysis is shown separately for REACH RI (top) and REACH LC (bottom).

We first show the average level of the raw outcomes (all standardised to have mean of zero and standard deviation of one) across the intervention and control groups, together with the sample size these are based on, and the number of missing observations as compared with the total number of non-missing observations on each outcome. The missing observations result from imposing a common estimation sample and the dropping of one outlier. On the right hand side of each table, we thus show the estimated impact based on our preferred methodology accounting for baseline characteristics (FILM). Also shown are the sample sizes involved and the p-value for a two-sided test for a null hypothesis that the estimated impact is zero. As the post-test outcomes have been standardised to have a mean of 0 and standard deviation of 1 within the estimation sample, the coefficient on the treatment indicator variable in a regression can be interpreted as an effect size.

We now discuss our main results, starting with a focus on the primary outcome (NGRT). The estimated impact of REACH RI on NGRT test scores is large with an effect size of around 0.33. However, the estimated effect of REACH LC is even larger, with an effect size of over 0.5. Both estimates are statistically significant at the 1% level. They are also very similar to the raw differences in outcomes between treatment and control groups in each case. However, it should be noted that the difference between the two treatment groups is not statistically significant.²¹

Looking at the secondary outcomes, we see relatively small and statistically insignificant effects on the reading comprehension composite for both treatments. However, there is a statistically significant effect of both interventions on the reading accuracy composite, with a very similar estimated effect across both treatments (0.17 and 0.15).

Table 7: Analysis of the impact of reading interventions (phases combined, all outcomes standardised)

Outcome	Raw means				n in model (treatment, control)	Effect size	
	Treatment group		Control group			Estimated Impact (95% CI)	p-value
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)			
REACH RI							
NGRT (primary outcome)	70 (3)	-0.00 (-0.24; 0.23)	63 (4)	-0.26 (-0.55; 0.02)	133 (70; 63)	0.329*** (0.135; 0.523)	0.001
Reading Comprehension composite	61 (12)	-0.17 (-0.40; 0.07)	50 (17)	0.11 (-0.17; 0.38)	111 (61; 50)	-0.079 (-0.412; 0.254)	0.642
Reading Accuracy composite	70 (3)	0.04 (-0.19; 0.27)	63 (4)	-0.10 (-0.37; 0.18)	133 (70; 63)	0.167*** (0.054; 0.279)	0.004
REACH LC							
NGRT (primary outcome)	69 (3)	0.25 (0.05; 0.44)	63 (4)	-0.26 (-0.55; 0.02)	132 (69; 63)	0.506*** (0.336; 0.677)	<0.001
Reading Comprehension composite	65 (7)	0.08 (-0.19; 0.35)	50 (17)	0.11 (-0.17; 0.38)	115 (65; 50)	0.136 (-0.158; 0.431)	0.364

²¹ Two considerations help to account for the lack of a statistically significant difference between the two treatment groups can be seen through two different methods. First, the confidence intervals for the two treatments effects are clearly overlap each other in Table 7. Second, the estimated effect of REACH LC vs REACH RI are not statistically significant whether we use FILM or OLS (available from the authors on request).

Reading Accuracy composite	69 (3)	0.05 (-0.18; 0.28)	63 (4)	-0.10 (-0.37; 0.18)	132 (69; 63)	0.153*** (0.037; 0.269)	0.01
----------------------------	--------	-----------------------	--------	------------------------	--------------	----------------------------	------

Note: * indicates that the treatment effect is significant at the 10% level ** at the 5% level *** at the 1% level. 95% confidence intervals are reported in parentheses.

Covariates included are: age in months; gender; phase of intervention, NGRT, WIAT, TOWRE word efficiency and SWRT baseline scores

Given the difference in experimental conditions for phases one and two of the trial, we also seek to estimate effects of the two treatments separately for each phase. This is shown in Table 8a (phase one) and Table 8b (phase two). Looking at the primary outcome (NGRT), the results do look quite different across phases. This is particularly the case for REACH RI where we observe a small and not statistically significant effect in phase one and a large and statistically significant effect in phase two. For treatment two (REACH LC), the effect is consistently large and statistically significant across both phases (note that it is possible for the individual effects for both phases to be below that of the pooled estimate as the pooled estimate imposes additional restrictions, such as equal effects of covariates across phases). There is also some consistency in terms of the estimated effects on the secondary outcomes across phases. There is no evidence of an effect on the reading comprehension measure. However, there is evidence of a positive impact on the reading accuracy measure, though the results are larger and only statistically significant for phase two.

What could be driving the differences in the estimated impact of the treatments across phases? First, they could be driven by differences in the experimental conditions, with pupils in the first phase being slightly older, experiencing a six-week break in the middle of the intervention, and schools having less time to prepare for phase one. Second, they could be driven by the large imbalances in baseline characteristics observed within phase. Unfortunately, it is not possible to separate the relative role of these factors. However, the fact that the difference across phases is only seen for REACH RI makes us suspect that the imbalances are the more likely explanation. It is not clear why only one of the interventions should have such different impacts across phases.

Table 8(a): Analysis of the impact of REACH RI separately by phase, all outcomes standardised

Outcome	Raw means				n in model (treatment, control)	Effect size	
	REACH RI Group		Control group			Estimated Impact (95% CI)	p-value
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)			
Phase 1							
NGRT (primary outcome)	26 (3)	-0.15 (-0.52; 0.22)	25 (3)	-0.08 (-0.48; 0.33)	51 (26; 25)	0.045 (-0.411; 0.501)	0.847
Reading Comprehension composite	26 (3)	-0.30 (-0.69; 0.08)	23 (5)	0.01 (-0.48; 0.50)	49 (26; 23)	0.084 (-0.556; 0.723)	0.798
Reading Accuracy composite	26 (3)	0.09 (-0.26; 0.44)	25 (3)	0.03 (-0.43; 0.49)	51 (26; 25)	0.092 (-0.065; 0.249)	0.253
Phase 2							
NGRT (primary outcome)	44 (0)	0.06 (-0.26; 0.37)	38 (1)	-0.36 (-0.74; 0.01)	82 (44; 38)	0.390*** (0.165; 0.616)	0.001
Reading Comprehension composite	35 (9)	-0.17 (-0.51; 0.17)	27 (12)	0.10 (-0.27; 0.47)	62 (35; 27)	-0.090 (-0.581; 0.402)	0.721
Reading Accuracy composite	44 (0)	0.02 (-0.28; 0.31)	38 (1)	-0.17 (-0.52; 0.19)	82 (44; 38)	0.203*** (0.084; 0.322)	0.001

Table 8(b): Analysis of the impact of REACH LC separately by phase, all outcomes standardised

Outcome	Raw means				n in model (treatment, control)	Effect size Estimated Impact (95% CI)	p-value
	REACH LC Group		Control group				
	n (missing)	Mean (95% CI)	n (missing)	Mean (95% CI)			
Phase 1							
NGRT (primary outcome)	26 (2)	0.22 (-0.22; 0.66)	25 (3)	-0.08 (-0.48; 0.33)	51 (26; 25)	0.418* (-0.034; 0.870)	0.07
Reading Comprehension composite	25 (3)	-0.04 (-0.58; 0.50)	23 (5)	0.01 (-0.48; 0.50)	48 (25; 23)	0.226 (-0.429; 0.880)	0.499
Reading Accuracy composite	26 (2)	0.02 (-0.28; 0.31)	25 (3)	0.03 (-0.43; 0.49)	51 (26; 25)	0.104 (-0.037; 0.246)	0.149
Phase 2							
NGRT (primary outcome)	43 (1)	0.26 (0.04; 0.49)	38 (1)	-0.36 (-0.74; 0.01)	81 (43; 38)	0.487*** (0.306; 0.668)	<0.001
Reading Comprehension composite	40 (4)	0.08 (-0.26; 0.42)	27 (12)	0.10 (-0.27; 0.47)	67 (40; 27)	0.172 (-0.166; 0.511)	0.319
Reading Accuracy composite	43 (1)	0.13 (-0.17; 0.42)	38 (1)	-0.17 (-0.52; 0.19)	81 (43; 38)	0.164** (0.013; 0.314)	0.033

Note: * indicates that the treatment effect is significant at the 10% level ** at the 5% level *** at the 1% level. 95% confidence intervals are reported in parentheses.

Covariates included are: age in months; gender; NGRT, WIAT, TOWRE word efficiency and SWRT baseline scores

When the phases are combined, therefore, there is evidence of a positive impact of both REACH RI and REACH LC on NGRT reading scores, with a larger estimated impact of the latter (though this difference is not statistically significant). There is also evidence of a small impact on the reading accuracy composite score, but no evidence of an impact on the reading comprehension composite score. This is consistent with the process evaluation which shows that TAs found the comprehension element to be the most challenging aspect to deliver, which may have dampened the effect of the comprehension-specific work.

However, the estimated effects differ across phases when they are estimated separately. REACH RI only appears to have a positive impact on the primary outcome in phase one, though there is a consistent picture of a large positive impact of REACH LC across both phases. It is not clear whether the differences in the estimates across phases are driven by actual differences in the effects of the treatment across phase, or by the large imbalances within each phase at baseline. The fact that we cannot replicate the combined results for the two individual phases on their own provides some cause for concern. The randomisation was performed separately for each phase and the experimental conditions differed. One could therefore interpret them as separate trials. The fact that we are only able to estimate credible results by combining two separate phases (that individually showed clear evidence of imbalances in baseline characteristics) reduces the credibility and security of the results.

Robustness checks

To examine the robustness of the FILM estimates reported here, treatment effects were also computed using a range of alternative methods. Table D3 shows that the estimated effects on the primary and secondary outcomes have a similar magnitude across all the different methodologies which supports the robustness of the combined results presented here. In several cases, the FILM estimates have a higher level of statistical significance than the alternative approaches: this is a result

of the greater precision of FILM estimation (Blundell *et al.*, 2005). It should also be noted that alternative ways of accounting for the experimental design (random effects and fixed effects) produce very similar estimates of the impact of the interventions and very similar standard errors. This further supports the robustness of our results. The main exception is that the impact estimates are clearly lower using kernel matching and lose their statistical significance. The loss of statistical significance for the kernel method is mainly down to the higher standard errors which can only be robustly estimated using the bootstrap method.

In Table 9, we repeat our main impact estimates for those pupils we could link to the NPD and control for the additional characteristics included in the NPD. Reassuringly, this does not lead to a major change in the impact estimates for the combined sample of pupils across phases. There remains a positive and statistically significant impact of both treatments on the primary post-test outcome, with a larger impact for REACH LC. The point estimates are also quite similar to those shown in Table 7. There also remains a small positive effect on reading accuracy. The only real difference compared with our main results is for REACH LC where we now observe a positive and statistically significant impact on the reading comprehension composite score as well.

Table 9: Impact estimates with NPD controls

Outcome	REACH RI (T1)			REACH LC (T2)		
	Treatment effect	95% Confidence Interval	n	Treatment effect	95% Confidence Interval	n
NGRT (primary outcome)	0.259**	(0.041; 0.476)	120	0.453***	(0.211; 0.694)	116
Reading Comprehension	-0.097	(-0.437; 0.244)	103	0.261*	(-0.049; 0.572)	103
Reading Accuracy	0.126***	(0.031; 0.222)	120	0.132*	(-0.012; 0.276)	116

Note: * indicates that the treatment effect is significant at the 10% level ** at the 5% level *** at the 1% level. Covariates included are: age in months; gender; NGRT, WIAT comp, TOWRE word efficiency and SWRT baseline scores. In addition, controls for FSM, SEN, EAL Ethnicity, IDACI and KS2 scores are included.

Follow up analysis

Secondary outcomes were also collected nine months after the end of the interventions.²² By this point, the control group has also begun to receive REACH LC.

As already discussed, comparing the original treatment groups to the control group is problematic as the control group had already started to receive the REACH LC intervention. We can, however, use the follow-up analysis to compare how the differences between the two groups—those receiving the original reading intervention (REACH RI) and those receiving the reading intervention with language comprehension (REACH LC)—change over time. This will test whether differences attributable to the two different treatments persist over time or not.

At the original post-test (Table 7) we saw a larger effect of REACH LC than REACH RI on the reading comprehension measure (though the estimated impact and the difference between the two treatments were not statistically significant, and there was very little difference between the effects of either intervention on reading accuracy). In Table 10, however, we see that there are no large or statistically significant differences between the two interventions groups at the follow-up stage in terms of either the reading accuracy or reading comprehension composite measures. Combined with the results at post-test stage, this means that we consistently observe no differences between the two treatments in terms of the reading accuracy composite measure, and if there is a difference in terms of the effect on the reading comprehension composite measure, it is likely to be short-lived.

²² The NGRT is not shown as it was not administered at the nine months post-test.

Table 10: Follow-up effect estimates (REACH RI relative to REACH LC, nine months after original post-test)

Reading intervention (T1)			
Outcome	Treatment effect	95% Confidence Interval	n
Follow-up test after 9 months			
Reading Comprehension composite	0.06	(-0.244; 0.364)	127
Reading Accuracy composite	-0.001	(-0.169; 0.166)	127

Note: * indicates that the treatment effect is significant at the 10% level ** at the 5% level *** at the 1% level. Covariates included are: age in months; gender; NGRT, WIAT comp, TOWRE word efficiency and SWRT baseline scores. In addition, controls for FSM, SEN, EAL Ethnicity, IDACI and KS2 scores are included.

Cost

The project team have estimated that the cost of materials for delivering the programme is £486 per TA. These materials can be re-used each year and kept by the school.

One must add to this the (larger) cost of training, estimated at £500 per day for the trainer (£2,500 for the whole five days). There are no additional costs if the training is held in a school, however were it held at a hotel or training centre, there would be an additional cost of £28–£35 per day (£140–£175 across five days) for each delegate (including refreshments and lunch). These figures do not include any potential costs to schools of finding cover for the TA during their training.

The major ongoing cost of delivering the interventions is staff time. The intervention requires TAs to deliver three 35-minute one to one sessions with each pupil involved each week for 20 weeks. If three pupils were involved, this would equate to around five hours of a TA's time per week. The project and delivery team estimate that this staff time represents a cost of £275 per pupil involved in the intervention. This is likely to be an underestimate of the cost and scale of staff time involved as it excludes preparation time.

The table in the executive summary presents the EEF cost rating for each intervention. This is based on the *additional monetary cost* to a school of delivering the interventions. It does not include, for example, the costs of staff time for staff already employed in the school. The cost is calculated as the cost per pupil over a three year period. To calculate this it is therefore necessary to make an assumption about how many TAs would be trained at one time by the trainer, and how many children would receive the intervention a given year. For the trial, the training was delivered to at least two TAs at once on average, usually outside of schools, and an average of 11 children in each school were selected as eligible for the interventions. The cost rating is therefore based on these conditions. The EEF cost ratings are explained in Appendix E.

Summary

In summary, our preferred impact estimates suggest that both treatments had a positive impact on the primary post-test outcome of reading skills as measured by the NGRT, with the REACH LC treatment having a larger impact than the REACH RI treatment (though this latter difference is not statistically significant²³). There is evidence of a small positive impact of both interventions on reading accuracy,

²³ This can be seen in all the results presented in Tables 7–10, as the confidence intervals for the impacts of the two treatments overlap. In formal tests, the estimated effect of REACH LC vs REACH RI are not statistically significant whether we use FILM or OLS (available from the authors on request).

though no evidence of a positive impact on reading comprehension. The groups are also generally well-balanced at baseline, increasing the credibility of the results.

Unfortunately, the credibility of the results is somewhat reduced when we consider the two phases of the trial separately. This is an important robustness check for the results since the two phases could be interpreted as separate trials (the age of pupils were slightly different, they were run at different points in the school year, and levels of school preparation differed). When we look at the phases separately, we see that the treatment and control groups are poorly balanced at baseline, with some very large differences in pre-test outcomes within each phase. The estimated impacts of REACH RI also differ quite sharply across the two phases, though there is more evidence of a consistent positive impact of the REACH LC. The fact that we can only provide credible results of the impact of the two treatments by combining the two phases is not entirely satisfactory. The randomisation was performed separately for each phase and the groups should, in principle, be well-balanced within each individual phase. The fact that they are not is almost certainly down to the small sample size.

Our view is that the results are promising, particularly as the results for the NGRT are close to the effect sizes suggested in previous work (0.4 to 0.6 standard deviations) (Hatcher *et al.*, 2006). What is surprising, and a little disappointing, are the apparent effects of the programme on reading comprehension. The estimated impact of the reading intervention with language comprehension is larger than for the reading intervention alone, but the differences between the effects of the interventions are not statistically significant. Moreover, there is no impact of either intervention on the composite measure of reading comprehension. It is however important to remember the very large differences in pre-test outcomes within each individual phase means that all the results we present, including the statistically significant impacts on the NGRT and the null effects on reading comprehension, lack a certain degree of credibility.

Process evaluation

The process evaluation explored professionals' experiences of delivery and views on impact. Feedback and opinions were sought regarding the implementation and setting up of the programme, the delivery approach in schools, the impact of the programme, and its sustainability. Results are based on a survey filled out by TAs who consented to be re-contacted during a questionnaire at the start of the intervention, and upon in-depth interviews with members of staff from six case study schools. The TA survey was sent to all TAs that consented to be involved in the evaluation. In phase one, 12 were sent out and 7 returned; in phase two, 23 were sent out and 18 were returned. The in depth interviews were conducted with the TAs, a member of the school's senior leadership team such as the Deputy Headteacher, and the SENCo. Individuals were spoken to separately and only if they were involved with the programme in some manner. There was a further case study with one school that dropped out to explore the reasons for withdrawal from the intervention.

Implementation

Implementation of the programme within a school was evaluated and assessed in relation to the training and setting up the programme within the school.

TAs were generally very positive about the five days of training they received for the programme (see Appendix A for the training plan): all TAs responding to the survey said that the training was relevant. Most thought that the content taught was at least fairly easy to implement with pupils. All TAs thought the materials they received during the training were useful. Some TA comments on how the training could be improved were as follows:

- There were some concerns that the training felt a little choppy. Some TAs found that the training jumped around quite a bit between topics and methods of teaching.
- Many TAs were overwhelmed by the amount of material covered in the intervention when first confronted with it. However, most agreed that the mock sessions on the last day tied things together in a good way.
- Some TAs mentioned that it would have been helpful to focus more on the practicalities of delivery, or to have a video clip to remind them how to deliver the intervention in an efficient way.
- Most TAs agreed that a greater focus on the practical elements of delivering the intervention would have been helpful.
- TAs were concerned that there was no check on how they delivered the intervention once the training was over, and a few mentioned that a refresher session would have been helpful.

Ultimately, many TAs expressed a lack of confidence about delivering the intervention after the training was done. This is worrying, given that five days is a relatively long training period for an intervention of this type. They felt that a greater focus on the practical elements of delivery would have given them more confidence, especially regarding whether they were delivering the interventions in the correct manner.²⁴

In terms of setting up the intervention in the schools, a principal barrier to implementation was timetabling the sessions. The most common mention in the questionnaire as a barrier to successful delivery was the fact that teachers did not like their pupils being withdrawn from classes. The TA

²⁴ TAs in two of the case study schools said directly that they were unsure whether they were delivering the courses properly, and that a refresher would have been helpful. Another two expressed some concerns that they were the only TAs in their school delivering the intervention and therefore were unsure, because of a lack of feedback, whether they were implementing the intervention correctly.

survey suggested that in the majority of schools pupils were withdrawn from classes other than English.²⁵

Fidelity

We now consider factors that affected delivery, including how these impacted the fidelity of delivery and the ongoing support received by TAs.

The overall delivery of the programme was affected by school level and programme related factors. School level factors that emerged as helpful to the delivery of the programme are given below. Schools may need to be actively encouraged to develop these conditions prior to delivering the programme and, as such, programme delivery may benefit from early discussions with schools on how best to do this.

- **Support from senior staff.** The process evaluation clearly revealed that those TAs who felt supported within a school were the ones who found it easiest to deliver the intervention and enjoyed it the most. Those TAs who reported feeling a high level of support were more likely to say that their experience with the intervention had been very positive. They were also more likely to rank the intervention as having been more effective.
- **A dedicated area to deliver the intervention.** TAs reported higher levels of general satisfaction in schools where they had been given dedicated space. TAs who lacked their own classroom space regarded this as a problem and a barrier to delivering the intervention.
- **Having access to expertise outside of the school.** This was also seen as an important factor in delivering the intervention successfully: five TAs identified this as one of the three most important factors in the successful delivery of the intervention. Those TAs that used the email address in order to ask about specific aspects of the intervention found this a very helpful tool.

In terms of programme-related factors, TAs had the following concerns that the project team might want to address before rolling out the programme to a larger number of schools.

- **Insufficient material for higher-performing pupils.** Some TAs identified that those pupils progressing quickly would soon run out of books to read (which made the pupils bored).
- **A weak comprehension section.** Many TAs mentioned that the comprehension section was the weakest. TAs said that—due to this being the most difficult to deliver in an engaging way—it was often the case that pupils got bored and did not enjoy it. This made the TAs less confident in delivering this in general.
- **An insufficiently user-friendly teaching manual.** The manual received as part of the training was seen as good, but again, TAs thought that the amount of information it contained was “daunting” at first. Early in delivering the intervention—when TAs sometimes had to go back to the manual to remind themselves of specifics—they found that it was not well set up for easy reference. Almost all TAs agreed that more preparation time and reassurance from the trainers would have been helpful in building their confidence before starting to deliver the intervention. More straightforward exercises, and shorter information leaflets for each stage (to avoid having to search through the full manual), would have been helpful.

There were varying levels of fidelity to the programme delivery model, and this must be taken into account when interpreting the impact evaluation findings. Three sets of factors were consistently mentioned as challenges to implementation:

²⁵ Of the 24 TAs surveyed, 2 said that the pupils were mainly withdrawn from English, 4 answered that it varied across pupils, and the remainder replied that pupils were mostly withdrawn from subjects other than English.

1. **Timetabling.** The 35-minute lesson timing precludes running two sessions consecutively during a normal one-hour lesson. Some schools tried to get round this by taking two pupils for the full hour, doing the intervention with one while the other did homework or numeracy/literacy exercises, before switching the pupils. This meant that each pupil lost five minutes of the intervention.
2. **Session length insufficient for material.** This was the most commonly mentioned barrier to successful delivery. TAs expressed that it was not “practically possible” to do the session in the allotted time. TAs either had to cut the sessions slightly short or, more commonly as they often overran, take the pupil out of the class for the full hour and do a slightly longer session, or find something else for the pupil to do in the remaining time until the start of the next lesson. This sometimes increased the workload of the TAs.
3. **Variation in pupil ability.** Some TAs expressed concern that the pupils participating in the intervention were at different levels: it was felt that some of the more able pupils would not benefit much from the intervention. As such, while the sessions generally worked well, some of them had to be adapted in order to suit the particular level of the pupil concerned.

The process evaluation also investigated the potential impact on fidelity of the phasing of the intervention. Together, the case studies and survey findings suggest that schools participating in phase one were more likely to raise concerns about issues such as timetabling, having insufficient lesson time to deliver the sessions, and withdrawing pupils from lessons. While these issues were also raised in phase two, they were less persistent and seen as less of a problem. In the case study, TAs in phase two were, in general, better prepared for these issues and thus better able to plan for them. The schools seemed to have been better at communicating the specifics of the intervention in Phase 2 compared with Phase 1, highlighting how expectations about what the intervention means for the school and for TAs play a large part in the general satisfaction with the intervention. While these case study schools are not necessarily a representative sample, the results are consistent with the difficulties in recruiting schools and the tight time constraints in Phase 1.

Some TAs in phase one expressed some dissatisfaction with starting just before the summer, citing having to “start over” when the autumn term started again. However, in general this was not seen as a problem, and TAs did not feel that there were any negative effects on pupils resulting from having the intervention broken up by holidays.

Outcomes

The recurrent view was that the programme had a positive impact on pupils, and aided the professional development of staff delivering the intervention. TAs believed that there had been clear improvements in reading and comprehension skills, and in pupil vocabulary. In addition, almost all TAs mentioned that pupil confidence had increased. Pupils who previously would never put up their hand in class were now not afraid to read out loud in the class, and other teachers had told some TAs that some of the pupils were now more likely to volunteer to read in class.

Benefits for TAs included new skills in delivering interventions in general, and reading interventions in particular, plus extra recognition within their schools as a result of being responsible for the intervention. Nearly all the TAs surveyed (23 out of 24) said that they were using some of the techniques and methods taught in the intervention in their other work.²⁶

These questionnaires were completed towards the end of the interventions—February 2014 for phase one and June 2014 for phase two. This is concerning as it raises the possibility that the control group was benefitting from intervention techniques in the treatment period, biasing the treatment effect downwards. Unfortunately it is not possible to rule out these types of spillover effects.

²⁶ The specific question asked was: ‘To what extent, if any, have you applied any of the techniques or methods taught in the intervention in any of your other work?’.

The delivery of the programme through one to one sessions was viewed as key to its success. TAs felt that the pupils' vocabulary as well as their comprehension skills improved as a result of the intervention. One TA said that the individual session enabled her to "gain a real insight into their individual needs", which meant that she was able to tailor the intervention to the needs of that pupil and make real progress. Others reported that not all pupils reacted in the same way: some felt that boys enjoyed the intervention less than girls; others felt that, for weaker pupils, the comprehension part of the intervention worked less well.

Most TAs involved in the case studies reported some problems in delivering the comprehension element, with pupils becoming bored or harder to engage. As a result, some TAs lost some confidence in delivering this more challenging aspect of REACH. One TA reported running out of material when the pupil progressed in skill, and having to resort to reusing reading passages, which they felt added to the pupil's boredom. TAs in one school reported that they had started to break the comprehension element into smaller sections to help maintain pupil interest. However, one TA highlighted the comprehension element of the programme as fundamental to its success and felt that it had played an instrumental role in pupils' improved vocabulary and reading confidence. There may be some value in revisiting the way this part of the intervention is delivered, reviewing the materials available to TAs, and/or focusing more on this element in the training sessions to build TAs' confidence.

Sustainability

Virtually all TAs who had delivered the programme reported that they were using the *techniques and methods* with other pupils, or as part of other interventions. Only one of the 24 TAs surveyed reported not doing so. The case studies also suggest that the intervention *materials* are also being used in other situations, although a few mentioned they planned to simplify these before applying them to their other work. There is some evidence that the intervention could affect schools' approaches more generally: 18 of 24 TAs responding to the survey said it was likely that their school would continue to use similar techniques and methods with struggling readers, although only four of these thought it was 'very likely'. As highlighted elsewhere, TAs highlighted some inefficiencies in the delivery of the programme that suggest it would be delivered in different forms if continued (for example, by adapting the length of sessions or refining the targeting of pupils).

It is clear that the delivery of the intervention requires significant ongoing resource within schools, and that it will compete with other reading interventions to some extent. The case studies highlighted that the one to one delivery of the intervention was seen as critical to its success. They also highlighted the importance of ongoing support from within schools at a senior level to help with practical issues (such as timetabling and negotiating the withdrawal of pupils from other lessons) as well as selecting the right TAs for the job (schools preferred to have two experienced TAs involved in the delivery). TAs also valued the ongoing support provided by the University of Leeds to check they were delivering the programme correctly. As such, the sustainability of the programme could depend on continued ongoing support from schools. Whether schools continue to use the methods introduced during the programme will depend on the perceived benefit to pupils, relative to the alternative uses of time and resources.

Formative findings

From the case studies and the questionnaires, we can identify some key aspects of the interventions seen as leading to successful implementation:

- **adequate support** for TAs, including
 - help with timetabling,
 - help with scheduling; and
 - providing access to a dedicated teaching area;

- **using two or more TAs** to deliver the project, working closely together;
- **using more experienced TAs** to deliver the project (more experienced TAs were more positive about the intervention and generally thought that the delivery went better than those who were less experienced, and due to its relative complexity, both TAs and senior staff at the schools thought that the intervention needed more experienced TAs to deliver it successfully);
- **dedicated one to one sessions** allowing TAs to tailor the intervention to the specific needs of pupils; and
- **a flexible approach** on the part of TAs with regard to session timing, content and delivery (for example, responding to potential pupil boredom by breaking up the session).

It is also possible to pick out aspects which require improvement in case the intervention is rolled out more widely. TAs generally felt that the following aspects need to be focused on:

- The 35-minute session timings do not fit with most schools' lesson planning. Usually lessons are one hour, meaning that sessions either have to be shortened in order to fit in two per lesson, or drawn out, with the pupil taken out of a whole lesson.
- The training should focus more on the practical aspects of delivery, and the manual could be redesigned to provide a quick reference guide for TAs using it in 'real time'.
- TAs should be encouraged to share learnings, potentially through an online forum or messaging board.
- The comprehension section, in some cases, did not work very well—pupils became bored by this. This needs to be more varied, and segmented into shorter pieces.
- More preparation time needs to be built into the programme, especially in the beginning. Many TAs felt that the envisaged timings were unrealistic.

Control group activity

None of the TAs mentioned any issues about how the waitlist control worked. However, a couple of TAs expressed reservations about the appropriateness of withholding the intervention for some of the pupils in the control group who, they believed, needed it urgently. In future trials, ensuring that TAs understand the rationale and justification for using this type of waitlist will be important.

There remains an ongoing concern from the outcomes of the survey that some of the TAs were using the methods for other pupils (potentially including the control group) during the intervention period. However, there is no way of verifying or refuting this using the information available.

Conclusion

Key conclusions

1. Both REACH interventions had a positive effect on the reading skills of the pupils in the trial. These effects are unlikely to have occurred by chance.
2. Pupils receiving the reading intervention with language comprehension experienced the equivalent of about six months of additional progress on average. For pupils receiving the standard reading intervention the figure was about four months.
3. The evaluation did not provide any evidence that the interventions improved reading comprehension in particular, as opposed to other skills such as word recognition.
4. Staff reported that the interventions improved literacy, reading ability, and confidence. Staff views were more positive in schools where the interventions were delivered by experienced teaching assistants, supported by senior staff, and allocated a dedicated space for delivery.
5. Teaching assistants sometimes found the interventions challenging to deliver. In particular, many said they were not confident delivering the one to one sessions even after training, and some found that the reading comprehension elements sometimes failed to hold pupils' attention.

Interpretation

The objective of the evaluation was to estimate the impact of the two interventions on reading skills.

In general, although there are some important caveats noted above, the results are promising. There is a large and positive estimated effect of the interventions on reading skills as measured by the NGRT. The results are in line with previous estimates, look the same no matter how we control for baseline differences across pupils, and are largely unchanged when we control for additional characteristics about pupils from administrative data. Both treatments had a smaller positive effect on the secondary outcome of a reading accuracy composite score.

When comparing the impacts of the programme across the two treatments, the effect of the reading intervention supplemented with language comprehension (REACH LC) is larger than the impact of the standard reading intervention alone (REACH RI), but the difference is not statistically significant. Moreover, there is no evidence of any gain from either treatment in the reading comprehension composite score considered. This is surprising and a little disappointing given that the intervention was designed to target comprehension.

Unfortunately, the differences in the pre-test scores and characteristics of pupils in the treatment and control groups within each phase mean that we cannot have absolute confidence in the results. The scale and range of the differences lead us to worry that there could be unobservable differences driving the results. There is also no way to test the extent to which this is the case. Pooling the sample does improve the balance, but this acts to mask the problem within phases.

The results do, however, fit with the process evaluation as schools were generally very positive about the programme, and phase two schools that were interviewed generally felt better prepared than those schools that were interviewed from phase one. Schools also made some useful suggestions about how the programme could be improved. First, those involved generally thought that the training should incorporate more practical elements. Second, the 35-minute sessions were not well matched with the standard one-hour lessons that schools generally use. Third, preparation time for the programme is essential, and schools in the second phase were noticeably more prepared than those in the first phase. Last, the comprehension section, in some cases, did not work very well—pupils

became bored by this. This might need to be more varied and segmented into shorter pieces, particularly given the importance of reading comprehension that has been highlighted by the earlier work of the project team.

A final and important point to note is that the trial design does not allow us to identify the impact of the interventions separately from (1) the provision of one to one teaching time or (2) the increased time devoted to literacy as pupils were typically withdrawn from lessons other than English. This issue of interpretation is common to most trials of this type.

Limitations

The original protocol specified that 27 schools and 486 pupils were to be recruited across a single phase. This would have represented a relatively high-powered trial. However, difficulties in the recruitment of schools and pupils, combined with a tight time scale for commencing the intervention, resulted in (1) a sample size of less than half that originally intended, and (2) the trial being split into two phases starting in June 2013 and November 2013. The phasing of the trial is particularly problematic as pupils in the two phases faced different experimental conditions: the intervention was conducted at different ages; phase one pupils had a six-week break over the summer; and different amounts of time were available for schools to prepare for the intervention. The two phases were therefore, in many respects, separate trials.

When the two phases were combined, the two treatment groups and one control group were reasonably balanced. However, the balance is poor when considering each phase separately, particularly in terms of the primary pre-test outcome. The direction of the imbalance, almost certainly by chance, is in the opposite direction in each phase meaning that pooling the phases masks the problem. It is therefore potentially misleading to interpret the differences in the post-test outcomes between treatment and control groups as a genuine impact of the programme. Even though we can control for some of the differences in the baseline test scores and characteristics, the number and scale of baseline differences lead us to suspect that there are differences between the groups that we cannot observe and therefore cannot control for.

There are two main reasons why the groups are likely to be imbalanced within each phase, reasons which themselves reflect the limitations of the programme. First, some schools dropped out from the interventions with many citing high workload in trying to deliver the programme. If this dropout was non-random, then this could lead to imbalance among those that completed the programme. We do see some evidence of this: at randomisation, pupils were well-balanced in terms of the main screening test; they were less balanced amongst schools that actually completed the programme.

Second, the sample sizes are quite small, particularly for phase one. As a result, chance differences in the make-up of the treatment and control groups can easily occur, particularly when we are examining different dimensions of reading ability. The small sample sizes represent a further problem as they reduce the precision of our impact estimates.

Another potential limitation is the finding from the process evaluation that most TAs were using some of the techniques and methods taught in the intervention in their other work. This could reflect confidence in the potential effects of the interventions. However, it also raises the possibility that the control group were receiving some of the techniques in the treatment period: this would bias the treatment effect downwards. Unfortunately it is not possible to rule out these types of spillover effects as the survey was conducted after the trial.

In hindsight, a number of lessons can be learned. If this trial were to be repeated for a larger sample size (and we think the results would justify such a trial), a longer lead time is required to ensure sufficient schools are recruited and schools are well-prepared. Splitting schools across two phases

was a major undoing of the trial; it reduced sample sizes and led to a number of schools not being sufficiently well-prepared.

Further thought is also required about how the language comprehension element should be delivered to this age-group. Schools generally felt that this was the weakest component of the intervention and they found it difficult to deliver.

Another issue which arose was that TAs did not feel fully confident in delivering the intervention, even after five days of training.

Future research and publications

In our view, the results are promising enough to justify a future larger trial. This would allow more precise—and potentially more credible—estimates of the impact, as there could be a greater chance of ensuring balance across groups. Such a trial would need to give schools adequate preparation time, and respond to some of the feedback from schools on the content and delivery of the programme. The process evaluation reported improved confidence levels among pupils. It might therefore also be helpful to consider collecting other data about pupils, such as their self-confidence levels, or measures of locus of control. This could apply to other reading or literacy interventions. Just focusing on reading and comprehension outcomes might lead us to under-appreciate the results of reading programmes. Given that pupils were often taken out of subjects other than English to receive the intervention, one might be worried that the effects of the intervention simply reflect a greater amount of time focused on reading rather than the impact of these particular interventions. It could be interesting to examine the performance of participating pupils in other subjects to see if there are any negative consequences of this, or consider future trials where pupils are withdrawn from English classes (though schools may be unwilling to do the latter). Finally, the process evaluation identified the one to one delivery of the programme as fundamental to its success. In any future RCTs, it would therefore be useful to separate the impacts of the content of the programme from the method of delivery, by either varying the delivery method or maintaining the one to one delivery and testing multiple literacy programmes against one another.

The evaluation team may seek to publish the results of this evaluation in academic journals. However, further thought will be required to consider the academic contribution of such an article, particularly given the large imbalance across groups within phases. We may seek to combine or compare the results with those found in other trials (such as the LIT programme) or seek to make a methodological contribution by examining a number of further ways to mitigate the impact of the imbalances.

References

- Blundell, R., Dearden, L. and Sianesi, B. (2005) 'Evaluating the effects of education on earnings: models, methods and results from the National Child Development Survey'. *Journal of the Royal Statistical Society Series A*, Royal Statistical Society, 168: 3, 473–512.
- Bruhn, M. and McKenzie, D. (2009) 'In Pursuit of Balance: Randomization in Practice in Development Field Experiments'. *American Economic Journal: Applied Economics*, 1: 4, 200–232.
- Clarke, P., Snowling, M. J., Truelove, E., and Hulme, C. (2010) 'Ameliorating children's reading comprehension difficulties: A randomised controlled trial'. *Psychological Science*, 21: 1106–1116.
- Crawford C. and Cribb, J. (2015) 'The link between childhood reading skills and adult outcomes: analysis of a cohort of British children'. IFS Briefing Note (BN169).
- Foster, H. and the National Foundation for Educational Research (2008) *Single Word Reading Test 6–16*. London: GL Assessment.
- GL Assessment (2010) *New Group Reading Test (NGRT)*. London: GL Assessment.
- Hatcher, P. J., Goetz, K., Snowling, M. J., Hulme, C., Gibbs, S. and Smith, G. (2006) 'Evidence for the effectiveness of the Early Literacy Support Programme'. *British Journal of Educational Psychology* (2006a), 73: 351–367.
- Morgan, K. L. and Rubin, D. B. (2012) 'Rerandomization to improve covariate balance in experiments'. *Annals of Statistics*, 40: 2, 1263–1282.
- Nation, K. and Snowling, M. J. (1997) 'Assessing reading difficulties: the validity and utility of current measures of reading skill'. *British Journal of Educational Psychology*, 67: 359–370.
- Rashotte, C., Torgesen, J. K. and Wagner, R. (1999) *Test of Word Reading Efficiency (TOWRE)*. London: Pearson Assessment.
- Raudenbush, S.W. and Bryk, A. S. (2002). *Hierarchical Linear Models (Second Edition)*. Thousand Oaks: Sage Publications.
- Scott, N. W., McPherson, G. C., Ramsay, C. R. and Campbell, M. K. (2002) 'The method of minimization for allocation to clinical trials. a review'. *Control Clinical Trials*, 23: 662–674.
- Snowling, M. J., Stothard, S. E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K. and Hulme, C. (2009). *YARC: York Assessment of Reading for Comprehension Passage Reading*. Swindon: GL Publishers.
- Wechsler, D. (2005) *Wechsler Individual Achievement Test 2nd Edition (WIAT II)*. London: Pearson Assessment.
- Wooldridge, J. (2010) *Econometric Analysis of Cross Section and Panel Data*. Second Edition, MIT Press.

Appendix A: Project team training plan

Day 1	<ul style="list-style-type: none"> • Introduction to the Project • Rationale of a Randomised Controlled Trial (RCT) and protocols • Background to the intervention programmes, Reading Intervention (RI) and Reading for Meaning (Readme) • Research, evidence base for both interventions and establishing cause and effect • Introduction to the 2 websites RI and Readme • Learning to Read – 2 dimensions: decoding and language comprehension • Cues for learning to read with an in depth look at phonological awareness and decoding
Day 2	<ul style="list-style-type: none"> • Assessment and building a profile of the learner for RI • Introduction to the range of assessments for RI • Learning how to make a Running Record and miscue analysis • Establishing level of reading book for RI and what is meant by an ‘instructional text’ • Knowledge of high frequency sight words and links to Letters and Sounds (DCSF 2007) • Phonological awareness assessments including Sound Linkage test and non word reading • Free writing assessment • Building a profile of the learner and identifying key starting points • Introduction to the RI manual and focus on section 2
Day 3	<ul style="list-style-type: none"> • What is reading comprehension? • Revisit Simple View of Reading (Gough & Tunmer, 1984) • Modelled think aloud • Annotated think aloud • Introduce Construction Integration model (Kintsch & Rawson, 2005)

	<ul style="list-style-type: none"> • Session structure • Recap Oral Language components - Vocabulary, RT, Figurative Language, Narrative • Teaching principles • Vocabulary (Tier 2 words, Vocabulary prompts, Role play) • Figurative Language (Using Jokes, Riddles and Idioms - Identifying teaching points) • Reciprocal teaching (Clarification, Summarisation, Prediction, Question Generation) • Narrative (Creating story maps to organise the key elements of the story: Events, Characters, Settings, Times, Phrases, Words) • Intervention materials • Record keeping
Day 4	<ul style="list-style-type: none"> • Using the Learner Profile and establishing a starting point for RI • Delivering the elements of the programme • Phonological Development • Strategies and Activities to support the learner • Resources to Support the Programme including range of reading books
Day 5	<ul style="list-style-type: none"> • Sound Linkage materials, starting point and delivery of the activities • Case study and session planning • Focus on sections 3 and 4 of the RI manual • Final recap and time for questions

Appendix B: Consent forms and information packs

Appendix B.1 – Initial contact email

Email to Head teacher/ SENCO/ Head of Year 7

We are inviting selected schools in the Leeds area to participate in a large scale research project for Year 7 students. This research project is funded by the Educational Endowment Foundation (<http://educationendowmentfoundation.org.uk/>).

Project Information

The READING for CompreHension (REACH) Project funded by the Education Endowment Foundation. The team of researchers involved in the project include Professor Maggie Snowling (St John's College Oxford), Professor Charles Hulme (University College London), Dr Paula Clarke (University of Leeds) and Glynnis Smith (Educational Consultant). The research project will focus on students currently in Year 7 who entered the school/academy with English below level 4. The project will evaluate the effectiveness of two approaches to supporting reading skills.

- The first will be the Reading Intervention (RI) approach which involves - reading easy and instructional level books, letter-sound work, phoneme awareness activities, phonological linkage training, writing sentences and spelling. The RI programme is one of the most effective interventions for addressing reading difficulties. It has been used successfully to accelerate progress in reading in Cumbria and North Yorkshire. North Yorkshire results this past year (2011) indicate an average gain of 10 months reading progress over 10 weeks (www.interventionsforliteracy.org.uk).
- The second intervention will combine the RI approach with comprehension activities including multiple context vocabulary training, figurative language work and reciprocal teaching (clarification, summarisation, prediction and question generation). The comprehension activities will be based on those used in the York Reading for Meaning project (www.readingformeaning.co.uk) which was trialled with children in Years 4-5 in 20 schools in York and North Yorkshire and generated significant gains in reading comprehension following 20 weeks of intervention.

REACH is a funded research project that covers the cost of training two teaching assistants, the teaching materials needed including a range of books and the delivery of the two programmes over twenty weeks. The delivery of the two programmes will take place from April 2013 through to December 2013.

At this stage we are asking schools to register an interest. To do this, we need a contact name for the school with job title and an email address. Any additional contact details would be helpful. Further information will be available and we are happy to discuss any questions you may have.

Best wishes,

Dr Paula Clarke

Appendix B.2 – Head teacher’s information pack

Information Sheet for Schools

You will be given a copy of this information sheet.

*Title of Project: **The REAding for CompreHension (REACH) Project***

This study has been approved by the University of Leeds AREA Faculty Research Ethics Committee (Project ID Number):

Name Dr Paula Clarke

Work Address School of Education, University of Leeds LS2 9JT

Contact Details Tel: 01133439410 Email: p.j.clarke@leeds.ac.uk

We would like to invite (INSERT SCHOOL NAME) to participate in this research project.

Details of Study:

A Randomised Controlled Trial (RCT) of two evidence-based interventions to improve the reading skills of pupils following transition into secondary school.

Researchers based at UCL, London, University of Oxford and the University of Leeds are carrying out a research project that evaluates the impact of two different programmes of structured reading intervention funded by the Education Endowment Foundation (EEF). The two programmes are:

1. The Reading Intervention (RI) approach which involves, reading easy and instructional level books, letter-sound work, phoneme awareness activities, phonological linkage training, writing sentences and spelling. The RI approach has been used successfully to accelerate progress in reading in Cumbria and North Yorkshire. North Yorkshire results this past year (2011) indicate an average gain of 10 months reading progress over 10 weeks
www.interventionsforliteracy.org.uk.
2. An approach which combines RI with comprehension activities (RI+C) including multiple context vocabulary training, figurative language work and reciprocal teaching (clarification, summarisation, prediction and question generation). The comprehension activities will be based on those used in the York Reading for Meaning project (www.readingformeaning.co.uk) which was trialled with children in Years 4-5 in 20 schools in York and North Yorkshire and generated significant gains in reading comprehension following 20 weeks of intervention.

A condition of the funding provided by EEF is that the effectiveness of these interventions in improving children’s literacy skills must also be estimated by an independent evaluation team. The team for this intervention comprises researchers from the Institute for Fiscal Studies and Ipsos MORI.

What will be in the study involve?

In early 2013, we will ask you to identify the children in your school who entered secondary school with English below Level 4. We will then contact the parents of these children with information about the project and a consent form to be signed and returned if they would like their child to take part. We will ask you to nominate a member of school staff to be a point of contact for the parents to respond to any questions which they might have about the project. Parents will also be given the research team details and can contact us at any point.

A trained researcher from the University will visit your school and screen the children, for whom consent has been obtained, using a standardised single word reading accuracy measure. The 18 children in each school who obtain the lowest scores on this measure will then be selected to take part in the main trial.

Once the children have been selected, the parents will be notified through a letter. Those parents whose children have not been selected will be reassured that the training and materials that are part of the project will be available for the school to use with their children after the project has finished, so should benefit their child in the longer term. The evaluation team will also contact you to request the children's Unique Pupil Numbers (UPNs) so that they can link to their education records held by the Department for Education, which will be used as part of the independent evaluation. The selected children will then be randomly allocated to one of three groups:

Group 1 This group will receive 20 weeks of the RI programme starting in the summer term of Year 7 through to the autumn term of Year 8.

Group 2 This group will receive 20 weeks of the RI+CI programme starting in the summer term of Year 7 through to the autumn term of Year 8.

Group 3 A 'waiting list' intervention group will receive the programme deemed to have been most successful in the main trial (RI or RI+C) starting in the spring term of Year 8 through to the summer term of Year 8.

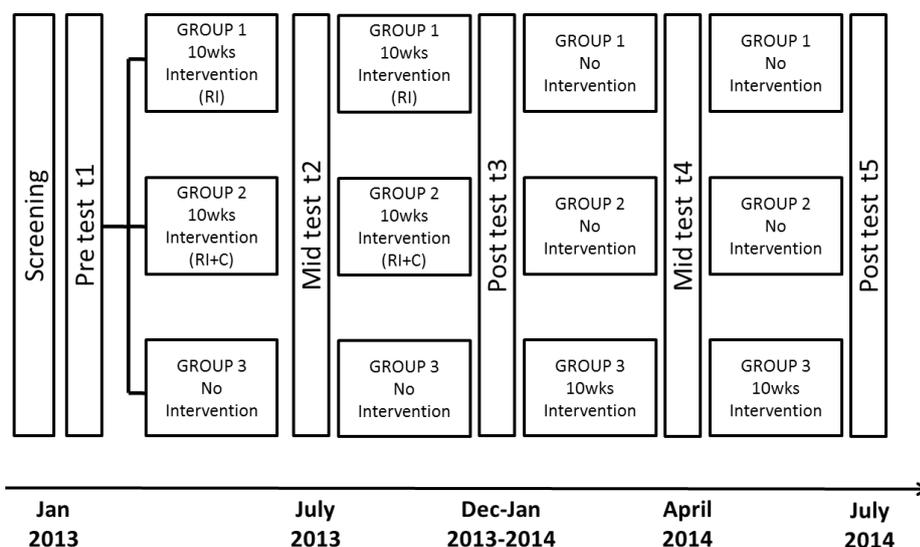
When receiving intervention the children will take part in three 35 minute sessions per week over 20 weeks. Care will be taken to ensure that the timetabling of these sessions varies across the 20 weeks so that children will not always be missing the same lessons to take part in the intervention.

How will progress be assessed?

Researchers from the University will assess all three groups of children on standardised measures of reading and language skills at five key points (t1-t5) in the trial. The researchers will be fully trained in the administration of the assessments and have enhanced CRB checks.

At t1, t3, and t5 a full battery of tasks will be used. At points t2 and t4 a reduced battery will be administered. We anticipate for the full battery assessments to take approximately 1 hour per child and for the reduced battery 15 minutes per child. In addition to this a computer administered standardised reading test will also be given at each time point; this takes approximately 45 minutes. We are also interested in monitoring the impact of reading intervention on children's attitudes to learning and perceptions of their own learning. To do this, at each time point, we will ask the children to complete a short questionnaire, in which rating scales will be used to indicate attitudes and perceptions. This is a bespoke questionnaire which is currently in development, we anticipate it should take no longer than 10 minutes to complete.

Project Timeline



Who will administer the Intervention Programmes?

Each school will be required to identify two teaching assistants to deliver the programmes. The teaching assistants will receive 5 days training in March/April 2013 (dates and venue to be confirmed). The training will be delivered by the research team who have extensive experience in running training for similar projects. In addition to training on how to administer the programmes each teaching assistant will receive a full teaching pack containing session by session guidelines, general teaching principles, photocopiable resources, and progress monitoring sheets. The research team will provide email and telephone support to the teaching assistants throughout the project at mutually convenient times.

We are interested in researching the experiences of the teaching assistants on the project and documenting their professional development. To this end we will give the teaching assistants a questionnaire to complete at each time point which we anticipate should take no longer than 20 minutes to complete. The independent evaluation team may also contact you and your teaching assistants separately to discuss the project in greater detail.

Funding

The funding from the EEF covers:

- Delivery of the interventions in the main part of the trial
- Teaching assistant training
- Travel expenses
- Teaching packs
- Book boxes
- All researcher costs
- All assessment materials

Your school will be given the teaching packs and book boxes to keep for future use.

Please note: The funding from the EEF does not cover the delivery of intervention to the 6 children in the waiting control group. Your school will be required to cover the costs of this.

Confidentiality and Data Protection

We would like to assure you that information from the study will be kept strictly confidential. Children's results will never be identified by name. All children are free to withdraw from the study at any time.

Both the research team and the independent evaluation team will produce a report after the project is completed to summarise our main findings; this will be available, on request, to head teachers, teaching assistants and parents. The data we obtain will be used in conference presentations, journal publications, book chapters and future grant applications. Please be assured that at no point will the data be identifiable; personal details will only be held by the research team, and neither personal details nor school information will be presented.

Data will be stored in accordance with the Data Protection Act 1998. All electronic data will be stored on password protected computers and will be identified using code numbers. Paper data (e.g. questionnaires and reading test forms) will be stored in locked filing cabinets. Two lists (one as back up) of the code numbers and corresponding names will be stored in locked filing cabinets away from the data.

What happens next?

If you would like your school to participate in this study please complete the consent form attached and return it to the address at the top of this information pack. If you would like any more information about this research then please contact me on 01133439410 or p.j.clarke@leeds.ac.uk.

On receipt of the signed consent form we will sign it and then return a copy to you along with a university-school agreement which outlines the responsibilities of the head teacher, the teaching assistants and the members of the research team. This will need to be signed by all parties before the interventions begin.

Best wishes and thank you,

Dr Paula Clarke, Lecturer in Education

Appendix B.3 – Head teacher’s consent form

Consent FormThe **REA**Ding for **CompreH**ension (**REACH**) Project

Head teacher name

School name

	Please initial
I give permission for our school to take part in this study.	
I have been given information about the purpose of this research and the opportunity to ask questions about it.	
I understand that data will be treated confidentially and referred to by code number on all data sheets (electronic & paper).	
I agree for the data collected to be used in relevant future research.	
I understand that names will not be linked with the research materials, and participants will not be identified or identifiable in the report or reports that result from the research.	
I understand that I can ask for my school’s data to be removed from the project’s database at any time by contacting the research team.	
I understand that I can withdraw my school from the project at any time by contacting the research team.	

Signed (Head teacher) **Date**.....**Signed** (Lead researcher) **Date**.....

Once this has been signed by all parties you will receive a copy of the signed and dated information pack and consent form. Please keep this for your records. We will keep our copy in a locked filing cabinet.

Appendix B.4 – Parent’s information pack

Information Sheet for Parents/Carers

The **RE**ADING for **CompreH**ension (**REACH**) Project

How effective are two different methods to supporting reading skills?

Researchers based at UCL, London, University of Oxford and the University of Leeds are carrying out a research project to test the benefits of two different reading interventions for children in secondary school. The results are being independently evaluated by a team from the Institute for Fiscal Studies and Ipsos MORI. This leaflet provides you with more details about the project.

What will be in the study involve?

Your child’s school have suggested that your child could benefit from taking part in this study. If you give permission for them to take part your child will firstly complete a short reading test given by a researcher from our team. Based on the test scores we will then select the 18 children with the lowest reading scores to take part in the study. We will send you a letter to let you know if your child is selected for the study.

If your child is selected then they will be randomly allocated to one of three groups:

Group 1 - This group will receive 20 weeks of the Reading Intervention (RI) programme starting in the summer term of Year 7 through to the autumn term of Year 8. The RI programme involves, reading easy and instructional level books, letter sound work, phoneme awareness activities, phonological linkage training, writing sentences and spelling.

Group 2 - This group will receive 20 weeks of the Reading Intervention + Comprehension (RI+C) programme starting in the summer term of Year 7 through to the autumn term of Year 8. The RI+C programme combines RI activities with comprehension work, including vocabulary training, figurative language, and reciprocal teaching (clarification, summarisation, prediction and question generation).

Group 3 - A ‘waiting list’ intervention group will receive regular classroom teaching when the other two groups are receiving intervention. After the other two groups have finished this group will then receive the programme deemed to have been most successful in the main trial (RI or RI+C) starting in the spring term of Year 8 through to the summer term of Year 8.

When receiving intervention the children will complete three 35 minute 1:1 sessions per week. The children will be taken out of their regular classroom activities for these sessions. The timetabling of the sessions will vary to ensure that the children do not always miss the same lessons.

How will my child’s progress be checked?

A researcher from our team will visit the school at regular intervals to measure children’s progress. There will be five assessment time points (t1-t5) and a range of standardised reading and language tests will be used. These tests will involve reading some information and answering some questions about it, reading lists of words and nonwords and providing definitions of words. The researchers will be fully trained in the administration of the assessments and have up to date enhanced CRB checks. At t1, t3, and t5 a full set of tests (taking approximately 1 hour) will be used. At point’s t2 and t4 a reduced set (taking approximately 15 minutes) will be given. A computerised reading test, which involves reading information and answering multiple choice questions, will also be given at each time point; this takes approximately 45 minutes. We are also interested in monitoring the impact of reading intervention on children’s attitudes to learning and perceptions of their own learning. To do this, at each time point, we will ask the children to complete a short questionnaire which should take approximately 10 minutes to complete.

We will also be seeking permission from your child’s school to access the information that you share with them about your child, such as their date of birth and ethnicity, as well as their Key Stage test results. This will form an

integral part of the evaluation process and will enable the funders of the study to continue to follow your child's progress after the programme has ended.

Who will be teaching my child in the Intervention Programme?

A trained teaching assistant will be teaching your child during the intervention sessions. They will receive 5 days of training and will be supervised by the research team who will be in regular email and telephone contact with them. The teaching assistants will be using teaching manuals, activities and books developed and selected by the research team based on current evidence for best practice in supporting reading skills.

Confidentiality

We would like to assure you that information from the study will be kept strictly confidential. We will need to share your child's test results and other personal information with the evaluation team, but this information will be transferred and stored securely using password protected files. All children are free to withdraw from the study at any time. You are able to withdraw your child's data at any time up until the completion of the project in December 2014. Both we and the evaluation team will produce a report after the project is completed to summarise our main findings; this will be available, on request, to head teachers, teaching assistants and parents. The data we obtain will be used in conference presentations, journal publications, book chapters and future grant applications. Please be assured that at no point will the data be identifiable; personal details and school information will not be presented.

Data will be stored in accordance with the Data Protection Act 1998. All electronic data will be stored on password protected computers and will be identified using code numbers. Paper data (e.g. questionnaires and reading test forms) will be stored in locked filing cabinets. Two lists (one as back up) of the code numbers and corresponding names will be stored in locked filing cabinets away from the data.

What happens next?

(insert name) is the main in-school contact regarding this study and they are available if you would like to find out more about the project. Alternatively you are very welcome to contact the lead researcher at the University of Leeds, Dr Paula Clarke, on 01133439410 or **p.j.clarke@leeds.ac.uk**

If you would like your child to be involved in this study then please complete the consent form attached and return it to the (insert school contact name) by (insert date).

Best wishes and thank you,

Dr Paula Clarke,

Lecturer in Education

Appendix B.5 – Parent consent form

Please complete this form after you have read the Information Sheet

Title of Project: The **RE**ADING for **CompreH**ension (**REACH**) Project

This study has been approved by the University of Leeds AREA Faculty Research Ethics Committee (Project ID Number):

	Please initial
I have read and understood the information given about The RE ADING for CompreH ension (REACH) Project, and would like my child to take part.	
I have read the Information Sheet, and understand what the study involves.	
I have been given information about the purpose of this research and the opportunity to ask questions about it.	
I understand that data will be treated confidentially and referred to by code number on all data sheets (electronic & paper).	
I agree for the data collected to be used in relevant future research.	
I understand that names will not be linked with the research materials, and children will not be identified or identifiable in the report or reports that result from the research.	
I understand that I can ask for my child's data to be removed from the project's database at any time by contacting the research team.	
I understand that I can withdraw my child from the project at any time by contacting the research team.	

Signed (Parent/Carer) **Date**.....

PRINT NAME (Parent/Carer)

PRINT NAME OF CHILD

Signed (Lead researcher) **Date**.....

You will receive a copy of the signed and dated information pack and consent form. Please keep this for your records. We will keep our copy in a locked filing cabinet.

Appendix B.6 – Teaching assistants information letter

Dear (insert name),

We are interested in researching the experiences and documenting the professional development of the teaching assistants on The **RE**ADING for **C**ompre**H**ension (**REACH**) Project. To this end we have developed a questionnaire to be completed at each of the five assessment time points in the study. The questionnaire includes a mixture of rating scales and open ended questions. We anticipate it should take no longer than 20 minutes to complete.

You are under no obligation to take part in this piece of research, if you do not wish to take part this will not impact on your involvement in The **RE**ADING for **C**ompre**H**ension (**REACH**) Project in any way. If you decide to participate then you are able to withdraw your data at any time up until the completion of the project in December 2014. You may also leave questions on the questionnaire blank if you wish. With your permission, the independent evaluation team may also contact you to discuss your responses to this questionnaire in further detail.

We would like to assure you that information from the questionnaires will be kept strictly confidential. Responses are identified by code number only. Data will be stored in accordance with the Data Protection Act 1998. All electronic data will be stored on password protected computers and will be identified using code numbers. The paper questionnaires will be stored in locked filing cabinets. Two lists (one as back up) of the code numbers and corresponding names will be stored in locked filing cabinets away from the data.

If you have any questions about this piece of research please contact Dr Paula Clarke on 01133439410 or **p.j.clarke@leeds.ac.uk**

Best wishes and thank you,

Dr Paula Clarke,

Lecturer in Education

Appendix B.7 – Teaching assistant consent form

Please complete this form after you have read the Information Sheet

Title of Project: The **RE**ADING for **CompreH**ension (**REACH**) Project – Documenting the experiences and professional development of Teaching Assistants.

This study has been approved by the University of Leeds AREA Faculty Research Ethics Committee (Project ID Number):

	Please initial
I have read and understood the information given about The RE ADING for CompreH ension (REACH) Project – Documenting the experiences and professional development of Teaching Assistants, and would like to take part.	
I have read the information letter, and understand what the study involves.	
I have been given information about the purpose of this research and the opportunity to ask questions about it.	
I understand that data will be treated confidentially and referred to by code number on all data sheets (electronic & paper).	
I agree for the data collected to be used in relevant future research.	
I understand that names will not be linked with the research materials, and I will not be identified or identifiable in the report or reports that result from the research.	
I understand that I can ask for my data to be removed from the study at any time by contacting the research team.	
I understand that I can withdraw from the study at any time by contacting the research team.	

Signed..... **Date**.....

PRINT NAME

Signed (Lead researcher) **Date**.....

You will receive a copy of the signed and dated information letter and consent form. Please keep this for your records. We will keep our copy in a locked filing cabinet.

Appendix B.8 – University-School agreement

TO BE COMPLETED AFTER READING THE PROJECT INFORMATION PACK

School Name:

Research: The **REA**dIng for **CompreH**ension (**REACH**) Project

The Research Team will:

Testing:

- Ensure that all staff carrying out assessments are trained and have enhanced CRB clearance.
- Collect and analyse all data from the project.
- On request, provide head teachers with the screening data before the intervention begins and all progress data after the final post test has been completed.

Training:

- Run a 4 day training course and a 1 day top up training course for the teaching assistants delivering the interventions.
- Provide ongoing support to the teaching assistants through email and telephone contact.

Expenses:

- Provide schools with teaching assistant expenses invoices.

Intervention:

- Provide parental consent forms for the study.

Resources:

- Provide a teaching pack to accompany each intervention programme.
- Provide a set of books for each teaching assistant.

Communication:

- Send out regular updates on the progress of the project through newsletters.
- Disseminate research findings.

The School will:

Testing:

- Allow time for each testing phase and liaise with the research team to find appropriate dates and times for the testing to take place.

Training:

- Release teaching assistants so that they can attend the four day training course.
- Release teaching assistants so that they can attend the one day top up training.

Expenses:

- Complete and return all expense claim invoices in good time.

Intervention:

- During the intervention weeks allow teaching assistants to have the allocated amount of preparation time.
- To collect the required parental consent forms for the study.
- Allocate time for the intervention sessions to take place.
- Provide an appropriate area in school for the delivery of the intervention programmes.

Resources:

- Will allow the teaching assistant to have access to school resources needed to support the intervention programmes.
- Ensure that all resources provided are retained for the sole use of the project, until project completion in December 2014.
- Use the resources and intervention manuals provided by the research team as they wish after the completion of the project.
- Cover the costs of delivering the intervention to the 6 children in the waiting control group.

Communication:

- Inform the governor for special educational needs of the project.
- Ensure shared understanding and support of all school staff to the project and the personnel involved.
- To nominate a contact person as a point of information for parents/carers seeking more information about the project.

The Teaching Assistant will:**Training:**

- The teaching assistant will attend all training sessions and tutorials.

Expenses:

- Teaching assistants will claim back travel expenses once a term.

Intervention:

- Run sessions according to the manual instructions.

Communication:

- The teaching assistant will complete all necessary paperwork and submit it to the research team on time.

We commit to The **RE**ading for **CompreH**ension (**REACH**) Project as detailed above and in the project information pack.

On behalf of the research team:

Project Manager (PJC): _____

Date: _____

On behalf of the School:

Head teacher: _____

Teaching assistant (1): _____

Teaching assistant (2): _____

Date: _____

Please sign both of the copies, retaining one and returning the other to Dr Paula Clarke, School of Education, University of Leeds, LS2 9JT.

Appendix B.9 – Children’s information sheets and consent forms

PART 1

To be given with a verbal explanation from member of the research team once selected to participate in screening phase

Information Sheet for Children

The **REA**ding for **CompreH**ension (**REACH**) Project

What is the REACH project about?

The project aims to find out the best ways to help secondary school children to be able to read well and to be able understand what they are reading. It is a large study being carried out by researchers at three different universities and involves nearly 500 children from lots of different secondary schools.

Why have I been taken out of class?

You have been taken out of class to complete a short reading test, in which you will be asked to read some words out loud to me. Your school have selected you to take part and your parent/guardian has given their permission for you to be involved.

Who will see my test score?

The researchers will see your score and it will be shared with your head teacher if they ask to see it. No one else will be allowed to see your score. It will be stored on our computer and on the paper test sheet using a code number rather than your name to keep it secure and confidential.

What happens after I have finished the test?

We will use your score to work out whether or not you might be suitable to take part in the next stage of the project which will involve individual reading sessions with a teaching assistant using methods developed by the REACH research team. If you are chosen then you will be given more information to help you to decide whether or not you would like to have these sessions.

What if I change my mind about taking part?

If you no longer wish to be involved in this project you are free to leave at any time. You just need to let (insert school contact person name) know and we will take you out of the project. If during the reading test you decide you would like to stop then that is fine too just let me know and you can go back to class.

Do you have any questions?

CONSENT SLIP

I have read and understood the information given about the first stage of the **REACH** Project and would like to take part.

Signed..... **Date**.....

PRINT NAME

PART 2

To be given with a verbal explanation from teaching assistant once selected to participate in main trial

Information Sheet for Children

The READING for CompreHension (REACH) Project

What is the REACH project about?

The project aims to find out the best ways to help secondary school children to be able to read well and to be able understand what they are reading. It is a large study being carried out by researchers at three different universities and involves nearly 500 children from lots of different secondary schools.

Why have I been chosen to take part in the project?

You have been chosen because your score on the reading test you recently completed suggests that you might benefit from some individual reading sessions using the methods developed by the REACH research team.

What will the reading sessions be like?

If you decide that you would like to take part in stage two of the project you will either begin reading sessions in the summer term of Year 7 or in the spring term of Year 8 (this will be decided randomly). The sessions will take place three times a week and will last 35 minutes each. They will run for 20 weeks so you will have 60 sessions in total.

The sessions are designed to be lively and fun and will include some of the following activities:

- Reading short stories and information books
- Listening to sounds in words
- Making sounds and connecting them together
- Writing sentences
- Learning new and interesting words
- Telling jokes and thinking about why they are funny
- Listening to stories

These activities should help you to develop your reading skills so that you can access and enjoy greater range of books.

When will the reading sessions be?

You will need to miss some lessons to complete the reading sessions. We will try to have the sessions on different days and at different times to make sure that you do not always miss the same things. I will arrange this with your teachers. It is hoped that missing lessons will not disadvantage you in anyway; rather the reading sessions should help you in your lessons to be able to read and understand more.

What else will I be asked to do as part of the project?

The research team would like to find out about how much the reading sessions improve reading skills. This will be very useful information which could help lots of other pupils in secondary schools after this project has finished. To collect this information the research team will visit your school five times during the project and ask you to complete some tasks. At time points 1, 3 and 5 these will take about an hour to complete. At time points 2 and 4 these will take 15 minutes. You will also be given a reading activity to complete on the computer at each time point which will take about 45 minutes.

Who will see my test scores?

The researchers will see your scores and they will be shared with your head teacher at the end of the project if they ask to see them. No one else will be allowed to see your scores. They will be stored on our computer and on paper test sheets using a code number rather than your name to keep them secure and confidential.

What if I change my mind about taking part?

If you no longer wish to be involved in this project you are free to leave at any time. You just need to let me or (insert school contact person name) know and we will take you out of the project. If during the reading tests you decide you would like to stop then that is fine too just let the researcher know and you can go back to class.

Do you have any questions?

CONSENT SLIP

I have read and understood the information given about stage two of the **REACH** Project and would like to take part.

Signed..... **Date**.....

PRINT NAME

Appendix C: Teaching assistant survey

Teaching Assistant Survey on the Reading Intervention and Reading for Meaning Intervention

AN INDEPENDENT AND CONFIDENTIAL SURVEY

Any information you provide will be treated in the strictest confidence and will not be attributed to you personally. The results of this survey will primarily be used to evaluate the process and perceived impact of the Reading Intervention and Reading for Meaning Intervention, but may also be used in wider research and analysis projects about the intervention.

- The questionnaire should be completed by the teaching assistant at the school carrying out the intervention. You can complete the questionnaire on your computer in Word.
- Please read the instructions for answering each question carefully. Most questions ask you to "TICK ONE BOX ONLY". You can tick the boxes on your computer by clicking them.
- If you mark the wrong box, simply click in the box again to untick it.
- Please check you have answered all the questions that you should have answered
- Once you have completed the questionnaire please email it back to reachprocessevaluation@ipsos.com

The first few questions are about you.

Q1



Name of school **(Please write below)**

Q2



How many years have you been working at this school for? **(Write number in box)**

Q3

How many years have you been a teaching assistant for?
(Write number in box)



How many pupils did you personally teach in sessions as part of the intervention?

(Write number in box)

Q5



Overall, how positive or negative has your experience been with the delivery of the Reading Intervention?

(Please tick one box only)

Very positive

Fairly positive

Neither

Fairly negative

Very negative

Don't know

Q6



Have there been any negative aspects of your experience of delivering the Reading Intervention? If so, please write the details of these below.

(Please write in below)

Q7



Have there been any positive aspects of your experience of delivering the Reading Intervention? If so, please write the details of these below.

(Please write in below)

Q8



How likely, or unlikely, are you to recommend that other teaching assistants in other schools participate in a similar intervention?
(Please tick one box only)

- | | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Very likely | Fairly likely | Neither | Fairly unlikely | Very unlikely | Don't know |
| <input type="checkbox"/> |

Q9



How easy or difficult did you find it to implement the training you received for the sessions with the pupils?
(Please tick one box only)

- | | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Very easy | Fairly easy | Neither | Fairly difficult | Very difficult | Don't know |
| <input type="checkbox"/> |

Q10



How relevant, if at all, was the training you received for the sessions that you conducted with the pupils?
(Please tick one box only)

- | | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Very relevant | Fairly relevant | Neither | Not very relevant | Not at all relevant | Don't know |
| <input type="checkbox"/> |

Q11



How useful, if at all, were the materials you received in the training in helping you to deliver the intervention to the pupils?
(Please tick one box only)

- | | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Very useful | Fairly useful | Neither | Not very useful | Not at all useful | Don't know |
| <input type="checkbox"/> |

Q12



How supported, if at all, did you feel in your role in delivering the intervention within the school?
(Please tick one box only)

- | | | | | | |
|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| Very supported | Fairly supported | Neither | Not very supported | Not at all supported | Don't know |
| <input type="checkbox"/> |

-
-
-

Q13

Which, if any, of the below did you find as a barrier at your school in delivering the intervention?
(Please tick as many boxes that apply)

- Insufficient lesson time to deliver the sessions
- Teachers did not like their pupils being withdrawn from classes
- Pupils did not like being withdrawn from their regular classes
- Parents did not like their child being withdrawn from their regular classes
- Pupils were not engaged in the sessions
- Parents were unsupportive of the intervention
- There were not enough classroom spaces for the sessions
- There was not enough equipment provided by the school for the sessions
- Did not feel properly supported by senior staff in delivering the intervention
- Did not feel there was enough guidance and support once delivering the intervention
- There were no barriers
- Other (tick and write below)

For the next few questions, please refer to your session plans you created for the pupils over the last 20 weeks and consider the classes that pupils were withdrawn from.

Q14

How were the classes which the pupils were withdrawn from generally chosen?
(Please tick as many as apply)

- Choosing subjects the pupils were better at
- To deliberately give a mix of lessons the pupil missed
- At random
- To fit with my schedule
- Choosing subjects that were seen as less important for their education/ exams
- I was not involved with the arrangements at all
- Other

Q15

Please rank your answers from Q14 in order of what was most important when deciding which classes to withdraw pupils from with 1 being the most important consideration.

	Choosing subjects the pupils were better at
	To deliberately give a mix of lessons the pupil missed
	At random
	To fit with my schedule
	Choosing subjects that were seen as less important for their education/ exams
	I was not involved with the arrangements at all
	Other

Don't know

Q16 Which lessons were the pupils withdrawn from?
(Please tick one box only)

Mostly Maths

Mostly English

Mostly Sciences

Mostly Humanities

Mostly other subjects

It differed too much between pupils

Don't know



Q17 Please rank the top 3, with 1 being the most important, of what you found to be most useful in delivering the intervention.

	The senior staff in the school being supportive
	Other staff/teachers in the school being supportive
	The pupils were well engaged in the sessions
	Having access to expertise outside of the school
	Having access to expertise within the school
	The techniques taught in the training
	The materials taught in the training
	Having access to facilities within the school
	Other (Please write below)

Q18  On a scale of 1-10, with 1 being the least effective and 10 being the most effective, how effective did you think the intervention was in your school at raising reading and comprehension levels?
(Please tick one box only by clicking in it)

Least Effective											Most Effective
	1	2	3	4	5	6	7	8	9	10	
	<input type="checkbox"/>										

Q19

If the pupils who took part in the intervention had not taken part, do you think they would have been able to achieve the same outcomes?

(Please tick one box only)

Yes, definitely	Yes, maybe	Unsure	No, probably not	No, definitely not	Don't know
<input type="checkbox"/>					

For the next couple of questions please look back over the session plans again regarding the questions on pupil engagement and rating of the sessions.

Q20

In your opinion, to what extent were the pupils engaged in the sessions?

(Please tick one box only)

Highly engaged	Fairly engaged	Neither	Not very engaged	Not engaged at all	Too varied to tell	Don't know
<input type="checkbox"/>						

Q21



In your opinion, to what extent did the pupils enjoy the sessions?

(Please tick one box only)

Enjoyed them a lot	Enjoyed them somewhat	Neither	Did not enjoy them much	Did not enjoy them at all	Too varied to tell	Don't know
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>				

Q22

To what extent, if at all, have you applied any of the techniques and methods taught in the intervention to any of your other work?

(Please tick one box only)

- To a great extent
- To some extent
- Very little
- Not at all

Q23



Once the trial is over, how likely or unlikely do you think it is that the school will continue to use similar methods and techniques with pupils that are struggling with their reading?

(Please tick one box only)

Very likely

Fairly likely

Neither

Fairly unlikely

Very unlikely

Don't know

Q24



What, if any, are the barriers to sustaining the intervention beyond the lifetime of the funded trial?

(Please write in below)

Q25

Ipsos MORI may wish to speak to you further in the next six months about yours and the school's experiences of delivering the intervention.

Would it be OK for them to contact you to invite you to take part?

Yes

Please enter your phone number: _____

No

This is the end of the questionnaire. Thank you very much for completing it. Please return the completed version via email to reachprocessevaluation@ipsos.com

Appendix D: Robustness Checks

Table D1: Comparison of Baseline Characteristics – Phase 1

Variable	30-week (T1)		20-week (T2)		Control	Total N
	Mean (sd)	T1 –C (effect size)	Mean (sd)	T2 –C (effect size)	Mean (sd)	
Characteristics considered at randomisation						
% Female	0.500 (0.510)	-0.04	0.500 (0.510)	-0.04	0.520 (0.510)	77
Age (months)	146.885 (3.593)	0.234	147.538 (3.337)	0.415	146.040 (3.867)	77
SWRT at baseline	33.115 (6.575)	-0.143	32.577 (6.476)	-0.223	34.080 (7.308)	77
Additional baseline tests						
NGRT at baseline	240.769 (45.733)	-0.351	252.462 (49.125)	-0.126	259.040 (60.885)	77
WIAT at baseline	106.231 (15.570)	-0.092	101.192 (19.614)	-0.386	107.800 (15.722)	77
YARC at baseline	7.619 (2.376)	-0.042	7.588 (3.043)	-0.054	7.722 (2.109)	56
TOWRE baseline	79.038 (9.031)	0.141	75.192 (9.940)	-0.247	77.640 (10.700)	77
Additional characteristics from the National Pupil Database						
% Eligible for Free School Meals	0.26 1(0.449)	-0.213	0.304 (0.470)	-0.119	0.360 (0.490)	71
% English as an Additional Language	0.304 (0.470)	0.057	0.130 (0.344)	-0.348	0.280 (0.458)	71
% SEN (Statement or School Action Plus)	0.783 (0.422)	0.387	0.652 (0.487)	0.111	0.600 (0.500)	71
% Not White British	0.435 (0.507)	0.072	0.261 (0.449)	-0.287	0.400 (0.500)	71
KS2 English Points	2.927 (0.959)	0.17	2.757 (0.683)	-0.038	2.788 (0.781)	70
KS2 Maths Points	3.616 (0.739)	0.127	3.677 (0.720)	0.207	3.520 (0.839)	70
Total Sample Size	26		26		25	77
Total Sample with NPD data	23		23		25	71

Note: * indicates that the difference in means (TX-C) is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in square brackets.

Table D2: Comparison of Baseline Characteristics – Phase 2

Variable	30-week (T1)		20-week (T2)		Control	Total N
	Mean (sd)	T1 –C (effect size)	Mean (sd)	T2 –C (effect size)	Mean (sd)	
Characteristics considered at randomisation						
% Female	0.386 (0.493)	-0.017	0.395 (0.495)	0.001	0.395 (0.495)	125
Age (months)	139.977 (4.190)	0.149	139.814 (3.838)	0.107	139.395 (3.702)	125
SWRT at baseline	31.682 (9.114)	0.099	31.814 (8.787)	0.114	30.789 (9.373)	125
Additional baseline tests						
NGRT at baseline	223.795 (62.653)	0.026	241.674 (50.545)	0.337	222.316 (57.589)	125
WIAT at baseline	100.886 (16.880)	-0.14	101.674 (16.981)	-0.096	103.368 (19.686)	125
YARC at baseline	6.452* (1.895)	-0.583	6.967 (1.671)	-0.352	7.750 (2.982)	85
TOWRE baseline	72.045 (9.447)	-0.107	74.907 (10.982)	0.15	73.237 (13.033)	125
Additional characteristics from the National Pupil Database						
% Eligible for Free School Meals	0.279 (0.454)	-0.116	0.351 (0.484)	0.038	0.333 (0.479)	113
% English as an Additional Language	0.163 (0.374)	-0.26	0.270 (0.450)	-0.006	0.273 (0.452)	113
% SEN (Statement or School Action Plus)	0.558 (0.502)	-0.097	0.595 (0.498)	-0.023	0.606 (0.496)	113
% Not White British	0.209 (0.412)	-0.277	0.297 (0.463)	-0.08	0.333 (0.479)	113
KS2 English Points	3.093* (0.500)	-0.54	3.088* (0.607)	-0.549	3.406 (0.587)	111
KS2 Maths Points	3.410 (0.859)	-0.331	3.419 (0.568)	-0.318	3.657 (0.790)	110
Total Sample Size	44		43		38	125
Total Sample with NPD data	43		37		33	113

Note: * indicates that the difference in means (TX-C) is significant at the 10% level ** at the 5% level *** at the 1% level. Standard deviations are reported in square brackets.

Table D3: Alternative treatment effect estimates for all pupils

	Methodology					
	(1) Raw	(2) OLS	(3) Random Effects	(4) Fixed Effects	(5) FILM	(6) Kernel Matching
<u>REACH Reading Intervention</u>						
NGRT – Primary Outcome	0.258	0.348***	0.373***	0.402***	0.329***	0.228
	[0.155]	[0.103]	[0.113]	[0.118]	[0.099]	[0.174]
Reading comprehension (secondary composite outcome)	-0.272*	-0.101	-0.098***	-0.048***	-0.079	-0.136
	[0.152]	[0.142]	[0.149]	[0.151]	[0.170]	[0.194]
Reading accuracy (secondary composite outcome)	0.135	0.167**	0.175***	0.185***	0.167***	0.131
	[0.160]	[0.059]	[0.062]	[0.064]	[0.058]	[0.108]
<u>REACH Reading Intervention with Language Comprehension</u>						
NGRT – Primary Outcome	0.509***	0.473***	0.486***	0.476***	0.506***	0.268
	[0.148]	[0.102]	[0.092]	[0.091]	[0.087]	[0.165]
Reading comprehension (secondary composite outcome)	-0.03	0.104	0.123***	0.090***	0.136	0.286
	[0.148]	[0.109]	[0.110]	[0.128]	[0.150]	[0.215]
Reading accuracy (secondary composite outcome)	0.148	0.119*	0.135**	0.136**	0.153***	0.008
	[0.119]	[0.058]	[0.057]	[0.058]	[0.059]	[0.128]
<p>Note: * indicates that the difference in means is significant at the 10% level ** at the 5% level *** at the 1% level. Standard errors are clustered at the school level and are reported in square brackets. All outcomes are standardised within the estimation sample prior to estimation. Columns (2)-(6) control for the following covariates: age in months, gender, cohort and pre-test scores in the NGRT, Single Word Reading Test, WIAT reading comprehension test and TOWRE measure of word reading efficiency. Column (2) controls for covariates using Ordinary Least Squares. Column (3) allows for a school effect that is uncorrelated with covariates. Column (4) estimates a fixed effect for each school. Column (5) allows the treatment effect to linearly interact with the treatment. Column (6) uses kernel propensity score matching to balance the samples.</p>						

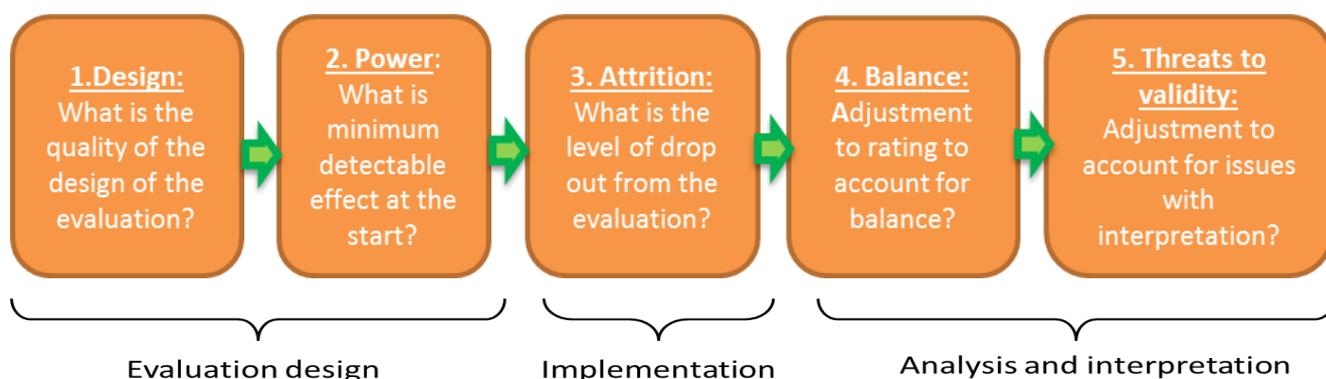
Appendix E: Cost rating

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
£	<i>Very low:</i> less than £80 per pupil per year.
£ £	<i>Low:</i> up to about £200 per pupil per year.
£ £ £	<i>Moderate:</i> up to about £700 per pupil per year.
£ £ £ £	<i>High:</i> up to £1,200 per pupil per year.
£ £ £ £ £	<i>Very high:</i> over £1,200 per pupil per year.

Appendix F: Security classification of trial findings

25 February 2016 completed by Camilla Nevill



Rating	1. Design	2. Power (MDES)	3. Attrition	4. Balance	5. Threats to validity
5	Fair and clear experimental design (RCT)	< 0.2	< 10%	Well-balanced on observables	No threats to validity
4	Fair and clear experimental design (RCT, RDD)	< 0.3	< 20%	↓	↓
3	Well-matched comparison (quasi-experiment)	< 0.4	< 30%	↓	↓
2	Matched comparison (quasi-experiment)	< 0.5	< 40%	↓	↓
1	Comparison group with poor or no matching	< 0.6	< 50%	↓	↓
0	No comparator	> 0.6	> 50%	Imbalanced on observables	Significant threats

The final security rating for this trial is 2 . This means that the conclusions have moderate to low security.

The trial was designed as an efficacy trial and could achieve a maximum of 4 . Attrition among pupils was high at 29%.

Importantly, the original design had to be changed because of delays in recruiting schools, meaning that the trial was run in two separate phases with slightly different experimental conditions, rather than as a single trial, as planned. Whilst balance was achieved when the data from the 2 phases was combined, this was not the case for the individual phases.

Finally, the process evaluation suggested that some participating TAs used some of the intervention techniques they had learned when teaching pupils from the comparison group. These pupils were not supposed to receive the REACH interventions, and the fact that they did makes it harder interpret the findings.

You may re-use this document/publication (not including logos) free of charge in any format or medium, under the terms of the Open Government Licence v2.0.

To view this licence, visit www.nationalarchives.gov.uk/doc/open-government-licence/version/2 or email: psi@nationalarchives.gsi.gov.uk

Where we have identified any third party copyright information you will need to obtain permission from the copyright holders concerned. The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

This document is available for download at www.educationendowmentfoundation.org.uk



Education
Endowment
Foundation

The Education Endowment Foundation

9th Floor, Millbank Tower

21–24 Millbank

London

SW1P 4QP

www.educationendowmentfoundation.org.uk